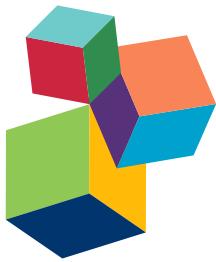


BRAIN OSCILLATIONS IN HUMAN COMMUNICATION

EDITED BY: Johanna Rimmele, Joachim Gross, Sophie Molholm and Anne Keitel

PUBLISHED IN: Frontiers in Human Neuroscience





frontiers

Frontiers Copyright Statement

© Copyright 2007-2018 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88945-458-7

DOI 10.3389/978-2-88945-458-7

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

BRAIN OSCILLATIONS IN HUMAN COMMUNICATION

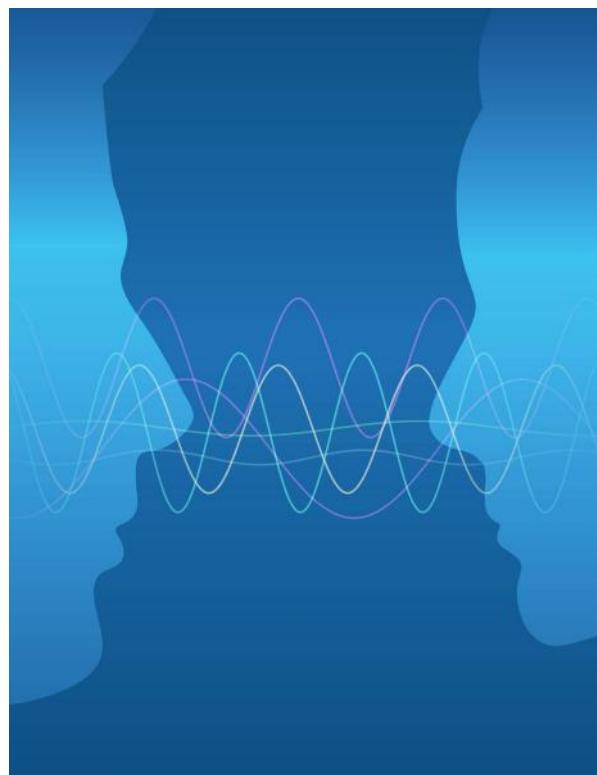
Topic Editors:

Johanna Rimmele, Max Planck Institute for Empirical Aesthetics, Germany

Joachim Gross, University of Münster, Germany

Sophie Molholm, Albert Einstein College of Medicine, United States

Anne Keitel, University of Glasgow, United Kingdom



Oscillating brains. Image: Felix Bernoulli.

Brain oscillations, or neural rhythms, reflect widespread functional connections between large-scale neural networks, as well as within cortical networks. As such they have been related to many aspects of human behaviour. An increasing number of studies have demonstrated the role of brain oscillations at distinct frequency bands in cognitive, sensory and motor tasks. Consequentially, those rhythms also affect diverse aspects of human communication. On the one hand, this comprises verbal communication; a field where the understanding of neural mechanisms has

seen huge advances in recent years. Speech is inherently organised in a quasi-rhythmic manner. For example, time scales of phonemes and syllables, but also formal prosodic aspects such as intonation and stress, fall into distinct frequency bands. Likewise, neural rhythms in the brain play a role in speech segmentation and coding of continuous speech at multiple time scales, as well as in the production of speech. On the other hand, human communication involves widespread and diverse nonverbal aspects where the role of neural rhythms is far less understood. This can be the enhancement of speech processing through visual signals, thought to be guided via brain oscillations, or the conveying of emotion, which results in differential modulations of brain rhythms in the observer. Additionally, body movements and gestures often have a communicative purpose and are known to modulate sensorimotor rhythms in the observer.

This Research Topic of Frontiers in Human Neuroscience highlights the diverse aspects of human communication that are shaped by rhythmic activity in the brain. Relevant contributions are presented from various fields including cognitive and social neuroscience, neuropsychiatry, and methodology. As such they provide important new insights into verbal and non-verbal communication, pathological changes, and methodological innovations.

Citation: Rimmeli, J., Gross, J., Molholm, S., Keitel, A., eds. (2018). Brain Oscillations in Human Communication. Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-458-7

Table of Contents

- 06 Editorial: Brain Oscillations in Human Communication**
Johanna M. Rimmeli, Joachim Gross, Sophie Molholm and Anne Keitel
- Perception of Spoken and Written Language**
- 10 Multi-Scale Entrainment of Coupled Neuronal Oscillations in Primary Auditory Cortex**
M. N. O'Connell, A. Barczak, D. Ross, T. McGinnis, C. E. Schroeder and P. Lakatos
- 26 Exploring the Role of Brain Oscillations in Speech Perception in Noise: Intelligibility of Isochronously Retimed Speech**
Vincent Aubanel, Chris Davis and Jeesun Kim
- 37 The Role of High-Level Processes for Oscillatory Phase Entrainment to Speech Sound**
Benedikt Zoefel and Rufin VanRullen
- 49 A Predictive Coding Perspective on Beta Oscillations during Sentence-Level Language Comprehension**
Ashley G. Lewis, Jan-Mathijs Schoffelen, Herbert Schriefers and Marcel Bastiaansen
- Production of Language**
- 55 Auditory cortical deactivation during speech production and following speech perception: an EEG investigation of the temporal dynamics of the auditory alpha rhythm**
David Jenson, Ashley W. Harkrider, David Thornton, Andrew L. Bowers and Tim Saltuklaroglu
- 72 Withholding planned speech is reflected in synchronized beta-band oscillations**
Vitória Piai, Ardi Roelofs, Joost Rommers, Kristoffer Dahlslätt and Eric Maris
- Social Cognition and Interpersonal Communication**
- 82 Electrocorticographic Activation within Human Auditory Cortex during Dialog-Based Language and Cognitive Testing**
Kirill V. Nourski, Mitchell Steinschneider and Ariane E. Rhone
- 97 The coordination dynamics of social neuromarkers**
Emmanuelle Tognoli and J. A. Scott Kelso
- Non-Verbal Communication**
- 113 The Functional Role of Neural Oscillations in Non-Verbal Emotional Communication**
Ashley E. Symons, Wael El-Deredy, Michael Schwartze and Sonja A. Kotz

- 127 *Synchronization by the hand: the sight of gestures modulates low-frequency activity in brain responses to continuous speech***

Emmanuel Biau and Salvador Soto-Faraco

Pathological Changes to Communication in Autism and Schizophrenia

- 133 *Atypical coordination of cortical oscillations in response to speech in autism***

Delphine Jochaut, Katia Lehongre, Ana Saitovitch, Anne-Dominique Devauchelle, Itsaso Olasagasti, Nadia Chabane, Monica Zilbovicius and Anne-Lise Giraud

- 145 *The Oscillopathic Nature of Language Deficits in Autism: From Genes to Language Evolution***

Antonio Benítez-Burraco and Elliot Murphy

- 162 *Bridging the Gap between Genes and Language Deficits in Schizophrenia: An Oscillopathic Approach***

Elliot Murphy and Antonio Benítez-Burraco

Methodological Considerations and Innovations

- 177 *Flicker-Driven Responses in Visual Cortex Change during Matched-Frequency Transcranial Alternating Current Stimulation***

Philipp Ruhnau, Christian Keitel, Chrysa Lithari, Nathan Weisz and Toralf Neuling

- 190 *Interpretations of Frequency Domain Analyses of Neural Entrainment: Periodicity, Fundamental Frequency, and Harmonics***

Hong Zhou, Lucia Melloni, David Poeppel and Nai Ding



Editorial: Brain Oscillations in Human Communication

Johanna M. Rimmeli^{1*}, Joachim Gross², Sophie Molholm³ and Anne Keitel^{4*}

¹ Department of Neuroscience, Max Planck Institute for Empirical Aesthetics (MPG), Frankfurt am Main, Germany, ² Institut für Biomagnetismus und Biosignalanalyse, Universitätsklinikum Münster, Münster, Germany, ³ Departments of Pediatrics and Neuroscience, Albert Einstein College of Medicine, Bronx, NY, United States, ⁴ Centre for Cognitive Neuroimaging, University of Glasgow, Glasgow, United Kingdom

Keywords: speech perception and production, non-verbal communication, brain rhythms, neurobiology of language, MEG source analysis, communication disorders

Editorial on the Research Topic

Brain Oscillations in Human Communication

This Research Topic featured 15 articles from a wide range of research areas related to human communication. All contributions focus on rhythmic brain activity as opposed to, for example, event related potentials or functional imaging approaches. Rhythmic brain activity has been shown to be of immense importance for the temporal coordination of neural activity and, consequently, for all aspects of cognition and behaviour (Buzsaki and Draguhn, 2004; Wang, 2010). In this editorial, we summarise the research on rhythmic brain activity in language and communication that appeared in this Research Topic (see Table 1).

PERCEPTION OF SPOKEN LANGUAGE

Low-frequency (delta, theta) neuronal oscillations (and the coupling with gamma-band oscillations) have been shown to have a crucial role in speech perception, particularly in the segmentation of the continuous acoustic stream into linguistically meaningful units (Giraud and Poeppel, 2012; Ding and Simon, 2014; Ding et al., 2016; Keitel et al., in review). By now, many studies have shown that low-frequency neuronal activity in auditory cortex tracks the slow energy fluctuations in the speech acoustics (Luo and Poeppel, 2007; Gross et al., 2013; Rimmeli et al., 2015; Keitel et al., 2017). The functional relevance of this neuronal phase alignment, however, and whether it actually reflects the entrainment of endogenous neuronal oscillations to a speech signal (as opposed to event related potentials) are controversial.

O'Connell et al. shed light on the tonotopy of multiscale neuronal entrainment in auditory cortex (A1) by using single cell recordings in macaques. They provide evidence for multiscale entrainment to clicks presented at regular intervals in the gamma and delta range. Particularly, neurons in the region of 11–16 kHz on the tonotopic maps of A1 aligned their excitability to the attended sounds, while the remaining part of A1 showed response suppression. The findings suggest a function of cortical entrainment to a rhythmic stimulation in the selection and processing of attended sounds (O'Connell et al., 2014), which might be at play in speech perception.

A crucial question is which aspects of the speech signal trigger the low-frequency phase alignment to the speech acoustics that are involved in speech comprehension. Aubanel et al. investigated this by isochronously re-timing temporally distorted speech with anchor points either at syllable onsets (linguistic cues) or at amplitude envelope peaks (acoustic cues). They showed that speech comprehension benefits most from linguistically motivated cues, that is, a re-timing at stressed syllable onsets.

The extent to which low-frequency neuronal phase alignment can be modulated by higher-level processes is controversial (Haegens and Golumbic, 2017; Teng et al., 2017). By reviewing

OPEN ACCESS

Edited and reviewed by:

Mikhail Lebedev,
Duke University, United States

*Correspondence:

Johanna M. Rimmeli
johanna.rimmeli@aesthetics.mpg.de
Anne Keitel
anne.keitel@glasgow.ac.uk

Received: 22 January 2018

Accepted: 24 January 2018

Published: 07 February 2018

Citation:

Rimmeli JM, Gross J, Molholm S and Keitel A (2018) Editorial: Brain Oscillations in Human Communication. *Front. Hum. Neurosci.* 12:39.
doi: 10.3389/fnhum.2018.00039

TABLE 1 | Papers in this issue categorized along methodological and conceptual dimensions.

Authors	Type	Group	Method	Modality	Domain
Aubanel et al.	Original research	Healthy human	Behavioural	Auditory	Speech perception
Benítez-Burraco and Murphy	Hypothesis and Theory	ASD	All	Auditory Visual	Language perception
Biau and Soto-Faraco	Perspective				
Jenson et al.	Original research	Healthy human	EEG	Auditory	Speech perception, production
Jochaut et al.	Original research	ASD	EEG, fMRI	Auditory	Speech perception
Lewis et al.	Perspective	Healthy human	All, MEG	Auditory Motor	Speech processing
Murphy and Benítez-Burraco	Hypothesis and Theory	Schizophrenia	All	Auditory Visual	Language perception
Nourski et al.	Original research	Epilepsy	ECoG, electrode arrays	Auditory Visual Motor	Speech perception, production
O'Connell et al.	Original research	Macaque	Electrode array	Auditory	Speech perception
Piai et al.	Original research	Healthy human	MEG	Auditory	Speech production
Ruhnau et al.	Original research	Healthy human	tACS	Visual	Flicker
Symons et al.	Review	Healthy human	EEG, MEG	Emotion Auditory Visual	Emotion Non verbal communication
Tognoli and Kelso	Hypothesis and Theory	Healthy human	EEG, MEG	Auditory Motor	Social cognition, communication
Zhou et al.	Hypothesis and Theory	Healthy human	EEG, MEG, LFP	Auditory	Speech, music perception
Zoefel and VanRullen	Review	Healthy human	All	Auditory	Speech perception

ASD, autism spectrum disorder; EEG, electroencephalography; ECoG, electrocorticography; fMRI, functional magnetic resonance imaging; LFP, local field potential; MEG, magnetoencephalography; tACS, transcranial alternating current stimulation.

previous literature, Zoefel and VanRullen discuss the contribution of high-level processes to the cortical low-frequency entrainment to speech. Particularly, based on high-level modulations of speech processing, as well as the finding of entrainment in the absence of low-level rhythmical cues (Zoefel and VanRullen, 2015, 2016) they suggest that the theta phase alignment indicates oscillatory entrainment and not mere stimulus-driven responses caused by a rhythmic stimulation (cf. Keitel et al., 2014). The authors discuss the functional role of cortical entrainment with respect to speech intelligibility and provide theoretical considerations to integrate stimulus driven effects and top-down modulations of speech processing into a unified model.

Lewis et al. focus on a specific aspect of high-level processes in spoken language comprehension. They review current research on the role of beta-band oscillations in sentence processing. This research suggests a function either in sentence-level top-down predictions about the up-coming linguistic input, or in indicating the maintenance of the current processing mode, relevant for deriving the meaning of a sentence (Lewis and Bastiaansen, 2015; Lewis et al., 2015). In this review, they additionally present preliminary magnetoencephalography (MEG) data, supporting the latter interpretation.

SPEECH PRODUCTION AND NEURONAL OSCILLATIONS

Although cortical oscillations have been shown to be involved in sensorimotor coordination (Arnal and Giraud, 2012), which is vital for speech production, the exact function of oscillations in speech production is little understood. Jenson et al. investigated the temporal dynamics of alpha-band activity in the auditory posterior dorsal stream during speech production and perception, using a novel analysis method (combining independent component analysis and event related spectral perturbations) in electroencephalography (EEG) recordings.

Together with previous findings by Jenson et al. (2014), these results show the temporal dynamics of the anterior and posterior dorsal stream during speech perception and production. In sum, their findings suggest a crucial role of alpha oscillations in sensorimotor interactions that allow monitoring the speech production through efference copies from the motor system.

The ability to control our speech output and withhold planned speech is critical during communication, as we need to time the turn-taking of the interacting partners (Wilson and Wilson, 2005). In an MEG study, Piai et al. investigated neuronal activity during the withholding of planned speech. They provide evidence for a two-fold mechanism. First, alpha-band desynchronisation in occipital brain areas might indicate the task-specific allocation of attention during the withholding of speech. Second, increased frontal beta-band activity during the withholding of speech most likely indicates the maintenance of the current motor or cognitive state, i.e., maintaining the planned verbal response (Engel and Fries, 2010; Piai et al., 2015; Rimmele et al., in review).

SOCIAL COGNITION AND COMMUNICATION

In a more natural communication setting, Nourski et al. tested speech production and perception during a conversation between epilepsy patients and an instructor, using electrocorticography (ECoG) recordings. They found no difference in high gamma activity in the auditory core cortex when listening to self-produced speech vs. the speech of others. However, gamma-activity was reduced in non-core areas when participants listened to their own speech. The findings indicate that signals from self-produced speech are differentiated from speech of others at higher non-core auditory processing areas, and high gamma oscillations play a role in these processes (Nourski, 2017).

Tognoli and Kelso approach social cognition and communication beyond the individual brain, by pursuing the hypothesis that social interaction results in the phase-locking

and coupling of neuronal activity across brains (e.g., Dumas et al., 2010). They theoretically underpin a neuromarker approach to social cognition, review findings from dual-EEG recordings (Tognoli et al., 2007a,b), and discuss them in the context of previous research on social cognition. In sum, neuromarkers in the alpha, mu, kappa and phi bands seem to be differentially involved in simultaneous action and perception processes (e.g., tango dancing) and the alternating perception and production of social behaviour (e.g., imitating someone).

OSCILLATIONS IN NON-VERBAL COMMUNICATION

For successful interpersonal communication, it is crucial to detect and identify emotional expressions from auditory, visual, and audiovisual information (Jessen and Kotz, 2011; Kotz et al., 2013). Here, Symons et al. review the available literature focussing on oscillatory mechanisms. They conclude that theta- and gamma-band synchronisation most consistently reflect the processing of emotional expressions across sensory modalities (e.g., Knyazev, 2007; Luo et al., 2008). On the other hand, oscillations in the delta-, alpha-, and beta-bands have also been implied in the processing of other's emotions, but their role is less consistent across modalities and tasks.

Another important non-verbal aspect of speech is the use of gestures (Hubbard et al., 2009; Biau and Soto-Faraco, 2013). Based on the previous finding that hand gestures phase-reset ongoing neural oscillations (Biau et al., 2015), Biau and Soto-Faraco discuss the role of beat gestures in audiovisual speech processing. They conclude that beat gestures promote a cross-modal phase reset at important word onsets, which might facilitate the segmentation of the speech stream.

PATHOLOGICAL CHANGES TO COMMUNICATION IN AUTISM AND SCHIZOPHRENIA

Many psychological and neurological conditions also affect the production or perception of language (e.g., Uhlhaas and Singer, 2006). Jochaut et al. investigated the response to continuous speech in individuals with and without autism spectrum disorder (ASD), using concurrent EEG and fMRI. They report anomalies of theta and gamma oscillations in the left auditory cortex in ASD participants, as well as altered functional connectivity between auditory and other language cortices. Furthermore, the theta/gamma coupling predicted verbal impairment as well as ASD symptoms.

Benítez-Burraco and Murphy take a different approach to language deficits in ASD. In their theoretical article, they

propose to relate genetics to the pathophysiology of ASD by studying oscillatory mechanisms for language processing in the autistic brain. They note that candidate genes for ASD are overrepresented among the genes that played a role in the evolution of language and brain oscillations, thereby bringing together these different methodological approaches.

In a second theoretical article, Murphy and Benítez-Burraco (2016) take a similar approach toward understanding the relationship between genes and language deficits in schizophrenia. Here, they suggest that the language deficits in schizophrenia seem to be rooted in the evolutionary processes that brought about modern language. This evolutionary account and the common oscillatory profiles of language deficits in schizophrenia and ASD are further described in a recent article (Murphy and Benítez-Burraco, 2016).

METHODOLOGICAL CONSIDERATIONS AND INNOVATIONS

This Research Topic also featured predominantly methodological articles that could advance research into communication processes. Ruhnau et al. used a novel combination of concurrent transcranial alternating current stimulation (tACS) and frequency tagging. They carefully tease apart source-level tACS effects and steady-state responses (SSRs) in the MEG, by using a new method to reconstruct sources of SSRs that are unaffected by the strong tACS artifact (Neuling et al., 2015). Frequency tagging is a potentially fruitful approach for studying speech processing (Buiatti et al., 2009), and this proof-of-principle study opens up new possibilities to combine frequency tagging, tACS, and MEG.

Finally, Zhou et al. discuss the implications and limitations of the often used Fourier analysis to study low-frequency neural entrainment. They conclude that true low-frequency entrainment results in a peak in the power spectrum at the fundamental frequency (the lowest frequency produced by an oscillation), and describe how the phenomenon of higher harmonics can be interpreted.

In conclusion, the studies included here review, highlight, and specify the role of distinct cortical oscillations in practically all processes related to human communication.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

JG was supported by the Wellcome Trust (Joint Senior Investigator Grant, No 098433).

REFERENCES

- Arnal, L. H., and Giraud, A. L. (2012). Cortical oscillations and sensory predictions. *Trends Cogn. Sci.* 16, 390–398. doi: 10.1016/j.tics.2012.05.003

- Biau, E., and Soto-Faraco, S. (2013). Beat gestures modulate auditory integration in speech perception. *Brain Lang.* 124, 143–152. doi: 10.1016/j.bandl.2012.10.008
- Biau, E., Torralba, M., Fuentemilla, L., de Diego Balaguer, R., and Soto-Faraco, S. (2015). Speaker's hand gestures modulate speech perception

- through phase resetting of ongoing neural oscillations. *Cortex* 68, 76–85. doi: 10.1016/j.cortex.2014.11.018
- Buiatti, M., Peña, M., and Dehaene-Lambertz, G. (2009). Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. *Neuroimage* 44, 509–519. doi: 10.1016/j.neuroimage.2008.09.015
- Buzsaki, G., and Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science* 304, 1926–1929. doi: doi: 10.1126/science.1099745
- Ding, N., Melloni, L., Zhang, H., Tian, X., and Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nat. Neurosci.* 19, 158–164. doi: 10.1038/nn.4186
- Ding, N., and Simon, J. Z. (2014). Cortical entrainment to continuous speech: functional roles and interpretations. *Front. Hum. Neurosci.* 8:311. doi: 10.3389/fnhum.2014.00311
- Dumas, G., Nadel, J., Soussignan, R., Martinerie, J., and Garnero, L. (2010). Inter-brain synchronization during social interaction. *PLoS ONE* 5:e12166. doi: 10.1371/journal.pone.0012166
- Engel, A. K., and Fries, P. (2010). Beta-band oscillations—signalling the status quo? *Curr. Opin. Neurobiol.* 20, 156–165. doi: 10.1016/j.conb.2010.02.015
- Giraud, A. L., and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* 15, 511–517. doi: 10.1038/nn.3063
- Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., et al. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biol.* 11:e1001752. doi: 10.1371/journal.pbio.1001752
- Haegens, S., and Golumbic, E. Z. (2017). Rhythmic facilitation of sensory processing: a critical review. *Neurosci. Biobehav. Rev.* pii: S0149-7634(17)30500-6. doi: 10.1016/j.neubiorev.2017.12.002
- Hubbard, A. L., Wilson, S. M., Callan, D. E., and Dapretto, M. (2009). Giving speech a hand: Gesture modulates activity in auditory cortex during speech perception. *Hum. Brain Mapp.* 30, 1028–1037. doi: 10.1002/hbm.20565
- Jenson, D., Bowers, A. L., Harkrider, A. W., Thornton, D., Cuellar, M., and Saltuklaroglu, T. (2014). Temporal dynamics of sensorimotor integration in speech perception and production: independent component analysis of EEG data. *Front. Psychol.* 5:656. doi: 10.3389/fpsyg.2014.00656
- Jessen, S., and Kotz, S. A. (2011). The temporal dynamics of processing emotions from vocal, facial, and bodily expressions. *Neuroimage* 58, 665–674. doi: 10.1016/j.neuroimage.2011.06.035
- Keitel, A., Ince, R. A., Gross, J., and Kayser, C. (2017). Auditory cortical delta-entrainment interacts with oscillatory power in multiple fronto-parietal networks. *Neuroimage* 147, 32–42. doi: 10.1016/j.neuroimage.2016.11.062
- Keitel, C., Quigley, C., and Ruhnau, P. (2014). Stimulus-driven brain oscillations in the alpha range: entrainment of intrinsic rhythms or frequency-following response? *J. Neurosci.* 34, 10137–10140. doi: 10.1523/JNEUROSCI.1904-14.2014
- Knyazev, G. G. (2007). Motivation, emotion, and their inhibitory control mirrored in brain oscillations. *Neurosci. Biobehav. Rev.* 31, 377–395. doi: 10.1016/j.neubiorev.2006.10.004
- Kotz, S. A., Kalberlah, C., Bahlmann, J., Friederici, A. D., and Haynes, J. D. (2013). Predicting vocal emotion expressions from the human brain. *Hum. Brain Mapp.* 34, 1971–1981. doi: 10.1002/hbm.22041
- Lewis, A. G., and Bastiaansen, M. (2015). A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. *Cortex* 68, 155–168. doi: 10.1016/j.cortex.2015.02.014
- Lewis, A. G., Wang, L., and Bastiaansen, M. (2015). Fast oscillatory dynamics during language comprehension: Unification versus maintenance and prediction? *Brain Lang.* 148, 51–63. doi: 10.1016/j.bandl.2015.01.003
- Luo, H., and Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54, 1001–1010. doi: 10.1016/j.neuron.2007.06.004
- Luo, Q., Mitchell, D., Cheng, X., Mondillo, K., McCaffrey, D., Holroyd, T., et al. (2008). Visual awareness, emotion, and gamma band synchronization. *Cereb. Cortex* 19, 1896–1904. doi: 10.1093/cercor/bhn216
- Murphy, E., and Benítez-Burraco, A. (2016). Language deficits in schizophrenia and autism as related oscillatory connectopathies: an evolutionary account. *Neurosci. Biobehav. Rev.* 83, 742–764. doi: 10.1016/j.neubiorev.2016.07.029
- Neuling, T., Ruhnau, P., Fuscà, M., Demarchi, G., Herrmann, C. S., and Weisz, N. (2015). Friends, not foes: magnetoencephalography as a tool to uncover brain dynamics during transcranial alternating current stimulation. *Neuroimage* 118, 406–413. doi: 10.1016/j.neuroimage.2015.06.026
- Nourski, K. V. (2017). Auditory processing in the human cortex: An intracranial electrophysiology perspective. *Laryngoscope Investig. Otolaryngol.* 2, 147–156. doi: 10.1002/lio2.73
- O'Connell, M. N., Barczak, A., Schroeder, C. E., and Lakatos, P. (2014). Layer specific sharpening of frequency tuning by selective attention in primary auditory cortex. *J. Neurosci.* 34, 16496–16508. doi: 10.1523/JNEUROSCI.2055-14.2014
- Piai, V., Roelofs, A., Rommers, J., and Maris, E. (2015). Beta oscillations reflect memory and motor aspects of spoken word production. *Hum. Brain Mapp.* 36, 2767–2780. doi: 10.1002/hbm.22806
- Rimmele, J. M., Zion Golumbic, E., Schroger, E., and Poeppel, D. (2015). The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene. *Cortex* 68, 144–154. doi: 10.1016/j.cortex.2014.12.014
- Teng, X., Tian, X., Doelling, K., and Poeppel, D. (2017). Theta band oscillations reflect more than entrainment: behavioral and neural evidence demonstrates an active chunking process. *Eur. J. Neurosci.* doi: 10.1111/ejn.13742. [Epub ahead of print].
- Tognoli, E., Lagarde, J., DeGuzman, G. C., and Kelso, J. A. S. (2007a). The phi complex as a neuromarker of human social coordination. *Proc. Natl. Acad. Sci. U.S.A.* 104, 8190–8195. doi: 10.1073/pnas.0611453104
- Tognoli, E., Magne, C., de Guzman, G. C., and Kelso, J. A. S. (2007b). “Brain rhythms underlying intentional social coordination,” in *Society for Neuroscience Itinerary Planner* (San Diego, CA: Society for Neuroscience).
- Uhlhaas, P. J., and Singer, W. (2006). Neural synchrony in brain disorders: relevance for cognitive dysfunctions and pathophysiology. *Neuron* 52, 155–168. doi: 10.1016/j.neuron.2006.09.020
- Wang, X. J. (2010). Neurophysiological and computational principles of cortical rhythms in cognition. *Physiol. Rev.* 90, 1195–1268. doi: 10.1152/physrev.00035.2008
- Wilson, M., and Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychon. Bull. Rev.* 12, 957–968. doi: 10.3758/BF03206432
- Zoefel, B., and VanRullen, R. (2015). Selective perceptual phase entrainment to speech rhythm in the absence of spectral energy fluctuations. *J. Neurosci.* 35, 1954–1964. doi: 10.1523/JNEUROSCI.3484-14.2015
- Zoefel, B., and VanRullen, R. (2016). EEG oscillations entrain their phase to high-level features of speech sound. *Neuroimage* 124, 16–23. doi: 10.1016/j.neuroimage.2015.08.054

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Rimmele, Gross, Molholm and Keitel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Multi-Scale Entrainment of Coupled Neuronal Oscillations in Primary Auditory Cortex

M. N. O'Connell¹, A. Barczak¹, D. Ross¹, T. McGinnis¹, C. E. Schroeder^{1,2}
and P. Lakatos^{1,3*}

¹ Cognitive Neuroscience and Schizophrenia Program, Nathan Kline Institute, Orangeburg, NY, USA, ² Department of Psychiatry, Columbia College of Physicians and Surgeons, New York, NY, USA, ³ Department of Psychiatry, NYU School of Medicine, New York, NY, USA

Earlier studies demonstrate that when the frequency of rhythmic tone sequences or streams is task relevant, ongoing excitability fluctuations (oscillations) of neuronal ensembles in primary auditory cortex (A1) entrain to stimulation in a frequency dependent way that sharpens frequency tuning. The phase distribution across A1 neuronal ensembles at time points when attended stimuli are predicted to occur reflects the focus of attention along the spectral attribute of auditory stimuli. This study examined how neuronal activity is modulated if only the temporal features of rhythmic stimulus streams are relevant. We presented macaques with auditory clicks arranged in 33 Hz (gamma timescale) quintets, repeated at a 1.6 Hz (delta timescale) rate. Such multi-scale, hierarchically organized temporal structure is characteristic of vocalizations and other natural stimuli. Monkeys were required to detect and respond to deviations in the temporal pattern of gamma quintets. As expected, engagement in the auditory task resulted in the multi-scale entrainment of delta- and gamma-band neuronal oscillations across all of A1. Surprisingly, however, the phase-alignment, and thus, the physiological impact of entrainment differed across the tonotopic map in A1. In the region of 11–16 kHz representation, entrainment most often aligned high excitability oscillatory phases with task-relevant events in the input stream and thus resulted in response enhancement. In the remainder of the A1 sites, entrainment generally resulted in response suppression. Our data indicate that the suppressive effects were due to low excitability phase delta oscillatory entrainment and the phase amplitude coupling of delta and gamma oscillations. Regardless of the phase or frequency, entrainment appeared stronger in left A1, indicative of the hemispheric lateralization of auditory function.

OPEN ACCESS

Edited by:

Johanna Maria Rimmele,
University Medical Center
Hamburg-Eppendorf, Germany

Reviewed by:

Oded Ghita,
Boston University, USA
Molly J. Henry,
University of Western Ontario,
Canada

*Correspondence:

P. Lakatos
plakatos@nki.rfmh.org

Received: 20 August 2015

Accepted: 17 November 2015

Published: 09 December 2015

Citation:

O'Connell MN, Barczak A, Ross D, McGinnis T, Schroeder CE and Lakatos P (2015) Multi-Scale Entrainment of Coupled Neuronal Oscillations in Primary Auditory Cortex. *Front. Hum. Neurosci.* 9:655.
doi: 10.3389/fnhum.2015.00655

INTRODUCTION

The most fundamental organizing principle of the auditory system at lower hierarchical levels of acoustic information processing is a faithful spatial representation of the auditory receptor surface in the cochlea (Schreiner and Winer, 2007). One of the likely reasons for the topographical organization of auditory information (tonotopy) across several earlier

processing stages is that just like in signal processing (e.g., EEG analysis, photo or music editing), information can be best manipulated (e.g., filtered or sharpened) at high resolutions to enhance desired features. Once the information is compressed, at higher processing stages, only cruder aspects can be manipulated. Since frequency representation is condensed to a large degree already at the level of the second cortical processing stage in belt auditory cortex (Recanzone et al., 2000), theoretically, any refinement in the frequency composition of the auditory environment should ideally take place before this stage, in primary auditory cortex (A1) and subcortical structures.

Several studies have found that along with amplifying responses to task-relevant stimulus frequencies, attention also suppresses responses in neuronal ensembles tuned to the “ignored region” of the frequency spectrum (Fritz et al., 2003, 2005, 2007; Da Costa et al., 2013; Lakatos et al., 2013). This essentially represents a spectral filter mechanism that sharpens the frequency tuning of A1 by modulating auditory information across topographically organized neuronal ensembles. Two recent studies have found that when band limited attended auditory stimuli (pure tones) are presented rhythmically, a temporal filter component is superimposed on this spectral filter, in that frequency tuning will only be sharpened and attended frequency content will only be amplified at specific times, when relevant stimuli are predicted to occur (Lakatos et al., 2013; O’Connell et al., 2014). The mechanism of this spectrotemporal filter is the entrainment of ongoing neuronal oscillatory activity that represents spontaneous excitability fluctuations of the local neuronal ensemble. Oscillatory entrainment by the attended stimuli results in a predictive excitability modulation across all of A1. A key to the mechanism of the spectral filter component is that neuronal oscillations are entrained in counterphase across A1 neuronal ensembles tuned to relevant vs. irrelevant frequency content: while the excitability of neuronal ensembles tuned to attended frequencies is up-regulated preceding the predicted occurrence of stimuli to amplify responses, the excitability of neuronal ensembles around this region across most of A1 is down-regulated, as a means to suppress irrelevant inputs that temporally coincide with attended stimuli.

Contrasting with this counterphase entrainment in A1, an earlier study investigating the processing of rhythmic visual stimuli found that in the neuronal ensembles of the primary visual cortex (V1), ongoing oscillations were always entrained to their high excitability phases by attended stimuli (Lakatos et al., 2008). A likely reason for the difference between entrainment effects across topographically organized neuronal ensembles in A1 and V1 is that while the auditory studies used pure tones which excite only a subset of the tonotopically organized neuronal ensembles in A1, the visual study used flashes that were not confined in space, and thus activated a large proportion of the retinotopically organized V1 neuronal ensembles. Also, while in the auditory tasks the topographically mapped feature, the frequency of the stimuli, was task relevant, in the visual task, the topographically mapped feature, the spatial location of the flash was task irrelevant.

Based on these earlier results, the main hypothesis tested here was that if subjects attend to broadband auditory stimuli, whose frequency content is task irrelevant, ongoing oscillations in most auditory neuronal ensembles will be entrained to their high excitability, depolarizing phases, in order to predictively amplify incoming inputs. We also hypothesized that as a consequence, the overall effect of attention to these stimuli will be a response enhancement across most A1 sites independent of tonotopy. To test this, we presented auditory click-trains, since clicks have a broad frequency spectrum. Five clicks were used to create a click train (the standard stimuli), the same five clicks formed rarely occurring deviant, or target stimuli, and the only difference between the standard and deviant stimuli was in their temporal structure, rendering the frequency content task irrelevant. Contrary to our hypothesis, effects of attention in the auditory task differentiated across the tonotopic gradient in A1. Only neuronal ensembles tuned to around 11–16 kHz had a strong tendency to entrain to their high excitability phases on multiple timescales. The more common effect, observed over the remainder of A1 sites, was response attenuation, due to the predictive suppression of neuronal excitability on at least one of the task structure related timescales (intra and inter click-train) in most A1 sites. Additionally, we found evidence that entrainment was left lateralized, indicating that similar to humans (for a review, see Giraud and Poeppel, 2012), auditory cortical function might be lateralized even at the level of primary auditory cortex in non-human primates.

MATERIALS AND METHODS

Subjects

In the present study, we analyzed the electrophysiological data recorded during 48 total penetrations of area A1 of the auditory cortex from two female macaques (*Macaca mulatta*; 22 penetrations from macaque A and 26 from macaque K) weighing 4–7 kg, who had been prepared surgically for chronic awake electrophysiological recordings. Prior to surgery, each animal was adapted to a custom fitted primate chair and to the recording chamber. All procedures were approved in advance by the Animal Care and Use Committee of the Nathan Kline Institute.

Surgery

Preparation of subjects for chronic awake intracortical recording was performed using aseptic techniques, under general anesthesia, as described previously in Schroeder et al. (1998). The tissue overlying the calvarium was resected and appropriate portions of the cranium were removed. The neocortex and overlying dura were left intact. To provide access to the brain and to promote an orderly pattern of sampling across the surface of the auditory areas, cilux recording chambers (Crist Instruments) were positioned normal to the cortical surface of the superior temporal plane for orthogonal penetration of area A1, as determined by a pre-implant MRI. Together with socketed Plexiglas bars (to permit painless head restraint), they

were secured to the skull with orthopedic screws and embedded in bone cement. A recovery time of minimum 6 weeks was allowed before the animal was head restrained and we began data collection.

Electrophysiology

Animals sat in a primate chair in a dark, isolated, electrically shielded, sound-attenuated chamber with head fixed in position, and were monitored with infrared cameras. Laminar profiles of neuroelectric activity were obtained simultaneously from left and right hemisphere auditory cortices using two linear array multi-contact electrodes (23 contacts, 100 μm intercontact spacing). Multielectrodes were inserted acutely through guide tube grid inserts, lowered through the dura into the brain, and positioned such that the electrode channels would span all layers of the cortex (Figure 3), which was determined by inspecting the laminar response profile to binaural broadband noise bursts. Neuroelectric signals were impedance matched with a pre-amplifier (10 \times gain, bandpass dc–10 kHz) situated on the electrode, and after further amplification (500 \times) they were recorded continuously in a 0.01–8000 Hz passband digitized at a sampling rate of 20 kHz and precision of 16-bits using custom made software in Labview. The signal was split into the field potential (0.1–300 Hz) and multiunit activity (MUA; 300–5000 Hz) range by zero phase shift digital filtering. MUA data was also rectified in order to improve the estimation of firing of the local neuronal ensemble (Legatt et al., 1980). One-dimensional current source density (CSD) profiles were calculated from the local field potential profiles using a three-point formula for the calculation of the second spatial derivative of voltage (Freeman and Nicholson, 1975). The advantage of CSD analysis is that CSD signals are not affected by volume conduction like the local field potentials, and they also provide a more direct index of the location, direction, and density of the net transmembrane current flow (Mitzdorf, 1985; Schroeder et al., 1998). At the beginning of each experimental session, after refining the electrode position in the neocortex, we established the best frequency (BF) of the recording site using a “suprathreshold” method (Steinschneider et al., 1995; Lakatos et al., 2005a). The method entails presentation of a stimulus train consisting of 100 random order occurrences of a broadband noise burst and pure tone stimuli presented at 50 dB loudness with frequencies ranging from 353.5 Hz–32 kHz in half octave steps (duration: 100 ms, rise/fall time: 5 ms, stimulus onset *asynchrony* (SOA) = 624.5 ms). Auditory stimuli for tonotopy and for the behavioral task were generated at 100 kHz sampling rate in Labview using a multifunction data acquisition device (National Instruments DAQ USB-6259), and presented through SA1 stereo amplifiers coupled to FF1 free field speakers (Tucker-Davis Technologies). Loudness was calibrated using measurements made with an ACO Pacific PS9200/4012 calibrated microphone system.

Behavioral Task and Stimuli

Using an auditory task with broadband stimuli, the goal of the present set of experiments was to examine the effect of

engagement on the entrainment of neuronal oscillations on multiple time-scales and on auditory responses. We presented the subjects rhythmic streams of click-trains (Figure 1A): the click-trains consisted of five clicks (40 or 50 dB SPL loudness), generated by driving the speakers with five 0.1 ms square waves that were arranged 30.3 ms apart. The click-trains were repeated every 624.5 ms (constant SOA). In this rhythmic stream of standard, frequently presented click-trains, deviant click-trains occurred at 2–6 s random time intervals. Deviant click-trains only differed in their temporal structure: the third click was delayed by 15–30.3 ms depending on the subject's performance which we tried to keep between 60–80% correct. To engage the monkeys in detecting deviant or target click-trains, in the beginning of training, 0.25–1 ml juice reward was delivered to them simultaneously with each deviant through a tube. The tube was positioned such that the monkeys had to stick out their tongue in order to get the juice. Licking was monitored using a simple contact detector circuit (Slotnick, 2009), the output of which was continuously recorded together with the timing of standard and deviant tones for offline analyses via a multifunction data acquisition device (National Instruments DAQ USB-6259) in Labview. In this phase of training the third click in the deviant click-trains was shifted by 30.3 ms corresponding to a missing third click. After two sessions, the juice reward was omitted on every 10th deviant. If the monkeys licked on these deviants without a paired juice reward, signaling that they were engaged in the auditory task, we omitted the reward on 20% of the deviants, and also gradually decreased the shift of the third click when the monkey's performance increased to around 80%. For one of the subject's the shift was decreased to 15 ms in the last experiments, while for the other monkey the shift was never below 25 ms. We only analyzed data related to standard stimuli that preceded deviants on which the subjects licked (whether or not the deviants were paired with juice). Further, we only analyzed CSD and MUA data related to standards that followed the deviant by a minimum of four stimulus positions, to avoid artifacts related to licking and to ensure that subjects re-engaged in the task (deviants could not occur for 2 s following a deviant/target).

Besides the engaged, auditory task condition, we recorded data during the presentation of the same stimuli in a passive condition, when the juicer was removed, and the subjects had no auditory or other task, but were quietly sitting in the recording chamber. Following the passive condition, we also recorded 3–5 min of spontaneous neuroelectric activity in the absence of stimuli presented.

Data Analysis

Data were analyzed offline using native and custom-written functions in Matlab (Mathworks, Natick, MA, USA). After selective averaging of the CSD and MUA responses to the tones presented in the suprathreshold tonotopy paradigm, recording sites were functionally defined as belonging to AI or belt auditory cortices based on the sharpness of frequency tuning, the inspection of the tonotopic progression across adjacent sites,

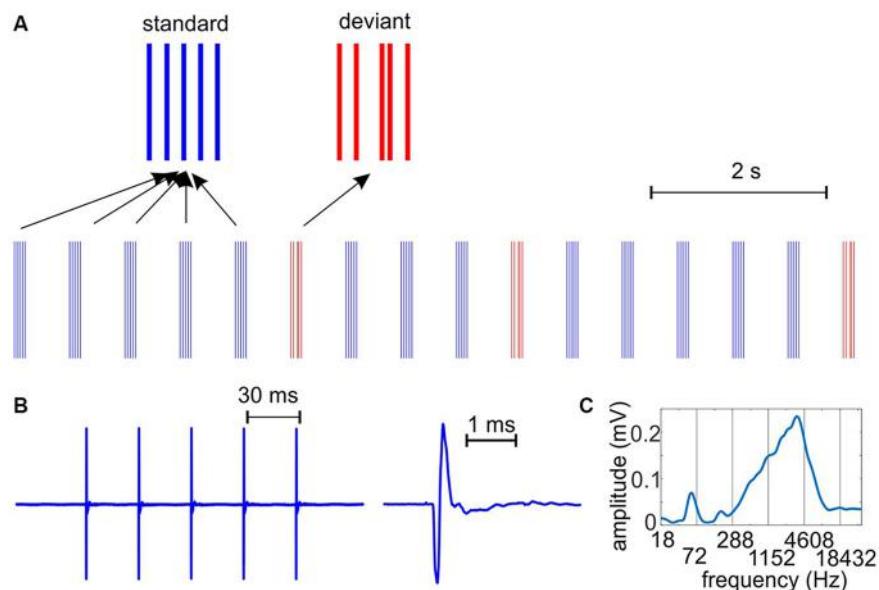


FIGURE 1 | The auditory stimulus stream. **(A)** We presented auditory clicks with a hierarchically organized rhythmic structure on two time-scales: trains of five clicks had a repetition rate of 33 Hz, while the click-trains occurred at a rate of 1.6 Hz. Deviant (target in the engaged condition) stimuli had a different temporal structure: the third click was shifted in time. Standard and deviant click trains presented were the same in passive and engaged conditions. **(B)** To the left, a standard click-train recorded with a microphone and digitized at 100 kHz. To the right, the waveform of an individual “click” as emitted by the speakers. The duration of the bipolar pulse is approximately 0.3 ms. **(C)** Spectrogram of the click train shows a prominent peak around 3.5 kHz, corresponding to the 0.3 ms “wavelength”, and smaller peaks at harmonics of the repetition rate.

and relative sensitivity to pure tones vs. broad-band noise of equivalent intensity (Merzenich and Brugge, 1973; Rauschecker et al., 1997; Lakatos et al., 2005a). In the present study only recordings obtained from area A1 were analyzed. At the end of each animal's experimental participation, functional assignment of the recording sites was confirmed histologically (Schroeder et al., 2001).

Utilizing the BF-tone related laminar CSD profile, the functional identification of the supragranular, granular and infragranular cortical layers in area A1 (**Figure 3**) is straightforward based on our earlier studies (Schroeder et al., 1998, 2001; Lakatos et al., 2005a, 2007). In the present study, we focused the analyses of ongoing and event related neuronal activity on the supragranular CSD with largest BF tone related activation (sink), and the MUA averaged across all layers. The reason for this selection is that both ongoing and entrained oscillatory activity are most prominent in the supragranular layer (Lakatos et al., 2005b, 2007, 2008), and they appear to reflect synchronous excitability fluctuations of the local neuronal ensembles across all layers, as evidenced by synchronous MUA amplitude fluctuation across the layers (O'Connell et al., 2011). Also, dominant delta frequency neuronal activity in all cortical layers is largely coherent with supragranular delta oscillatory activity, with varying but stable phase differences across cortical depths (Lakatos et al., 2005b, 2013; O'Connell et al., 2011, 2014).

To determine MUA response onset latencies, the MUA averaged across all cortical layers was used, and response onset was defined as the earliest significant [>2 standard

deviation (SD) units] deviation of the averaged waveforms from their baseline (-50 – 0 ms), that was maintained for at least 5 ms.

For the analysis of ongoing and event related delta and gamma oscillatory activity, instantaneous power and phase in single trials were extracted by wavelet decomposition (Morlet wavelet) with 345 logarithmically spaced frequency steps ranging from 0.5–55 Hz. Oscillatory amplitudes were measured in spontaneous recordings and also in data recorded during stimulus presentation. In both cases, a continuous wavelet transform was performed on the entire recording, but in the latter case, only time-points during and following the presentation of standard tones (see above) were averaged. To characterize delta and gamma phase distributions related to stimulus presentations (trials), the wavelet transformed data were normalized (unit vectors), the data at corresponding time-points relative to each stimulus onset were averaged, and the length (modulus) of the resulting vector was computed (e.g., Lakatos et al., 2007). The value of the mean resultant length, also called inter-trial coherence (ITC) ranges from 0–1; higher values indicate that the observations (oscillatory phase at a given time-point across trials) are clustered more closely around the mean (i.e., phase distribution is biased) than lower values. Phase distributions were evaluated statistically using circular statistical methods. Significant deviation from uniform (random) phase distribution was tested with Rayleigh's uniformity test. Pooled phase distributions were compared by a nonparametric test for the equality of circular means (Fisher, 1993; Rizzuto et al., 2006).

Independent of their waveform shape (frequency composition in the frequency domain), cyclically occurring events like the suprathreshold, “evoked type” response waveforms can artificially bias phase measures at the frequency that corresponds to the stimulus presentation rate (Lakatos et al., 2013; Zoefel and Heil, 2013). Since in some cases, visual inspection revealed a clear “evoked type” transient waveform in the supragranular CSD in response to the click-train (Figure 3, right traces), similar to our earlier studies in the case of responses to pure tones, we applied a linear interpolation to the single trials in the time interval of the evoked-type auditory response (5–150 ms), and determined delta phases in the interpolated data at click-train onset. For the same reason, we determined gamma phases at the time of the fourth click (90.9 ms) rather than at click-train onset, plus gamma entrainment most likely only develops after the third click.

RESULTS

We analyzed neuroelectric data recorded in 48 total A1 sites from two macaque monkeys (see Figure 2D for BF distribution).

Ongoing and event related neuronal activity was recorded with linear array multielectrodes which spanned all cortical layers at each A1 recording site. To be able to directly compare simultaneous activity of left and right hemisphere A1 neuronal ensembles, the majority of the data (42 sites in 21 experiments) was obtained via simultaneous left and right A1 recordings targeting regions tuned to similar frequencies. To minimize the effects of volume conduction and more precisely define local laminar transmembrane current flow profiles (Freeman and Nicholson, 1975; Mitzdorf, 1985; Schroeder et al., 1998), we calculated one dimensional CSD from the field potentials and carried out most of our analyses on the CSD waveforms and concomitant MUA.

Auditory stimulus-related neuronal activity was recorded in two conditions in separate trial blocks: either the monkeys were attending to frequently repeating standard stimuli in order to detect deviants that differed from standards in their temporal structure (engaged), or they were passively listening to the same stimuli (passive). The SOA was a constant 624.5 ms in both conditions, corresponding to the average wavelength of dominant delta frequency oscillations in the ongoing neuronal

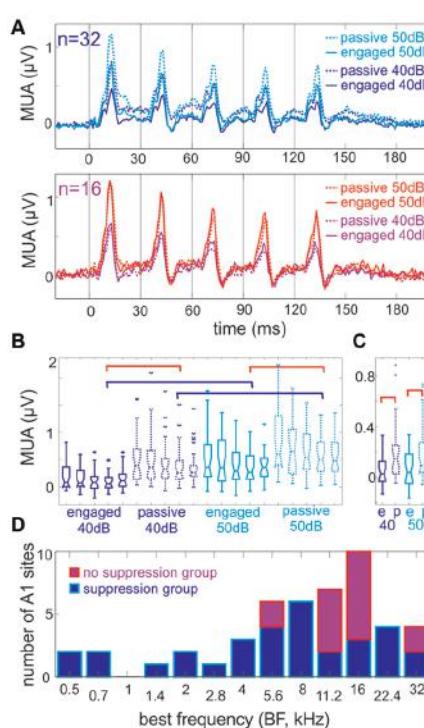


FIGURE 2 | Engagement generally suppresses responses to click-trains. (A) Multiunit responses to click-trains. Top traces show averaged responses to click-trains presented at 40 and 50 dB loudness (dark vs. light blue), in engaged and passive trial blocks (solid vs. dotted traces), recorded in Auditory cortex (A1) sites where engagement resulted in a significant suppression of the Multiunit activity (MUA) response in the 6–160 ms time range. The bottom traces are pooled responses from A1 sites where engagement resulted in either no amplitude change (12 out of 16) or a significant enhancement of MUA in the same time range. (B) Boxplots show the amplitude of MUA responses of the suppressive group to the five clicks (8–14 ms after the onset of each click) presented at 40 and 50 dB loudness in engaged and passive trial blocks. The blue brackets indicate significant differences due to change in loudness, while red brackets indicate significant differences due to engagement in responses to the fourth click (Tukey's test, $p < 0.01$). For clarity, not all significant differences are indicated by brackets. (C) Boxplots show the amplitude of “baseline” MUA (−10 to 5 ms around the onset clicks No 2–5) in different trial blocks (same as in A and B). Only engagement had a significant effect on baseline MUA indicated by the red brackets (Tukey's test, $p < 0.01$), change in loudness did not. (D) The best frequency (BF) of sites with significant engagement related suppression (blue) and no significant suppression (violet). Note that the latter group of sites is concentrated in the high frequency tuned region of A1.

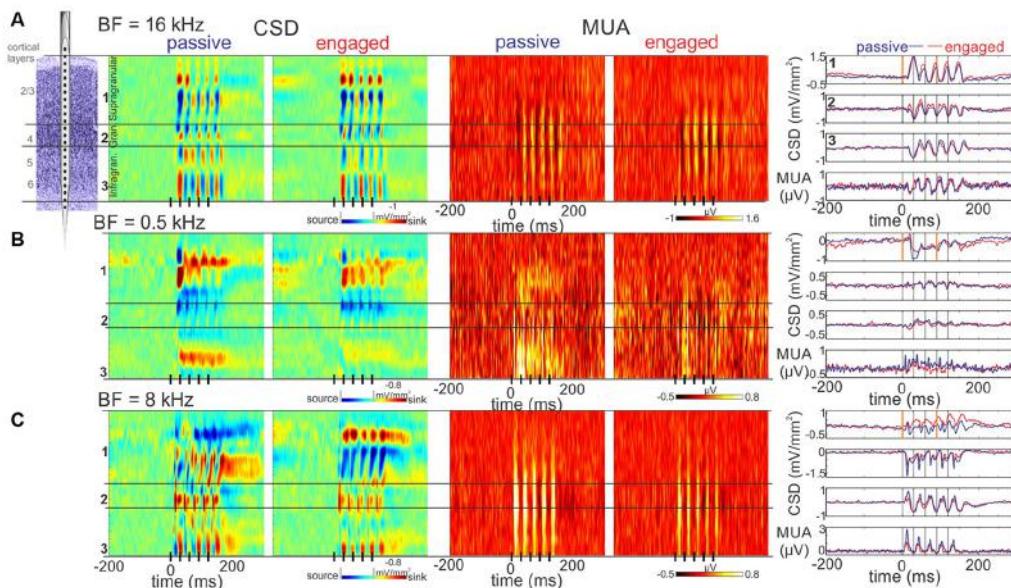


FIGURE 3 | Representative laminar response profiles. **(A)** Color-maps to the left display laminar Current source density (CSD) profiles in response to click-trains from an A1 site with a BF of 16 kHz (no suppression group). The boundaries of the supragranular, granular and infragranular layers are marked by black horizontal lines. The click-train is represented by black lines on the x-axis. Note that there appear to be sink (red)-source (blue) pairs representing active and associated passive currents in each of the layers. Color-maps to the right display the concomitant laminar MUA responses, which show a somewhat larger amplitude response in the engaged condition. The upper three traces to the right of the color-maps show CSD responses from the supragranular, granular and infragranular electrode locations marked by numbers to the left of the CSD maps (for the selection of these see “Materials and Methods” Section). The apparently largest response amplitude difference in passive vs. engaged condition is at the supragranular electrode location. Note that at stimulus onset, the baseline is negative-trending (especially in the engaged condition), and at the onset of the fourth click (orange line at 90.9 ms) the gamma frequency waveform (SSR) in between is also negative-trending. The bottom traces display the MUA averaged across all layers recorded in passive vs. engaged conditions. **(B)** Same as **(A)**, but from a low BF A1 site with significant engagement related suppression (suppression group). As opposed to **(A)**, the supragranular CSD in the engaged condition is positive trending at 0 ms, indicating an opposite phase low frequency excitability modulation. The slope of the SSR at 90.9 ms is negative trending. **(C)** Same as **(A)**, but from a relatively high BF A1 site with significant engagement related suppression. While similar to **(A)**, the baseline is negative trending, the SSR waveform is positive trending at the onset of the fourth click. This latter effect appears much stronger in the engaged condition. Note that in all three sites, MUA at stimulus onset is oppositely trending to the supragranular CSD, indicating that similar to what previous studies found, a negative CSD trend signals increasing, while a positive CSD trend signals decreasing excitability.

activity of primary auditory cortex (Lakatos et al., 2005b). Standard auditory stimuli used in the experiments consisted of five clicks arranged at regular, 30.3 ms time intervals (corresponding to 33 Hz, thus we named the click-trains gamma quintets), while deviant stimuli (targets in the engaged condition) differed in that the third click was shifted towards the fourth (shift range = 15–30.3 ms, Figure 1). The 33 Hz repetition rate of the five clicks corresponds to the gamma frequency range of the EEG. This stimulus arrangement resulted in a hierarchically organized rhythmic stimulus structure on two coupled time scales [i.e., delta (1.6 Hz) and gamma (33 Hz), that was designed to examine whether the entrainment of ongoing neuronal oscillations can occur simultaneously in multiple frequency bands, a mechanism that was proposed as one of the cornerstones of speech perception and analysis (Schroeder et al., 2008; Ghizzi, 2011; Giraud and Poeppel, 2012)].

The Effect of Engagement on Responses to Click-Trains

To assess the general effect of engagement in the task on auditory responses, we statistically compared MUA response amplitudes

to the gamma quintets in engaged vs. passive conditions within each experiment. Since we presented stimuli at two different loudness levels in both conditions (40 and 50 dB), we determined the effect of engagement separately for these. Across all recording sites, response onset latency to 50 dB attended click-trains was on average 7.41 ms (SD = 0.92 ms) and did not differ significantly between the active and passive conditions (Wilcoxon signed rank, $p = 0.239$). Since previous studies have shown that response onset varies across differently tuned regions in A1 (Mendelson et al., 1997; Kaur et al., 2004; Lakatos et al., 2005a; O’Connell et al., 2011), we tested this by comparing response onsets in A1 regions with best frequencies (BF) of 8 kHz or lower to response onsets in neuronal ensembles tuned to higher frequencies ($BF > 8$ kHz). As predicted, response onset latencies in A1 sites tuned to lower frequencies were significantly longer than those in sites tuned to higher frequencies (17 vs. 31 recording sites, Wilcoxon rank sum, $p = 0.0409$). Interestingly, when we tested whether response onset latency was significantly different across left and right hemispheres (23 vs. 25 recording sites), we found that left hemisphere latencies were significantly shorter, albeit only on average by 0.74 ms (Wilcoxon rank sum, $p = 0.0227$).

Since, as described above, there was no significant difference in response onsets across different behavioral conditions, we measured MUA response amplitude averaged across all cortical layers in the timeframe from earliest response onset (6 ms) until 40 ms post-stimulus (160 ms). When we statistically compared response amplitudes within all experiments in the engaged vs. passive conditions (Wilcoxon rank sum with Bonferroni correction), we found that for both 40 and 50 dB click-trains, engagement resulted in significant response suppression in most A1 sites [$n = 26$ (54%) in the case of 40 dB and $n = 32$ (66%) in the case of 50 dB click-trains]. The upper traces in **Figure 2A** (upper panel) show the averaged responses of sites that showed significant engagement related response suppression to the stimulus trains presented at either loudness ($n = 32$, “suppression group”). The rest of the sites either showed no engagement-related response modulation at any loudness ($n = 12$), or a significant response enhancement ($n = 3$ in the case of 40 dB and $n = 2$ in the case of 50 dB trains). Since enhancement only occurred in a fraction of our experiments (0.08%), we pooled these sites with the “no response amplitude change” ones and termed them “no suppression group” ($n = 16$), the averaged responses of which are shown in the lower panel of **Figure 2A**. By observing the averaged responses of the suppression group (**Figure 2A**, upper), we noted that the effect of loudness only appears to affect the transient parts of the response, thus we performed quantitative analyses to verify this notion. Indeed, while blue brackets in **Figure 2B** denote significant differences between 40 and 50 dB fourth click related transient response amplitudes in the 8–14 ms post-click timeframe within the same attentional condition (i.e., passive or engaged), there is no significant difference between 40 and 50 dB related pre-click (−10 to −5 ms) amplitudes (**Figure 2C**). However, the effect of engagement is observable across the whole “response timeframe”: red brackets in **Figure 2B** denote significant differences between passive and engaged fourth click related transient responses at both stimulus intensities, and red brackets in **Figure 2C** denote significantly different pre-click (or inter-click) MUA amplitudes in passive vs. engaged conditions. We also noted that in these averaged responses, the effect of engagement on the transient responses to clicks corresponds to a 10 dB decrease in loudness (there was no significant difference between responses to 40 dB clicks in the passive and responses to 50 dB clicks in the engaged condition, $p > 0.05$, Kruskal-Wallis test with Tukey's test).

To determine whether there is a relationship between the tuning of the neuronal ensembles and effect of engagement on click-train related responses, we sorted the recording sites according to their BF, which is displayed in **Figure 2D**. It is apparent that the “no suppression” group of sites mostly had BFs of 11 or 16 kHz, and never BFs lower than 5.6 kHz. As opposed to this, neuronal ensembles with engagement related suppression occurred in regions tuned to both high and low frequencies along the tonotopic axis. These two groups of sites were relatively evenly distributed across both hemispheres (Jarque-Bera test, $p = 0.071$, 10 vs. 6 non-suppressive sites in left vs. right hemispheres).

In an attempt to categorize the responses based on the effect of engagement, we examined the responses in laminar CSD profiles and at first, did not notice any apparent pattern. In fact, we were puzzled by the variability in engagement related effects. **Figure 3** illustrates this by showing the CSD response profiles of three differently tuned A1 sites: one from the non-suppressive and two from the suppressive group. The CSD profiles of the non-suppressive site ($BF = 16$ kHz) are highly similar in the passive vs. engaged conditions (**Figure 3A**), with slightly higher averaged CSD response amplitudes in the supragranular layers, as illustrated by the traces to the right that show the CSD of selected electrode contacts from different cortical layers. Additionally, the baseline appears less “flat” (more tilted in the CSD traces) in the engaged condition in the supragranular layers. The MUA response of this particular site is enhanced in the engaged condition across all cortical layers. The next site (**Figure 3B**) is tuned to low frequencies ($BF = 0.5$ kHz), and as the laminar MUA profiles show, the MUA response to click-trains is suppressed across all layers in the engaged condition. Compared to the first site, the laminar CSD response in the passive condition appears overall larger in amplitude, with maybe a slight polarity difference in the supragranular layers. Similar to the first site, the baseline appears more tilted on the selected supragranular channel in the engaged condition (CSD traces), although in the opposite direction. The third site shown in (**Figure 3C**) was tuned to 8 kHz, and since this site also belongs to the suppressive group, the MUA response appears attenuated across all layers in the engaged condition. The most apparent difference between the laminar CSD response profiles in the two conditions is that in the supragranular layers, the source over sink pattern in the passive condition appears flipped to sink over source in the engaged condition in this third site.

The Pattern of Delta and Gamma Frequency Entrainment Across A1

To quantify the observed CSD differences between the two conditions, we measured the mean phase and phase consistency (inter-trial phase coherence or ITC) of supragranular neuronal activity at the delta and gamma frequencies that corresponded to the repetition rates across and within click-trains (1.6 and 33 Hz respectively). Our reasoning was that several previous studies have shown that modulating the phase and/or strength of oscillatory entrainment can modulate responses to attended tones (Lakatos et al., 2008, 2013; O'Connell et al., 2014). Thus, assuming that the phases measured reflect the phase of entrained oscillatory activity as opposed to evoked type, *de novo* generated neuronal activity, the pattern of phase alignments could reveal a potential mechanism of response suppression in the engaged condition. To verify this assumption, we compared the amplitudes of delta and gamma band neuronal activity in data that were recorded in the absence of stimulation (spontaneous activity) to delta and gamma amplitudes measured in the passive and engaged conditions.

Figure 4A shows the spectrograms of supragranular neuronal activity (CSD) in the absence of stimulation and in different

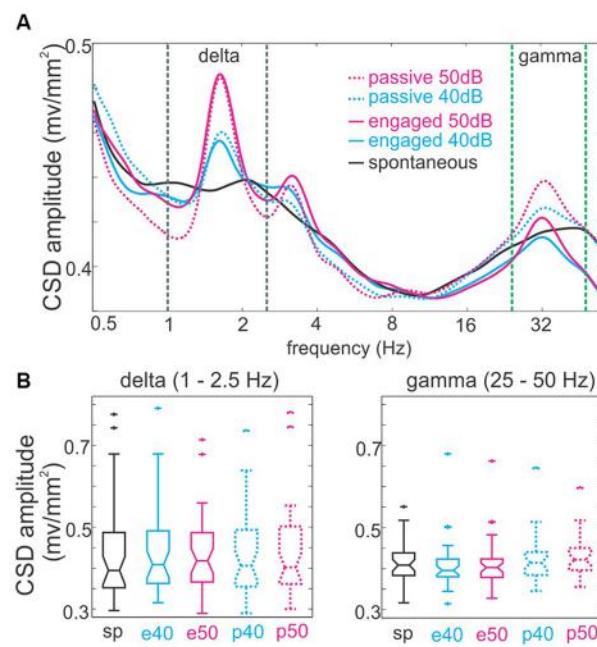


FIGURE 4 | The amplitude of spontaneous and auditory stimulus stream related delta and gamma band neuronal activity. (A) The spectra averaged across all experiments ($n = 48$) show that the presentation of click-train streams results in discrete peaks that correspond to the repetition of clicks within click-trains (33 Hz) and the rate at which click-trains occurred in the stream (1.6 Hz). These peaks are larger in the case of louder streams. The dotted lines mark the delta (black, 1–2.5 Hz) and gamma (green, 25–50 Hz) frequency bands. **(B)** Boxplots show pooled amplitudes of neuronal activity averaged in the delta (left) and gamma (right) frequency bands.

task conditions during the presentation of click-trains. While it is obvious that at both delta and gamma stimulation rates, the amplitude spectrum of neuronal activity is “peaked” compared to the spontaneous spectrum, note that this is paired with lower amplitudes around the peak in the auditory stimulus stream related spectra resulting in no significant net amplitude change in the delta and gamma bands (Kruskal-Wallis test, $p = 0.9519$ and $p = 0.1549$ respectively, **Figure 4B**). Rather, the peaks most likely represent a reorganization of oscillatory activity to match relevant temporal scales that results in a concentration of energy at the frequencies that correspond to the repetition rates of stimuli. In other words, the peaks mostly signal neuronal activity that is less variable in frequency in the delta and gamma bands, which has been shown to be characteristic of oscillatory entrainment (Lakatos et al., 2013; Zoefel and Heil, 2013). Nevertheless, we cannot exclude the possibility that evoked type activity contributes to the measured spectra. As a matter of fact it is likely, especially in the case of 50 dB click-trains: the harmonic at double the stimulation rate (~ 3.2 Hz) can be a strong indication of evoked type activity that “distorts” the sinusoidal waveform that is characteristic of entrainment. Previous studies indicate that the amplitude ratio of evoked type (added) neuronal activity to the ongoing neuronal activity determines the “distorting” effect of evoked responses on phase measurements of the ongoing neuronal oscillations (Lakatos et al., 2013). Since based on the spectra, this ratio is very small in the case of 40 dB click-trains (on average 1.00 in the delta and 0.99 in the gamma range for 40 dB, and 1.03 in both the delta

and gamma frequency ranges for 50 dB, with the gamma ratio difference significant (Wilcoxon signed rank test, $p < 0.001$), we only analyzed delta and gamma phases related to the lower intensity stimuli in engaged vs. passive conditions. To further minimize confounding effects of the response evoked by the onset of the click-train, as in previous studies (Lakatos et al., 2013; O'Connell et al., 2014), we applied linear interpolation to the data in the 5–150 ms timeframe before we measured delta phases (see “Materials and Methods” Section). Furthermore, while we measured delta phase at stimulus onset (0 ms), gamma phases were measured at the time of the fourth click (90.9 ms in non-interpolated data) to get a more reliable estimate of the entrained gamma phase. **Figure 5A** displays the histograms of mean delta (top) and gamma (bottom) phases across all experiments in the engaged (left) and passive (right) task conditions. It is apparent that the phase distributions are bimodal in most cases (except gamma phases in the passive task condition). One group of phases peaks between $-\pi$ and 0, on the upward deflection of the neuronal oscillation, while the other group is centered on the downward deflection. Our previous studies analyzing supragranular CSD oscillations in the same laminar position have provided evidence that while the upward deflection corresponds to the low excitability, or hyperpolarizing phase of cortical neuronal oscillations, the downward deflection corresponds to the high excitability, depolarizing phase (e.g., Lakatos et al., 2005b). This was determined indirectly by analyzing fluctuations in the level of spontaneous (incidental) neuronal ensemble firing and gamma oscillatory amplitudes, both of which are highest

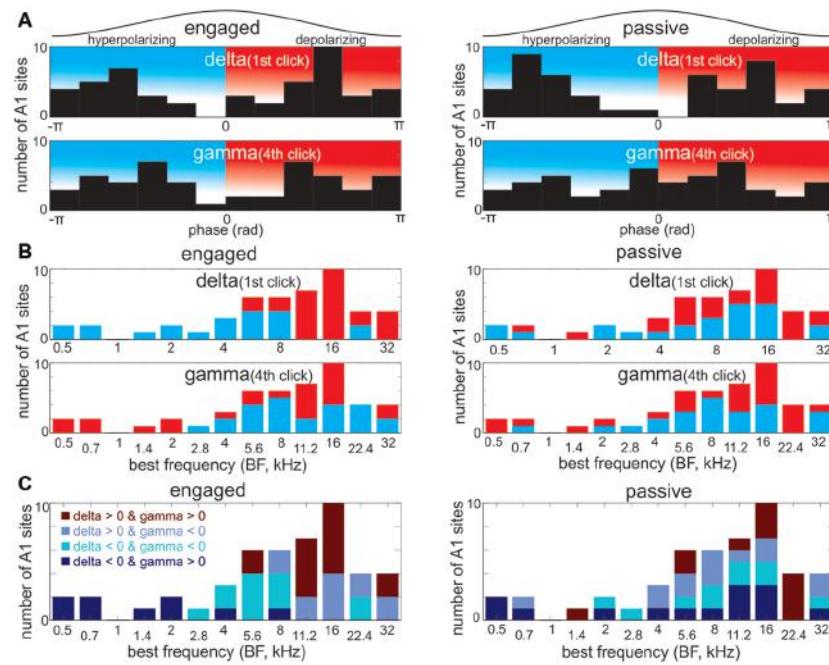


FIGURE 5 | The distribution of pre-stimulus delta and gamma oscillatory phases across differently tuned A1 regions. (A) Histograms show the distribution of delta phases measured at the onset of the first click (0 ms) and gamma phases measured at the onset of the fourth click (90.9 ms) in the engaged and passive conditions. Phases were measured at the delta and gamma frequencies that correspond to the stimulus onset asynchrony (SOA) between and repetition rate within click-trains (1.6 and 33 Hz respectively). Black traces above the histograms represent one oscillatory cycle. Therefore, the histograms illustrate where in the delta/gamma cycles the first/fourth clicks tend to occur. It is apparent that both delta and gamma phases in the engaged, and delta phases in the passive condition have a bimodal distribution, indicating that click related inputs tend to fall on either of two preferential phases either the upward deflection (hyperpolarizing phase) or downward deflection (depolarizing phase). (B) The BF distribution of sites with hyperpolarizing and depolarizing pre-stimulus phases. Note that sites with depolarizing pre-stimulus delta phases in the engaged condition have high BFs. Also note that the gamma phases in high frequency tuned sites are mixed, while low frequency sites have depolarizing gamma phases at the onset of the fourth click. Importantly, the BF distribution of sites with depolarizing and hyperpolarizing delta and gamma phases is markedly different only in the engaged condition. (C) Theoretically, delta and gamma phases can combine in four different ways, and as the histograms show, they indeed do. Note that the distribution of sites with depolarizing delta and gamma phases ($\delta > 0$ and $\gamma > 0$) in the engaged condition appears to be very similar to the BF distribution of sites with no significant engagement related suppression (Figure 2D).

on the depolarizing phase of ongoing oscillations. To verify this in our current data, we first grouped MUA and gamma oscillatory amplitude (measured in the 25–50 Hz band) averaged across all layers into two bins based on the phase of delta oscillatory activity: MUA and gamma amplitude during delta phases from $-\pi$ to 0 (upward deflection) fell into one bin, while the rest (during delta phases from 0 to $-\pi$, the downward deflection) were put into the second bin. We found that even though the difference between bins was on average very small (0.8% difference for MUA and 2.9% for gamma), both MUA and gamma frequency laminar activity was significantly larger during the downward slope of the supragranular delta oscillation (Wilcoxon signed rank, both $p < 0.0001$), confirming that this is indeed the high excitability or depolarizing phase of delta (similar to Figure 4A of Lakatos et al., 2005b). Similarly, although not as highly significant, we found that MUA was significantly higher in amplitude during the depolarizing phase of gamma band oscillatory activity (Wilcoxon signed rank, $p = 0.015$). This is physiologically plausible, because while negative trending values in the CSD represent net inward transmembrane current which signals a depolarization of the local neuronal ensemble (hence the name depolarizing phase), positive trending values signal

net outward current and thus hyperpolarization (hyperpolarizing phase). Therefore, we decided to use these two phase bins to tag delta and gamma phases (Figures 5A,B).

To examine the relationship between the frequency tuning of A1 neuronal ensembles and the phase of delta and gamma entrainment, we created bar graphs with the phase of oscillations color coded (Figure 5B, red = depolarizing, blue = hyperpolarizing phase). We found that in the engaged condition, attended click-trains in sites with higher BFs entrained delta oscillations to their depolarizing phase, while in sites tuned to lower frequencies delta oscillations were entrained to their hyperpolarizing, low excitability phase. The BF distribution of the sites entraining to the click trains with a depolarizing delta phase was significantly different from that of the sites entraining with a hyperpolarizing delta phase (Wilcoxon rank sum test, $p < 0.001$, depolarizing delta phase median $BF = 16$ kHz, hyperpolarizing delta phase median $BF = 4$ kHz). The mean phase of gamma oscillatory activity during the engaged condition showed a more complicated pattern: in sites tuned to ≤ 2 kHz and to 11–16 kHz we measured depolarizing phases, while hyperpolarizing gamma phases appeared biased towards sites tuned to frequencies surrounding 11–16 kHz, however there was

no significant difference in the BF distribution of sites exhibiting depolarizing or hyperpolarizing pre-stimulus gamma phases (Wilcoxon rank sum test, $p = 0.2$). In the passive condition, for both delta and gamma pre-stimulus phases, there was no obvious BF distribution pattern (Wilcoxon rank sum test, both p values > 0.2).

Next, we examined how the phases of delta and gamma entrainment “combine” in each site. One possibility is that, for example, the depolarizing phase of delta always co-occurs with the depolarizing phase of gamma in one group of sites, and the hyperpolarizing phases of the entrained oscillations combine in another group of sites. However, two other combinations are theoretically possible: hyperpolarizing gamma phases combined with depolarizing delta, and vice versa. In fact, when we looked at the combination of delta and gamma phases, we found that all four possible combinations occur. Furthermore, these seem to be grouped in sites tuned to similar frequencies in the engaged condition (**Figure 5C**): e.g., while hyperpolarizing delta and depolarizing gamma phases co-occur in regions tuned to 2 kHz and below, depolarizing delta and gamma phases co-occur overwhelmingly in regions tuned to 11–16 kHz (11 out of 17 sites, or 64.7%). We noted that the BF distribution of sites in this latter group shows remarkable similarity to the BF distribution of sites in the no-suppression group (**Figure 2D**). When we compared the “depolarizing delta-gamma” ($n = 15$) and no-suppression group of sites ($n = 16$), we found that 12 of the sites were indeed the same.

This indicates that the majority of sites in the depolarizing group show either no response amplitude change or a response enhancement in the engaged vs. passive condition. It also follows that sites in the other three phase combination groups belong to the group of sites with significant engagement related MUA response suppression. To verify this and to uncover any multiscale entrainment specific differences, we pooled MUA responses based on delta-gamma phase combination into four groups, and compared response amplitudes in the later portion (60–160 ms) of the response in the engaged vs. passive conditions (**Figure 6**). We found that as predicted, with the exception of the depolarizing delta-gamma group, engagement resulted in significant response suppression (Wilcoxon signed rank test: depolarizing delta-gamma group: $p = 0.118$, all other groups: $p < 0.05$).

While in some cases, delta and gamma phase combinations differ across engaged and passive conditions (**Figure 5C**), mean delta and gamma phase in the majority of sites does not change (e.g., sites tuned to 0.5 kHz). Thus we asked whether phase consistency across single trials was different in engaged vs. passive conditions, since it has been shown that engagement results in a stronger phase reset and entrainment of ongoing neuronal activity (Lakatos et al., 2009, 2013; O'Connell et al., 2014), and a stronger enforcement of suppressive phase patterns via entrainment could in theory result in significant suppression of responses. We found that as expected, both delta (1.6 Hz) ITC measured at stimulus onset, and gamma (33 Hz) ITC measured at the time of the fourth click in the click-train was significantly greater in the engaged condition (Wilcoxon signed rank, both $p < 0.001$). While delta and gamma phase consistency was

significant in all sites in the engaged condition, in the passive condition, delta and gamma ITC was not always significant (Rayleigh test, Bonferroni corrected $p < 0.0005$ in 21 sites for delta and 25 sites for gamma). This most likely indicates that the passive condition is a mixed ignore/attend condition, since we did not employ a selective attention task where the animals had to ignore the click trains in order to attend to an alternate stimulus stream. The fact that in most cases in the passive condition, delta and gamma phases were not significantly biased also explains why mean phase distributions across differently tuned sites appear less congregated around sites with similar BFs (**Figures 5B,C**), since the mean of non-significant phase distributions can be considered random. Thus it is not phase distribution *per se* that is different between passive and engaged conditions but the strength of entrainment.

Engagement Related Hemispheric Asymmetry

Recent human research suggests that the processing of auditory stimuli structured at different timescales is hemispherically asymmetric (for a review, see Giraud and Poeppel, 2012). We designed our stimuli in part to mimic the multi-temporal scale organization aspect of vocalizations, and thus, we were interested in the question of whether there is evidence of hemispheric asymmetry in the entrainment of fast and slow oscillations. To determine this, we pooled our delta and gamma ITC measures according to task condition and hemisphere. As **Figure 7A** shows, we found that delta ITC related to click-train streams in the engaged condition was significantly greater in left hemisphere sites than delta ITC in either left or right hemisphere sites in the passive condition (Tukey's test, $p < 0.01$). Importantly, left delta ITC in the engaged condition was also greater than right delta ITC in the same condition. These data indicate that there is a hemispheric asymmetry in the strength of delta entrainment due to engagement. A similar trend is apparent for gamma ITC, however left hemisphere gamma ITC in the engaged condition was only significantly larger than right hemisphere gamma ITC in the passive condition (Tukey's test, $p < 0.01$). When we compared oscillatory amplitudes across hemispheres and task conditions, at the delta and gamma frequencies that corresponded to the SOA (1.6 Hz) and the repetition rate of click in the gamma quintets (33 Hz), there was a trend towards a similar leftward bias but no significant effects (Kruskal-Wallis test, $p > 0.05$; **Figure 7B**). Taken together these data indicate that left A1 exhibits greater stimulus structure related delta and gamma band oscillatory activity, and that engagement in the task enhances hemispheric differences in the oscillatory neuronal activity of the supragranular layers.

Importantly, as our results above foreshadowed (**Figure 4**), larger delta and gamma amplitudes related to click-trains in the left A1 are not due to larger evoked responses. As the spontaneous and stimulus-related spectra in **Figure 7C** show, the difference between spontaneous and auditory stimulus-related delta amplitudes at the rate of stimulation is actually larger in the right hemisphere indicating that perhaps in the right hemisphere evoked activity contributes more substantially to the delta peak in

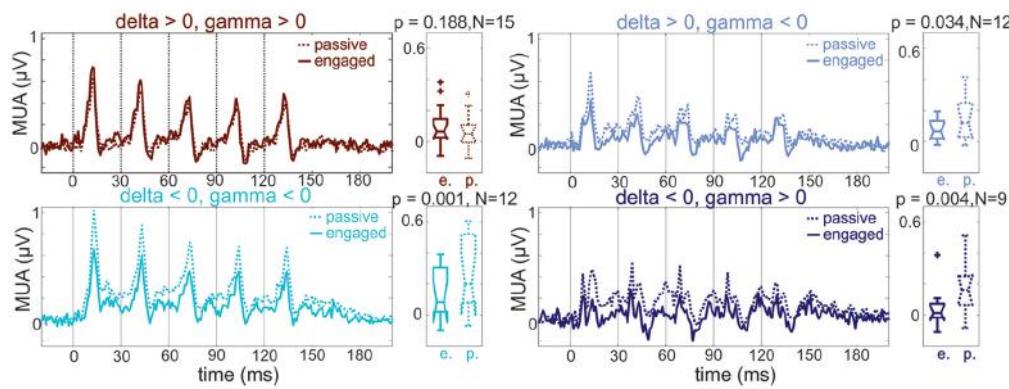


FIGURE 6 | MUA responses in A1 regions with differing pre-stimulus phases. Pooled MUA responses of sites with different pre-stimulus delta-gamma phase combinations (same groups of sites and same color coding as sites in the engaged condition of **Figure 5C**). Boxplots to the right show the MUA response amplitudes across sites averaged in the 60–160 ms time interval. While the first group of sites shows no engagement related suppression, the other three groups of sites do. Note that the patterns of suppression seem to match the delta-gamma phase combinations remarkably well: while the suppression is maximal at the times of the individual click related responses in the depolarizing delta—hyperpolarizing gamma group (upper right panel, light blue), the suppression is more sustained in the groups where pre-stimulus delta is in a hyperpolarizing phase (lower panels, cyan and dark blue). Wilcoxon signed rank test, $p < 0.01$, N = number of A1 sites in each group.

the spectrum of stimulus-related activity. Contrary to this, in the left hemisphere there is no net amplitude change between the two conditions, indicating that most likely entrainment is responsible for the delta peak at the stimulation rate.

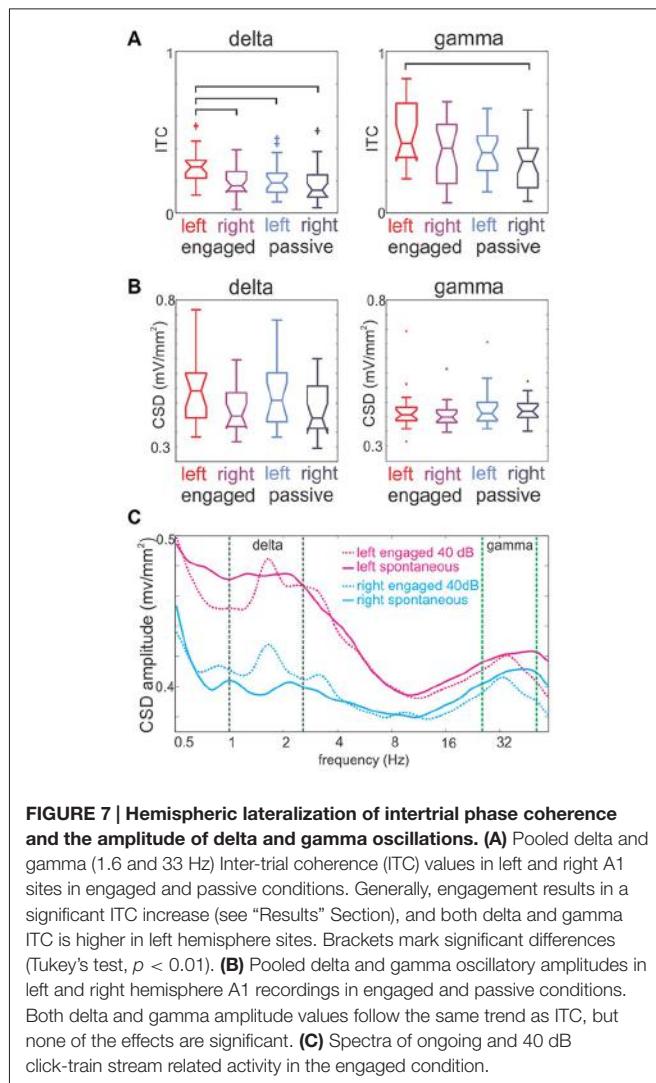
DISCUSSION

Our main hypothesis was that when the broadband frequency spectrum of attended stimuli is irrelevant for an auditory task, ongoing oscillatory activity across all of A1 would be entrained by the auditory stimuli so that its high excitability, depolarizing phase would be aligned to the stimuli's onset, maximally amplifying auditory responses. However, to our surprise we found that even though all of A1 entrained its ongoing neuronal oscillations to the temporal structure of attended stimuli on two timescales, delta and gamma, the net effect of entrainment on auditory responses was mostly suppressive. Both the pattern of entrainment and engagement related response amplitude modulation differentiated sites across the tonotopic map in A1: ongoing oscillations of most neuronal ensembles within the 11–16 kHz region of A1 were entrained by the stimuli to their high excitability phases when monkeys engaged in the auditory task, and responses to task relevant stimuli in these sites were either enhanced or not significantly modulated compared to responses in the passive condition. In contrast, in most neuronal ensembles outside of this A1 region, either delta (for sites further from the 11–16 kHz region) or gamma (for sites closer to the 11–16 kHz region) oscillations were entrained to their low excitability phases in the engaged condition (see **Figure 5B**), which co-occurred with significant response suppression compared to the passive condition. Congruent with a more organized pattern of entrained delta and gamma phases, stimulus timing related delta and gamma phase consistency (ITC) were both significantly larger in the engaged compared to the passive condition. Taken together, our findings indicate that

neuronal ensembles tuned to the higher frequency portion of the audible spectrum play a central role in the sensory representation and processing of relevant broadband transient sounds, like auditory clicks. Additionally we found that engagement-related oscillatory entrainment on both slow and fast time scales was stronger in left hemisphere A1 sites, albeit only delta frequency effects were significant.

Mechanisms of Predictive Response Suppression

Our result of oscillatory entrainment across multiple frequency bands is in line with a previous study (Henry et al., 2014), but how does the multiscale entrainment of delta and gamma band oscillations result in a net suppressive effect in most neuronal ensembles? Previous studies provide a wealth of evidence, which we verified in the present study, that both delta and gamma oscillations have depolarizing (or high-excitability) and hyperpolarizing (or low excitability) phases (for a review, see Young and Eggermont, 2009). Our results show that in about half of the A1 sites examined, delta and gamma band oscillations were entrained by the clicks to the former, while the other half to the latter phase (**Figure 5A**). On a first hunch, this should result in an equal distribution of response enhancement and suppression across sites, which is not what we found (out of the 48 sites we recorded from, 16 sites showed no suppression while 32 sites exhibited suppression). The reason for this is twofold: first, hyperpolarizing and depolarizing phases of entrained delta and gamma oscillations are not always paired; we found that they co-occur in all four possible combinations. Second, delta and gamma oscillations are phase amplitude coupled, meaning that the phase (i.e., high/low excitability) of a lower frequency oscillation determines the amplitude (large/small) of a higher frequency band oscillation (Buzsáki et al., 2003; Lakatos et al., 2005b, 2008; Canolty et al., 2006), as shown by the significantly smaller gamma amplitude on the hyperpolarizing phases of delta oscillations in



our data. Now let us consider the four phase combinations taking into account phase amplitude coupling. Diagrams in **Figure 8** show the predicted excitability of neuronal ensembles entraining with the different delta-gamma phase combinations at click onset. If the depolarizing phases of delta and gamma co-occur, since delta is in a high excitability phase, the amplitude of gamma oscillations will be large, and as they are also being entrained to their high excitability phase by the clicks, this should result in a high excitability state of the neuronal ensemble when stimuli are predicted to occur, and thus enhanced response amplitudes (e.g., **Figure 8A**). However, if gamma oscillations entrained to their hyperpolarizing phases ride on the depolarizing phase of delta, gamma amplitude will still be large, but the hyperpolarizing phase of gamma will negate the depolarizing effect of delta, resulting in a net hyperpolarized state of the local neuronal ensemble in short, precisely timed temporal windows of low excitability centered on the clicks (e.g., **Figure 8B**), which should result in transient response suppression (**Figures 8B, 6**, upper right panel). In the remaining two categories of sites, since delta is entrained to its hyperpolarizing phase by the click

trains, and thus gamma amplitudes will be low, the phase of gamma does not play an effective role in modulating excitability (**Figures 8C,D**). Therefore, the net effect should be long time-scale predictive suppression related to the hyperpolarizing phase of delta. In support of this, a visual inspection of the MUA responses in **Figure 6** indicates that while mainly the transient part of the click-train related responses is suppressed in sites with depolarizing delta and hyperpolarizing gamma entrainment (**Figure 6**, upper right panel), suppressive effects appear much more tonic (with a longer time-constant) in hyperpolarizing delta sites (**Figure 6**, lower two panels).

The Effect of Engagement on Neuronal Activity in A1

Contrasting task engaged and passive conditions, like in the present study, is often used in animal studies to investigate behavioral state related changes in neuronal activity. Regardless of sensory modality, a common finding in rodent studies when comparing responses to stimuli in engaged vs. passive states is that responses are suppressed in the active behavioral condition (Fanselow and Nicollelis, 1999; Castro-Alamancos, 2004; Crochet and Petersen, 2006; Ferezou et al., 2006; Otazu et al., 2009). This is usually interpreted as a sharpening or refinement of the sensory input. Our main finding is in line with these previous studies in that the overall effect of engagement in the task is response suppression. While not tested quantitatively, our data suggest that, at least when stimuli are broadband, engagement related MUA suppression is biased towards lower frequency tuned A1 neuronal ensembles (**Figure 2D**). This differs from the results of a previous study in rat auditory cortex, which found no tonotopic organization of engagement related suppression (Otazu et al., 2009). Our data also reveal a candidate mechanism for the engagement related sharpening of the sensory representation: the modulation of subthreshold neuronal ensemble activity via the alignment of rhythmic excitability fluctuations to the temporal structure of relevant auditory stimuli. This alignment, the entrainment of neuronal oscillations occurs in the passive condition as well (similar to Lakatos et al., 2005b), but to a significantly lesser degree, and in a less organized pattern.

The Importance of Broadband Transient Sounds in Auditory Processing

In one of our earlier studies, we proposed that the brain “models” the spectrotemporal properties of selectively attended auditory stimuli and stimulus streams in the form of temporally evolving phase patterns arranged in space across topographically organized A1 neuronal ensembles (Lakatos et al., 2013). This in turn forms the basis for enhancing and stabilizing the representation of attended auditory information at the expense of irrelevant, background auditory stimuli. The present study, however, found that the physical frequency spectrum of the auditory click is represented in a “distorted” form, since its representation is mostly enhanced in high, while suppressed in low frequency regions of A1. Thus we speculate that it is possible that sharp transients, like clicks or formant transitions, represent a special category of auditory stimuli for which preserving an

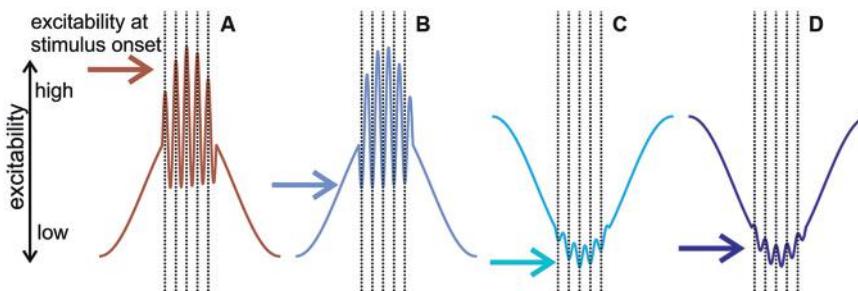


FIGURE 8 | The predicted effect of delta–gamma phase amplitude coupling and multiscale entrainment to various phases on neuronal excitability.

(A) Diagram of neuronal excitability when delta and gamma are both entrained to their depolarizing phases: depolarizing delta results in large amplitude gamma, and as gamma is also at its depolarizing phase at the time of the stimuli, this results in a high excitability state when stimuli are expected to occur. (B) Depolarizing delta/hyperpolarizing gamma: the amplitude of gamma is still large but as it is entrained to its hyperpolarizing phase, this results in a relatively low neuronal excitability state at the onset of stimuli, which should lead to transient MUA response suppression. In (C,D) delta is entrained to its hyperpolarizing phase and therefore gamma amplitudes are small, thus gamma phase (hyperpolarizing in (C) and depolarizing in (D)) does not play a major role in neuronal excitability modulation). Therefore, we predict sustained MUA suppression due to the hyperpolarizing phase of delta in both scenarios.

accurate frequency representation is less important. Rather, the main role of these “acoustic edges” could be to orchestrate the coherent multiscale entrainment of neuronal oscillations across differently tuned A1 neuronal ensembles, thereby setting up a spatiotemporal excitability pattern that is ideal for the parsing and processing of relevant auditory content mainly contained in the lower frequency spectrum e.g., speech (Fletcher, 1948; Peele et al., 2013). In this theoretical framework, acoustic edges would form the temporal context that enables the most efficient processing of the acoustic content by modulating ongoing neuronal oscillations.

Broadband transient sounds are common features of speech in humans (e.g., stop consonants) and conspecific vocalizations in monkeys (May et al., 1989; Wang et al., 1995). Aside from communication sounds, they also occur frequently in the acoustic environment, in which case they mostly indicate something alerting requiring quick action (e.g., the snap of a twig). Thus, in theory, it would be advantages to process these sounds via a fast dedicated auditory processing hierarchy of neuronal ensembles. Indeed, there are neurons in the posteroventral cochlear nucleus that are specialized in responding to broadband transients, called octopus cells. The main function of the octopus cells appears to be the integration of the cochlear activation via the summation of orderly dendritic synaptic activation, which compensates for the traveling wave delay of the cochlea (Rhode et al., 1983; Golding et al., 1999; Oertel et al., 2000; McGinley et al., 2012). These cells fire extremely fast and are very precise temporally (Rhode and Smith, 1986). Their output is transmitted via a separate ascending pathway mainly to the contralateral ventral nucleus of the lateral lemniscus, a pathway which appears to be much more prominent in humans (Adams, 1997). Interestingly, it has also been shown that octopus cells integrate cochlear inputs over about 1/3 of the audible spectrum (Oertel et al., 1990; Golding et al., 1995, 1999), which does correspond to the BF “spread” of the non-suppressive group in our data. Thus, we hypothesize that the group of non-suppressive sites in A1 that are tuned to 11–16 kHz might form the first cortical stage of the ascending auditory

pathway specialized in rapidly processing broadband transient sounds.

Besides rapid alerting, this “transient specialized system” together with the modulation of ongoing neuronal activity across A1 could play a crucial role in the processing of complex acoustic patterns like communication sounds. As suggested earlier, one role could be parsing (Ghitza and Greenberg, 2009; Buzsáki, 2010; Ghitza, 2011; Giraud and Poeppel, 2012): entrained delta/theta phase related suppression of the neuronal ensembles processing lower frequency speech sounds (<5 kHz) would be an efficient mechanism to segment continuous speech, while gamma, which is “nested” in delta/theta assists in the modulation of excitability and thus stimulus processing on shorter timescales (e.g., phonemic scale).

Expanding on the “gamma nested in delta/theta” model (Ghitza, 2011; Giraud and Poeppel, 2012), we propose that complementary to parsing, the other main role of transients in speech might be to prepare lower frequency cortical areas for the processing of band-limited sounds. This would be important since speech has regularly and predictably interchanging broadband transients (i.e., consonants) and more “tonal” elements whose main spectral energy is band limited and is usually below 5 kHz (i.e., vowels). Our results provide evidence that this could be achieved by resetting and entraining ongoing oscillatory activity to their low excitability phases in most A1 areas (i.e., outside the 11–16 kHz region): as a consequence, the high excitability, depolarizing phase of ongoing oscillations will be centered on acoustic elements positioned between sharp transients (acoustic edges). In fact, this notion is in line with the findings of a recent behavioral study where the ability of subjects to detect a 1 kHz tone modulated in a fashion that was antiphasic to the amplitude-modulated broadband noise stream that preceded it (Hickok et al., 2015). This “antiphasic oscillation of perceptibility” is most likely due to the fact that the broadband noise (like click) entrained oscillations in high frequency regions of auditory cortex to high-, but low frequency regions to low-excitability phases. Thus the detection of a low frequency tone would be enhanced at the trough, not the peak

of the amplitude modulated noise. Based on our previous studies (Lakatos et al., 2013; O'Connell et al., 2014), lower frequency speech elements (vowels) could also reset delta and gamma oscillations to their depolarizing phases in low frequency regions and to their opposite, hyperpolarizing phases in high frequency regions. This in turn would prepare these areas for an upcoming high frequency element or sharp transient. Therefore, we hypothesize that within a syllable, which occurs at a delta/theta rate (Greenberg et al., 2003), and is usually constructed of a consonant (high frequency element) and a vowel (low frequency element; Poeppel, 2003), counterphase entrained oscillations in the delta/theta band across all of A1 are reset twice, once by high frequency (consonant related) and once by low frequency (vowel related) inputs. This would provide a highly adaptive very precise dual timing mechanism for the synchronization of neuronal oscillations to attended speech that is thought to be a key element of speech processing and perception (see below), which should be especially helpful in noisy environments, like at a cocktail party. An everyday observation in support of the above hypothesis is that it is close to impossible to make out someone's speech over the phone (which transmits acoustic signals only below 5 kHz) when background noise is high or when multiple people are speaking. We speculate that the presence of noise results in deterioration in performance not only due to missing acoustics masked by the noise (see Appendix in Ghitza, 2011; also Shamir et al., 2009), but also as a result of the missing half of the temporal context contained in the high frequencies of the auditory spectrum, which prevents the precise alignment of nested oscillations to speech. A similar mechanism could explain aging related deficits in speech comprehension, which manifests stronger in the presence of environmental noise, since ageing often results in high frequency hearing loss (reviewed by Pichora-Fuller and Souza, 2003).

Evidence for Hemispheric Functional Lateralization in Non-Human Primates

Both delta and gamma oscillations, along with theta have been proposed to be important in the processing of speech and species specific communication (Schroeder et al., 2008; Ghitza, 2011; Giraud and Poeppel, 2012). We found that multiscale oscillatory entrainment at these rates shows greater phase consistency at the time attended auditory stimuli occur in left A1, indicating a stronger involvement of left hemisphere oscillatory activity. This finding provides support for the functional asymmetry of left and right auditory systems at the level of their first cortical processing stage, primary auditory cortex, and possibly indicates that the precursor of left hemisphere association with speech is present in non-human primates. Previous studies in monkeys provide behavioral (Petersen et al., 1978; Ghazanfar et al., 2001), ablation (Heffner and Heffner, 1984) and neuroimaging (Poremba et al., 2004) evidence for hemispheric lateralization for the processing of species specific communication. Anatomical studies in new and old world monkeys also found evidence for a leftward asymmetry (Heilbronner and Holloway, 1988; Gannon et al., 1998, 2008). Nevertheless, since the results relating to functional asymmetry are scarce, the hypothesized functional

lateralization of auditory processing is still unresolved in non-human primates. To our knowledge, our study is the first to provide electrophysiological evidence for such hemispheric lateralization in monkeys. Importantly, our left and right measures are directly comparable since most data was recorded simultaneously in left and right primary auditory cortices.

We found that the hemispheric asymmetry in the strength of delta and gamma entrainment only became significant in the engaged condition. Additionally, we found no indication of an asymmetry in the strength of delta entrainment in a previous study (O'Connell et al., 2014), where monkeys were presented with a rhythmic stream of pure tones and were performing a frequency deviant detection task. Thus it appears that in macaques, functional asymmetry becomes apparent when spectrotemporally more complex stimuli are used and the subjects are engaged in a task where these stimuli are relevant. In fact, human studies that show indications for functional asymmetry using electrophysiological recordings and/or neuroimaging utilized similar spectrotemporally complex, rhythmic stimuli (Boemio et al., 2005; Jamison et al., 2006; Giraud et al., 2007; Obleser et al., 2008; Morillon et al., 2010), even though in some of these studies stimuli were presented in a passive condition. We speculate that in humans, hemispheric asymmetry might be structurally more solidified via evolution, which is why functional differences can be revealed even in a passive state. Another difference between human findings and our results in monkeys is that while in our data, entrainment on both long and short time-scales was left lateralized, most human studies find that slower (delta-theta) modulations of neuronal activity related to the temporal structure of the acoustic input are lateralized to the right hemisphere (e.g., Luo and Poeppel, 2012). We speculate that one reason for the difference might be stronger evoked type responses at the rate of stimulation in the right hemisphere, which would bias both neuroimaging and electrophysiological measurements. Although a thorough analysis of this proposition is beyond the scope of the present study, the spectrograms in **Figure 7C** do provide some support for this notion: while spontaneous low frequency oscillatory amplitude is smaller in right hemisphere recordings, the amplitude increase related to auditory stimuli in the delta band is larger. In the future, it will be important to conduct human experiments with near threshold auditory stimuli so that the effect of evoked type responses would be negligible. Nevertheless, despite some discrepancies, our results provide support that, similar to humans, relevant spectrotemporally complex rhythmic stimuli are processed asymmetrically by the left and right hemispheres. This result suggests that the functional-anatomical precursor to the machinery that enables speech perception and production might be present in non-human primates, at least at lower cortical stages.

CONCLUSION

Our findings indicate that attended broadband stimuli organized on multiple timescales (i.e., repetitive click-trains) result in a multi-scale entrainment of ongoing oscillations across all of A1, and that the phases of entrainment of low and high frequency

oscillations are independent of each other. Nonetheless, the intricate combination of low and high excitability phases in differently tuned neuronal ensembles results in a predominantly suppressive effect on auditory responses to click-trains, except in a subset of high frequency neuronal ensembles of A1. We hypothesize that the opposite sign excitability modulation of high vs. low frequency representation related to broadband transients could set the stage for the predictive processing of alternating high vs. low frequency elements of complex acoustic stimuli like speech. In this theoretical framework, oscillatory alignment to speech would be supported by a highly adaptive dual timing mechanism: both high (broadband) and low frequency elements

would reset counterphase oscillations across all of A1 within the same delta/theta cycle at time points separated by only a half delta/theta cycle. Additionally, evidence of superior phase consistency of entrained oscillations in left A1 provides support for functional hemispheric asymmetry even at the earliest auditory cortical processing stage and remarkably even in non-human primates.

FUNDING

This research project was funded by NIH grants RO1DC012947 and RO1DC011490.

REFERENCES

- Adams, J. C. (1997). Projections from octopus cells of the posteroventral cochlear nucleus to the ventral nucleus of the lateral lemniscus in cat and human. *Aud. Neurosci.* 3, 335–350.
- Boemio, A., Fromm, S., Braun, A., and Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat. Neurosci.* 8, 389–395. doi: 10.1038/nn1409
- Buzsáki, G. (2010). Neural syntax: cell assemblies, synapsembles and readers. *Neuron* 68, 362–385. doi: 10.1016/j.neuron.2010.09.023
- Buzsáki, G., Buhl, D. L., Harris, K. D., Csicsvari, J., Czéh, B., and Morozov, A. (2003). Hippocampal network patterns of activity in the mouse. *Neuroscience* 116, 201–211. doi: 10.1016/s0306-4522(02)00669-3
- Canolty, R. T., Edwards, E., Dalal, S. S., Soltani, M., Nagarajan, S. S., Kirsch, H. E., et al. (2006). High gamma power is phase-locked to theta oscillations in human neocortex. *Science* 313, 1626–1628. doi: 10.1126/science.1128115
- Castro-Alamancos, M. A. (2004). Dynamics of sensory thalamocortical synaptic networks during information processing states. *Prog. Neurobiol.* 74, 213–247. doi: 10.1016/j.pneurobio.2004.09.002
- Crochet, S., and Petersen, C. C. (2006). Correlating whisker behavior with membrane potential in barrel cortex of awake mice. *Nat. Neurosci.* 9, 608–610. doi: 10.1038/nn1690
- Da Costa, S., van der Zwaag, W., Miller, L. M., Clarke, S., and Saenz, M. (2013). Tuning in to sound: frequency-selective attentional filter in human primary auditory cortex. *J. Neurosci.* 33, 1858–1863. doi: 10.1523/JNEUROSCI.4405-12.2013
- Fanselow, E. E., and Nicolelis, M. A. (1999). Behavioral modulation of tactile responses in the rat somatosensory system. *J. Neurosci.* 19, 7603–7616.
- Ferezou, I., Bolea, S., and Petersen, C. C. (2006). Visualizing the cortical representation of whisker touch: voltage-sensitive dye imaging in freely moving mice. *Neuron* 50, 617–629. doi: 10.1016/j.neuron.2006.03.043
- Fisher, N. I. (1993). *Statistical Analysis of Circular Data*. New York: Cambridge University Press.
- Fletcher, H. (1948). The perception of speech and its relation to telephony. *Science* 108:682. doi: 10.1121/1.1906605
- Freeman, J. A., and Nicholson, C. (1975). Experimental optimization of current source-density technique for anuran cerebellum. *J. Neurophysiol.* 38, 369–382. doi: 10.1016/0006-8993(75)90494-1
- Fritz, J., Elhilali, M., and Shamma, S. (2005). Active listening: task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex. *Hear. Res.* 206, 159–176. doi: 10.1016/j.heares.2005.01.015
- Fritz, J. B., Elhilali, M., and Shamma, S. A. (2007). Adaptive changes in cortical receptive fields induced by attention to complex sounds. *J. Neurophysiol.* 98, 2337–2346. doi: 10.1152/jn.00552.2007
- Fritz, J., Shamma, S., Elhilali, M., and Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat. Neurosci.* 6, 1216–1223. doi: 10.1038/nn1141
- Gannon, P. J., Holloway, R. L., Broadfield, D. C., and Braun, A. R. (1998). Asymmetry of chimpanzee planum temporale: humanlike pattern of Wernicke's brain language area homolog. *Science* 279, 220–222. doi: 10.1126/science.279.5348.220
- Gannon, P. J., Kheck, N., and Hof, P. R. (2008). Leftward interhemispheric asymmetry of macaque monkey temporal lobe language area homolog is evident at the cytoarchitectural, but not gross anatomic level. *Brain Res.* 1199, 62–73. doi: 10.1016/j.brainres.2007.12.041
- Ghazanfar, A. A., Smith-Rohrberg, D., and Hauser, M. D. (2001). The role of temporal cues in rhesus monkey vocal recognition: orienting asymmetries to reversed calls. *Brain Behav. Evol.* 58, 163–172. doi: 10.1159/000047270
- Ghitza, O. (2011). Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Front. Psychol.* 2:130. doi: 10.3389/fpsyg.2011.00130
- Ghitza, O., and Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica* 66, 113–126. doi: 10.1159/000208934
- Giraud, A. L., Kleinschmidt, A., Poeppel, D., Lund, T. E., Frackowiak, R. S., and Laufs, H. (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron* 56, 1127–1134. doi: 10.1016/j.neuron.2007.09.038
- Giraud, A. L., and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* 15, 511–517. doi: 10.1038/nn.3063
- Golding, N. L., Ferragamo, M. J., and Oertel, D. (1999). Role of intrinsic conductances underlying responses to transients in octopus cells of the cochlear nucleus. *J. Neurosci.* 19, 2897–2905.
- Golding, N. L., Robertson, D., and Oertel, D. (1995). Recordings from slices indicate that octopus cells of the cochlear nucleus detect coincident firing of auditory nerve fibers with temporal precision. *J. Neurosci.* 15, 3138–3153.
- Greenberg, S., Carvey, H., Hitchcock, L., and Chang, S. (2003). Temporal properties of spontaneous speech-a syllable-centric perspective. *J. Phonetics* 31, 465–485. doi: 10.1016/j.wocn.2003.09.005
- Heffner, H. E., and Heffner, R. S. (1984). Temporal lobe lesions and perception of species-specific vocalizations by macaques. *Science* 226, 75–76. doi: 10.1126/science.6474192
- Heilbronner, P. L., and Holloway, R. L. (1988). Anatomical brain asymmetries in New World and Old World monkeys: stages of temporal lobe development in primate evolution. *Am. J. Phys. Anthropol.* 76, 39–48. doi: 10.1002/ajpa.1330760105
- Henry, M. J., Herrmann, B., and Obleser, J. (2014). Entrained neural oscillations in multiple frequency bands comodulate behaviour. *Proc. Natl. Acad. Sci. U S A* 111, 14935–14940. doi: 10.1073/pnas.1408741111
- Hickok, G., Farahbod, H., and Saberi, K. (2015). The rhythm of perception: Entrainment to acoustic rhythms induces subsequent perceptual oscillation. *Psychological Science* 26(7), 1006–1013. doi: 10.1177/0956797615576533
- Jamison, H. L., Watkins, K. E., Bishop, D. V., and Matthews, P. M. (2006). Hemispheric specialization for processing auditory nonspeech stimuli. *Cereb. Cortex* 16, 1266–1275. doi: 10.1093/cercor/bhj068
- Kaur, S., Lazar, R., and Metharate, R. (2004). Intracortical pathways determine breadth of subthreshold frequency receptive fields in primary auditory cortex. *J. Neurophysiol.* 91, 2551–2567. doi: 10.1152/jn.01121.2003
- Lakatos, P., Chen, C. M., O'Connell, M. N., Mills, A., and Schroeder, C. E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron* 53, 279–292. doi: 10.1016/j.neuron.2006.12.011

- Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., and Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* 320, 110–113. doi: 10.1126/science.1154735
- Lakatos, P., Musacchia, G., O'Connell, M. N., Falchier, A. Y., Javitt, D. C., and Schroeder, C. E. (2013). The spectrotemporal filter mechanism of auditory selective attention. *Neuron* 77, 750–761. doi: 10.1016/j.neuron.2012.11.034
- Lakatos, P., O'Connell, M. N., Barczak, A., Mills, A., Javitt, D. C., and Schroeder, C. E. (2009). The leading sense: supramodal control of neurophysiological context by attention. *Neuron* 64, 419–430. doi: 10.1016/j.neuron.2009.10.014
- Lakatos, P., Pincze, Z., Fu, K. M., Javitt, D. C., Karmos, G., and Schroeder, C. E. (2005a). Timing of pure tone and noise-evoked responses in macaque auditory cortex. *Neuroreport* 16, 933–937. doi: 10.1097/00001756-200506210-00011
- Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., and Schroeder, C. E. (2005b). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *J. Neurophysiol.* 94, 1904–1911. doi: 10.1152/jn.00263.2005
- Legatt, A. D., Arezzo, J., and Vaughan, H. G. Jr. (1980). Averaged multiple unit activity as an estimate of phasic changes in local neuronal activity: effects of volume-conducted potentials. *J. Neurosci. Methods* 2, 203–217. doi: 10.1016/0165-0270(80)90061-8
- Luo, H., and Poeppel, D. (2012). Cortical oscillations in auditory perception and speech: evidence for two temporal windows in human auditory cortex. *Front. Psychol.* 3:170. doi: 10.3389/fpsyg.2012.00170
- May, B., Moody, D. B., and Stebbins, W. C. (1989). Categorical perception of conspecific communication sounds by Japanese macaques, *Macaca fuscata*. *J. Acoust. Soc. Am.* 85, 837–847. doi: 10.1121/1.397555
- McGinley, M. J., Liberman, M. C., Bal, R., and Oertel, D. (2012). Generating synchrony from the asynchronous: compensation for cochlear traveling wave delays by the dendrites of individual brainstem neurons. *J. Neurosci.* 32, 9301–9311. doi: 10.1523/JNEUROSCI.0272-12.2012
- Mendelson, J. R., Schreiner, C. E., and Sutter, M. L. (1997). Functional topography of cat primary auditory cortex: response latencies. *J. Comp. Physiol. A* 181, 615–633. doi: 10.1007/s003590050145
- Merzenich, M. M., and Brugge, J. F. (1973). Representation of the cochlear partition of the superior temporal plane of the macaque monkey. *Brain Res.* 50, 275–296. doi: 10.1016/0006-8993(73)90731-2
- Mitzdorf, U. (1985). Current source-density method and application in cat cerebral cortex: investigation of evoked potentials and EEG phenomena. *Physiol. Rev.* 65, 37–100.
- Morillon, B., Lehongre, K., Frackowiak, R. S., Ducorps, A., Kleinschmidt, A., Poeppel, D., et al. (2010). Neurophysiological origin of human brain asymmetry for speech and language. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18688–18693. doi: 10.1073/pnas.1007189107
- Obleser, J., Eisner, F., and Kotz, S. A. (2008). Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *J. Neurosci.* 28, 8116–8123. doi: 10.1523/JNEUROSCI.1290-08.2008
- O'Connell, M. N., Barczak, A., Schroeder, C. E., and Lakatos, P. (2014). Layer specific sharpening of frequency tuning by selective attention in primary auditory cortex. *J. Neurosci.* 34, 16496–16508. doi: 10.1523/JNEUROSCI.2055-14.2014
- O'Connell, M. N., Falchier, A., McGinnis, T., Schroeder, C. E., and Lakatos, P. (2011). Dual mechanism of neuronal ensemble inhibition in primary auditory cortex. *Neuron* 69, 805–817. doi: 10.1016/j.neuron.2011.01.012
- Oertel, D., Bal, R., Gardner, S. M., Smith, P. H., and Joris, P. X. (2000). Detection of synchrony in the activity of auditory nerve fibers by octopus cells of the mammalian cochlear nucleus. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11773–11779. doi: 10.1073/pnas.97.22.11773
- Oertel, D., Wu, S. H., Garb, M. W., and Dizack, C. (1990). Morphology and physiology of cells in slice preparations of the posteroverentral cochlear nucleus of mice. *J. Comp. Neurol.* 295, 136–154. doi: 10.1002/cne.902950112
- Otazu, G. H., Tai, L. H., Yang, Y., and Zador, A. M. (2009). Engaging in an auditory task suppresses responses in auditory cortex. *Nat. Neurosci.* 12, 646–654. doi: 10.1038/nn.2306
- Peele, J. E., Gross, J., and Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cereb. Cortex* 23, 1378–1387. doi: 10.1093/cercor/bhs118
- Petersen, M. R., Beecher, M. D., Zoloth, S. R., Moody, D. B., and Stebbins, W. C. (1978). Neural lateralization of species-specific vocalizations by Japanese macaques (*Macaca fuscata*). *Science* 202, 324–327. doi: 10.1126/science.99817
- Pichora-Fuller, M. K., and Souza, P. E. (2003). Effects of aging on auditory processing of speech. *Int. J. Audiol.* 42(Suppl. 2), 11–16. doi: 10.3109/14992020309074638
- Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as “asymmetric sampling in time”. *Speech Communication* 41, 245–255. doi: 10.1016/S0167-6393(02)00107-3
- Poremba, A., Malloy, M., Saunders, R. C., Carson, R. E., Herscovitch, P., and Mishkin, M. (2004). Species-specific calls evoke asymmetric activity in the monkey's temporal poles. *Nature* 427, 448–451. doi: 10.1038/nature02268
- Rauschecker, J. P., Tian, B., Pons, T., and Mishkin, M. (1997). Serial and parallel processing in rhesus monkey auditory cortex. *J. Comp. Neurol.* 382, 89–103. doi: 10.1002/(sici)1096-9861(19970526)382:1<89::aid-cne6>3.0.co;2-g
- Recanzone, G. H., Guard, C., and Phan, M. L. (2000). Frequency and intensity response properties of single neurons in the auditory cortex of the behaving macaque monkey. *J. Neurophysiol.* 83, 2315–2331.
- Rhode, W. S., Oertel, D., and Smith, P. H. (1983). Physiological response properties of cells labeled intracellularly with horseradish peroxidase in cat ventral cochlear nucleus. *J. Comp. Neurol.* 213, 448–463. doi: 10.1002/cne.902130408
- Rhode, W. S., and Smith, P. H. (1986). Encoding timing and intensity in the ventral cochlear nucleus of the cat. *J. Neurophysiol.* 56, 261–286.
- Rizzuto, D. S., Madsen, J. R., Bromfield, E. B., Schulze-Bonhage, A., and Kahana, M. J. (2006). Human neocortical oscillations exhibit theta phase differences between encoding and retrieval. *Neuroimage* 31, 1352–1358. doi: 10.1016/j.neuroimage.2006.01.009
- Schreiner, C. E., and Winer, J. A. (2007). Auditory cortex mapmaking: principles, projections and plasticity. *Neuron* 56, 356–365. doi: 10.1016/j.neuron.2007.10.013
- Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., and Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends Cogn. Sci.* 12, 106–113. doi: 10.1016/j.tics.2008.01.002
- Schroeder, C. E., Lindsley, R. W., Specht, C., Marcovici, A., Smiley, J. F., and Javitt, D. C. (2001). Somatosensory input to auditory association cortex in the macaque monkey. *J. Neurophysiol.* 85, 1322–1327.
- Schroeder, C. E., Mehta, A. D., and Givre, S. J. (1998). A spatiotemporal profile of visual system activation revealed by current source density analysis in the awake macaque. *Cereb. Cortex* 8, 575–592. doi: 10.1093/cercor/8.7.575
- Shamir, M., Ghitta, O., Epstein, S., and Kopell, N. (2009). Representation of time-varying stimuli by a network exhibiting oscillations on a faster time scale. *PLoS Comput. Biol.* 5:e1000370. doi: 10.1371/journal.pcbi.1000370
- Slotnick, B. (2009). A simple 2-transistor touch or lick detector circuit. *J. Exp. Anal. Behav.* 91, 253–255. doi: 10.1901/jeab.2009.91-253
- Steinschneider, M., Reser, D., Schroeder, C. E., and Arezzo, J. C. (1995). Tonotopic organization of responses reflecting stop consonant place of articulation in primary auditory cortex (A1) of the monkey. *Brain Res.* 674, 147–152. doi: 10.1016/0006-8993(95)00008-e
- Wang, X., Merzenich, M. M., Beitel, R., and Schreiner, C. E. (1995). Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. *J. Neurophysiol.* 74, 2685–2706.
- Young, C. K., and Eggermont, J. J. (2009). Coupling of mesoscopic brain oscillations: recent advances in analytical and theoretical perspectives. *Prog. Neurobiol.* 89, 61–78. doi: 10.1016/j.pneurobio.2009.06.002
- Zoefel, B., and Heil, P. (2013). Detection of near-threshold sounds is independent of EEG phase in common frequency bands. *Front. Psychol.* 4:262. doi: 10.3389/fpsyg.2013.00262

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 O'Connell, Barczak, Ross, McGinnis, Schroeder and Lakatos. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Exploring the Role of Brain Oscillations in Speech Perception in Noise: Intelligibility of Isochronously Retimed Speech

Vincent Aubanel*, Chris Davis and Jeesun Kim

MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Penrith, NSW, Australia

A growing body of evidence shows that brain oscillations track speech. This mechanism is thought to maximize processing efficiency by allocating resources to important speech information, effectively parsing speech into units of appropriate granularity for further decoding. However, some aspects of this mechanism remain unclear. First, while periodicity is an intrinsic property of this physiological mechanism, speech is only quasi-periodic, so it is not clear whether periodicity would present an advantage in processing. Second, it is still a matter of debate which aspect of speech triggers or maintains cortical entrainment, from bottom-up cues such as fluctuations of the amplitude envelope of speech to higher level linguistic cues such as syntactic structure. We present data from a behavioral experiment assessing the effect of isochronous retiming of speech on speech perception in noise. Two types of anchor points were defined for retiming speech, namely syllable onsets and amplitude envelope peaks. For each anchor point type, retiming was implemented at two hierarchical levels, a slow time scale around 2.5 Hz and a fast time scale around 4 Hz. Results show that while any temporal distortion resulted in reduced speech intelligibility, isochronous speech anchored to P-centers (approximated by stressed syllable vowel onsets) was significantly more intelligible than a matched anisochronous retiming, suggesting a facilitative role of periodicity defined on linguistically motivated units in processing speech in noise.

OPEN ACCESS

Edited by:

Johanna Maria Rimmeli,
Max-Planck-Institute for Empirical
Aesthetics, Germany

Reviewed by:

Nai Ding,
Zhejiang University, China
Keith B. Doelling,
New York University, USA

*Correspondence:

Vincent Aubanel
v.aubanel@westernsydney.edu.au

Received: 29 February 2016

Accepted: 10 August 2016

Published: 31 August 2016

Citation:

Aubanel V, Davis C and Kim J (2016)
Exploring the Role of Brain Oscillations
in Speech Perception in Noise:
Intelligibility of Isochronously Retimed
Speech.
Front. Hum. Neurosci. 10:430.
doi: 10.3389/fnhum.2016.00430

1. INTRODUCTION

Human speech perception is remarkably robust to the auditory perturbations commonly encountered in everyday communicative situations, such as the babble noise in the background of a busy café. Indeed, under such conditions, recognition rates that can be achieved outstrip those attained by the best currently available automatic methods. While much is known about the role of spectral cues in recognizing speech in noise, and about how speech intelligibility is affected by energetic masking, much less is known about the role that temporal factors play.

Temporal factors are important for the comprehension process because perception of temporal properties underlies the human ability to focus attention on the target source and to segregate it from competing sources. Recently, interest in temporal mechanisms in speech recognition has been bolstered by proposals concerning the role of brain oscillations in sensory perception in general (Buzsaki and Draguhn, 2004) and speech processing in particular (Giraud and Poeppel, 2012b).

Specifically, it has been proposed that cortical oscillations underpin mechanisms of sensory selection (Schroeder and Lakatos, 2009) and the parsing of important elements of the speech signal for further decoding (Giraud and Poeppel, 2012a).

Brain oscillations are fluctuations of local field potentials that can be measured using electrophysiological techniques and reflect the excitability of population of neurons. Increased excitability translates into increased dynamic range for coding information by those neurons ensembles. Although cortical oscillations arise due to basic physiological mechanisms (endogenous rhythms), it has been shown that these oscillations are not fixed (even single neurons can oscillate at different frequencies Hutcheon and Yarom, 2000); moreover, cortical oscillations entrain to the fluctuations of external stimuli (Ahissar et al., 2001; Lakatos et al., 2005, 2008; Aiken and Picton, 2008). It is this latter property that has motivated proposals that phases in the processing efficiency are interactively aligned to important phases in the unfolding of spoken information (Large and Jones, 1999; Schroeder and Lakatos, 2009; Besle et al., 2011; Peelle and Davis, 2012). Such a coupling would result in a net increase in processing efficiency for those parts of speech that are crucial to decode. This idea has received experimental support from findings that theta phase patterns track spoken sentences, with the degree of tracking correlated with speech intelligibility (Ahissar et al., 2001; Luo and Poeppel, 2007; Nourski et al., 2009; Peelle et al., 2013).

In addition to a general alignment of cortical and speech oscillations, there is an enticing link between the timing of cortical oscillatory activity and the duration of specific spoken linguistic units. For example, the time scale of delta-theta band activity (2–8 Hz) corresponds to the average frequency of the syllabic rate, i.e., in English stressed syllables occur at a rate below 4 Hz (Greenberg et al., 2003) and across languages the syllable rate is between 5 and 8 Hz (Pellegrino et al., 2011). Gamma band activity (25–70 Hz) has an approximate correspondence with the duration of subphonemic elements. Further, frequency nesting, or the dependence of higher frequency power on the phase of a lower frequency band is a pervasive phenomenon in brain oscillations (Canolty and Knight, 2010) and has consistently been observed between the delta and gamma band during speech processing (Schroeder and Lakatos, 2009; Ding and Simon, 2014). This hierarchical relationship bears a striking similarity to the nesting of linguistic units in speech, i.e., syllables are composed of phonemes, and words are composed of syllables (although strict inclusion of a smaller unit into a larger unit is not always the case, as seen, for example, in French syllabification). Taken together, these results have led researchers to propose that brain oscillations act as an active parser of speech, packaging the continuous stream of information into units of different granularity for further processing (Giraud and Poeppel, 2012a).

Although the linking of cortical oscillations and speech processing provides a rich framework for understanding the efficiency of speech perception, a number of important details remain unclear. For example, it is not clear which speech units and properties provide the basis for cortical tracking. The syllable has been proposed as a central unit with the critical-band amplitude envelopes playing a vital role (Greenberg et al., 2003; Luo and Poeppel, 2007; Ghitza, 2013; Ghitza

et al., 2013). However, it should be noted that the syllable rate is not robustly encoded in the fluctuations of the amplitude envelope (Cummins, 2012). That is, although in most languages, and for carefully articulated forms of speech, a syllable consists of a vocalic nucleus in which the amplitude peak stands out in relation to surrounding consonants, this characterization does not always hold in the more commonly encountered conversational speech forms. This is due to reduction phenomena where vocalic nuclei can be deleted (Meunier and Espesser, 2011), and voiced consonants can be louder than the vocalic nucleus. Therefore, in continuous speech, the mapping of amplitude peaks to a syllable or stressed syllable is not as straightforward. Perhaps a more useful conceptualization of the speech cues that may be involved in driving cortical entrainment (and parsing) is that of P-centers, or perceptually defined moments of occurrence of word onsets (Morton et al., 1976). While P-centers bear a close connection with the syllabic description of an utterance, they may constitute an appropriate level of description for determining cortical entrainment to speech, in that they provide a perceptually grounded parsing of beats of an utterance.

Another uncertainty in connecting cortical oscillations and speech processing concerns whether the entrainment of cortical oscillations is driven primarily by the bottom-up physical characteristics of the stimulus or whether top-down control is important. On the one hand, oscillatory activity could be the result of a direct bottom-up response to the physical patterning of the stimulus, since the speech amplitude envelope has been found to correlate with the cortical response (Ahissar et al., 2001). On the other hand, the speech envelope is often obscured by noise (Houtgast and Steeneken, 1985), which would require an active phase resetting mechanism to realign cortical tracking (perhaps with salient acoustic events playing a role, Doelling et al., 2014). Moreover, it has been shown that entrainment still occurs even in the absence of amplitude fluctuations (Zoefel and VanRullen, 2015a) or marked auditory events signaling onsets of complex auditory rhythmic patterns (Chapin et al., 2010; Barczak et al., 2015) and that phase locking is affected by intelligibility even though stimuli have the same amplitude envelope (Peelle et al., 2013). These demonstrations have led some researchers to propose that top-down factors must be taken into account when considering the cortical tracking of speech (Obleser et al., 2012; Peelle et al., 2013).

In summary, the oscillation perspective is an important one but a number of prominent issues need to be clarified. The main issue we address in the current study concerns the basis for tracking the speech signal and can be summed up by this question: Is tracking based purely on a low-level physical property, such as amplitude envelope, or do top-down linguistic factors have a role to play? To test this we determined how timing modifications that imposed periodicity on naturally produced speech (i.e., making speech isochronous) affect intelligibility. Here, we manipulate the basis of timing modification and take intelligibility as an index of speech processing efficiency. Given the proposed link between intelligibility and cortical entrainment, we interpret the results in the context of the neurocognitive framework of brain oscillations.

In setting out to test the influence of imposed periodicity on speech perception we chose to examine speech perception in noise. One reason to test in noise was so that correct word identification performance would be well away from ceiling levels; so providing a chance to readily observe the effects of any manipulation. More importantly, testing such a manipulation in a noise environment is well motivated by theory. That is, although cortical tracking can be maintained in noise (Ding and Simon, 2013; Ding et al., 2014) it is likely that with an increasing noise level, more frequent phase resetting would be required due to the increased sparsity of available speech information and the concomitant increase in uncertainty about the speech sequence being processed. If there is some cost associated with this online adjustment, then a stimulus with isochronous periodic characteristics should attract less cost since its phase would be easier to predict. Here, the effect of isochrony on intelligibility would be greatest when the anchor point used for the temporal modification coincides with what is important for cortical tracking. Another possibility is that in the absence of predictable cues for phase resetting, the oscillatory system engages into a default set of “idle” frequencies at values typically observed during speech perception, which may provide a processing benefit when the high excitability phases align with important speech information.

In the current study, we transformed naturally timed speech to an isochronous form, and compared it with the baseline of unmodified speech and a matched transformed condition of anisochronous speech (see Section 2.2.1). To test a bottom-up account of cortical tracking we used peaks in the amplitude-envelope as anchor-points to render speech isochronous. That is, we hypothesized that if cortical entrainment is mainly driven by salient acoustic cues then making those cues regular should lead to intelligibility benefits in noise. To test whether cortical tracking used higher-level acoustic cues defined in linguistic terms, we used syllable onsets as anchor points for the isochronous modification.

We also assessed the effect of the isochronous modification at two levels of timing corresponding to two distinct frequencies. Entrainment has been observed in the delta-theta range spanning 2–8 Hz, which encompasses the average frequency of two hierarchical metrical levels in speech: that of the stressed syllable and that of the syllable. These two timing levels were chosen for the syllable-based transformation, and were matched in frequency for the amplitude envelope-based isochronous transformation.

Selecting the stressed syllable onset as an anchor point provides a way of operationalizing the concept of perceptual beat, or P-center, since the latter tend to be located near the onset of vowels in stressed syllables (Allen, 1972a,b; Morton et al., 1976). Note that in employing syllable and stressed syllable onsets as anchor points we are not necessarily proposing that oscillatory mechanisms need knowledge of syllable and stress boundaries but simply that these are sensitive to the perceptual beat or rhythm of an utterance.

In Section 2, we present the speech material used and detail the isochronous transformation, introducing a temporal distortion metric that defines the anisochronous transformation. Listener results are presented in Section 3, and discussed in Section 4.

2. MATERIALS AND METHODS

2.1. Stimuli

One hundred and ninety sentences from the Harvard set (Rothauser et al., 1969) were spoken by a female native Australian English talker in her mid-twenties. The sentences had at least five keywords and were mildly predictable, such as in *Large size in stockings is hard to sell*. Sentences were individually segmented and automatically forced-aligned into words and phonemes. Phoneme boundaries were manually checked and corrected. Syllables were individually coded as stressed or unstressed based on a dictionary lookup¹ of lexical stress and manually adjusted for sentence level stress patterns and particular production of the talker.

The amplitude envelope of speech was computed by taking the root mean square of the waveform amplitude values for adjacent 16 ms frames and the final value was taken as the running average over 7 frames. Peak values were selected iteratively by selecting the maximum of the envelope, marking the surrounding 80 ms to prevent from subsequent selection, and repeating the process until no maximum value with a surrounding region could be determined.

2.2. Experimental Design

2.2.1. Isochronous Transformation

The isochronous transformation (hereafter: *iso*) operates on each sentence by locally compressing or expanding contiguous speech regions so that these regions have an identical duration. For the N anchor points $a_1 \dots a_N$ identifying the boundaries of speech regions in a sentence, the time scale function τ is defined as a step function that associates a time scale factor to each sample n :

$$\tau(n) = \begin{cases} 1 & \text{if } n < a_1 \\ \frac{a_{i+1} - a_i}{d} & \text{if } a_i < n \leq a_{i+1} \text{ with } 1 \leq i < N \\ 1 & \text{if } n \geq a_N \end{cases} \quad (1)$$

where $d = \frac{a_N - a_1}{N}$ is the mean duration of the sequence of speech regions to transform. With this definition, the timing of speech portions preceding the first anchor point and following the last anchor point remains unchanged, and so does the total duration of the speech regions. An example of the time scale function is seen in **Figure 1D**. The time scale factors are applied to the speech signal using WSOLA (Demol et al., 2005), a non-uniform time scaling algorithm that achieves high naturalness by adjusting local time scale factors according to sound class while preserving accurate timing.

We then define the temporal distortion metric δ which quantifies the amount of elongation and compression applied to a sentence, as the root mean square of the log-transformed time scale factors:

$$\delta = \sqrt{\frac{1}{N} \sum_{n=1}^N \log(\tau(n))^2}. \quad (2)$$

This metric enables us to further define an anisochronous transformation (*aniso*) as a counterpart of the isochronous

¹<http://dictionary.cambridge.org/dictionary/british/>

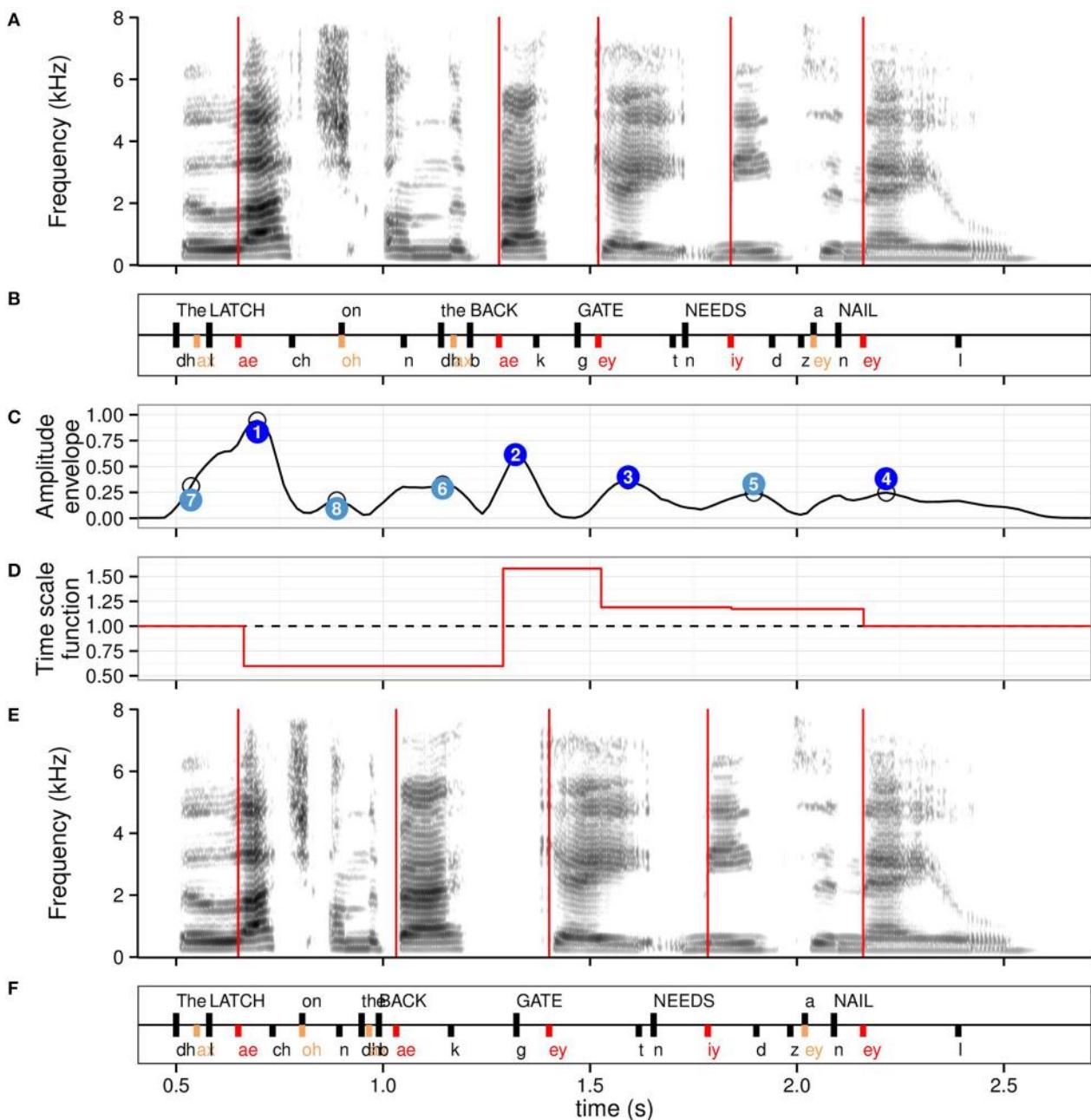


FIGURE 1 | Anchor points used for the isochronous modification, with the associated time scale function for the stressed syllable level for the example sentence *The latch on the back gate needs a nail*. **(A)** Spectrogram of the naturally timed sentence, with time instants of stressed syllables onsets overlaid in red. **(B)** Annotation in words (scored keywords in capital letters) and phonemes. Stressed syllables are in red, unstressed syllables in orange, remaining phonemes, and word onsets in black. **(C)** Amplitude envelope in normalized units with original peaks identified with empty circles. Peaks adjusted for sentence-level decay are shown in numbered circles with numbers indicating decreasing value order. Dark blue: four highest peaks; light blue: remaining peaks up to height. Note the relative agreement in timing between stressed syllables and low number of amplitude envelope peaks anchor points. **(D)** Time scale function for the stressed syllables anchor points. Values < 1 indicate compression, values > 1 elongation. **(E)** Spectrogram of resulting isochronous sentence at the stressed syllable level, with time instants of isochronous stressed syllables onsets overlaid in red. **(F)** Resulting isochronous annotation.

one, which applies identical amount of temporal distortion as the latter but which does not result in equal duration of speech regions. This condition is important in our design as it will help to disentangle the effects of periodicity and

temporal distortion in performance. Here, the anisochronous transformation was operationally defined by applying the time scale factors obtained for the isochronous transformation in a reverse order, i.e., $\tau_{aniso}[1, 2, \dots, N] = \tau_{iso}[N, N-1, \dots, 1]$.

This way, anisochronously retimed anchor points will have locations which are perceptually unpredictable from neither the original nor the isochronously retimed anchor point locations.

2.2.2. Anchor Points

Two types of anchor points were defined to assess the effect of isochronous retiming: the *syllable* and the *amplitude envelope* anchor points. For each of the anchor point type, retiming was implemented at a slow and a fast time scale (syllable: *stressed* and *all syllables*; amplitude envelope: *low* and *high number of peaks*, respectively).

The stressed syllable anchor points (*str*) were taken as the onset of the nuclear vowel of stressed syllables in the sentence. This time point is considered to carry the perceptual beat of a syllable (Allen, 1975; Port, 2003) and is a good approximation of the P-center (Morton et al., 1976; Scott, 1993) which exact location varies as a function of the length of the consonant cluster of the syllable onset (Patel et al., 1999; Cummins, 2015). Using stressed syllables as anchor points for the isochronous transformation results in a form of speech that is similar to that obtained in a speech cycling task where talkers are asked to repeat a sentence in the presence of a regular timing beat (Cummins and Port, 1998). The all syllable anchor points (*syl*) were obtained by selecting vowel onsets of all syllables in the sentence.

These two hierarchical levels were matched for the amplitude envelope anchor points. For the current material of sentences with a mean duration of 2.10 s ($SD = 0.25$), low number of peaks (*loN*) anchor points were empirically selected as the first 4–5 peaks of the amplitude envelope. The exact number of peaks was chosen as the number of peaks that led to the least temporal distortion (see Equation 2). Similarly, high number of peaks (*hiN*) anchor points were selected as the first 7–8 highest peaks, whichever led to the least temporal distortion. Prior to peak selection, the amplitude decay that occurs naturally in production between the beginning and ending of a sentence was

compensated for by applying a correction factor to the values of the envelope peaks. The decay was estimated by the slope a of the regression line modeling the amplitude values of the eight highest envelope peaks, and the value of the peaks were adjusted according to:

$$y_{i,\text{adj}} = y_i - at_i + a \frac{t_N - t_1}{2} \quad (3)$$

where $y_{i,\text{adj}}$ is the adjusted amplitude value of peak i ($i = 1, \dots, N; N = 8$), y_i is the amplitude value of peak i and t_i the time instant of peak i . The resulting regression line modeling the adjusted peak values has a null slope, and the resulting ordering of the adjusted peak values is normalized with reference to amplitude decay. Original and adjusted peak values are represented in **Figure 1C**.

Figure 1 shows the two types of anchor points at two hierarchical levels for an example sentence, along with the resulting isochronous retiming using the stressed syllable anchor points (condition *iso.str*).

Figure 2 summarizes the average inter-anchor point frequency and the temporal distortion across the four transformation conditions over the 190 sentences of the corpus. Low number of peaks anchor points had an average frequency of 2.37 Hz, not significantly different from stressed syllables anchor points (2.43 Hz). High number of peaks anchor points were paced at a lower frequency than all syllables anchor points (3.33 and 4.56 Hz respectively). Low number of peaks anchor points selection led to the greatest temporal distortion while stressed syllables anchor points were the one leading to the least temporal distortion. Both high number of peaks and all syllables anchor points led to similar temporal distortion.

The two types of anchor points being different in nature, with one being linguistically grounded and the other signal based, there is the possibility that they affect the timing of phonetic segments in different ways. We examined the phoneme-level temporal distortion by computing the temporal distortion

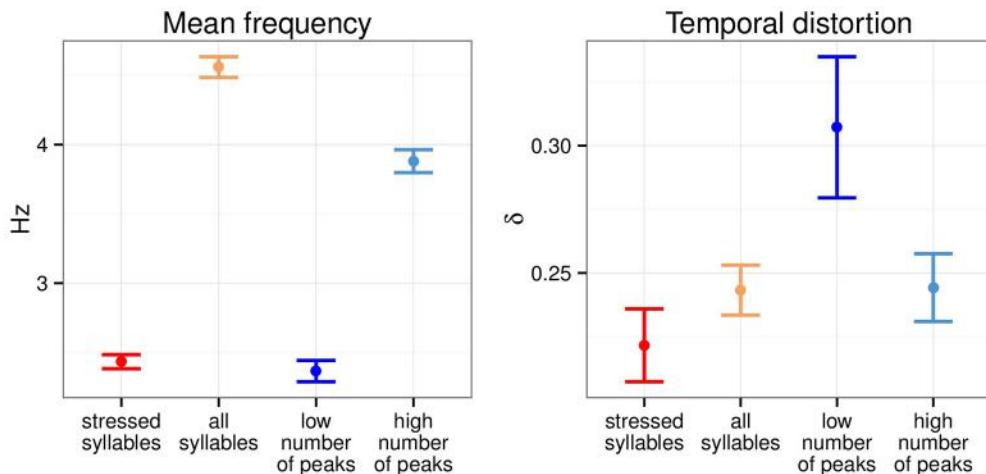


FIGURE 2 | Mean inter-anchor point frequency and temporal distortion across four temporal modification conditions. Errorbars, here and elsewhere, show 95% confidence intervals ($N = 190$).

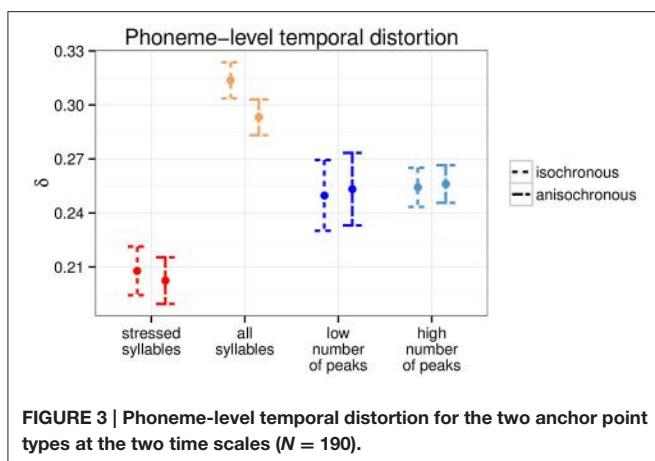


FIGURE 3 | Phoneme-level temporal distortion for the two anchor point types at the two time scales ($N = 190$).

(Equation 2) over successive phoneme units instead of every sample. **Figure 3** shows the phoneme-level temporal distortion for the two types of anchor points. Both isochronous and anisochronous modifications led to comparable phoneme-level temporal distortion [all $p > 0.5$, for individual Welch two sample t -tests per anchor point type], to the exception of the all syllable level, where the isochronous modification resulted in significant greater phoneme-level temporal distortion than the anisochronous one [$t_{(377.93)} = 2.85, p < 0.01$].

The effect of the two types of anchor points were tested in two separate experiments. Experiment I tested the effect of the syllable based anchor points and Experiment II that of the amplitude envelope based anchor points. Both had five conditions: an unmodified naturally timed speech condition (*unmod*) and four other conditions obtained by crossing the transformation polarity with the two metrical levels of the anchor points (Experiment I: *unmod*, *iso.str*, *aniso.str*, *iso.syl*, *aniso.syl*; Experiment II: *unmod*, *iso.loN*, *aniso.loN*, *iso.hiN*, *aniso.hiN*). Example stimuli can be found in Supplementary Materials.

2.3. Participants and Procedure

Participants were recruited from the undergraduate population of Western Sydney University and through personal acquaintances. University students received course credit for participation while other participants did not receive any remuneration. All participants provided informed consent and reported normal hearing. All research procedures were approved by the Human Research Ethics Committee of Western Sydney University under the reference H9495. Thirty participants took part in Experiment I. Four participants were discarded following performance-based exclusion criteria detailed in Section 2.3.1, leaving twenty-one females and five males with mean age of 20.9 ($SD = 6.3$). A different cohort of thirty participants were recruited for Experiment II and one participant was excluded following the same exclusion criteria as in Experiment I, leaving twenty-four females and five males with mean age of 20.9 ($SD = 5.9$) for analysis.

Participants were tested individually and sat in a sound attenuated booth in front of a computer screen, where there

were presented with online instructions. Both experiments had an identical setup: sentences mixed with noise were presented in blocks and the participants had to type what they heard. The experiments were self-paced, and participants could take a break after the third block out of five. Stimuli were presented over Beyerdynamic DT 770 Pro 80 Ohm closed headphones at a fixed level. Sentences were mixed with speech-shaped noise (SSN) at a fixed signal-to-noise ratio of -3 dB SNR. SSN was constructed by filtering white noise with 200 LPC coefficients taken from the long-term average speech spectrum computed on a concatenation of all sentences of the corpus. RMS energy of sentence-plus-noise mixtures were individually adjusted to a fixed value of 0.04. Each experiment took 45 min to complete on average.

In both experiments, sentences were blocked in five sets of thirty-six sentences. Block order was determined by a latin square design. Sentences were randomly distributed across the five conditions for each participant so that each participant heard each sentence only once and each sentence could be heard in different condition across participants. Within each block, sentences were ordered from low to high anchor points frequency, in order to minimize perceived rhythmic change from trial to trial. The remaining ten sentences were presented as practice, two at the beginning of each block, and were not used for scoring.

2.3.1. Scoring

In both experiments, typed sentences were scored by counting the correct keywords per sentence. Keywords were determined from the sentence orthographic form by excluding a list of function words, such as "a," "the," "for," "in." Original orthographic sentence and typed responses were parsed into a canonical form to account for homophones and spelling mistakes, and the proportion of matching words was computed for each typed sentence.

Data from participants who scored less than 20% in at least one condition or did not provide a response for at least 50% of the sentences across any condition were discarded from the dataset.

3. RESULTS

3.1. Listeners' Performance

Figure 4 shows the proportion of keywords correctly identified for Experiments I and II. Temporally unmodified speech in noise was best recognized in both experiments. Stressed-syllable based retimed speech was better recognized than all-syllable based retimed speech (Experiment I) while both metrical levels had similar intelligibility reduction in Experiment II. Crucially, isochronous retiming led to better recognition for syllable based transformations (Experiment I) but not for amplitude based ones (Experiment II).

For each experiment, we evaluated the effect of the five conditions on intelligibility with a generalized linear mixed model applied to individual word counts. We used *condition* with five levels as the fixed effect, and intercept for subjects and sentence as random effects. *P*-values were obtained by

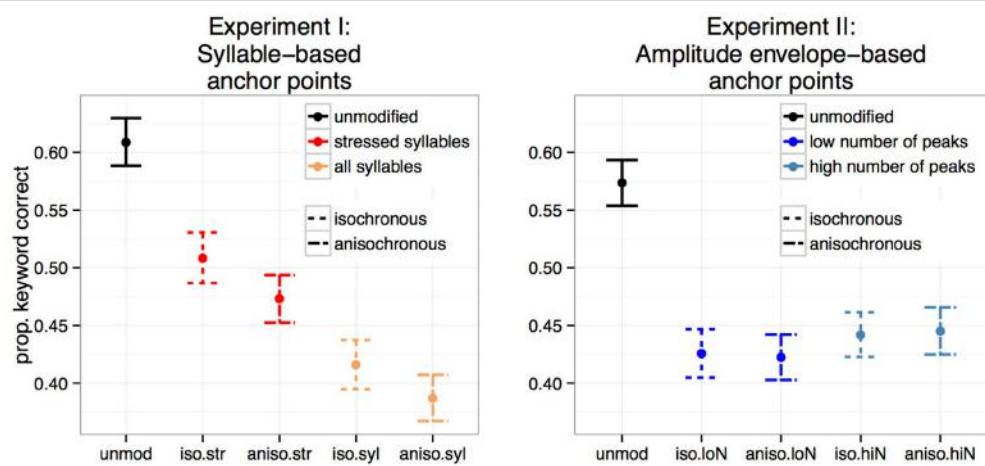


FIGURE 4 | Mean proportion of correctly identified keyword per sentence over participants for the five conditions. Experiment I: syllable-based anchor points ($N = 26$). Experiment II: amplitude envelope-based anchor points ($N = 29$).

conducting simultaneous tests for general linear hypotheses, specifying a matrix of contrast across the conditions (function `glht()` of the *lme4* R package, Bates et al., 2015). Random effects standard deviation of subject and sentence were 0.40 and 1.10 respectively for Experiment I and 0.36 and 1.10 respectively for Experiment II. Results of individual comparisons are given in **Table 1** for both experiments.

As shown by comparisons 1–4 in **Table 1**, all transformed speech conditions resulted in significantly poorer intelligibility than unmodified speech, for both experiments. Next, testing the isochronous modification against the anisochronous modification separately for each anchor point type revealed that when stressed syllables are taken as anchor points, isochronous speech is more intelligible than anisochronous speech (Experiment I, comparison 5). A tendency for this effect is observed when all syllables are taken as anchor points (Experiment I, comparison 6). In contrast, when applied to anchor points defined on the amplitude envelope, the isochronous transformation did not result in intelligibility changes, for any of the low and high number of peaks (Experiment II, comparisons 5 and 6 respectively). This effect was also observed when collapsing identical anchor point types within each experiment: in Experiment I, isochronous speech was more intelligible than anisochronous speech regardless of the anchor point type (Experiment I, comparison 7) and in Experiment II, intelligibility of isochronous speech was not distinguishable from intelligibility of anisochronous speech when collapsing across anchor point types (Experiment II, comparison 7). Finally, in both experiments, the choice of anchor points had a clear net effect on intelligibility, with transformations anchored on stressed syllable being significantly more intelligible than that implemented on all syllables (Experiment I, comparison 8), and transformations anchored on high number of peaks being slightly but significantly more intelligible than transformations anchored on low number of amplitude envelope peaks (Experiment II, comparison 8).

3.2. Sentence Intelligibility

As shown in **Figure 2**, the choice of anchor point type resulted in marked differences in inter-anchor points frequency and temporal distortion. We examined the relation of these metrics with transformed sentences intelligibility. Unmodified sentences were not analyzed as these metric do not apply to them. **Table 2** shows the correlation between transformed sentences intelligibility and temporal distortion on one hand, and with mean frequency on the other hand.

Table 2a shows that transformed sentences intelligibility was negatively correlated with temporal distortion. This correlation applied across the board for all transformation conditions, with a highest value for low number of peaks isochronous, a condition that had highest temporal distortion, and that also lead to a low intelligibility score. Within an identical anchor point type condition, isochronous and anisochronous transformed sentences had similar correlation coefficients, probably owing to the fact that sentences did indeed have identical distortion factors.

Also displayed in **Table 2b** is the result that retimed sentences' mean frequency was positively correlated with sentence intelligibility, with a notable exception of sentences that were retimed according to stressed syllables, where mean retimed sentence frequency did not explain any intelligibility variation.

Further analysis was conducted to evaluate the correlation between temporal distortion and mean frequency of retimed sentences. Only isochronous sentences were analyzed as anisochronous sentences have identical temporal distortion and mean frequency to isochronous ones. A negative correlation for sentences retimed using all syllables as anchor points was found ($r = -0.28, p < 0.001$) and similarly with any of the amplitude envelope-based anchor points (low number of peaks: $r = -0.33, p < 0.001$; high number of peaks: $r = -0.51, p < 0.001$). However, no correlation was found for sentences retimed at the stressed syllable level ($p = 0.25$).

TABLE 1 | Output of generalized linear mixed models fitted separately for each experiment.

n	Experiment I					Experiment II				
	Comparison	Estim.	z	p	Sig.	Comparison	Estim.	z	p	Sig.
1	<i>iso.str, unmod</i>	-0.545	-12.16	< 0.001	***	<i>iso.loN, unmod</i>	-0.847	-19.87	< 0.001	***
2	<i>aniso.str, unmod</i>	-0.722	-16.06	< 0.001	***	<i>aniso.loN, unmod</i>	-0.821	-19.18	< 0.001	***
3	<i>iso.syl, unmod</i>	-1.017	-22.40	< 0.001	***	<i>iso.hiN, unmod</i>	-0.713	-16.78	< 0.001	***
4	<i>aniso.syl, unmod</i>	-1.127	-24.68	< 0.001	***	<i>aniso.hiN, unmod</i>	-0.666	-15.67	< 0.001	***
5	<i>iso.str, aniso.str</i>	0.177	4.00	< 0.001	***	<i>iso.loN, aniso.loN</i>	-0.027	-0.63	0.965	
6	<i>iso.syl, aniso.syl</i>	0.110	2.43	0.092	.	<i>iso.hiN, aniso.hiN</i>	-0.047	-1.12	0.771	
7	<i>iso, aniso</i>	0.287	4.54	< 0.001	***	<i>iso, aniso</i>	-0.074	-1.24	0.700	
8	<i>syl, str</i>	-0.878	-13.80	< 0.001	***	<i>hiN, loN</i>	0.290	4.82	< 0.001	***

Each numbered line shows a comparison, its estimate, the z-value and associated p-value, and visual indication of signficativity.

TABLE 2 | Correlation between intelligibility scores of transformed sentences and a. Temporal distortion; b. Mean frequency for both Experiments across all subjects.

Experiment I					Experiment II				
Anchor point	Transf.	r	p	Sig.	Anchor point	Transf.	r	p	Sig.
a. Temporal distortion									
Stressed syllables (str)	<i>iso</i>	-0.19	0.000	***	Low number (<i>loN</i>)	<i>iso</i>	-0.39	0.000	***
	<i>aniso</i>	-0.20	0.000	***		<i>aniso</i>	-0.31	0.000	***
All syllables (syl)	<i>iso</i>	-0.26	0.000	***	High number (<i>hiN</i>)	<i>iso</i>	-0.25	0.000	***
	<i>aniso</i>	-0.31	0.000	***		<i>aniso</i>	-0.22	0.000	***
b. Mean frequency									
Stressed syllables (str)	<i>iso</i>	-0.03	0.382		Low number (<i>loN</i>)	<i>iso</i>	0.20	0.000	***
	<i>aniso</i>	0.03	0.439			<i>aniso</i>	0.18	0.000	***
All syllables (syl)	<i>iso</i>	0.15	0.000	***	High number (<i>hiN</i>)	<i>iso</i>	0.23	0.000	***
	<i>aniso</i>	0.25	0.000	***		<i>aniso</i>	0.24	0.000	***

Each cell displays the anchor point type, the polarity of the transformation, the Pearson's product moment correlation coefficient with its associated p-value, and a visual indication of signficativity.

4. DISCUSSION

In this study, we evaluated the effect on speech intelligibility of imposing isochronous timing on sentences presented in stationary noise. The idea that isochronous timed speech may be more intelligible in noise is based on the proposal that when speech information is degraded, an isochronous rhythm will make the tracking of important speech time instants, identified here as anchor points, more reliable.

Two approaches for implementing isochronous transformations were contrasted. The first used linguistically defined anchor points, namely, stressed syllables, and all syllables; the second used amplitude envelope peaks with the number of peaks matched to the frequency range of stressed syllables (low number of peaks: 4–5 peaks, approx. 2.5 Hz) and all syllables (high number of peaks: 7–8 peaks, approx. 4 Hz, see also Figure 2).

We found that sentences where the isochronous retiming used stressed syllables as anchor points were significantly more intelligible than those having a matched anisochronous

transformation, and we also observed a tendency for a benefit from the use of all syllables anchor points. In contrast, isochronous transformations based on the peaks in the speech amplitude envelope had no effect on intelligibility. Before discussing these differential intelligibility benefits, we will first consider the finding that any departure from the natural speech timing resulted in a decrease of intelligibility.

4.1. Intelligibility Decrease for Retimed Speech

The result that retimed speech was less intelligible than naturally timed speech in stationary noise accords with the results of a previous study of speech intelligibility using nonlinear retimed speech (Aubanel and Cooke, 2013). However, it does appear to be at odds with the idea that when the speech signal is degraded by noise, isochronous timing between the important speech regions should promote processing efficiency. That natural speech is best recognized can be explained when considering speech perception processes as broadly divided into two stages, as in the Tempo

model proposed by Ghitza (2011). In this approach, the first stage involves the registration and perception of the speech signal, the second stage involves the use of that information to access stored representations. Under this scheme, we suggest that in noise, isochronous syllable-based timing assists in the first stage, but that any advantage gained over natural speech is more than offset by the advantage that natural timing of speech has in the second stage. In what follows, we consider processing at each stage in turn.

In regards to the first stage of processing, it was found that compared to an anisochronous control, an imposed isochrony boosted intelligibility only when the retiming anchor points occurred at the onset of the stressed syllable. Here we propose that while the oscillatory tracking system is highly flexible, its ability to phase reset to the salient properties of ongoing speech breaks down under energetic masking conditions. In the absence of continuity cues, one interpretation is that cortical oscillations revert to a default state with an “idle” fixed frequency in the theta range typical when listening to speech. When the phase of this oscillatory mechanism aligns with regularly occurring stressed syllable onsets, increased processing efficiency results in intelligibility increase.

The finding that stressed syllable onsets are an important temporal cue in English speech perception fits with theories of speech processing that highlight the importance of the syllabic level (Greenberg et al., 2003; Luo and Poeppel, 2007; Giraud and Poeppel, 2012a; Peele and Davis, 2012; Ghitza, 2013). Indeed, converging results from different fields of research point to the importance of syllable onsets in speech perception, e.g., the onset timing of vowels is mostly preserved in spontaneous compared to laboratory speech (Greenberg, 1999), the perceptual beat associated with stressed syllable, or P-center, is located in the vicinity of the syllable onset and supports meter representation (Port, 2003). Given the results of the current study, one could hypothesize that cortical oscillations track P-centers, despite the fact that the latter do not strictly align with salient acoustic cues such as amplitude envelope peaks.

It is in regard to the second stage of processing for which we propose that the advantage for natural timed speech accrued. That is, naturally timed utterances have an advance at the recognition stage where spoken representations are accessed. This is because processing at this stage makes reference to the listener’s knowledge about timing statistics learned through exposure to spoken language, and used in production. This knowledge of regular timing patterns enables listeners to make precise and minute predictions about the ongoing speech stream (Pickering and Garrod, 2013). However, in the case where speech has been retimed (isochronously or not) the signal will not match these predictions and recognition will suffer. This notion that a mismatch between the input and an expected timing profile impairs recognition is supported by the current analysis showing that intelligibility correlates negatively with the amount of temporal distortion applied to the sentences.

It is the nature of the temporal distortion implemented here, with applies an alternation of compression and expansion within the same sentence, that may be most harmful for breaking timing expectations. Uniform transformations to speech timing are less

detrimental to recognition as the perceptual system is thought to adjust future predictions based on the ongoing speaking rate. As Dilley and Pitt (2010) showed, the identification of a speech target depends on the speaking rate of the preceding context, and a mismatch between the speaking rate of the target and preceding context leads to the reinterpretation of the target to match the preceding context.

It should be noted here that the timing statistics which together make up the speech rhythm of a particular language do not usually result in periodicity. Similarly, the literature does not posit a fixed frequency for oscillation but a frequency range. If periodicity were a necessary dimension of successful speech communication then talkers would use this form of speech. On an informational theoretic angle, microvariation in timing also allow the encoding of information—a perfectly regular “carrier”-like speech rhythm would be impoverished in information, and would limit suprasegmental encoding. We instead propose that periodicity could have a facilitatory effect that would be exploited in situation where the input is corrupted by noise, and the minute temporal adjustments to track the ongoing speech stream is impaired.

The idea that in noise, cortical sampling falls back to a default sampling period has interesting implications for the design of a follow-up study. In the current one, the period of isochrony was derived from the average occurrence of anchor points, which, for stressed syllables, ranged from 2.38 to 2.48 Hz (see Figure 2). Sentences were presented in increasing order of inter-anchor point frequency but no explicit indication of this frequency was given, nor was any explicit indication of the beginning of the first period provided. That is, no attempt was made to manipulate initial phase resetting or subsequent phase tracking and indeed, none of the participants reported periodicity in the stimuli. Given this, one could hypothesize that providing explicit cues for onset and periodicity may result in a greater intelligibility increase for isochronous timing, potentially due to increased alignment between brain oscillations and stimulus characteristics.

4.2. Relative Contribution of Syllabic Information vs. Amplitude Envelope to Cortical Tracking

In addition to finding support for isochrony anchored to the stressed syllable (Experiment I), an important finding was that the isochronous modification anchored to the peaks in the amplitude envelope did not improve intelligibility compared to anisochronous retiming (Experiment II). At a methodological level, this null result confirms the validity of the anisochronous modification (i.e., reversing the polarity of any retiming does not in itself degrade intelligibility) and provides additional support for the positive result obtained for the linguistically based modifications. Note that we do not directly compare the listeners’ performance across Experiment I and II as they used different cohorts of participants and had different intelligibility baselines for unmodified speech.

The null result in Experiment II is interesting in the light of the many studies that have proposed that the fluctuations of the amplitude envelope of speech play a crucial role in driving

cortical entrainment (Ahissar et al., 2001; Luo and Poeppel, 2007; Aiken and Picton, 2008; Giraud and Poeppel, 2012a). In this regard, we point out that our results may be specific to the timing of important information in speech recognition in noise and as such do not call into question the importance of the amplitude envelope for speech recognition in general.

Both syllable and amplitude envelope based transformations led to modifications to the timing of the amplitude envelope. But while the time instants of amplitude envelope peaks were derived from the controlled retiming of syllable onsets in the syllable based retiming conditions, they were directly manipulated in the amplitude envelope retiming conditions. Interestingly, this direct control of the timing of amplitude envelope peaks did not reveal an isochronous advantage, suggesting that if periodicity may facilitate processing through alleviating the need of phase resetting, then amplitude envelope peaks were not the appropriate cues for phase resetting. In fact, it would be more correct to conclude that regularly occurring amplitude peaks misinform upcoming predictions, at least as much as irregular ones.

More specifically, it seems more plausible that amplitude envelope peaks act as second-order temporal cues in signaling neighboring linguistically meaningful units, and that it is the latter that constitute the primary temporal cues for speech recognition, and the support for phase resetting. This result is all the more relevant considering that amplitude envelope peaks are more more audible than syllable onsets in the type of speech shape noise employed here, and therefore the periodicity should be more readily accessed by the oscillatory system if it would be purely acoustically driven. In all, the results of this study support an increasingly shared view that if oscillations track incoming speech, top-down linguistic cues may play a stronger role than bottom-up acoustic cues (Obleser et al., 2012; Gross et al., 2013; Zoefel and VanRullen, 2015b; Ding et al., 2016).

Amplitude envelope peaks are usually assumed to constitute the critical time instants for amplitude envelope tracking, for example, in Ghitza (2012), the location of audible pulses indicating rhythm are placed at amplitude peaks. Our results nuance that account, indicating that P-centers, which are perceptually defined and linguistically informed temporal cues, could constitute a more appropriate level of description of the cues that drive cortical entrainment, at least when speech is presented in noise. While we contrasted two types of cues in the

current study, other cues such as consonantal onsets or amplitude acceleration could provide further insight in the nature of the cues that support tracking.

English (including the Australian variety studied here) is traditionally considered as a stress-timed language as opposed to syllable- or mora-timed languages, although this distinction, which classically makes a hypothesis of constant duration of the corresponding units, has not been robustly verified empirically (Lehiste, 1977; Dauer, 1983; White and Mattys, 2007; Nolan and Jeon, 2014; Cummins, 2015). Nevertheless, in our study, isochronous retiming based on stressed syllable was the transformation resulting in the minimum temporal distortion, and also the condition that led to greater intelligibility compared to an all-syllable isochronous retiming. Apart from providing a weak support for stressed-syllable based isochrony in English, this result promotes the hypothesis that for syllable-timed languages, an opposite pattern of isochronous retiming benefit may be observed, with a greater benefit for isochronous retiming at the all-syllable level as opposed to higher metrical units.

A final point is that the current study employed behavioral measures of intelligibility to assess the effect of imposed periodicity in noise. A future line of research will be concerned with the evaluation of electrophysiological measures associated with this type of isochronous stimuli.

AUTHOR CONTRIBUTIONS

VA, CD, and JK designed research; VA performed research; VA analyzed data; VA, CD, and JK wrote the paper.

ACKNOWLEDGMENTS

The authors thank Lauren Smith for her help in organizing recruitment and testing and acknowledge support of the Australian Research Council under grant agreement DP130104447.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fnhum.2016.00430>

REFERENCES

- Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., and Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc. Natl. Acad. Sci. U.S.A.* 98, 13367–13372. doi: 10.1073/pnas.201400998
- Aiken, S. J., and Picton, T. W. (2008). Human cortical responses to the speech envelope. *Ear Hear.* 29, 139–157. doi: 10.1097/AUD.0b013e31816453dc
- Allen, G. D. (1972a). The location of rhythmic stress beats in English: an experimental study I. *Lang. Speech* 15, 72–100.
- Allen, G. D. (1972b). The location of rhythmic stress beats in English: an experimental study II. *Lang. Speech* 15:179.
- Allen, G. D. (1975). Speech rhythms: its relation to performance universals and articulatory timing. *J. Phon.* 3, 75–86.
- Aubanel, V., and Cooke, M. (2013). “Information-preserving temporal reallocation of speech in the presence of fluctuating maskers,” in *Proceeding of Interspeech* (Lyon), 3592–3596.
- Barczak, A., O’Connel, M. N., McGinnis, T., Ross, D., and Lakatos, P. (2015). “Entrainment of neuronal oscillations by repetitive complex auditory patterns,” in *Poster J44, Proceeding of Society for Neuroscience* (Chicago, IL).
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Besle, J., Schevon, C. A., Mehta, A. D., Lakatos, P., Goodman, R. R., McKhann, G. M., et al. (2011). Tuning of the human neocortex to the temporal dynamics of attended events. *J. Neurosci.* 31, 3176–3185. doi: 10.1523/JNEUROSCI.4518-10.2011
- Buzsaki, G., and Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science* 304, 1926–1929. doi: 10.1126/science.1099745

- Canolty, R. T., and Knight, R. T. (2010). The functional role of cross-frequency coupling. *Trends Cogn. Sci.* 14, 506–515. doi: 10.1126/science.1099745
- Chapin, H. L., Zanto, T., Jantzen, K. J., Kelso, S. J. A., Steinberg, F., and Large, E. W. (2010). Neural responses to complex auditory rhythms: the role of attending. *Front. Psychol.* 1:224. doi: 10.3389/fpsyg.2010.00224
- Cummins, F. (2012). Oscillators and syllables: a cautionary note. *Front. Psychol.* doi: 10.3389/fpsyg.2012.00364
- Cummins, F. (2015). "Rhythm and speech," in *The Handbook of Speech Production*, ed M. A. Redford (Chichester: John Wiley & Sons), 158–177.
- Cummins, F., and Port, R. F. (1998). Rhythmic constraints on stress timing in English. *J. Phon.* 26, 145–171. doi: 10.1006/jpho.1998.0070
- Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *J. Phon.* 11, 51–62.
- Demol, M., Verhelst, W., Struyve, K., and Verhoeve, P. (2005). "Efficient non-uniform time-scaling of speech with WSOLA," in *International Conference on Speech and Computers (SPECOM)* (Patras), 163–166.
- Dilley, L. C. and Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychol. Sci.* 21, 1664–1670. doi: 10.1177/0956797610384743
- Ding, N., Chatterjee, M., and Simon, J. Z. (2014). Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *Neuroimage* 88, 41–46. doi: 10.1016/j.neuroimage.2013.10.054
- Ding, N., Melloni, L., Zhang, H., Tian, X., and Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nat. Neurosci.* 19, 158–164. doi: 10.1038/nn.4186
- Ding, N., and Simon, J. Z. (2013). Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. *J. Neurosci.* 33, 5728–5735. doi: 10.1523/JNEUROSCI.5297-12.2013
- Ding, N., and Simon, J. Z. (2014). Cortical entrainment to continuous speech: functional roles and interpretations. *Front. Hum. Neurosci.* 8:311. doi: 10.3389/fnhum.2014.00311
- Doelling, K. B., Arnal, L. H., Ghitza, O., and Poeppel, D. (2014). Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage* 85, 761–768. doi: 10.1016/j.neuroimage.2013.06.035
- Ghitza, O. (2011). Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Front. Psychol.* 2:130. doi: 10.3389/fpsyg.2011.00130
- Ghitza, O. (2012). On the role of theta-driven syllabic parsing in decoding speech: intelligibility of speech with a manipulated modulation spectrum. *Front. Psychol.* 3:238. doi: 10.3389/fpsyg.2012.00238
- Ghitza, O. (2013). The theta-syllable: a unit of speech information defined by cortical function. *Front. Psychol.* 4:138. doi: 10.3389/fpsyg.2013.00138
- Ghitza, O., Giraud, A.-L., and Poeppel, D. (2013). Neuronal oscillations and speech perception: critical-band temporal envelopes are the essence. *Front. Hum. Neurosci.* 6:340. doi: 10.3389/fnhum.2012.00340
- Giraud, A.-L., and Poeppel, D. (2012a). Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* 15, 511–517. doi: 10.1038/nn.3063
- Giraud, A.-L., and Poeppel, D. (2012b). "Speech Perception from a Neurophysiological Perspective," in *The Human Auditory Cortex*, eds D. Poeppel, T. Overath, A. N. Popper, and R. R. Fay (New York, NY: Springer), 225–260.
- Greenberg, S. (1999). Speaking in shorthand—A syllable-centric perspective for understanding pronunciation variation. *Speech Commun.* 29, 159–176. doi: 10.1016/S0167-6393(99)00050-3
- Greenberg, S., Carvey, H., Hitchcock, L., and Chang, S. (2003). Temporal properties of spontaneous speech—a syllable-centric perspective. *J. Phon.* 31, 465–485. doi: 10.1016/j.wocn.2003.09.005
- Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., et al. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biol.* 11:e1001752. doi: 10.1371/journal.pbio.1001752
- Houtgast, T., and Steeneken, H. J. M. (1985). A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *J. Acoust. Soc. Am.* 77, 1069. doi: 10.1121/1.392224
- Hutcheon, B., and Yarom, Y. (2000). Resonance, oscillation and the intrinsic frequency preferences of neurons. *Trends Neurosci.* 23, 216–222. doi: 10.1016/S0166-2236(00)01547-2
- Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., and Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* 320, 110–113. doi: 10.1126/science.1154735
- Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., and Schroeder, C. E. (2005). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *J. Neurophysiol.* 94, 1904–1911. doi: 10.1152/jn.00263.2005
- Large, E. W., and Jones, M. R. (1999). The dynamics of attending: how people track time-varying events. *Psychol. Rev.* 106:119. doi: 10.1037/0033-295X.106.1.119
- Lehiste, I. (1977). Isochrony reconsidered. *J. Phon.* 5, 253–263.
- Luo, H., and Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54, 1001–1010. doi: 10.1016/j.neuron.2007.06.004
- Meunier, C., and Espesser, R. (2011). Vowel reduction in conversational speech in French: the role of lexical factors. *J. Phon.* 39, 271–278. doi: 10.1016/j.wocn.2010.11.008
- Morton, J., Marcus, S., and Frankish, C. (1976). Perceptual centers (P-centers). *Psychol. Rev.* 83:405.
- Nolan, F., and Jeon, H.-S. (2014). Speech rhythm: a metaphor? *Philos. Trans. R. Soc. B Biol. Sci.* 369:20130396. doi: 10.1098/rstb.2013.0396
- Nourski, K. V., Reale, R. A., Oya, H., Kawasaki, H., Kovach, C. K., Chen, H., et al. (2009). Temporal envelope of time-compressed speech represented in the human auditory cortex. *J. Neurosci.* 29, 15564–15574. doi: 10.1523/JNEUROSCI.3065-09.2009
- Obleser, J., Herrmann, B., and Henry, M. J. (2012). Neural oscillations in speech: don't be enslaved by the envelope. *Front. Hum. Neurosci.* 6:250. doi: 10.3389/fnhum.2012.00250
- Patel, A. D., Lofqvist, A., and Naito, W. (1999). "The acoustics and kinematics of regularly timed speech: a database and method for the study of the p-center problem," in *Proceeding of ICPhS* (San Francisco), 405–408.
- Peele, J. E., and Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Front. Psychol.* 3:320. doi: 10.3389/fpsyg.2012.00320
- Peele, J. E., Gross, J., and Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cereb. Cortex* 23, 1378–1387. doi: 10.1093/cercor/bhs118
- Pellegrino, F., Coupé, C., and Marsico, E. (2011). Across-language perspective on speech information rate. *Language* 87, 539–558. doi: 10.1353/lan.2011.0057
- Pickering, M. J., and Garrod, S. (2013). An integrated theory of language production and comprehension. *Behav. Brain Sci.* 36, 329–347. doi: 10.1017/S0140525X12001495
- Port, R. F. (2003). Meter and speech. *J. Phon.* 31, 599–611. doi: 10.1016/j.wocn.2003.08.001
- Rothauser, E. H., Chapman, W. D., Guttman, N., Hecker, M. H. L., Nordby, K. S., Silbiger, H. R., et al. (1969). IEEE recommended practice for speech quality measurements. *IEEE Trans. Audio Acoust.* 17, 225–246.
- Schroeder, C. E., and Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci.* 32, 9–18. doi: 10.1016/j.tins.2008.09.012
- Scott, S. K. (1993). *P-Centers in Speech: An Acoustic Analysis*. PhD thesis, University College London, London, UK.
- White, L., and Mattys, S. (2007). Calibrating rhythm: first language and second language studies. *J. Phon.* 35, 501–522. doi: 10.1016/j.wocn.2007.02.003
- Zoefel, B., and VanRullen, R. (2015a). Selective perceptual phase entrainment to speech rhythm in the absence of spectral energy fluctuations. *J. Neurosci.* 35, 1954–1964. doi: 10.1523/JNEUROSCI.3484-14.2015
- Zoefel, B., and VanRullen, R. (2015b). The role of high-level processes for oscillatory phase entrainment to speech sound. *Front. Hum. Neurosci.* 9:651. doi: 10.3389/fnhum.2015.00651
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Copyright © 2016 Aubanel, Davis and Kim. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.**



The Role of High-Level Processes for Oscillatory Phase Entrainment to Speech Sound

Benedikt Zoefel^{1,2*} and Rufin VanRullen^{1,2}

¹ Université Paul Sabatier, Toulouse, France, ² Centre de Recherche Cerveau et Cognition (CerCo), CNRS, UMR5549, Pavillon Baudot CHU Purpan, Toulouse, France

Constantly bombarded with input, the brain has the need to filter out relevant information while ignoring the irrelevant rest. A powerful tool may be represented by neural oscillations which entrain their high-excitability phase to important input while their low-excitability phase attenuates irrelevant information. Indeed, the alignment between brain oscillations and speech improves intelligibility and helps dissociating speakers during a “cocktail party”. Although well-investigated, the contribution of low- and high-level processes to phase entrainment to speech sound has only recently begun to be understood. Here, we review those findings, and concentrate on three main results: (1) Phase entrainment to speech sound is modulated by attention or predictions, likely supported by top-down signals and indicating higher-level processes involved in the brain’s adjustment to speech. (2) As phase entrainment to speech can be observed without systematic fluctuations in sound amplitude or spectral content, it does not only reflect a passive steady-state “ringing” of the cochlea, but entails a higher-level process. (3) The role of intelligibility for phase entrainment is debated. Recent results suggest that intelligibility modulates the behavioral consequences of entrainment, rather than directly affecting the strength of entrainment in auditory regions. We conclude that phase entrainment to speech reflects a sophisticated mechanism: several high-level processes interact to optimally align neural oscillations with predicted events of high relevance, even when they are hidden in a continuous stream of background noise.

OPEN ACCESS

Edited by:

Johanna Maria Rimmele,
University Medical Center
Hamburg-Eppendorf, Germany

Reviewed by:

Christian Keitel,
University of Glasgow, UK
Elana Zion Golumbic,
Bar Ilan University, Israel

*Correspondence:

Benedikt Zoefel
zoefel@cerco.ups-tlse.fr

Received: 30 July 2015

Accepted: 16 November 2015

Published: 02 December 2015

Citation:

Zoefel B and VanRullen R (2015) The Role of High-Level Processes for Oscillatory Phase Entrainment to Speech Sound. *Front. Hum. Neurosci.* 9:651.
doi: 10.3389/fnhum.2015.00651

Keywords: EEG, oscillation, phase, entrainment, high-level, speech, auditory, intelligibility

PHASE ENTRAINMENT AS A TOOL FOR INPUT GATING

In virtually every situation of our life, the brain has to cope with an enormous amount of incoming information, only a fraction of which is essential for the scene’s interpretation or resulting behavior. Clearly, the brain must have evolved strategies to deal with this vast influx, and both amplification of relevant input and suppression of irrelevant information will be critical for survival. Based on recent research, one prominent tool for the described purpose are neural oscillations, assumed to reflect cyclic changes in the excitability of groups of neurons (Buzsáki and Draguhn, 2004; Rajkai et al., 2008; Mazzoni et al., 2010). These *endogenous* fluctuations in neural excitability *per se* might seem without function at first glance, as long as they are *passive* and unrelated to the environment (**Figure 1A**). However, as previous studies showed, both on a theoretical (Schroeder et al., 2008, 2010;

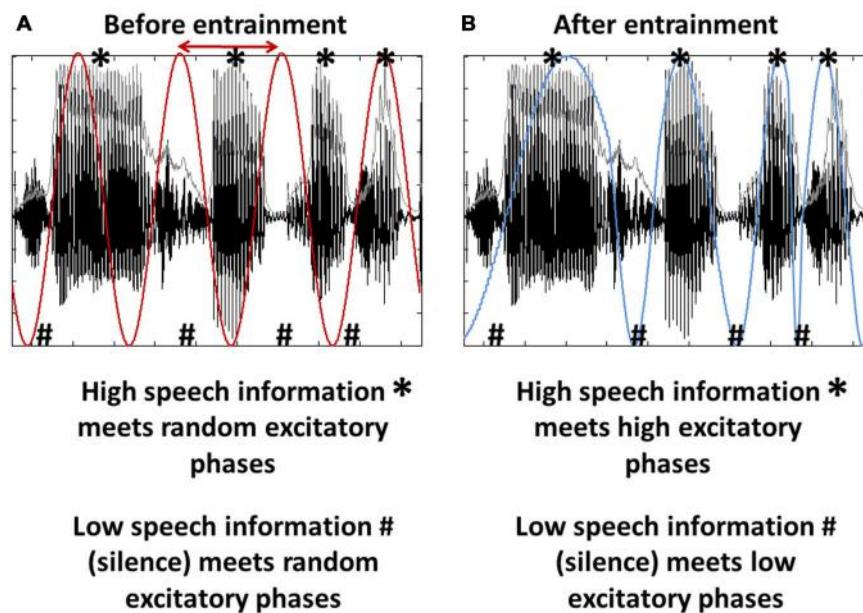


FIGURE 1 | Entrainment as a tool for input gating. **(A)** Brain oscillations (red) are unrelated to the stimulus input, here a segment of speech sound. Note that both oscillation and speech sound are rhythmic (~4 Hz) and that the speech input consists of phases of high (*) and low (#) informational content. Both phase and frequency (the latter to a certain extent; Ghitza, 2013, 2014) of the oscillations can be adjusted to match the input rhythm (red arrow), a phenomenon called phase entrainment. **(B)** Phase entrainment results in an alignment of the oscillation's high and low excitability phases (blue) with the input's high and low informational content. It can thus be used as a tool for input gating.

Schroeder and Lakatos, 2009; Ghitza, 2011; Morillon et al., 2015) and experimental level (Lakatos et al., 2005, 2008, 2013; Stefanics et al., 2010; Besle et al., 2011; Henry and Obleser, 2012; Henry et al., 2014; Morillon et al., 2014; Nozaradan, 2014; O’Connell et al., 2014; Arnal et al., 2015; Park et al., 2015), these oscillations might become an interesting tool when introducing the possibility that they can be *controlled* by the brain. By using the low and high excitability phases of those oscillations, the brain might *actively* “decide” what part of the incoming information should be amplified (the information coinciding with the oscillation’s high excitability phase) and what part should be suppressed (the information coinciding with the oscillation’s low excitability phase; **Figure 1B**). This phenomenon, the synchronization of an oscillatory system (here: brain oscillations) with external input has been termed *phase entrainment* (Schroeder and Lakatos, 2009). Of course, this kind of “input gating” can only be exploited functionally if the input is (1) rhythmic (i.e., predictable), (2) has a relatively stable frequency that the brain can entrain to, and (3) alternates between low and high informational content. Interestingly, one of the most salient stimuli in everyday life fulfills these criteria: speech sound. Although only considered “pseudo-rhythmic” (Cummins, 2012; but see Ghitza, 2013), the frequency of the speech envelope (roughly defined as the sum of energy across sound frequencies at a given point in time; shown as gray line in **Figure 1**) is relatively stable between 2 and 8 Hz and phases of low phonetic information (e.g., the silence between syllables) rhythmically alternate with phases of high phonetic information.

Indeed, the number of studies reporting an adaptation of neural oscillations to the envelope of speech sound is increasing continuously (Ding and Simon, 2012a,b, 2013, 2014; Peelle and Davis, 2012; Zion Golumbic et al., 2012, 2013b; Ding et al., 2013; Gross et al., 2013; Horton et al., 2013; Peelle et al., 2013; Power et al., 2013; Steinschneider et al., 2013; Doelling et al., 2014; Millman et al., 2015; Park et al., 2015). But not only speech sound is able to evoke an entrainment of neural oscillations, even simple stimuli, such as pure tones, have been found to produce phase entrainment (Stefanics et al., 2010; Besle et al., 2011; Gomez-Ramirez et al., 2011; Zoefel and Heil, 2013). Furthermore, rhythmic fluctuations in stimulus amplitude (which are present in both trains of pure tones and speech sound) introduce fluctuations at a level of auditory processing as low as the cochlea, a notion that is obviously not compatible with phase entrainment as an active or “high-level” process. Similar concerns have been raised by several authors in the last years (Obleser et al., 2012; Zion Golumbic et al., 2012; Ding et al., 2013; Peelle et al., 2013; Zoefel and Heil, 2013; Ding and Simon, 2014; VanRullen et al., 2014). Based on these concerns, it might be argued that a mere “following” of stimulus amplitude (leading to a series of evoked potentials) and the entrainment of endogenous neural oscillations might be completely different processes with different types of underlying mechanisms. Most studies investigating phase entrainment did not differentiate these components and might have measured a mix of evoked and entrained responses. For the sake of simplicity, and because it is not straightforward to disentangle the two, we will call both

processes “phase entrainment” throughout this manuscript, to describe an experimentally observable metric without assuming one or the other underlying process. However, we dedicated the last paragraph of Section “Phase Entrainment to High-Level Features of Speech Sound” to this issue, in which the controversy “evoked vs. entrained” is discussed in more detail.

The issues outlined in the previous paragraph lead to the fact that the role of high-level processes for phase entrainment to speech sound is far from clear. Nevertheless, significant progress has been made within the last decade, and the aim of this review is to summarize the obtained results in a systematic way. The scope of this review is not a summary of existing literature showing an alignment between brain oscillations and speech sound, as comprehensive reviews have been published recently (Peelle and Davis, 2012; Zion Golumbic et al., 2012; Ding and Simon, 2014). Rather, we will focus on high-level processes that can modulate or even underlie this alignment. Critically, it is necessary to differentiate between (i) high-level *modulations* of phase entrainment and (ii) high-level *entrainment*: In (i), phase entrainment can be produced as a “following” response to a low-level rhythmic stimulus sequence (potentially in early brain areas, as early as the cochlea); however, the entrainment is modulated by high-level processes that include attention or predictions. In this review, low-level features of speech are defined as stimulus amplitude and spectral content, as those two properties can passively entrain the lowest level of auditory processing and evoke steady-state-potential-like (ASSR; Galambos et al., 1981) fluctuations in the cochlea. In contrast to (i), high-level *entrainment* (ii) represents phase entrainment that can be observed even in the absence of systematic fluctuations of low-level properties. In this case, a simple “following” of stimulus amplitude is not possible anymore. Thus, it is the process of phase entrainment *itself* that operates on a higher level, as a certain level of processing is required in order to adjust to the rhythm of high-level features. Convincing results have been obtained in the last years for both types of high-level processes, and we will address them in separate sections. We conclude this review with a section dedicated to the role of intelligibility for phase entrainment to speech sound, as the influence of semantic information on the brain’s adjustment to speech is currently a strongly debated topic.

HIGH-LEVEL MODULATIONS OF PHASE ENTRAINMENT TO SPEECH SOUND

Certain cognitive processes, such as attention, expectation or interpretation, are often considered “high-level” functions of the human brain, as they require, for instance, evaluation, selection, and the comparison of the actual stimulation with experience (Lamme and Spekreijse, 2000; Gilbert and Li, 2013; Peelen and Kastner, 2014). A modulation of phase entrainment to speech sound by those cognitive processes would argue for phase entrainment being a process that is not restricted to a purely sensory mechanism, but rather the active gating mechanism (or “active sensing”; Schroeder et al., 2010) that was explained above.

Indeed, there is accumulating evidence for phase entrainment critically relying on attentional processes: one example is based on the so-called “cocktail party effect” (Cherry, 1953), describing a situation of several competing speakers, one of which has to be selected within the “noise” of the other, potentially distracting, speakers.

Several recent studies have shown a relation between the “cocktail party effect” and phase entrainment (the theoretical background is shown in **Figure 2A** and underlined by experimental results in **Figure 2B**). In Kerlin et al. (2010), two different speech streams were presented to the participants, one to each ear, and they were asked to selectively attend one of those two competing streams. They found that the representation of the attended speech stream in the delta/theta range (~2–8 Hz; the dominant frequency range of the speech envelope) of the electroencephalogram (EEG) signal was enhanced compared to that of the unattended stream. In other words, phase-locking between the EEG signal and the speech envelope of the attended stream was stronger than that between the EEG signal and the unattended stream. A similar paradigm was used in the studies by Ding and Simon (2012a), Horton et al. (2013) and Zion Golumbic et al. (2013b) in magnetoencephalographic (MEG), EEG and intracranial recordings in human subjects, respectively. All studies confirmed the finding that the phase of delta/theta brain oscillations “tracks” the envelope of speech sound, and that this “tracking” is enhanced when the speech is attended in a multi-speaker scenario. Interestingly, all studies reported that even the unattended speech signal is still represented (albeit weakly) in lower-level auditory cortices (i.e., regions closely related to sensory processing). However, as shown in the work by Zion Golumbic et al. (2013b), this unattended signal is “lost” in higher-level (e.g., frontal) regions. Ding and Simon (2012a) demonstrated that only the representation of the attended (and not the unattended) speech envelope varies as a function of stimulus intensity. This finding is important, because it suggests that attended and unattended inputs are processed separately in the brain, and that the alignment between neural phase and speech rhythm is used to form individual “auditory objects” (for a review on this notion, see Simon, 2015). In line with the notion of phase entrainment as an “amplifier-attenuator mechanism” (see “Phase Entrainment as a Tool for Input Gating”), Horton et al. (2013) reported cross-correlations between speech envelope and EEG signal for both attended and unattended streams, but with opposite signs, suggesting that phase entrainment is indeed used to amplify one stream while the other is attenuated. Finally, it has been shown in several studies that the speech envelope can be reconstructed (i.e., it can be identified which stimulus the listener is attending) in multi-speaker (Ding and Simon, 2012a; Zion Golumbic et al., 2013b; O’Sullivan et al., 2015) or noisy environments (Ding and Simon, 2013) by using the delta/theta phase of neural oscillations (but also their gamma power; Mesgarani and Chang, 2012; Zion Golumbic et al., 2013b). It is possible that in those kind of situations, where one speech stream has to be actively extracted from a noisy environment, attention is of particular importance for phase entrainment to speech sound, whereas clear speech can

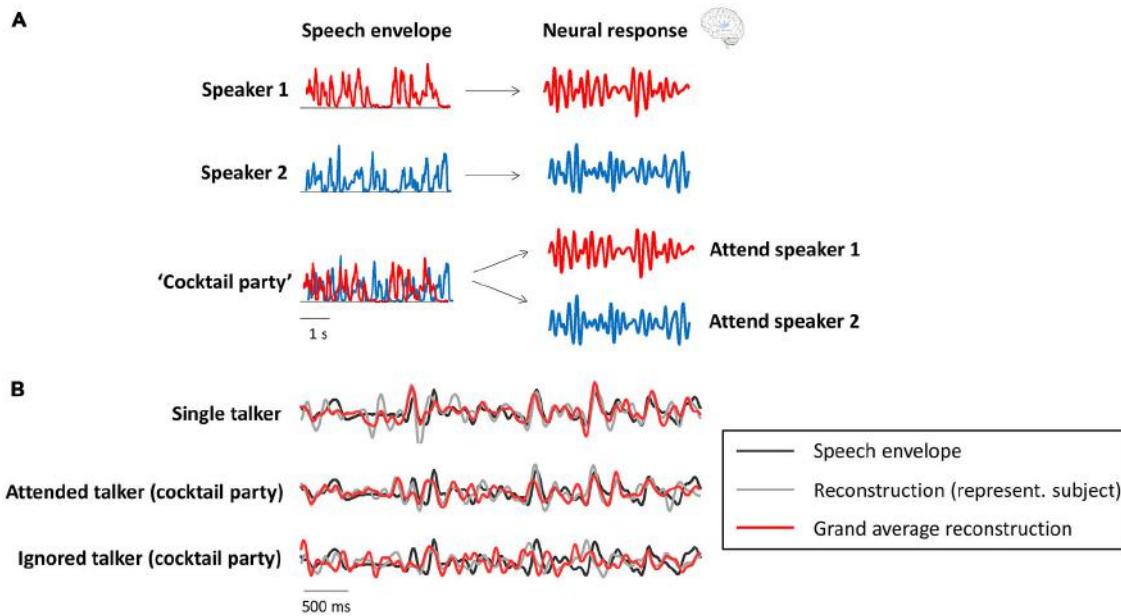


FIGURE 2 | Neural oscillations as a tool for attentional selection during a “cocktail party”. **(A)** Theoretical background (modified with permission from Zion Golumbic et al., 2012, copyright Elsevier). Recorded neural activity in the delta/theta band (right column) aligns with the speech envelope (left column) of the respective speaker (red and blue), when presented separately. In a multi-speaker scenario (“cocktail party”), the recorded data will reflect the attended, but not necessarily (or to a smaller degree) the unattended speech envelope. **(B)** Actual data (modified with permission from Zion Golumbic et al., 2013b, copyright Elsevier) confirms the theoretical background. The speech envelope reconstructed from the recorded data (gray: single subject; red: averaged across subjects) strongly resembles the speech envelope (black) of the attended, but not the unattended speaker.

be processed largely independently of attention (Wild et al., 2012).

Not only attention can be considered a high-level process: predictions reflect a comparison between present and previous experiences and its projection to the future and must therefore involve high-level functions of the brain (Friston, 2005; Arnal and Giraud, 2012). Indeed, it has been shown that predictions do influence phase entrainment to speech sound. For instance, in the “cocktail party” scenario described above, Zion Golumbic et al. (2013a) paired the auditory speech input with the speaker’s face and found that phase entrainment to the speech envelope was significantly enhanced by this visual input. Similar results were obtained by Arnal et al. (2011) using congruent and incongruent audiovisual stimuli (syllables) and by Luo et al. (2010) when subjects were watching audiovisual movies. A common interpretation of these findings is that, due to the slight delay between visual and auditory components of a conversation (the visual input preceding the auditory one), the former can be used to predict the timing of speech sound, thus enabling a better alignment between the oscillatory phase and speech envelope (Arnal et al., 2009, 2011; Zion Golumbic et al., 2013a; Perrodin et al., 2015; for a review, summarizing several existing theories, see Peelle and Sommers, 2015). A phase-reset of neural oscillations in primary auditory cortex by visual input seems to be an important underlying mechanism (Thorne and Debener, 2014; Mercier et al., 2015; Perrodin et al., 2015). Although this would indicate an involvement of low or intermediate hierarchical levels, we emphasize here that

a purely low-level mechanism is insufficient to explain many findings reported in the literature. For instance, introducing an additional delay between visual and auditory input disrupts the benefits of additional visual information for speech processing and incongruent visual information (which would result in a similar phase-reset as congruent information, assuming a purely low-level process) does not result in enhanced phase entrainment (e.g., Crosse et al., 2015; for a review, see Peelle and Sommers, 2015) but instead generates an increased neural response associated with error processing (Arnal et al., 2011). Finally, using a McGurk paradigm (McGurk and MacDonald, 1976; van Wassenhove et al., 2005) were able to show a correlation between the amount of prediction conveyed by the preceding visual input for the upcoming speech and the latency of speech processing. Together, these results speak for a mechanism that is tailored to speech-specific processing (Crosse et al., 2015) and against a purely low-level mechanism. The timing of the cross-modal phase-reset seems to have evolved in such a way that oscillations in the auditory system arrive at their high excitability phase exactly when the relevant auditory input is expected to be processed (Lakatos et al., 2009; Thorne and Debener, 2014). Finally, recent research suggests that not only the visual, but also the motor system plays a critical role for an efficient adjustment of excitability fluctuations in auditory cortex to expected upcoming events (Fujioka et al., 2012; Doelling et al., 2014; Morillon and Schroeder, 2015; Morillon et al., 2015). For instance, it has been suggested that the motor system possesses its own representation of expected auditory

events and can therefore prepare oscillations in auditory cortex for relevant upcoming stimuli (Arnal and Giraud, 2012; Arnal, 2012). This mechanism might underlie recent findings describing an enhanced segregation of relevant and irrelevant auditory events in the presence of rhythmic tapping (Morillon et al., 2014).

Not an experimental, but rather an analytical proof of high-level processes involved in phase entrainment was provided by two recent studies (Fontolan et al., 2014; Park et al., 2015). Fontolan et al. (2014) used Granger causality (Granger, 1969), applied on data recorded intracranially in human subjects, to demonstrate that information reflected in the phase of low-frequency oscillations in response to speech sound travels in top-down direction from higher-order auditory to primary auditory regions, where it modulates the power of (gamma) oscillations at higher frequencies. Park et al. (2015) analyzed their data, recorded with MEG, using transfer entropy measures (Schreiber, 2000). They were able to show that frontal and motor areas can modulate the phase of delta/theta oscillations in auditory cortex (note that the spatial resolution in this study was lower than for intracranial recordings). It is thus unclear whether these delta/theta oscillations correspond to those in higher-order auditory or primary auditory cortices described in Fontolan et al., 2014). Importantly, these top-down signals were correlated with an enhanced phase entrainment to speech sound when tracking of forward vs. backward speech was compared, indicating that higher-level processes can directly control the alignment between neural oscillations and speech sound.

The results described in this section strongly support the view that phase entrainment is a tool for attentional selection (Schroeder and Lakatos, 2009), filtering out irrelevant input and enhancing the representation of the attended stimulus in the brain. Predictions, potentially reflected by top-down mechanisms, help “designing” this filter by providing the timing for the alignment of “good” and “bad” phases of the oscillation to predicted relevant and irrelevant stimuli, respectively. This mechanism would not only help selecting relevant input in a noisy background, but also parse the speech signal at the same time: here, one cycle of the aligned oscillation would represent one segment of information (or “chunk”; Ghitza, 2011, 2013, 2014; Doelling et al., 2014) that is analyzed by means of faster oscillations (Giraud and Poeppel, 2012; Luo and Poeppel, 2012; for reviews, see Peelle and Davis, 2012; Ding and Simon, 2014). Thus, phase entrainment could function as a means of discretization (equivalent ideas are mentioned by Peelle and Davis, 2012; Zion Golumbic et al., 2012), similar to “perceptual cycles” commonly observed in vision (VanRullen et al., 2014).

PHASE ENTRAINMENT TO HIGH-LEVEL FEATURES OF SPEECH SOUND

In the previous section, we have seen that high-level mechanisms of the brain, related to attention or prediction, clearly contribute to phase entrainment to speech sound. However, it should be noted that this contribution may just be *modulatory*: high-level mechanisms could merely *influence* a process, namely

phase entrainment, that *itself* might rely on purely low-level processes. Indeed, speech sound consists of large fluctuations in low-level properties (i.e., stimulus amplitude and spectral content) that might evoke systematic fluctuations in neural activity already at the earliest level of auditory processing: the cochlea. These fluctuations in neural activity accompanying changes in the speech envelope would be indistinguishable from an active entrainment response. It is therefore necessary to construct stimuli without systematic fluctuations in those low-level properties in order to prove genuine high-level entrainment. In a recent publication (Zoefel and VanRullen, 2015b), we were able to construct such stimuli and we review the most important findings in this section, together with supporting results from other studies. **Figure 3** shows the idea underlying stimulus construction in Zoefel and VanRullen (2015b). In everyday speech sound (**Figure 3A**), spectral energy (color-coded) clearly differs between different phases of the speech envelope. In the view of a single cochlear cell, this sound would periodically alternate between weak (e.g., at phase $\pm \pi$, which is the trough of the speech envelope) and strong excitation (e.g., at phase 0, which is the peak of the speech envelope). Consequently, at a larger scale, we would measure an oscillatory pattern of neural activity that strongly depends on envelope phase. This pattern, however, would only reflect the periodicity of the stimulation. Therefore, we constructed noise sound whose spectral energy was tailored to counterbalance spectral differences as a function of envelope phase of the original speech sound (for details of stimulus construction, see Zoefel and VanRullen, 2015b). This noise was mixed with the original speech and resulted in speech/noise sound that did, on average, not show those systematic differences in spectral content anymore (**Figure 3B**). Critically, as those stimuli remain intelligible, high-level features of speech (such as, but not restricted to, phonetic information) are still present and enable the listener to entrain to the speech sound that is now “hidden” inside the noise (note that the degree to which the speech is “hidden” in noise depends on the original envelope phase, with speech perceptually dominant at the original envelope peak, and noise perceptually dominant at the original envelope trough). We applied those stimuli in two studies: in the first (Zoefel and VanRullen, 2015b), a psychophysical study, we found that the detection of a short tone pip was significantly modulated (*p*-values shown in **Figure 4A**) by the remaining high-level features. Performance (**Figure 4B**) depended on the original envelope phase and thus differed between periods of dominant speech and noise. Note that speech and noise were spectrally matched; differences in performance could thus not be due to spectral differences between speech and noise, but rather due to the remaining high-level features that enable the listener to differentiate speech and noise. In the second study (Zoefel and VanRullen, 2015a), those stimuli were presented to listeners while their EEG was recorded. We found that EEG oscillations phase-lock to those high-level features of speech sound (**Figure 4C**), and the degree of entrainment (but not the phase relation between speech and EEG signal; see insets in **Figure 4C**) was similar to when the original everyday

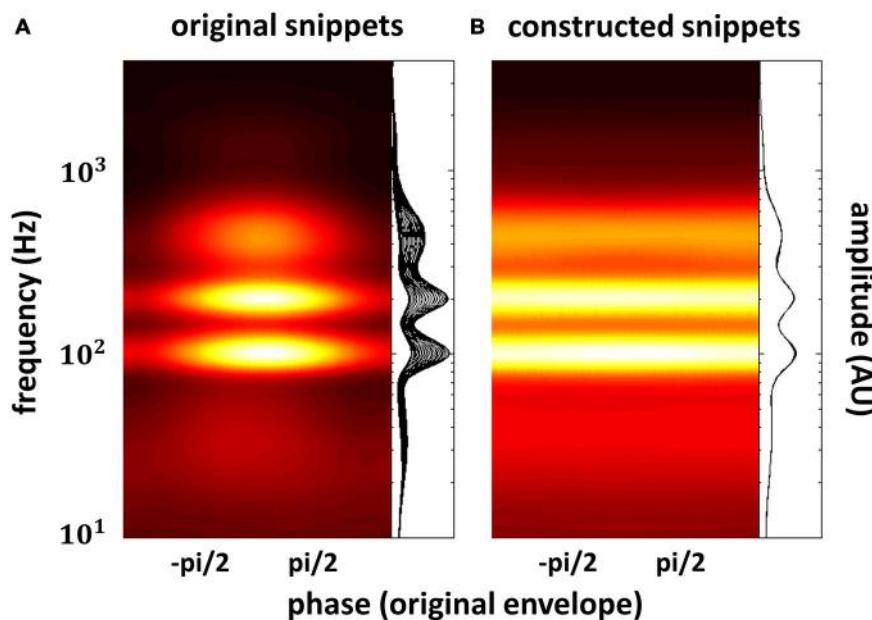


FIGURE 3 | Everyday speech sound (A) contains pronounced fluctuations in spectral energy (color-coded or shown as a family of curves; one curve for each phase bin of the speech envelope) that depend on the phase of the speech envelope. These (low-level) rhythmic fluctuations in energy *per se* might result in an apparent alignment between neural activity and speech envelope, as strong neural excitation (here at phase 0, due to high spectral energy) periodically alternates with weak neural excitation (here at phase $\pm \pi$, due to low spectral energy). Genuine high-level phase entrainment requires stimuli without those systematic fluctuations in spectral energy, as shown in (B). The construction of those stimuli has recently been reported (Zoefel and VanRullen, 2015b), and results obtained there are described in this review. Reproduced with permission from Zoefel and VanRullen (2015b).

speech was presented. These results suggest an entrainment of neural oscillations as the mechanism underlying our perceptual findings.

It is not only interesting to investigate phase entrainment to speech stimuli *without* potentially entraining low-level features, but also to speech stimuli *only* containing the latter. This was done in a study by Ding et al. (2013) that might be seen as complementary to the other two described in this section. In their study, noise-vocoding (Green et al., 2002) was used in order to design stimuli where spectro-temporal fine structure (which can be considered as high-level features) was strongly reduced, but the speech envelope was essentially unchanged. Those stimuli were presented either in noise or in quiet, and MEG was recorded in parallel. Ding et al. (2013) showed that, indeed, the reduction of spectro-temporal fine structure in noise-vocoded speech results in a decline in phase entrainment as compared to that in response to natural speech sound. This result suggests that oscillations do not merely (and passively) follow the slow fluctuations in low-level features of speech (e.g., the speech envelope), as they are present in both natural and noise-vocoded speech. Instead, phase entrainment to speech sound involves an additional adjustment to rhythmic changes in spectro-temporal fine structure. It is important to mention that the effect was only observed in noise (and not in quiet), stressing the idea that separating speech and noise might be one of the main functions of phase entrainment to speech sound (see “Phase Entrainment as a Tool for Input Gating” and “High-Level Modulations of

Phase Entrainment to Speech Sound”). Using similar stimuli as in Ding et al. (2013), Rimmele et al. (2015) both extended their findings and built a bridge to our section “High-Level Modulations of Phase Entrainment to Speech Sound”. In contrast to Ding et al. (2013), they presented natural and noise-vocoded speech concurrently and asked their subjects to attend one of them while ignoring the other. Interestingly, they were able to show that the enhanced “envelope tracking” for natural compared to noise-vocoded speech (as in Ding et al., 2013) is only present when the speech is attended. They interpret their results as evidence for a high-level mechanism (“linguistic processing”) that is only possible when speech is in the focus of the listener’s attention, and only when speech contains spectro-temporal fine structure (i.e., high-level features). Finally, no attentional modulation of phase entrainment was found for noise-vocoded speech which might be taken as evidence for a tracking of low-level features that does not depend on top-down processes (e.g., attention; see “High-Level Modulations of Phase Entrainment to Speech Sound”).

Taken together, the results reported in this section suggest that phase entrainment to speech sound is not only a reflection of fluctuations in low-level features of speech sound, but entails an adaption to phonetic information—and thus a genuine high-level process.

As briefly mentioned before, there is an ongoing debate which is directly related to the results presented in this section: as shown by Capilla et al. (2011), seemingly entrained

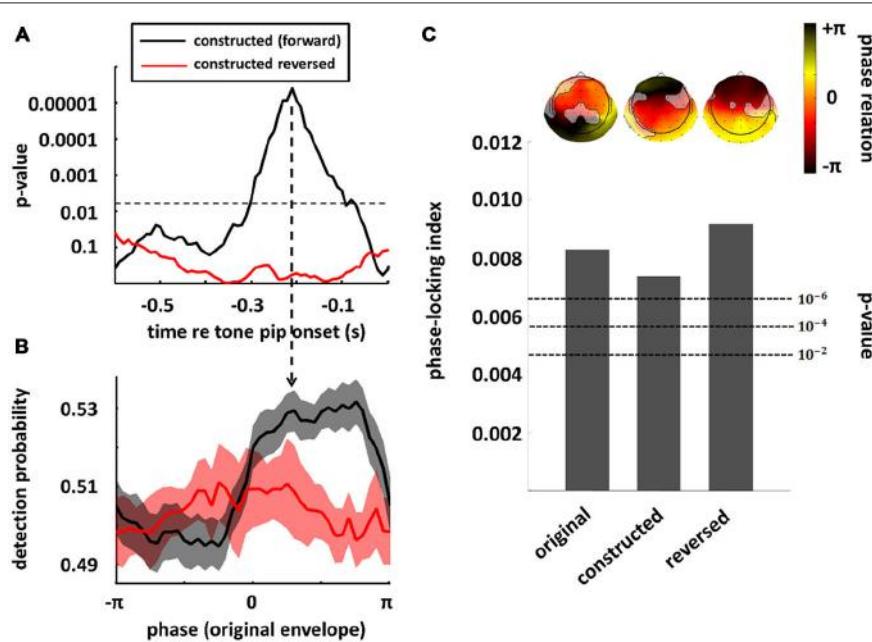


FIGURE 4 | Perception and neural oscillations entrain to high-level features of speech sound. Speech/noise stimuli without systematic fluctuations in amplitude or spectral content were constructed, whose high-level features are conserved and reflected by the original speech envelope (cf. **Figure 3**). In a psychophysical study (**A,B**), the detection of a tone pip was significantly modulated by those high-level features (black; in this plot, results obtained in the original experiment and a follow-up replication have been combined; both experiments are described in Zoefel and VanRullen, 2015b). The significance of this modulation is shown in (**A**) for different time lags relative to target onset, whereas the actual performance (at the time lag indicated by the vertical arrow in (**A**) is shown in panel (**B**). This effect was abolished when the speech/noise sound was reversed (red), indicating an important role of linguistic features (i.e., intelligibility) for behavioral consequences of the entrainment. In (**A**), the significance threshold is shown as a dashed line (corrected for multiple comparisons). In (**B**), standard error of mean (SEM) is shown by contours around the lines. When the same stimuli (and their original version of everyday speech) were presented in an EEG experiment (**C**) significant phase-locking between original speech envelope and EEG signal could be observed in all conditions (original, speech/noise sound and reversed speech/noise sound), suggesting that high-level features can entrain the phase of EEG oscillations, and do so even if the speech is unintelligible (note that acoustic high-level features remain present in the speech/noise sound, even when it is reversed, as the listener can still differentiate speech and noise). Bars show the average phase-locking across EEG channels, whereas the actual phase differences between EEG signal and original speech envelope, separately for each channel, are shown as insets above the bars (channels without significant entrainment are shaded out). *P*-values of phase entrainment, obtained by permutation tests, are shown as dashed lines. Note that, in contrast to the degree of entrainment which is comparable in all three conditions, the entrained phase does differ between everyday speech sound (original condition) and speech/noise sound in which systematic fluctuations in low-level features have been removed (constructed and constructed reversed conditions). Modified with permission from Zoefel and VanRullen (2015b) (**A,B**) and Zoefel and VanRullen (2015a), copyright Elsevier (**C**).

oscillations can be explained by a superposition of evoked responses (see also Keitel et al., 2014). Transferring this result to speech sound, it has been argued specifically that a phase-reset of neural oscillations by (e.g.) “acoustic edges” of speech might be an important mechanism underlying phase entrainment (Doelling et al., 2014; Howard and Poeppel, 2010)—assuming that these “edges” occur regularly in speech, a periodic sequence of phase-resets might thus be sufficient to explain the observed “phase entrainment”. This paragraph provides arguments against phase entrainment reflecting a purely passive mechanism, reflecting merely sequences of neural phase-resets or evoked potentials; however, we do emphasize here that most studies likely measure a mixture of evoked and entrained neural responses. As already outlined above, the two studies described in the first paragraph of this section (Zoefel and VanRullen, 2015a,b) support the notion that phase entrainment is more than a steady-state response to rhythmic stimulation: it can be observed even when the presented speech sound does not contain systematic fluctuations in amplitude or spectral content. Indeed, there are more

studies, using simpler, non-speech stimuli, that also support this conclusion. For instance, it has been found, for both vision (Spaak et al., 2014) and audition (Hickok et al., 2015), that behavioral performance fluctuates for several cycles after the offset of an entraining stimulus. A mere “following” of the stimulation would not produce these after-effects. Moreover, using entraining stimuli at threshold level, it has been shown that neural oscillations (as measured with EEG) entrain to the stimulation rate even when the stimulus is not perceived (e.g., in the case of several subsequent “misses”) and would therefore not evoke a strong neural response (Zoefel and Heil, 2013). Finally, in a clever experimental design, Herring et al. (2015) measured visual alpha oscillations (~8–12 Hz) after a single pulse of transcranial magnetic stimulation (TMS) that has previously been hypothesized to re-set (or entrain, in the case of multiple, rhythmic TMS pulses) endogenous oscillations (Thut et al., 2011). They then asked the question: how is the measured “alpha” modulated by attention? In the case of a simple evoked response (or “alpha-ringing”), the observed “alpha” would exhibit an increased amplitude when

attention is allocated to the visual domain; however, in the case of endogenous alpha, visual attention would decrease the alpha amplitude, as described already by Adrian (1944). Indeed, the latter is what Herring et al. (2015) observed. To conclude, although the issue remains open, there are promising first results suggesting that phase entrainment—to speech or other stimuli, including brain stimulation—is more than steady-state responses evoked by the rhythmic stimulation—it entails high-level processes and an adjustment of endogenous neural oscillations.

THE ROLE OF INTELLIGIBILITY FOR PHASE ENTRAINMENT TO SPEECH SOUND

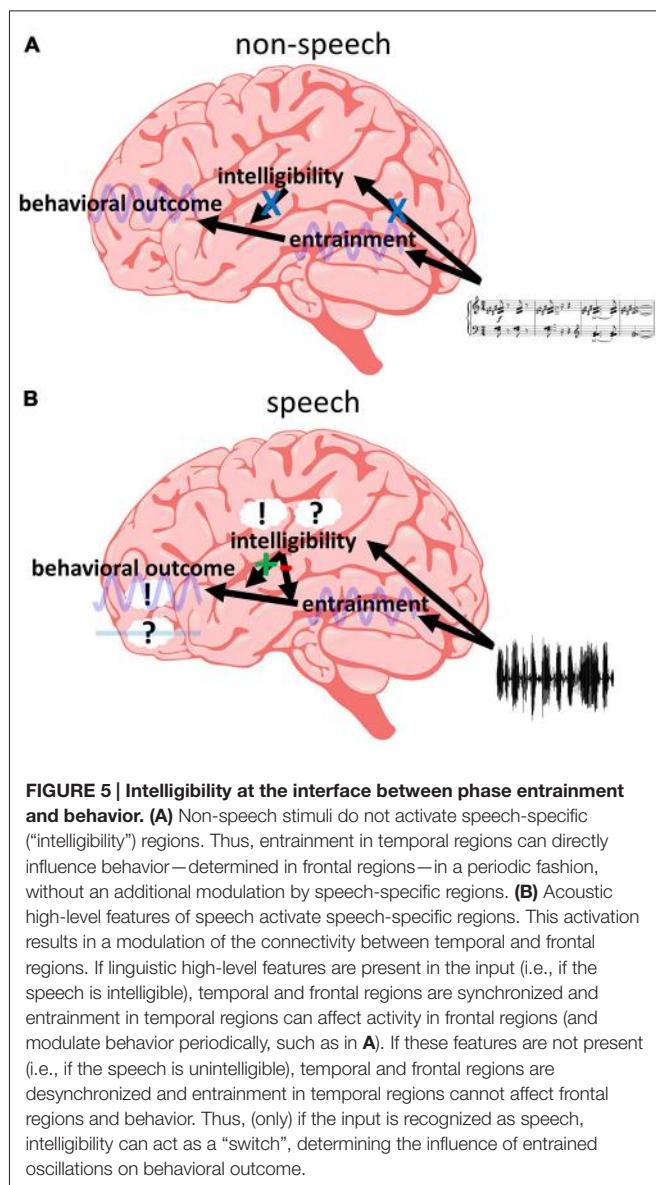
Of course, the ultimate goal of every conversation is to transmit information, and without intelligibility, this goal cannot be achieved. Thus, it is all the more surprising that the role of intelligibility for phase entrainment to speech is currently strongly debated. This controversy is due to seemingly contradictory results that have been published. On the one hand, both Ahissar et al. (2001) and Luo and Poeppel (2007) found a correlation between phase entrainment (i.e., alignment of delta/theta oscillations and speech envelope) and speech intelligibility, a finding that has been confirmed by recent studies (Ding et al., 2013; Doelling et al., 2014; Park et al., 2015). On the other hand, phase entrainment is not a phenomenon that is unique to speech sound and can also be found in response to much simpler stimuli, such as pure tones (Lakatos et al., 2005, 2008; Stefanics et al., 2010; Besle et al., 2011; Gomez-Ramirez et al., 2011; Zoefel and Heil, 2013). Also, the manipulation of speech intelligibility might destroy acoustic (i.e., non-semantic) properties of the sound that the brain actually entrains to (such as acoustic “edges”; Doelling et al., 2014), leading to a decline in phase entrainment and speech intelligibility at the same time, but without any relation between the two (Peelle and Davis, 2012; Millman et al., 2015). Moreover, several studies showed phase entrainment of neural oscillations to unintelligible speech sound (Howard and Poeppel, 2010; Peelle et al., 2013; Millman et al., 2015) suggesting that phase entrainment does not necessarily depend on intelligibility. The whole picture gets even more complicated, as, although phase entrainment to speech sound is possible even when the speech is unintelligible, it seems to be enhanced by intelligible speech in some (but not all) studies (Gross et al., 2013; Peelle et al., 2013; Park et al., 2015) and attention seems to be important for this enhancement (Rimmele et al., 2015). Further evidence that the role of intelligibility for phase entrainment is not trivial was reported in two of the studies described in the previous section. In Zoefel and VanRullen (2015b), it was found that perceptual entrainment to high-level features of speech sound is disrupted when the speech/noise sound is reversed (**Figures 4A,B**; red line) and this result was interpreted as a critical role of intelligibility for perceptual phase entrainment. On the other hand, in Zoefel and VanRullen (2015a), using the same reversed speech/noise stimuli, the observed EEG phase entrainment was similar to

that obtained in response to everyday speech and to (forward) speech/noise sound (**Figure 4C**), seemingly in contradiction to the behavioral results obtained in Zoefel and VanRullen (2015b).

How can we reconcile these studies, some of them clearly arguing against, and some for an important role of intelligibility for phase entrainment? Based on the current state of research, it is important to avoid overhasty conclusions and our interpretations have to remain speculative. Overall, phase entrainment seems to be a necessary, but not sufficient condition for speech comprehension. Speech intelligibility might not be possible without phase-locking, as we are not aware of any study reporting intelligible stimuli without oscillations (or perception) aligned to critical (low- and high-level) features of the speech sound. On the other hand, neural oscillations entrain to rhythmic structures (including reversed speech) even in the absence of intelligibility. It is clear that phase entrainment is a much more general phenomenon, and the brain might continuously scan its input for rhythmic patterns (indeed, popularity for auditory rhythms can be found in all cultures across the world and synchronization with rhythms—e.g., by clapping or dancing—is a general reaction to them). Once a rhythmic pattern has been detected, neural oscillations will align their phase to it (operating in the “rhythmic mode” described in Schroeder and Lakatos, 2009; see also Zoefel and Heil, 2013). Based on this notion, neural oscillations might always align to sound, as long as a rhythmic pattern can be detected (note that even the reversed speech/noise sound used in Zoefel and VanRullen, 2015a,b, contains a rhythmic pattern, as speech and noise can perceptually be differentiated). But what is the role of intelligibility? It is important to find a model that is at the same time parsimonious and can explain most results described in the literature. These findings are shortly summarized in the following:

1. Rhythmic non-speech stimuli, such as trains of pure tones, entrain neural oscillations (Lakatos et al., 2005; Besle et al., 2011; Gomez-Ramirez et al., 2011; Zoefel and Heil, 2013) and modulate behavior (Lakatos et al., 2008; Stefanics et al., 2010; Thut et al., 2012; Hickok et al., 2015).
2. Speech stimuli, both intelligible and unintelligible, entrain neural oscillations (Ahissar et al., 2001; Luo and Poeppel, 2007; Howard and Poeppel, 2010; Ding and Simon, 2012a; Ding et al., 2013; Peelle et al., 2013; Zion Golumbic et al., 2013b; Doelling et al., 2014; Millman et al., 2015; Park et al., 2015; Rimmele et al., 2015; Zoefel and VanRullen, 2015a).
3. The rhythm of speech only modulates behavior when speech is intelligible (Zoefel and VanRullen, 2015b).
4. Neural entrainment to intelligible speech might be increased when compared to unintelligible speech (Luo and Poeppel, 2007; Peelle et al., 2013; Doelling et al., 2014; Park et al., 2015; Rimmele et al., 2015). However, not all studies can confirm this result (Howard and Poeppel, 2010; Millman et al., 2015; Zoefel and VanRullen, 2015a).

One model that can potentially reconcile these findings is presented in **Figure 5**, and the different parts and implications of this model are discussed in the following. However, we



acknowledge that it is only one out of possibly several candidate models to explain the data available in the literature. Nevertheless, in our view, this model is currently the most parsimonious explanation for existing findings and we therefore focus our review on it. The first implication of our model is that different regions in the brain are “responsible” for different processes: Phase entrainment might be found throughout the whole auditory system, but most studies emphasize primary auditory cortex (A1; Lakatos et al., 2005, 2013; O’Connell et al., 2014) or early temporal regions (Gomez-Ramirez et al., 2011; Ding and Simon, 2012b; Zion Golumbic et al., 2013b). An influence of intelligibility is commonly related to regions specifically processing speech sound (Binder et al., 2000; Scott et al., 2000; Hickok and Poeppel, 2007; DeWitt and Rauschecker, 2012; Poeppel et al., 2012; Mesgarani et al., 2014). Finally, frontal regions are a likely candidate for behavioral outcome (Krawczyk, 2002; Coutlee and Huettel, 2012; Rushworth et al.,

2012; Romo and de Lafuente, 2013). In order to satisfy point (1), we assume that the entrainment in temporal regions can directly influence behavior as determined in frontal regions, as long as the entrainment is introduced by non-speech stimuli (**Figure 5A**). This results in a periodic modulation of performance as often described (Fiebelkorn et al., 2011; Vanrullen and Dubois, 2011; Landau and Fries, 2012; Thut et al., 2012; Song et al., 2014; Spaak et al., 2014; Zoefel and Sokoliuk, 2014; Hickok et al., 2015; note, however, that most studies report effects for the visual and not for the auditory system—it needs to be clarified whether this fact is biased by the number of studies investigating the visual system or whether there are genuine differences between the two systems). But not only non-speech stimuli can entrain temporal regions, the same is true for speech sound, irrespective of its intelligibility (point 2). However, speech intelligibility affects high-order auditory regions and they might directly influence the impact of temporal on frontal regions (**Figure 5B**). This notion is based on the increasing number of studies supporting the idea that the state of connectivity (or synchronization) between two (potentially distant) brain regions is crucial for perceptual outcome (Fries, 2005; Ruhnau et al., 2014; Weisz et al., 2014). Thus, speech intelligibility might modulate the state of connectivity between temporal and frontal regions. We hypothesize that speech-specific regions only become responsive if the input contains *acoustic* high-level (i.e., speech-specific) features of speech; otherwise these regions remain irrelevant and do not exhibit any modulatory effect on other regions or their connectivity. However, once the input is identified as speech (based on these acoustic features), *linguistic* features determine whether the modulatory effect is negative (desynchronizing temporal and frontal regions, resulting in no behavioral effect of the entrainment; in case of unintelligible speech) or positive (synchronizing temporal and frontal regions, resulting in a behavioral effect of the entrainment; in case of intelligible speech). This assumption satisfies point (3). In contrast to unintelligible speech, intelligible speech might result in an entrainment that also includes high-order (speech-specific) auditory regions. They might have to entrain to the speech sound in order to be able to synchronize temporal and frontal regions. That might be the reason that some studies show an increased entrainment for intelligible as compared to unintelligible speech whereas others do not (point 4). They might have captured the entrainment in those higher-level auditory regions—something which, due to the low spatial resolution in most EEG/MEG studies, is difficult to determine but could be resolved in future studies. More research is clearly needed: what are those behavioral variables that are differentially affected by intelligible and unintelligible speech? Where exactly are those brain regions hypothesized to be responsible for (or affected by) phase entrainment, for behavioral decisions and for the modulation of their relation by speech intelligibility? What are the mechanisms connecting these functional networks? Answering these questions has critical implications for our understanding of the brain’s processing of human speech and rhythmic input in general.

CONCLUSION

Recently, phase entrainment has attracted researchers' attention as a potential reflection of the brain's mechanism to efficiently allocate attentional resources in time (for a recent review, see, e.g., Frey et al., 2015). Nevertheless, the periodicity of the stimulation itself complicates this interpretation, as the brain might simply follow the rhythm of its input. In this review, we presented an increasing amount of evidence that speaks against a merely passive role of neural oscillations for phase entrainment to speech sound. Instead, the brain might constantly predict the timing of relevant and irrelevant events of speech sound, including acoustic high-level features, and actively align neural oscillations so that they efficiently

boost the current locus of attention in a noisy background. Linguistic high-level features, reflecting intelligibility, might play a modulatory, and speech-specific, role by determining the behavioral consequences of phase entrainment to speech sound.

ACKNOWLEDGMENTS

The authors thank Mitchell Steinschneider, Peter Lakatos, Daniel Pressnitzer, Alain de Cheveigné and Jesko Verhey for helpful comments and discussions. This study was supported by a Studienstiftung des deutschen Volkes (German National Academic Foundation) scholarship to BZ, and an ERC grant ("P-Cycles", number 614244) to RV.

REFERENCES

- Adrian, E. D. (1944). Brain rhythms. *Nature* 153, 360–362. doi: 10.1038/153360a0
- Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., and Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc. Natl. Acad. Sci. U S A* 98, 13367–13372. doi: 10.1073/pnas.201400998
- Arnal, L. H. (2012). Predicting "when" using the motor system's beta-band oscillations. *Front. Hum. Neurosci.* 6:225. doi: 10.3389/fnhum.2012.00225
- Arnal, L. H., Doelling, K. B., and Poeppel, D. (2015). Delta-beta coupled oscillations underlie temporal prediction accuracy. *Cereb. Cortex* 25, 3077–3085. doi: 10.1093/cercor/bhu103
- Arnal, L. H., and Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends Cogn. Sci.* 16, 390p–398. doi: 10.1016/j.tics.2012.05.003
- Arnal, L. H., Morillon, B., Kell, C. A., and Giraud, A.-L. (2009). Dual neural routing of visual facilitation in speech processing. *J. Neurosci.* 29, 13445–13453. doi: 10.1523/JNEUROSCI.3194-09.2009
- Arnal, L. H., Wyart, V., and Giraud, A.-L. (2011). Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nat. Neurosci.* 14, 797–801. doi: 10.1038/nrn.2810
- Besle, J., Schevon, C. A., Mehta, A. D., Lakatos, P., Goodman, R. R., McKhann, G. M., et al. (2011). Tuning of the human neocortex to the temporal dynamics of attended events. *J. Neurosci.* 31, 3176–3185. doi: 10.1523/JNEUROSCI.4518-10.2011
- Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., et al. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cortex* 10, 512–528. doi: 10.1093/cercor/10.5.512
- Buzsáki, G., and Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science* 304, 1926–1929. doi: 10.1126/science.1099745
- Capilla, A., Pazo-Alvarez, P., Darriba, A., Campo, P., and Gross, J. (2011). Steady-state visual evoked potentials can be explained by temporal superposition of transient event-related responses. *PLoS One* 6:e14543. doi: 10.1371/journal.pone.0014543
- Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.* 25, 975–979. doi: 10.1121/1.1907229
- Coutlee, C. G., and Huettel, S. A. (2012). The functional neuroanatomy of decision making: prefrontal control of thought and action. *Brain Res.* 1428, 3–12. doi: 10.1016/j.brainres.2011.05.053
- Crossette, M. J., Butler, J. S., and Lalor, E. C. (2015). Congruent visual speech enhances cortical entrainment to continuous auditory speech in noise-free conditions. *J. Neurosci.* 35, 14195–14204. doi: 10.1523/JNEUROSCI.1829-15.2015
- Cummins, F. (2012). Oscillators and syllables: a cautionary note. *Front. Psychol.* 3:364. doi: 10.3389/fpsyg.2012.00364
- DeWitt, I., and Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proc. Natl. Acad. Sci. U S A* 109, E505–E514. doi: 10.1073/pnas.1113427109
- Ding, N., Chatterjee, M., and Simon, J. Z. (2013). Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *Neuroimage* 88C, 41–46. doi: 10.1016/j.neuroimage.2013.10.054
- Ding, N., and Simon, J. Z. (2012a). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc. Natl. Acad. Sci. U S A* 109, 11854–11859. doi: 10.1073/pnas.1205381109
- Ding, N., and Simon, J. Z. (2012b). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *J. Neurophysiol.* 107, 78–89. doi: 10.1152/jn.00297.2011
- Ding, N., and Simon, J. Z. (2013). Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. *J. Neurosci.* 33, 5728–5735. doi: 10.1523/JNEUROSCI.5297-12.2013
- Ding, N., and Simon, J. Z. (2014). Cortical entrainment to continuous speech: functional roles and interpretations. *Front. Hum. Neurosci.* 8:311. doi: 10.3389/fnhum.2014.00311
- Doelling, K. B., Arnal, L. H., Ghitza, O., and Poeppel, D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage* 85, 761–768. doi: 10.1016/j.neuroimage.2013.06.035
- Fiebelkorn, I. C., Foxe, J. J., Butler, J. S., Mercier, M. R., Snyder, A. C., and Molholm, S. (2011). Ready, set, reset: stimulus-locked periodicity in behavioral performance demonstrates the consequences of cross-sensory phase reset. *J. Neurosci.* 31, 9971–9981. doi: 10.1523/JNEUROSCI.1338-11.2011
- Fontolan, L., Morillon, B., Liegeois-Chauvel, C., and Giraud, A.-L. (2014). The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. *Nat. Commun.* 5:4694. doi: 10.1038/ncomms5694
- Frey, J. N., Ruhnau, P., and Weisz, N. (2015). Not so different after all: the same oscillatory processes support different types of attention. *Brain Res.* 1626, 183–197. doi: 10.1016/j.brainres.2015.02.017
- Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn. Sci.* 9, 474–480. doi: 10.1016/j.tics.2005.08.011
- Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 815–836. doi: 10.1098/rstb.2005.1622
- Fujioka, T., Trainor, L. J., Large, E. W., and Ross, B. (2012). Internalized timing of isochronous sounds is represented in neuromagnetic β oscillations. *J. Neurosci.* 32, 1791–1802. doi: 10.1523/JNEUROSCI.4107-11.2012
- Galambos, R., Makeig, S., and Talmachoff, P. J. (1981). A 40-Hz auditory potential recorded from the human scalp. *Proc. Natl. Acad. Sci. U S A* 78, 2643–2647. doi: 10.1073/pnas.78.4.2643
- Ghitza, O. (2011). Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Front. Psychol.* 2:130. doi: 10.3389/fpsyg.2011.00130
- Ghitza, O. (2013). The theta-syllable: a unit of speech information defined by cortical function. *Front. Psychol.* 4:138. doi: 10.3389/fpsyg.2013.00138
- Ghitza, O. (2014). Behavioral evidence for the role of cortical θ oscillations in determining auditory channel capacity for speech. *Front. Psychol.* 5:652. doi: 10.3389/fpsyg.2014.00652

- Gilbert, C. D., and Li, W. (2013). Top-down influences on visual processing. *Nat. Rev. Neurosci.* 14, 350–363. doi: 10.1038/nrn3476
- Giraud, A.-L., and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* 15, 511–517. doi: 10.1038/nn.3063
- Gomez-Ramirez, M., Kelly, S. P., Molholm, S., Sehatpour, P., Schwartz, T. H., and Foxe, J. J. (2011). Oscillatory sensory selection mechanisms during intersensory attention to rhythmic auditory and visual inputs: a human electrocorticographic investigation. *J. Neurosci.* 31, 18556–18567. doi: 10.1523/JNEUROSCI.2164-11.2011
- Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37, 424–438. doi: 10.2307/1912791
- Green, T., Faulkner, A., and Rosen, S. (2002). Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous-interleaved-sampling cochlear implants. *J. Acoust. Soc. Am.* 112, 2155–2164. doi: 10.1121/1.1506688
- Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., et al. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biol.* 11:e1001752. doi: 10.1371/journal.pbio.1001752
- Henry, M. J., Herrmann, B., and Obleser, J. (2014). Entrained neural oscillations in multiple frequency bands comodulate behavior. *Proc. Natl. Acad. Sci. U.S.A.* 111, 14935–14940. doi: 10.1073/pnas.1408741111
- Henry, M. J., and Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proc. Natl. Acad. Sci. U.S.A.* 109, 20095–20100. doi: 10.1073/pnas.1213390109
- Herring, J. D., Thut, G., Jensen, O., and Bergmann, T. (2015). Attention modulates TMS-locked alpha oscillations in the visual cortex. *J. Neurosci.* 35, 14435–14447. doi: 10.1523/JNEUROSCI.1833-15.2015
- Hickok, G., Farahbod, H., and Saberi, K. (2015). The rhythm of perception: entrainment to acoustic rhythms induces subsequent perceptual oscillation. *Psychol. Sci.* 26, 1006–1013. doi: 10.1177/0956797615576533
- Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402. doi: 10.1038/nrn2113
- Horton, C., D'Zmura, M., and Srinivasan, R. (2013). Suppression of competing speech through entrainment of cortical oscillations. *J. Neurophysiol.* 109, 3082–3093. doi: 10.1152/jn.01026.2012
- Howard, M. F., and Poeppel, D. (2010). Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *J. Neurophysiol.* 104, 2500–2511. doi: 10.1152/jn.00251.2010
- Keitel, C., Quigley, C., and Ruhnau, P. (2014). Stimulus-driven brain oscillations in the alpha range: entrainment of intrinsic rhythms or frequency-following response? *J. Neurosci.* 34, 10137–10140. doi: 10.1523/JNEUROSCI.1904-14.2014
- Kerlin, J. R., Shahin, A. J., and Miller, L. M. (2010). Attentional gain control of ongoing cortical speech representations in a “cocktail party”. *J. Neurosci.* 30, 620–628. doi: 10.1523/JNEUROSCI.3631-09.2010
- Krawczyk, D. C. (2002). Contributions of the prefrontal cortex to the neural basis of human decision making. *Neurosci. Biobehav. Rev.* 26, 631–664. doi: 10.1016/S0149-7634(02)00021-0
- Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., and Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* 320, 110–113. doi: 10.1126/science.1154735
- Lakatos, P., Musacchia, G., O'Connell, M. N., Falchier, A. Y., Javitt, D. C., and Schroeder, C. E. (2013). The spectrotemporal filter mechanism of auditory selective attention. *Neuron* 77, 750–761. doi: 10.1016/j.neuron.2012.11.034
- Lakatos, P., O'Connell, M. N., Barczak, A., Mills, A., Javitt, D. C., and Schroeder, C. E. (2009). The leading sense: supramodal control of neurophysiological context by attention. *Neuron* 64, 419–430. doi: 10.1016/j.neuron.2009.10.014
- Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., and Schroeder, C. E. (2005). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *J. Neurophysiol.* 94, 1904–1911. doi: 10.1152/jn.00263.2005
- Lamme, V. A., and Spekreijse, H. (2000). Modulations of primary visual cortex activity representing attentive and conscious scene perception. *Front. Biosci.* 5, D232–D243. doi: 10.2741/lamme
- Landau, A. N., and Fries, P. (2012). Attention samples stimuli rhythmically. *Curr. Biol.* 22, 1000–1004. doi: 10.1016/j.cub.2012.03.054
- Luo, H., and Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54, 1001–1010. doi: 10.1016/j.neuron.2007.06.004
- Luo, H., Liu, Z., and Poeppel, D. (2010). Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biol.* 8:e1000445. doi: 10.1371/journal.pbio.1000445
- Luo, H., and Poeppel, D. (2012). Cortical oscillations in auditory perception and speech: evidence for two temporal windows in human auditory cortex. *Front. Psychol.* 3:170. doi: 10.3389/fpsyg.2012.00170
- Mazzoni, A., Whittingstall, K., Brunel, N., Logothetis, N. K., and Panzeri, S. (2010). Understanding the relationships between spike rate and delta/gamma frequency bands of LFPs and EEGs using a local cortical network model. *Neuroimage* 52, 956–972. doi: 10.1016/j.neuroimage.2009.12.040
- McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748. doi: 10.1038/264746a0
- Mercier, M. R., Molholm, S., Fiebelkorn, I. C., Butler, J. S., Schwartz, T. H., and Foxe, J. J. (2015). Neuro-oscillatory phase alignment drives speeded multisensory response times: an electro-corticographic investigation. *J. Neurosci.* 35, 8546–8557. doi: 10.1523/JNEUROSCI.4527-14.2015
- Mesgarani, N., and Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* 485, 233–236. doi: 10.1038/nature11020
- Mesgarani, N., Cheung, C., Johnson, K., and Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science* 343, 1006–1010. doi: 10.1126/science.1245994
- Millman, R. E., Johnson, S. R., and Prendergast, G. (2015). The role of phase-locking to the temporal envelope of speech in auditory perception and speech intelligibility. *J. Cogn. Neurosci.* 27, 533–545. doi: 10.1162/jocn_a_00719
- Morillon, B., Hackett, T. A., Kajikawa, Y., and Schroeder, C. E. (2015). Predictive motor control of sensory dynamics in auditory active sensing. *Curr. Opin. Neurobiol.* 31C, 230–238. doi: 10.1016/j.conb.2014.12.005
- Morillon, B., and Schroeder, C. E. (2015). Neuronal oscillations as a mechanistic substrate of auditory temporal prediction. *Ann. N Y Acad. Sci.* 1337, 26–31. doi: 10.1111/nyas.12629
- Morillon, B., Schroeder, C. E., and Wyart, V. (2014). Motor contributions to the temporal precision of auditory attention. *Nat. Commun.* 5:5255. doi: 10.1038/ncomms6255
- Nozaradan, S. (2014). Exploring how musical rhythm entrains brain activity with electroencephalogram frequency-tagging. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130393. doi: 10.1098/rstb.2013.0393
- O'Connell, M. N., Barczak, A., Schroeder, C. E., and Lakatos, P. (2014). Layer specific sharpening of frequency tuning by selective attention in primary auditory cortex. *J. Neurosci.* 34, 16496–16508. doi: 10.1523/JNEUROSCI.2055-14.2014
- O'Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., et al. (2015). Attentional selection in a cocktail party environment can be decoded from single-trial EEG. *Cereb. Cortex* 25, 1697–1706. doi: 10.1093/cercor/bht355
- Obleser, J., Herrmann, B., and Henry, M. J. (2012). Neural oscillations in speech: don't be enslaved by the envelope. *Front. Hum. Neurosci.* 6:250. doi: 10.3389/fnhum.2012.00250
- Park, H., Ince, R. A. A., Schyns, P. G., Thut, G., and Gross, J. (2015). Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Curr. Biol.* 25, 1649–1653. doi: 10.1016/j.cub.2015.04.049
- Peelen, M. V., and Kastner, S. (2014). Attention in the real world: toward understanding its neural basis. *Trends Cogn. Sci.* 18, 242–250. doi: 10.1016/j.tics.2014.02.004
- Peelle, J. E., and Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Front. Psychol.* 3:320. doi: 10.3389/fpsyg.2012.00320
- Peelle, J. E., Gross, J., and Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cereb. Cortex* 23, 1378–1387. doi: 10.1093/cercor/bhs118
- Peelle, J. E., and Sommers, M. S. (2015). Prediction and constraint in audiovisual speech perception. *Cortex* 68, 169–181. doi: 10.1016/j.cortex.2015.03.006
- Perrotin, C., Kayser, C., Logothetis, N. K., and Petkov, C. I. (2015). Natural asynchronies in audiovisual communication signals regulate neuronal multisensory interactions in voice-sensitive cortex. *Proc. Natl. Acad. Sci. U.S.A.* 112, 273–278. doi: 10.1073/pnas.1412817112

- Poeppel, D., Emmorey, K., Hickok, G., and Pylkkänen, L. (2012). Towards a new neurobiology of language. *J. Neurosci.* 32, 14125–14131. doi: 10.1523/JNEUROSCI.3244-12.2012
- Power, A. J., Mead, N., Barnes, L., and Goswami, U. (2013). Neural entrainment to rhythmic speech in children with developmental dyslexia. *Front. Hum. Neurosci.* 7:777. doi: 10.3389/fnhum.2013.00777
- Rajkai, C., Lakatos, P., Chen, C.-M., Pincze, Z., Karmos, G., and Schroeder, C. E. (2008). Transient cortical excitation at the onset of visual fixation. *Cereb. Cortex* 18, 200–209. doi: 10.1093/cercor/bhm046
- Rimmele, J. M., Zion Golumbic, E., Schröger, E., and Poeppel, D. (2015). The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene. *Cortex* 68, 144–154. doi: 10.1016/j.cortex.2014.12.014
- Romo, R., and de Lafuente, V. (2013). Conversion of sensory signals into perceptual decisions. *Prog. Neurobiol.* 103, 41–75. doi: 10.1016/j.pneurobio.2012.03.007
- Ruhnau, P., Hauswald, A., and Weisz, N. (2014). Investigating ongoing brain oscillations and their influence on conscious perception - network states and the window to consciousness. *Front. Psychol.* 5:1230. doi: 10.3389/fpsyg.2014.01230
- Rushworth, M. F. S., Kolling, N., Sallet, J., and Mars, R. B. (2012). Valuation and decision-making in frontal cortex: one or many serial or parallel systems? *Curr. Opin. Neurobiol.* 22, 946–955. doi: 10.1016/j.conb.2012.04.011
- Schreiber, T. (2000). Measuring information transfer. *Phys. Rev. Lett.* 85, 461–464. doi: 10.1103/physrevlett.85.461
- Schroeder, C. E., and Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci.* 32, 9–18. doi: 10.1016/j.tins.2008.09.012
- Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., and Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends Cogn. Sci.* 12, 106–113. doi: 10.1016/j.tics.2008.01.002
- Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H., and Lakatos, P. (2010). Dynamics of active sensing and perceptual selection. *Curr. Opin. Neurobiol.* 20, 172–176. doi: 10.1016/j.conb.2010.02.010
- Scott, S. K., Blank, C. C., Rosen, S., and Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123(Pt. 12), 2400–2406. doi: 10.1093/brain/123.12.2400
- Simon, J. Z. (2015). The encoding of auditory objects in auditory cortex: insights from magnetoencephalography. *Int. J. Psychophysiol.* 95, 184–190. doi: 10.1016/j.ijpsycho.2014.05.005
- Song, K., Meng, M., Chen, L., Zhou, K., and Luo, H. (2014). Behavioral oscillations in attention: rhythmic α pulses mediated through θ band. *J. Neurosci.* 34, 4837–4844. doi: 10.1523/JNEUROSCI.4856-13.2014
- Spaak, E., de Lange, F. P., and Jensen, O. (2014). Local entrainment of α oscillations by visual stimuli causes cyclic modulation of perception. *J. Neurosci.* 34, 3536–3544. doi: 10.1523/JNEUROSCI.4385-13.2014
- Stefanics, G., Hangya, B., Hernádi, I., Winkler, I., Lakatos, P., and Ulbert, I. (2010). Phase entrainment of human delta oscillations can mediate the effects of expectation on reaction speed. *J. Neurosci.* 30, 13578–13585. doi: 10.1523/JNEUROSCI.0703-10.2010
- Steinschneider, M., Nourski, K. V., and Fishman, Y. I. (2013). Representation of speech in human auditory cortex: is it special? *Hear. Res.* 305, 57–73. doi: 10.1016/j.heares.2013.05.013
- Thorne, J. D., and Debener, S. (2014). Look now and hear what's coming: on the functional role of cross-modal phase reset. *Hear. Res.* 307, 144–152. doi: 10.1016/j.heares.2013.07.002
- Thut, G., Miniussi, C., and Gross, J. (2012). The functional importance of rhythmic activity in the brain. *Curr. Biol.* 22, R658–R663. doi: 10.1016/j.cub.2012.06.061
- Thut, G., Schyns, P. G., and Gross, J. (2011). Entrainment of perceptually relevant brain oscillations by non-invasive rhythmic stimulation of the human brain. *Front. Psychol.* 2:170. doi: 10.3389/fpsyg.2011.00170
- van Wassenhove, V., Grant, K. W., and Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proc. Natl. Acad. Sci. U S A* 102, 1181–1186. doi: 10.1073/pnas.0408949102
- Vanrullen, R., and Dubois, J. (2011). The psychophysics of brain rhythms. *Front. Psychol.* 2:203. doi: 10.3389/fpsyg.2011.00203
- Vanrullen, R., Zoefel, B., and Ilhan, B. (2014). On the cyclic nature of perception in vision versus audition. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130214. doi: 10.1098/rstb.2013.0214
- Weisz, N., Wühle, A., Monittola, G., Demarchi, G., Frey, J., Popov, T., et al. (2014). Prestimulus oscillatory power and connectivity patterns predispose conscious somatosensory perception. *Proc. Natl. Acad. Sci. U S A* 111, E417–E425. doi: 10.1073/pnas.1317267111
- Wild, C. J., Yusuf, A., Wilson, D. E., Peele, J. E., Davis, M. H., and Johnsrude, I. S. (2012). Effortful listening: the processing of degraded speech depends critically on attention. *J. Neurosci.* 32, 14010–14021. doi: 10.1523/JNEUROSCI.1528-12.2012
- Zion Golumbic, E., Cogan, G. B., Schroeder, C. E., and Poeppel, D. (2013a). Visual input enhances selective speech envelope tracking in auditory cortex at a “cocktail party”. *J. Neurosci.* 33, 1417–1426. doi: 10.1523/JNEUROSCI.3675-12.2013
- Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., et al. (2013b). Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party”. *Neuron* 77, 980–991. doi: 10.1016/j.neuron.2012.12.037
- Zion Golumbic, E. M., Poeppel, D., and Schroeder, C. E. (2012). Temporal context in speech processing and attentional stream selection: a behavioral and neural perspective. *Brain Lang.* 122, 151–161. doi: 10.1016/j.bandl.2011.12.010
- Zoefel, B., and Heil, P. (2013). Detection of near-threshold sounds is independent of EEG phase in common frequency bands. *Front. Psychol.* 4:262. doi: 10.3389/fpsyg.2013.00262
- Zoefel, B., and Sokoliuk, R. (2014). Investigating the rhythm of attention on a fine-grained scale: evidence from reaction times. *J. Neurosci.* 34, 12619–12621. doi: 10.1523/JNEUROSCI.2134-14.2014
- Zoefel, B., and VanRullen, R. (2015a). EEG oscillations entrain their phase to high-level features of speech sound. *Neuroimage* 124, 16–23. doi: 10.1016/j.neuroimage.2015.08.054
- Zoefel, B., and VanRullen, R. (2015b). Selective perceptual phase entrainment to speech rhythm in the absence of spectral energy fluctuations. *J. Neurosci.* 35, 1954–1964. doi: 10.1523/JNEUROSCI.3484-14.2015

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Zoefel and VanRullen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A Predictive Coding Perspective on Beta Oscillations during Sentence-Level Language Comprehension

Ashley G. Lewis^{1,2,3}, Jan-Mathijs Schoffelen^{2,3}, Herbert Schriefers⁴ and Marcel Bastiaansen^{2,5*}

¹ Haskins Laboratories, New Haven, CT, USA, ² Neurobiology of Language Department, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands, ³ Center for Cognitive Neuroimaging, Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands, ⁴ Donders Center for Cognition, Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands, ⁵ Academy for Leisure, NHTV Breda University of Applied Sciences, Breda, Netherlands

OPEN ACCESS

Edited by:

Anne Keitel,
University of Glasgow, UK

Reviewed by:

Lars Meyer,
Max-Planck-Gesellschaft, Germany
Aneta Kielar,
Baycrest Hospital, Canada

***Correspondence:**

Marcel Bastiaansen
bastiaansen4.m@nhtv.nl

Received: 05 November 2015

Accepted: 18 February 2016

Published: 03 March 2016

Citation:

Lewis AG, Schoffelen J-M, Schriefers H and Bastiaansen M (2016) A Predictive Coding Perspective on Beta Oscillations during Sentence-Level Language Comprehension. *Front. Hum. Neurosci.* 10:85. doi: 10.3389/fnhum.2016.00085

Oscillatory neural dynamics have been steadily receiving more attention as a robust and temporally precise signature of network activity related to language processing. We have recently proposed that oscillatory dynamics in the beta and gamma frequency ranges measured during sentence-level comprehension might be best explained from a predictive coding perspective. Under our proposal we related beta oscillations to both the maintenance/change of the neural network configuration responsible for the construction and representation of sentence-level meaning, and to top-down predictions about upcoming linguistic input based on that sentence-level meaning. Here we zoom in on these particular aspects of our proposal, and discuss both old and new supporting evidence. Finally, we present some preliminary magnetoencephalography data from an experiment comparing Dutch subject- and object-relative clauses that was specifically designed to test our predictive coding framework. Initial results support the first of the two suggested roles for beta oscillations in sentence-level language comprehension.

Keywords: language comprehension, neural oscillations, beta, predictive coding, EEG, MEG

INTRODUCTION

Language comprehension requires the fast and efficient integration of information represented at a multitude of different levels and timescales (Jackendoff, 2007). This means that numerous different and often spatially distant brain regions have to interact quickly and dynamically in order to achieve even the most basic linguistic processing. It is therefore not surprising that oscillatory neural dynamics have been steadily receiving more attention as a robust and temporally precise signature of network activity related to language processing (e.g., Weiss and Mueller, 2012; Friederici and Singer, 2015; Lewis et al., 2015). We have recently suggested a role for beta and gamma oscillations in supporting sentence-level language comprehension (Lewis and Bastiaansen, 2015). In this article we zoom in on the role that beta oscillations play in our proposal, reviewing the available evidence old and new, and presenting some preliminary findings from an experiment designed to directly test one of our hypotheses.

THE PROPOSAL IN A NUTSHELL

Our proposal (for a detailed outline see Lewis and Bastiaansen, 2015) suggests that oscillatory neural activity in the beta frequency range (13–30 Hz) during sentence-level language comprehension reflects both the active maintenance/change of the underlying neurocognitive network (NCN; Bressler and Richter, 2014) responsible for the representation and construction of the current sentence-level meaning, and the top-down propagation of predictions from higher to lower levels of the cortical hierarchy based on that sentence-level meaning. When the language comprehension system actively maintains the current mode of processing, beta activity within the associated NCN should increase, while a change in the current mode of processing should result in a decrease in beta activity within that NCN (Engel and Fries, 2010; Lewis and Bastiaansen, 2015; Lewis et al., 2015). Similarly, for predictions about upcoming linguistic information with high levels of certainty, beta activity in the NCN should increase in a direction-specific manner (from higher to lower levels of the cortical processing hierarchy; Bastos et al., 2012; Friston et al., 2014). If there are cues in the linguistic input indicating that the current mode of processing is expected to change, the language comprehension system should place less emphasis on top-down predictions, which in turn should result in a decrease in top-down beta activity (Bastos et al., 2012; Friston et al., 2014; Lewis and Bastiaansen, 2015). Such a role for beta in top-down signaling of predictions based on a generative model within a predictive coding framework has been proposed outside of the domain of language comprehension (Bastos et al., 2012; Friston et al., 2014), and we simply apply these ideas to sentence-level comprehension. It may turn out that certain aspects of these two suggested roles for beta activity are complementary, while others are incompatible. The evidence reported here (see If the Evidence Fits ... and (If the Evidence Fits...) Test the Hypothesis) does not allow us to differentiate between them. We would like to make it explicit that we are not arguing for a relationship between beta activity and measure of word surprisal (cf. Levy, 2008), although it is entirely possible that such a relationship may exist.

Before moving on to examine evidence from previous literature, we think it is important to specify exactly what we mean by top-down predictions. In our opinion a clearer distinction has to be made between predictions at the cognitive level and predictions at the neural level. At the cognitive level, and for sentence-level language comprehension in particular, we consider prediction to refer to the activation of specific lexical information stored in long-term memory prior to the appearance of that information in the linguistic input stream (e.g., DeLong et al., 2005; Van Berkum et al., 2005; Szewczyk and Schriefers, 2013; see also Huettig, 2015 for discussion). On the other hand, within the domain of predictive coding implementations of hierarchical Bayesian inference in the brain, predictions are nothing more (but nothing less) than the neural activity at representational units at a ‘higher’ hierarchical level, that is propagated down to the error units at a hierarchically ‘lower’ level (Friston, 2005). This neural activity may sometimes directly correspond to prediction at the cognitive level, but

most often it will not, because cognitive predictions, and neural predictions generally operate on different timescales. Neural predictions are updated in an ongoing fashion based on numerous factors, including prediction errors sent up the cortical processing hierarchy. Predictions at the cognitive level likely involve evidence accumulation over time until some critical threshold is reached, after which lexical (or more generally long-term memory) pre-activation occurs. This lexical pre-activation may in turn serve as a neural prediction signal that influences activity at lower levels of the cortical hierarchy. Conflating prediction at the cognitive and at the neural level can often lead to confusion in discussions of predictive processing. Our proposal relating beta to top-down prediction refers to predictions at the neural level, but allows for the possibility that predictive processing at the cognitive level may drive these neural prediction signals.

IF THE EVIDENCE FITS ...

Next we turn our attention to the evidence supporting our proposed role for beta oscillations during sentence-level language comprehension. We start by briefly summarizing the evidence we have already reviewed elsewhere (Lewis and Bastiaansen, 2015; Lewis et al., 2015), and then move on to discuss one new piece of evidence.

There are by now a number of studies reporting that beta power is sensitive to both syntactic violations (Davidson and Indefrey, 2007; Bastiaansen et al., 2010; Pérez et al., 2012; Kielar et al., 2014) and semantic incongruities (Luo et al., 2010; Wang et al., 2012; Kielar et al., 2014). In all of these studies, beta power was higher following some target word for syntactically and semantically acceptable sentences compared to target words that resulted in a syntactic violation or a semantic incongruity. Similarly, Luo et al. (2010) showed that beta power was higher for rhythmically normal compared to abnormal target nouns in Chinese verb-noun pairs. In addition to grammatical violations, Pérez et al. (2012) showed that beta power following a target word was lower for the case of Spanish ‘Unagreement’ (where the sentence remains grammatical despite a mismatch between the grammatical person feature marking on the subject and that on the verb of a sentence) compared to grammatically legal target words. These studies all have in common that there is some cue in the linguistic input (e.g., syntactic violation, semantic incongruity, etc.) that indicates to the language comprehension system that the current representation of the sentence-level meaning is not correct and needs to be changed. We suggest that the result is a change in the NCN responsible for that representation, and that this leads to a decrease in beta power in that NCN (or in one or multiple nodes of that network). It may also result in the system assigning less value to top-down predictions as that information has proven unreliable, which would also result in a decrease in beta activity.

Another group of studies has shown that beta activity is higher when sentences are more syntactically demanding, but still grammatical (Weiss et al., 2005; Bastiaansen and Hagoort, 2006; Meyer et al., 2013). Bastiaansen and Hagoort (2006) reported

that beta power was higher for syntactically more demanding center-embedded compared to right-branching relative clauses. Meyer et al. (2013) showed that beta power was higher for long- compared to short-distance subject-verb agreement dependencies at the point in the sentences where the dependency could be resolved. Weiss et al. (2005) found higher beta coherence between frontal and posterior electrode sites for syntactically more complex object-relative (OR) compared to subject-relative (SR) clauses. We suggest that in all these cases the increased beta activity reflects the active maintenance of the current NCN configuration responsible for the construction and representation of the current sentence-level meaning. It may also indicate a greater reliance on top-down predictions based on that sentence-level meaning (i.e., the increased activity may be related to greater weighting of the top-down signal based on the current generative model), in order to actively try to integrate the new linguistic input into the current sentence-level meaning representation.

Bastiaansen et al. (2010) showed that beta power increased linearly over the course of syntactically legal sentences, but returned to baseline levels at the point of a syntactic violation within sentences. They also showed that lists of the same words contained in the sentences in random order (no syntactic structure) did not exhibit any increase in beta power over the course of presentation of the lists (see also Bastiaansen and Hagoort, 2015). We suggest that the gradual buildup of beta power over the course of sentences might be related to the gradually increasing activation of a NCN responsible for the construction and representation of the sentence-level meaning, and that this network becomes disengaged upon reaching a syntactic violation resulting in the decrease in beta power at that point. For random word lists no sentence-level meaning can be constructed, and thus beta power does not increase over the course of their presentation.

Finally, Magyari et al. (2014) presented participants with natural speech, where the ends of speaking turns were either highly predictable or unpredictable, and asked them to press a button when they thought a speaker's turn was about to end. They showed a decrease in beta power just before a button press in the highly predictable condition and an increase in beta power in the unpredictable condition. We suggest that the decrease in beta power in the predictable condition occurs because the language comprehension system anticipates that the current processing mode will have to change (from comprehending the sentence to giving a meta-linguistic judgment by making a button press). In the unpredictable condition the language comprehension system does not predict that the processing mode will change, and instead the current sentence-level meaning representation is actively maintained, resulting in the increased beta power in that condition.

There is one new beta finding that was not included in our previous reviews. Kielar et al. (2015) have followed up on their EEG study investigating syntactic violations and semantic anomalies compared to control sentences (Kielar et al., 2014) by adding conditions with auditory stimulus presentation (the original used only visual presentation), and by using a beamforming approach [in this case applied to magnetoencephalography (MEG) data] to obtain more precise

information about the spatial extent of their effects. They replicate the finding of higher beta (and alpha; see Kielar et al., 2015 for details) power for control sentences compared to both syntactic violations and semantic anomalies, this time for both the visual, and auditory input modalities. Furthermore, their source localization results (albeit computed for the broadband data in the alpha and beta frequency ranges combined; 8–30 Hz) implicated what are arguably the main nodes of the core language network (e.g., Hagoort, 2005, 2013; Hickok and Poeppel, 2007), namely left inferior frontal regions, left superior temporal cortex, and left angular and supramarginal gyri. Our suggestion that when a syntactic violation or semantic incongruity is encountered, decreased beta power reflects a change in the NCN responsible for the representation and construction of a sentence-level meaning holds here as well. However, this study makes an important next step by more precisely mapping out the cortical regions involved. In our opinion, the use of source reconstruction techniques with electrophysiological data is important in future language comprehension studies in order to gain more fine-grained insights into the spatial distribution of the cortical networks whose temporal dynamics are being investigated. At this stage we can only speculate that the critical cortical nodes comprising the NCN that supports sentence-level language comprehension include the core language regions mentioned above. Depending on the context in which language comprehension takes place, this network may interact with other cortical networks like the attention network (e.g., in case the listener/reader finds themselves in a particularly distracting environment) or the theory of mind network (e.g., when interacting with a conversation partner). Working out these details is one important avenue for future investigation.

(IF THE EVIDENCE FITS...) TEST THE HYPOTHESIS

So far, all evidence presented in favor of our hypothesis is based on a re-interpretation of the results of studies that were not specifically designed to test the hypothesis that beta power is related to the maintenance/change of the NCN responsible for representing a sentence-level meaning. Now we present some preliminary data from a MEG experiment that was designed to test this hypothesis. Participants read Dutch SR and OR clause sentences, where the input was identical up to an auxiliary verb presented at the end of the relative clause, disambiguating between the two relative clause types (see **Table 1** for example stimuli). The auxiliary could agree in grammatical number with either the referent in the matrix clause (SR) or with the referent in the relative clause (OR). There was no information in the linguistic input prior to the auxiliary that provided any indication about whether the sentence should be read as a subject- or an object-relative clause (the past participle did not bias the reader to have a preference for either of the possible referents). In all cases both referents were animate and the verb in the relative clause was not semantically biased toward having either of the two referents as a grammatical subject. Dutch readers show a clear preference for the SR reading of these sentences, which appears

TABLE 1 | Example materials used in preliminary experimental findings reported and their direct English translation (in italics).

Condition	Example materials
SR	Ongerust kijkt <u>de hardloper</u> , die de wandelaars in het park gegroet <u>heeft</u> , naar de regenwolken in de lucht. Anxiously looks <i>the runner</i> , that <i>the walkers in the park greeted has</i> , at the rainclouds in the sky.
OR	Ongerust kijkt de hardloper, die <u>de wandelaars</u> in het park gegroet <u>hebben</u> , naar de regenwolken in de lucht. Anxiously looks <i>the runner</i> , that <i>the walkers in the park greeted have</i> , at the rainclouds in the sky.

SR, subject-relative clause condition; OR, object-relative clause condition; auxiliary verb and referent that agrees with it in grammatical number are underlined.

Higher Beta Power for SR than OR Clauses

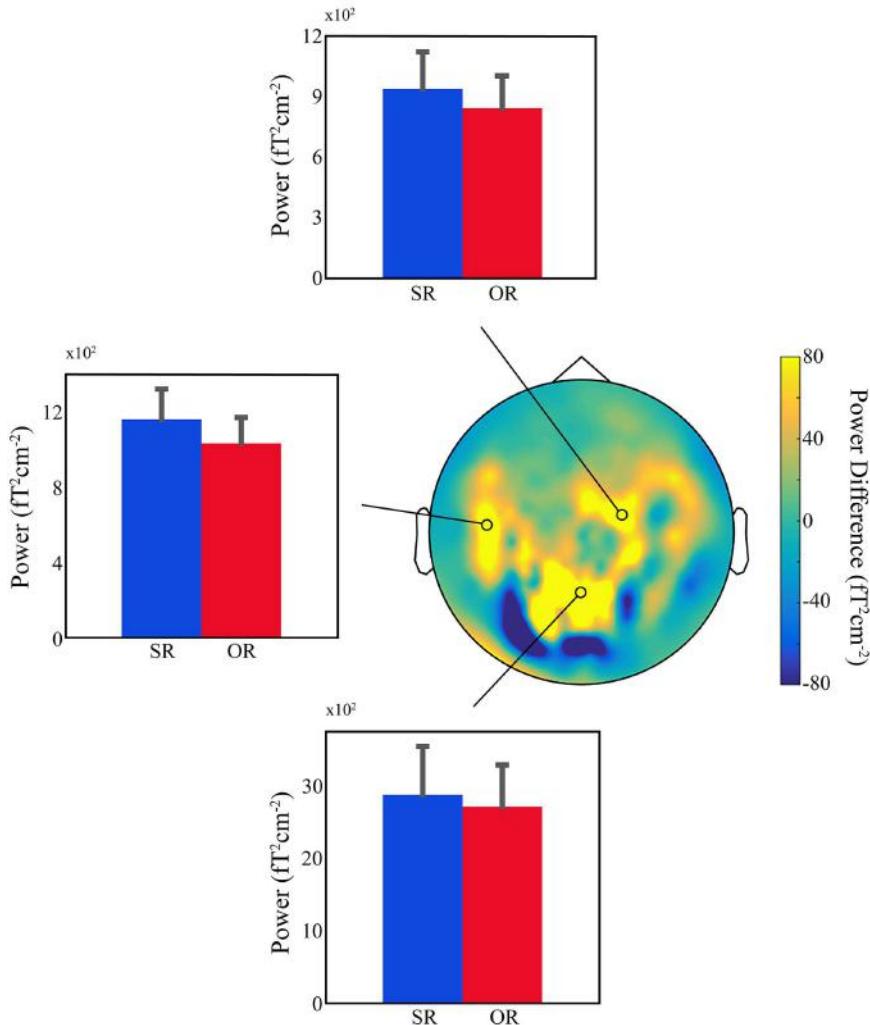


FIGURE 1 | Preliminary MEG findings. (Top, Middle Left, and Bottom) Bar plots showing the average beta power (12–16 Hz) for the subject-relative (SR) and object-relative (OR) conditions in a region between 750 and 1050 ms relative to the onset of the disambiguating auxiliary verb, and at three representative sensors. Error bars indicate standard error of the mean. (Middle Right) Topographical distribution of the difference in power (SR – OR) for the selected beta time-frequency range. Position of selected representative sensors indicated by black circles. Data presented were high-pass filtered above 0.1 Hz and artifacts related to power-line interference, superconducting quantum interference device (SQUID) jumps, muscle activity, eye-movements, eye-blanks, and cardiac activity were removed. The planar gradient representation of the data for each participant (25 in total – written ethical approval was obtained) was computed and a time-frequency decomposition was carried out using a series of Slepian tapers (Mitra and Pesaran, 1999), and a sliding-window approach in time steps of 20 ms and frequency steps of 2 Hz. Time windows of 500 ms and frequency smoothing of 4 Hz were employed. Data were then averaged over the time and frequency ranges of interest (see above) separately for each condition, and grand-averages across all participants were computed for comparison.

more frequently in Dutch corpora (Mak et al., 2002). This means that Dutch readers typically parse the sentence as a SR clause, and when encountering linguistic input indicating that it should instead be read as an OR clause they experience a disruption in processing, resulting in longer reading times (Mak et al., 2002, 2006, 2008). We hypothesized that this ‘unexpectedness’ of the OR reading should result in a decrease in beta power at the disambiguating auxiliary at the end of the relative clause compared to the SR condition (as an aside, one may wonder why this is not also predicted in the case of the center-embedded compared to right-branching relative clause sentences reported in Section “If the Evidence Fits . . .,” but in that case the center-embedded relative clause sentences were not unexpected). This is hypothesized because although the sentence is grammatical, there is a cue in the linguistic input (a mismatch between the grammatical number feature on the verb and the grammatical number feature on the expected referent), which indicates that the current representation of the sentence-level meaning (and therefore the underlying NCN) needs to change.

Figure 1 shows bar plots of the average power at selected MEG sensors in the beta frequency range (12–16 Hz), between 750 and 1050 ms relative to the onset of the disambiguating auxiliary, for the two conditions. There is a small but clear difference in lower-beta power over left temporal and right frontal regions (also present but much less clear over centro-parietal sensors; top, left middle, and bottom panels, respectively, in **Figure 1**), with higher beta power for SR compared to OR clauses, exactly as predicted. This provides further support for the idea that a decrease in beta power is related to the ‘unexpectedness’ of the incoming linguistic input, regardless of whether or not the sentence becomes ungrammatical or semantically anomalous. Added to the findings reviewed in Section “If the Evidence Fits . . .,” the available evidence suggests that upon encountering unexpected linguistic input the language comprehension system prepares for a change in the current mode of processing, and a change in the NCN responsible for representing the current sentence-level meaning. This change is reflected in a decrease in beta power in the underlying NCN (or in certain nodes of that NCN). We would like to emphasize that since this is only preliminary data it should only be considered tentative support for our hypothesis.

The beta power decrease may also reflect diminished ‘confidence’ in top-down predictions by the language comprehension system after encountering unexpected linguistic input. Our experiment does not directly address hypotheses about beta carrying top-down predictions, but it is possible that the local modulations of beta power do reflect such predictions. In order to directly test the hypothesis about top-down information in a predictive coding framework one

first needs to define the different hierarchical levels involved at the cognitive level (e.g., a unification component sending predictions down the hierarchy to a memory component; cf. Hagoort, 2005, 2013) and the cortical regions responsible for instantiating those cognitive components (e.g., left inferior frontal cortex, and left temporal cortex). Then a directional measure of oscillatory activity (e.g., Granger causality, dynamic causal modeling, or transfer entropy; Friston et al., 2013; Park et al., 2015) can be used to directly test whether or not beta activity is predominant from higher to lower levels of the cortical hierarchy (e.g., from left inferior frontal cortex to left temporal cortex).

CONCLUSION

In this article we have zoomed in on our proposed role for oscillatory activity in the beta frequency range in both the maintenance/change of the NCN underlying the construction and representation of a sentence-level meaning, and the propagation of top-down predictions to lower levels of the cortical processing hierarchy based on that sentence-level meaning. We reviewed old and new evidence supporting our proposed roles for beta, and presented some preliminary findings from an experiment designed to directly test one of our hypotheses. The results make a compelling case for beta as an index of maintenance/change of the current NCN underlying sentence-level meaning representation and construction. It will be important for future research to directly test the proposed role of beta in top-down predictions, to further specify which cortical nodes are incorporated into the NCN in different linguistic contexts, and to investigate the extent of overlap between the two potential roles for beta (maintenance and top-down predictions). Performing analyses at the level of cortical sources rather than at the sensor/electrode level will be an important part of this endeavor.

AUTHOR CONTRIBUTIONS

AL and MB conceived the structure of the article. AL, JS, HS, and MB wrote the manuscript. For the preliminary data presented, AL, JS, HS, and MB designed the experiment, AL collected the data, AL and JS analyzed the data.

ACKNOWLEDGMENTS

This work is partly supported by an NWO VIDI grant to JS (grant number 864.14.011), and an IMPRS PhD fellowship from the Max Planck Society to AL.

REFERENCES

- Bastiaansen, M. C. M., and Hagoort, P. (2006). Oscillatory neuronal dynamics during language comprehension. *Prog. Brain Res.* 159, 179–196. doi: 10.1016/S0079-6123(06)59012-0

- Bastiaansen, M. C. M., and Hagoort, P. (2015). Frequency-based segregation of syntactic and semantic unification during online sentence level language comprehension. *J. Cogn. Neurosci.* 27, 2095–2107. doi: 10.1162/jocn_a_00829
Bastiaansen, M. C. M., Magyari, L., and Hagoort, P. (2010). Syntactic unification operations are reflected in oscillatory dynamics during on-line

- sentence comprehension. *J. Cogn. Neurosci.* 22, 1333–1347. doi: 10.1162/jocn.2009.21283
- Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., and Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron* 76, 695–711. doi: 10.1016/j.neuron.2012.10.038
- Bressler, S. L., and Richter, C. G. (2014). Interareal oscillatory synchronization in top-down neocortical processing. *Curr. Opin. Neurobiol.* 31C, 62–66. doi: 10.1016/j.conb.2014.08.010
- Davidson, D. J., and Indefrey, P. (2007). An inverse relation between event-related and time-frequency violation responses in sentence processing. *Brain Res.* 1158, 81–92. doi: 10.1016/j.brainres.2007.04.082
- DeLong, K. A., Urbach, T. P., and Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat. Neurosci.* 8, 1117–1121. doi: 10.1038/nn1504
- Engel, A. K., and Fries, P. (2010). Beta-band oscillations—signalling the status quo? *Curr. Opin. Neurobiol.* 20, 156–165. doi: 10.1016/j.conb.2010.02.015
- Friederici, A. D., and Singer, W. (2015). Grounding language processing on basic neurophysiological principles. *Trends Cogn. Sci. (Regul. Ed.)* 19, 329–338. doi: 10.1016/j.tics.2015.03.012
- Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 815–836. doi: 10.1098/rstb.2005.1622
- Friston, K., Moran, R., and Seth, A. K. (2013). Analysing connectivity with Granger causality and dynamic causal modelling. *Curr. Opin. Neurobiol.* 23, 172–178. doi: 10.1016/j.conb.2012.11.010
- Friston, K. J., Bastos, A. M., Pinotsis, D., and Litvak, V. (2014). LFP and oscillations—what do they tell us? *Curr. Opin. Neurobiol.* 31C, 1–6. doi: 10.1016/j.conb.2014.05.004
- Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends Cogn. Sci. (Regul. Ed.)* 9, 416–423. doi: 10.1016/j.tics.2005.07.004
- Hagoort, P. (2013). MUC (Memory, Unification, Control) and beyond. *Front. Psychol.* 4:416. doi: 10.3389/fpsyg.2013.00416
- Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402. doi: 10.1038/nrn2113
- Huetting, F. (2015). Four central questions about prediction in language processing. *Brain Res.* 626, 118–135. doi: 10.1016/j.brainres.2015.02.014
- Jackendoff, R. (2007). A parallel architecture perspective on language processing. *Brain Res.* 1146, 2–22. doi: 10.1016/j.brainres.2006.08.111
- Kielar, A., Panamsky, L., Links, K. A., and Meltzer, J. A. (2014). Oscillatory responses to semantic and syntactic violations. *J. Cogn. Neurosci.* 26, 2840–2862. doi: 10.1162/jocn_a_00670
- Kielar, A., Panamsky, L., Links, K. A., and Meltzer, J. A. (2015). Localization of electrophysiological responses to semantic and syntactic anomalies in language comprehension with MEG. *Neuroimage* 105, 507–524. doi: 10.1016/j.neuroimage.2014.11.016
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition* 106, 1126–1177. doi: 10.1016/j.cognition.2007.05.006
- Lewis, A. G., and Bastiaansen, M. C. M. (2015). A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. *Cortex* 68, 155–168. doi: 10.1016/j.cortex.2015.02.014
- Lewis, A. G., Wang, L., and Bastiaansen, M. C. M. (2015). Fast oscillatory dynamics during language comprehension: unification versus maintenance and prediction? *Brain Lang.* 148, 51–63. doi: 10.1016/j.bandl.2015.01.003
- Luo, Y., Zhang, Y., Feng, X., and Zhou, X. (2010). Electroencephalogram oscillations differentiate semantic and prosodic processes during sentence reading. *Neuroscience* 169, 654–664. doi: 10.1016/j.neuroscience.2010.05.032
- Magyari, L., Bastiaansen, M. C. M., de Ruiter, J. P., and Levinson, S. C. (2014). Early anticipation lies behind the speed of response in conversation. *J. Cogn. Neurosci.* 26, 2530–2539. doi: 10.1162/jocn_a_00673
- Mak, W., Vonk, W., and Schriefers, H. (2006). Animacy in processing relative clauses: the hikers that rocks crush. *J. Mem. Lang.* 54, 466–490. doi: 10.1016/j.jml.2006.01.001
- Mak, W. M., Vonk, W., and Schriefers, H. (2002). The influence of animacy on relative clause processing. *J. Mem. Lang.* 47, 50–68. doi: 10.1006/jmla.2001.2837
- Mak, W. M., Vonk, W., and Schriefers, H. (2008). Discourse structure and relative clause processing. *Mem. Cogn.* 36, 170–181. doi: 10.3758/MC.36.1.170
- Meyer, L., Obleser, J., and Friederici, A. D. (2013). Left parietal alpha enhancement during working memory-intensive sentence processing. *Cortex* 49, 711–721. doi: 10.1016/j.cortex.2012.03.006
- Mitra, P. P., and Pesaran, B. (1999). Analysis of dynamic brain imaging data. *Biophys. J.* 76, 691–708. doi: 10.1016/S0006-3495(99)77236-X
- Park, H., Ince, R. A. A., Schyns, P. G., Thut, G., and Gross, J. (2015). Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Curr. Biol.* 25, 1649–1653. doi: 10.1016/j.cub.2015.04.049
- Pérez, A., Molinaro, N., Mancini, S., Barraza, P., and Carreiras, M. (2012). Oscillatory dynamics related to the Unagreement pattern in Spanish. *Neuropsychologia* 50, 2584–2597. doi: 10.1016/j.neuropsychologia.2012.07.009
- Szewczyk, J. M., and Schriefers, H. (2013). Prediction in language comprehension beyond specific words: an ERP study on sentence comprehension in Polish. *J. Mem. Lang.* 68, 297–314. doi: 10.1016/j.jml.2012.12.002
- Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., and Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 443–467. doi: 10.1037/0278-7393.31.3.443
- Wang, L., Jensen, O., van den Brink, D., Weder, N., Schoffelen, J., Magyari, L., et al. (2012). Beta oscillations relate to the N400m during language comprehension. *Hum. Brain Mapp.* 33, 2898–2912. doi: 10.1002/hbm.21410
- Weiss, S., and Mueller, H. M. (2012). “Too many betas do not spoil the broth”: the role of beta brain oscillations in language processing. *Front. Psychol.* 3:201. doi: 10.3389/fpsyg.2012.00201
- Weiss, S., Mueller, H. M., Schack, B., King, J. W., Kutas, M., and Rappelsberger, P. (2005). Increased neuronal communication accompanying sentence comprehension. *Int. J. Psychophysiol.* 57, 129–141. doi: 10.1016/j.ijpsycho.2005.03.013

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Lewis, Schoffelen, Schriefers and Bastiaansen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Auditory cortical deactivation during speech production and following speech perception: an EEG investigation of the temporal dynamics of the auditory alpha rhythm

David Jenson¹, Ashley W. Harkrider¹, David Thornton¹, Andrew L. Bowers² and Tim Saltuklaroglu^{1*}

¹ Department of Audiology and Speech Pathology, University of Tennessee Health Science Center, Knoxville, TN, USA

² Department of Communication Disorders, University of Arkansas, Fayetteville, AR, USA

OPEN ACCESS

Edited by:

Anne Keitel,
University of Glasgow, UK

Reviewed by:

Naomi S. Kort,
University of California,
San Francisco, USA

Daniel Callan,

Center for Information and Neural
Networks (CiNet), National Institute
of Information and Communications
Technology (NICT), Japan

*Correspondence:

Tim Saltuklaroglu,
Department of Audiology and
Speech Pathology, University
of Tennessee Health Science Center,
578 South Stadium Hall, Knoxville,
TN 37996, USA
tsaltuk@uthsc.edu

Received: 12 May 2015

Accepted: 14 September 2015

Published: 08 October 2015

Citation:

Jenson D, Harkrider AW, Thornton D,
Bowers AL and Saltuklaroglu T
(2015) Auditory cortical deactivation
during speech production and
following speech perception: an EEG
investigation of the temporal
dynamics of the auditory alpha
rhythm.

Front. Hum. Neurosci. 9:534.
doi: 10.3389/fnhum.2015.00534

Sensorimotor integration (SMI) across the dorsal stream enables online monitoring of speech. Jenson et al. (2014) used independent component analysis (ICA) and event related spectral perturbation (ERSP) analysis of electroencephalography (EEG) data to describe anterior sensorimotor (e.g., premotor cortex, PMC) activity during speech perception and production. The purpose of the current study was to identify and temporally map neural activity from posterior (i.e., auditory) regions of the dorsal stream in the same tasks. Perception tasks required “active” discrimination of syllable pairs (/ba/ and /da/) in quiet and noisy conditions. Production conditions required overt production of syllable pairs and nouns. ICA performed on concatenated raw 68 channel EEG data from all tasks identified bilateral “auditory” alpha (α) components in 15 of 29 participants localized to pSTG (left) and pMTG (right). ERSP analyses were performed to reveal fluctuations in the spectral power of the α rhythm clusters across time. Production conditions were characterized by significant α event related synchronization (ERS; $pFDR < 0.05$) concurrent with EMG activity from speech production, consistent with speech-induced auditory inhibition. Discrimination conditions were also characterized by α ERS following stimulus offset. Auditory α ERS in all conditions temporally aligned with PMC activity reported in Jenson et al. (2014). These findings are indicative of speech-induced suppression of auditory regions, possibly via efference copy. The presence of the same pattern following stimulus offset in discrimination conditions suggests that sensorimotor contributions following speech perception reflect covert replay, and that covert replay provides one source of the motor activity previously observed in some speech perception tasks. To our knowledge, this is the first time that inhibition of auditory regions by speech has been observed in real-time with the ICA/ERSP technique.

Keywords: sensorimotor integration, auditory alpha, EEG, dorsal stream, speech-induced suppression

Introduction

Human communication relies heavily on the functional integrity of the auditory system. Auditory cortical regions reside bilaterally in the temporal lobes, extending posteriorly along the superior temporal gyri (STG) to include primary and association regions. These regions allow humans to sense sounds and are particularly tuned to speech, providing spectro-temporal analysis of complex acoustic speech signals (Specht, 2014). Auditory association regions, such as the posterior superior temporal gyri (pSTG), comprise the posterior aspects of the dorsal stream, which function in tandem with anterior regions (e.g., the premotor cortex, PMC) of this network to facilitate sensorimotor integration (SMI) for speech (Hickok and Poeppel, 2000, 2004, 2007; Rauschecker, 2012). While there is clear evidence of dorsal stream activity in both speech perception and production (Numminen and Curio, 1999; Callan et al., 2000; Curio et al., 2000; Houde et al., 2002; Herman et al., 2013; Jenson et al., 2014), its temporal dynamics are still not well understood.

According to contemporary models of speech such as the Directions into Velocities of Articulators (DIVA; Tourville and Guenther, 2011; Guenther and Vladusich, 2012) and State Feedback Control (SFC; Hickok et al., 2011; Houde and Nagarajan, 2011), SMI for speech production is dependent upon the integrity of the dorsal stream network. When motor commands for speech are initiated, the PMC produces a corollary discharge (i.e., efference copy) containing an internal model of the expected sensory consequences of the movement (von Holst, 1954; Blakemore et al., 2000; Wolpert and Flanagan, 2001). The efference copy is sent from the PMC to higher order association regions for comparison with available acoustic reafferent feedback, delivered with the execution of the motor commands (Guenther, 2006; Tourville and Guenther, 2011; Guenther and Hickok, 2015). Any mismatch between prediction and reafference (i.e., an error signal) quickly results in corrective feedback sent to motor planning regions (e.g., PMC) for online updating of subsequent commands (Guenther et al., 2006; Houde and Chang, 2015). However, during continuous error-free unperturbed speech production, internally based predictions match the reafferent feedback, minimizing the need for corrective feedback. This accurate matching is thought to have a subtractive (i.e., canceling) effect, producing a net attenuation of activity in auditory regions, which is paramount to distinguishing our own speech from that of others (Blakemore et al., 2000; Wolpert and Flanagan, 2001). This suppression of predicted feedback is thought to enhance sensitivity to deviations from the intended production, facilitating online monitoring of speech (Niziolek et al., 2013; Sitek et al., 2013). This proposal is supported by evidence of lowered auditory thresholds to self-produced vs. externally produced sound (Reznik et al., 2014).

Evidence of speech-induced suppression (SIS; Curio et al., 2000; Sitek et al., 2013) has been demonstrated using various neuroimaging techniques. Positron electron tomography (PET) studies have shown reduced STG activation during speech

production compared to listening to playback of one's own speech (Frith et al., 1991; Hirano et al., 1997; Wise et al., 1999). Similarly, in ERP studies, the amplitude of the N100/M100 response has been found to be reduced in normal overt speech compared to replay tasks (Curio et al., 2000; Houde et al., 2002; Greenlee et al., 2011; Chang et al., 2013), and when compared to speech under altered auditory feedback conditions (Heinks-Maldonado et al., 2005; Behroozmand and Larson, 2011; Kort et al., 2014). Using electrocorticography (ECoG), Chang et al. (2013) found suppression of the pSTG in speaking vs. listening conditions. Taken together, these PET, ERP, and ECoG studies support the deactivation of posterior dorsal stream auditory regions via efference copy during normal speech production in accord with DIVA (Tourville and Guenther, 2011; Guenther and Vladusich, 2012) and SFC (Hickok et al., 2011; Houde and Nagarajan, 2011) models. While the pSTG appears to be the primary site of posterior dorsal stream activity in speech, some studies have reported similar activity in the posterior MTG (Christoffels et al., 2007; Herman et al., 2013; Bowers et al., 2014). Functional magnetic resonance imaging (fMRI) has also produced results that are consistent with these models, identifying suppression in the pSTG during overt speech production (Christoffels et al., 2011). However, some fMRI studies have produced conflicting results. For example, Reznik et al. (2014) reported enhanced responses in the auditory cortex (i.e., pSTG) to self-generated sounds contrasted with externally-generated sounds, interpreting this enhancement as evidence of efference copy improving perceptual sensitivity. Other studies have reported auditory suppression to self-generated stimuli in anterior (i.e., medial) locations of the STG while observing enhancement in posterior regions (Agnew et al., 2013). These mixed findings have been interpreted as representing two functionally distinct and spatially differentiated processes (Agnew et al., 2013; Chang et al., 2013; Houde and Chang, 2015), resolvable with the superior spatial resolution of fMRI.

Though there is ample evidence for SIS around posterior dorsal regions in speech production, a better understanding of its functional role is likely to be achieved by temporally mapping activity within the dorsal stream regions in reference to speech events. Increased temporal precision also may enhance understanding of the functional role of dorsal stream activity observed during speech perception. While dorsal stream activity is not typically observed during "passive" listening tasks (Scott et al., 2009; Szenkovits et al., 2012; Bowers et al., 2013), it has been reported in a variety of more challenging "active" perception tasks, such as discrimination of foreign phonemes (Callan et al., 2006), segmentation (Burton et al., 2000; LoCasto et al., 2004), and discrimination of speech in noise (Bowers et al., 2013, 2014). These mixed findings leave unanswered questions regarding the extent to which auditory-to-motor mapping functionally supports accurate perception vs. being merely a by-product of increased processing. One way to address these questions is by examining the timing of dorsal stream activity relative to stimulus presentation. Early activity may be indicative of predictive coding (Sohoglu et al., 2012), in which early motor representations are used

to constrain the analysis of incoming sensory information to aid accurate discrimination; a form of experience-based Constructivist hypothesis testing (Stevens and Halle, 1967; Callan et al., 2010; Skipper, 2014). In contrast, late activity following stimulus offset might reflect covert replay (Burton et al., 2000; Rogalsky et al., 2008; Jenson et al., 2014), in which stimuli are covertly rehearsed in working memory to facilitate discrimination (Baddeley, 2003). Thus, fine-grained temporal data are critical to addressing the functional role of dorsal stream activity in speech perception.

The excellent temporal and spectral detail inherent to neural oscillations makes them a prime candidate for evaluating the dynamics of neural activity in anterior and posterior regions of the dorsal stream (e.g., PMC, pSTG). These oscillations originate from local synchrony between the action potentials of neurons, giving rise to neuronal assemblies with periodic variations in their activity levels (Schnitzler and Gross, 2005; Buszaki, 2006). Fluctuations in spectral power relative to baseline within frequency bands can be measured as relative increases (event related synchronization, ERS) and decreases (event related desynchronization, ERD) in activity, respectively. The “gating by inhibition” hypothesis (Jensen and Mazaheri, 2010) proposes that spectral power within the alpha (α) band (8–13 Hz) can be interpreted as a measure of cortical activation. In support of this hypothesis, α spectral power has been shown to be inversely correlated with the fMRI BOLD signal (Laufs et al., 2003; Brookes et al., 2005; Scheeringa et al., 2009; Mayhew et al., 2013), leading to the interpretation of α ERS and ERD as indices of cortical inhibition and disinhibition, respectively (Klimesch et al., 2007; Weisz et al., 2011).

α oscillations are ubiquitous across the brain, having been implicated in the modulation of cortical activity found in both attention (Muller and Weisz, 2012; Frey et al., 2014) and working memory tasks (Jokisch and Jensen, 2007; van Dijk et al., 2010). A growing body of evidence also points to the existence of an independent auditory α rhythm distinct from other known α generators (Tiihonen et al., 1991; Lehtela et al., 1997; Weisz et al., 2011). Tiihonen et al. (1991) identified a magnetoencephalographic (MEG) α rhythm that demonstrated ERD during auditory stimulation that was not modulated by opening the eyes or clenching the fist, concluding that this was a distinct auditory α rhythm (Tiihonen et al., 1991; Lehtela et al., 1997). Subsequent investigation has implicated this auditory α rhythm in top-down attentional control during dichotic listening (Muller and Weisz, 2012), neural excitability and stimulus detection (Weisz et al., 2014), and auditory hallucinations (Weisz et al., 2011). These studies demonstrate the utility of auditory α oscillations to the investigation of cognitive processes underlying speech.

The temporal precision, economy, and non-invasive nature of electroencephalography (EEG) makes it well suited for capturing oscillatory activity from SMI in speech perception and production (Cuellar et al., 2012; Bowers et al., 2013; Jenson et al., 2014). However, historically EEG analysis has been limited by poor spatial resolution due to volume conduction (the fact that each channel contains information from multiple neural sources)

and its susceptibility to contamination by movement artifact. Recently, independent component analysis (ICA) has offered an effective means of overcoming these limitations. ICA is a method of blind source separation that decomposes complex mixtures of non-neural (i.e., artifact) and neural EEG signals into temporally independent and spatially fixed components (Stone, 2004; Onton et al., 2006). In effect, ICA provides a means of both separating muscle movement from neural activity and reliably identifying cortical sources of activity. Independent EEG components can be further decomposed across time and frequency via event related spectral perturbations (ERSP) to reveal patterns of ERS/ERD that characterize regional neural activity in cognitive and motor tasks. Identification of auditory α rhythm components can be followed by ERSP analysis to better understand auditory activity across the time course of speech production and perception tasks.

This ICA/ERSP analysis is well established in perception tasks (Lin et al., 2007, 2011; McGarry et al., 2012) and has more recently been applied to speech perception, examining changes in spectral activity in the sensorimotor μ rhythm components (Bowers et al., 2013, 2014; Jenson et al., 2014). Consistent with Callan et al. (2010) and constructivist interpretations of predictive coding and analysis-by-synthesis (Stevens and Halle, 1967), Bowers et al. (2013) found active syllable discrimination produced more robust mu (μ) ERD (indicating sensorimotor processing) than passive listening to syllables or discriminating between tones. In addition to identifying sensorimotor μ components, Bowers et al. (2014) reported bilateral components from posterior superior temporal lobes (pSTG) with characteristic α spectra, similar to those described by Muller and Weisz (2012). Though ICA clearly has demonstrated the capacity for identifying sources of neural activity in perceptual tasks, its application to motor tasks has been limited due to questions pertaining to its ability to accurately localize and estimate cortical activity within neural sources in the presence of competing myogenic activity (Oken, 1986; Shackman et al., 2009; McMenamin et al., 2010, 2011).

Recently, Jenson et al. (2014) used an ICA/ERSP technique to measure anterior dorsal stream activity in speech perception and production. Specifically, they identified μ components with characteristic α and beta (β ; ~20 Hz) peaks and related the changes in spectral power within these peaks to sensory (α) and motor (β) contributions to anterior dorsal stream activity in various tasks. Participants listened to passive noise, discriminated pairs of syllables with and without background noise, and performed overt productions of syllable pairs and tri-syllable nouns. ICA of concatenated raw EEG data from all (perception and production) tasks yielded independent left and right μ components localized to the PMC common to all conditions, supporting the use of ICA in speech production. The ERSP analysis revealed concurrent α and β ERD (reflecting sensory and motor processing) time-locked to muscle movement during overt production. The authors interpreted these findings as evidence of a normal continuous sensorimotor loop for speech production. Interestingly, this same pattern of concurrent α and β μ ERD was observed in the discrimination conditions

in the time period following acoustic offset. The authors cautiously interpreted this μ ERD in discrimination conditions as evidence of late covert rehearsal while the stimuli was being held in memory prior to a response (Burton et al., 2000; Baddeley, 2003). This interpretation supports the suggestion that similar anterior dorsal stream sensorimotor processes can be involved in covert and overt speech production (Gehrig et al., 2012; Ylinen et al., 2014). However, to achieve a better understanding of dorsal stream activity in speech perception and production, it is necessary to also examine the temporal dynamics of sensorimotor activity in the posterior aspects of the network.

The purpose of the current study is twofold. The first is to use ICA of raw EEG data to identify temporal lobe auditory components common to speech perception and production tasks. The second is to use ERSP analysis to provide high-resolution temporal and spectral information regarding the dynamics of this auditory α rhythm during speech perception and production. It is hypothesized that ICA will identify bilateral components with α spectra (~ 10 Hz) localized to auditory association regions, representing activity within posterior regions of the dorsal stream. The second hypothesis is that ERSP analysis of these components will reveal α ERS, representing reduced activity in posterior aspects of the dorsal stream (i.e., pSTG) by an efference copy while speech is being produced. Though the current study employs no connectivity measures, ERSPs from auditory components can be examined alongside those from anterior sensorimotor μ components reported in Jenson et al. (2014) to better understand dorsal stream activity in speech perception and production. Thus, the third hypothesis is that μ ERD and α ERS will be observed simultaneously reflecting synchronous complementary activity across anterior and posterior aspects of the dorsal stream. Observing this pattern of activity following speech perception will support the theory that dorsal stream activity in speech discrimination is characterized at least in part by covert replay.

Materials and Methods

Participants

Twenty-nine right-handed native English speakers were recruited from the audiology and speech pathology program at the University of Tennessee Health Science Center. Subjects (24 females, 5 males) had a mean age of 25.16 years (range 21–46) and no history of cognitive, communicative, or attentional disorders. The Edinburgh Handedness Inventory (Oldfield, 1971) was administered to establish handedness dominance for each subject. The Institutional Review Board for the University of Tennessee approved this study, and all subjects provided informed consent prior to participation.

Stimuli

Perception

Syllable stimuli (/ba/ and /da/) for the active perception conditions were generated with AT&T naturally speaking text-to-speech software, which utilizes synthetic analogs of a male

speaker. Syllable stimuli were combined to create syllable pairs such that half of the stimuli consisted of identical pairs (e.g., /ba ba/) and half of the stimuli contained different pairs (e.g., /da ba/). Syllable pairs were then low pass filtered at 5 kHz and normalized for root-mean-square (RMS) amplitude. Each syllable was 200 ms in duration and paired syllables were separated by 200 ms, yielding stimuli that were 600 ms from onset of the first syllable to offset of the second syllable.

One of the active perception conditions (discrimination in noise—Ndis) required subjects to discriminate syllable pairs embedded in white noise with a signal-to-noise ratio (SNR) of +4 dB. This condition was included as previous studies have reported that this SNR produces increased dorsal stream activity while allowing participants to accurately discriminate between the syllables (Binder et al., 2004; Osnes et al., 2011; Bowers et al., 2013, 2014). Another discrimination condition (quiet discrimination—Qdis) required participants to discriminate syllable pairs in the absence of background noise. In order to control for a discrimination response bias (Venezia et al., 2012), an equal number of different and identical syllable pairs were used in each discrimination condition. Discrimination responses were made using a button press. The stimulus used for the control (passive listening) condition was continuous white noise.

Production

Targets for speech production consisted of the same syllable pairings used in the discrimination conditions (e.g., /ba da/), as well as tri-syllable nouns initiated with either /b/ or /d/ and followed by a vowel (e.g., buffalo, daffodil). Visual stimuli for production were presented at the center of the visual field on Microsoft PowerPoint slides consisting of white text on a black background (Arial font) subtending a visual angle of 1.14°. The timelines for perception and production tasks are illustrated in Figure 1.

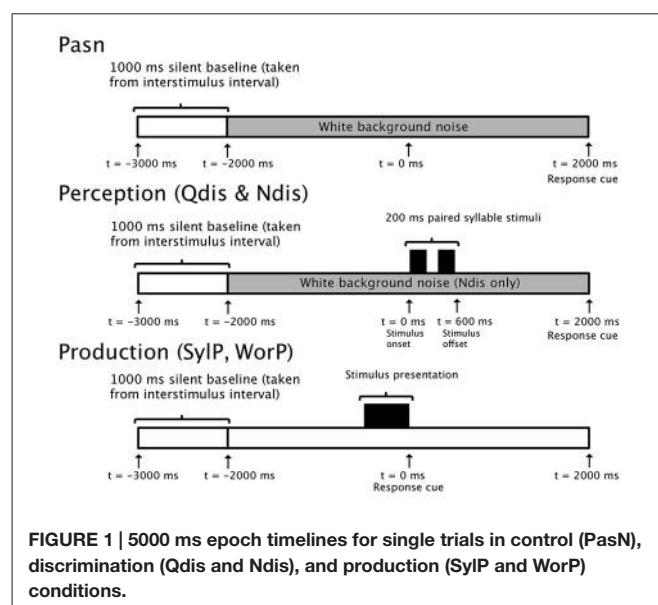


FIGURE 1 | 5000 ms epoch timelines for single trials in control (PasN), discrimination (Qdis and Ndis), and production (SyIP and WorP) conditions.

Design

The experiment consisted of a five condition, within subjects design. The conditions were designed to require gradually increased motoric demands, progressing from the perception of white noise to the production of tri-syllable nouns. The five conditions were:

1. Passive listening to white noise (PasN);
2. Discrimination of syllable pairs in quiet (Qdis);
3. Discrimination of syllable pairs in background noise (Ndis);
4. Overt production of syllable pairs (SylP); and
5. Overt production of tri-syllable nouns (WorP).

Condition 1 was a passive perception task, conditions 2–3 were active perception tasks, and conditions 4–5 were production tasks. The PasN condition required no discrimination, but was used as a control task for the Qdis and Ndis conditions. To control for the neural activity related to the button press response required in the Qdis and Ndis conditions, a button press response was used in the PasN condition. Conditions 2–4 used paired /ba/ and /da/ syllables. Qdis and Ndis required active discrimination of syllable pairs, while SylP required overt productions of /ba/ /da/ syllable pairs, respectively. The WorP condition required overt production of tri-syllable nouns initiated by a /b/ or /d/ and followed by a vowel. Stimuli for the WorP condition were selected from Blockcolsky et al. (2008).

Procedure

The experiment was conducted in an electrically and magnetically shielded, double-walled, soundproof booth. Participants were seated in a comfortable chair with their head and neck supported. Stimuli were presented and button press responses were recorded by a PC computer running Compumedics NeuroScan Stim 2, version 4.3.3. A button press response was embedded in the PasN condition for two reasons: (1) to control for anticipatory β suppression which has been previously reported in tasks requiring a button press response (Makeig et al., 2004; Graimann and Pfurtscheller, 2006; Hari, 2006) and (2) requiring a button press response in a condition with no active discrimination ensured that the subjects were attending to and engaged in the task. The response cue for all perception conditions was a 100 ms, 1000 Hz tone presented at the end of the trial epoch (2000 ms post stimulus). In the PasN condition, subjects were instructed to sit quietly, listen to the stimulus (i.e., white noise), and press the button when they heard the response cue. In the Qdis and Ndis conditions, subjects were instructed to press one of two buttons after hearing the response cue depending on whether the syllables were judged to be the same or different. Handedness of button press response was counterbalanced across all subjects and conditions. Discrimination accuracy was determined as percentage of trials correctly discriminated, and subjects who did not discriminate at a level significantly above chance were excluded from the analysis.

In the production conditions, visual stimuli were presented on a monitor (69.5×39 cm) placed 132 cm in front of the participant's chair. Visual stimuli (syllable pairs and words) remained on the screen for 1 s, and participants were instructed

to begin their production when the visual stimuli disappeared. Thus, stimulus offset was the response cue in the production conditions. In the SylP and WorP conditions, subjects were instructed to produce the syllable pairs in their normal speaking voice. All productions were complete within the 2000 ms window between the response cue and the end of the trial epoch. Each of the five conditions was comprised of 2 blocks of 40 trials each, yielding a total of 10 blocks (5 conditions \times 2 blocks). Order of block presentation was randomized for each subject.

EEG Acquisition

Whole head EEG data were acquired from 68 channels. These channels included two electromyography (EMG) and two electrocardiogram (EKG) electrodes. Data were recorded with an unlinked, sintered NeuroScan Quik Cap, based on the extended international standard 10–20 system (Jasper, 1958; Towle et al., 1993). All recording channels were referenced to the linked mastoid channels (M1, M2). The electro-oculogram was recorded by means of two electrode pairs placed above and below the orbit of the left eye (VEOL, VEOU) and on the medial and lateral canthi of the left eye (HEOL, HEOR) to monitor vertical and horizontal eye movement. The two EMG electrodes were placed above and below the lips to capture labial lip movement related to overt speech.

EEG data were recorded using Compumedics NeuroScan Scan 4.3.3 software in tandem with the Synamps 2 system. EEG data were band pass filtered (0.15–100 Hz) and digitized with a 24-bit analog to digital converter with a sampling rate of 500 Hz. Data collection was time locked to stimulus onset in the perception conditions, and to the response cue in the production conditions. The visual stimuli to be produced were displayed on the screen for 1 s prior to disappearing, which served as the response cue for production. Thus, time zero was defined as stimulus onset for the perception conditions, and stimulus offset served as time zero for the production conditions.

EEG Data Processing

Data processing and analysis were performed with EEGLAB 12 (Brunner et al., 2013), an open source MATLAB toolbox. Data were processed at the individual level and analyzed at both the individual and group level. The following steps were performed at each stage:

1. Individual processing/analysis:
 - (a) Preprocessing of 10 raw EEG files for each participant (5 conditions \times 2 blocks);
 - (b) ICA of preprocessed files across conditions for each participant; and
 - (c) Localization of all neural and non-neural dipoles for each independent component.
2. Group analysis:
 - (a) Two separate analyses using the STUDY module of EEGLAB 12; one study targeting neural components only ("in-head") and the other targeting neural and myogenic components ("all");

- (b) Components common across participants clustered by means of Principal Component Analysis (PCA);
- (c) Left and right STG clusters identified from the “in-head” STUDY analysis, while peri-labial EMG cluster was identified from the “all” STUDY;
- (d) Left and right STG clusters localized by equivalent current dipole (ECD) and current source density (CSD) methods; and
- (e) ERSP performed to yield time-frequency analysis of activity in STG and EMG clusters.

Analysis for Hypothesis 1

Data Preprocessing

Raw EEG data files from both blocks of each condition were appended to create one dataset per condition per participant, and then downsampled to 256 Hz to reduce the computational requirements of further processing steps. Trial epochs of 5000 ms (ranging from -3000 to +2000 ms around time zero) were extracted from the continuous EEG data. The data were then filtered from 3–34 Hz, which allowed for clear visualization of α and β bands, while filtering muscle artifact from surrounding frequency bands. All EEG channels were referenced to the mastoids (M1, M2) to remove common mode noise. Trials were visually inspected, and all epochs containing gross artifact (in excess of 200 μ V) were removed. Additionally, trials were rejected if the participant performed the discrimination incorrectly, or if the response latency exceeded 2000 ms. A minimum of 40 useable trials per subject per participant was required in order to ensure a successful ICA decomposition.

ICA

Following data preprocessing and prior to ICA analysis, data files for each participant were concatenated to yield a single set of ICA weights common to all conditions. This allowed for comparison of activity across conditions within spatially fixed components. The data matrix was decorrelated through the use of an extended Infomax algorithm (Lee et al., 1999). Subsequent ICA training was accomplished with the “extended runica” algorithm in EEGLAB 12 with an initial learning rate of 0.001 and the stopping weight set to 10⁻⁷. ICA decomposition yielded 66 ICs for each participant, corresponding to the number of recording electrodes (68 data channels–2 reference channels; M1, M2). Scalp maps for each component were generated by projecting the inverse weight matrix (W^{-1}) back onto the original spatial channel configuration.

After ICA decomposition, equivalent current dipole models (ECD) were generated for each component by using the boundary element model (BEM) in the DIPFIT toolbox, an open source MATLAB plugin available at sccn.ucsd.edu/eeglab/dipfit.html (Oostenveld and Oostendorp, 2002). Electrode coordinates conforming to the standard 10–20 configuration were warped to the head model. Automated coarse-fitting to the BEM yielded a single dipole model for each of the 1914 ICs (29 participants \times 66 ICs). Dipole localization entails a back projection of the signal to a potential source that could have generated the signal, followed by computing the best forward model from that hypothesized source that accounts for

the highest proportion of the scalp recorded signal (Delorme et al., 2012). The residual variance (RV) is the mismatch between the original scalp recorded signal and this forward projection of the ECD model. The RV can be interpreted as a goodness of fit measure for the ECD model.

STUDY (Group Level Analyses)

Group level analyses were performed in the EEGLAB STUDY module. The STUDY module allows for the comparison of ICA data across participants and conditions. The STUDY module also allows for the inclusion or exclusion of ICs based on RV and location (in head vs. outside head). Two different STUDY analyses were performed on participants’ ICA files containing dipole information. The “in head” analysis (neural) was limited to dipoles originating within the head, and the RV threshold was set to <20%.

In order to capture peri-labial EMG activity, a second STUDY (“all”) was performed, which included dipoles originating both within the head and outside the head. Additionally, the RV threshold was lifted to <50% to account for the fact that EMG activity inherently contains higher levels of RV. Peri-labial EMG activity was extracted from the “all” STUDY, while all neural data were analyzed within the “in head” study only.

PCA Clustering

In both of the STUDY analyses (“in head” and “all”), IC pre-clustering was performed based on commonalities of spectra, dipoles, and scalp maps. The K-means statistical toolbox was used to group similar components across participants based on the specified criteria via PCA. ICs from the “in head” analysis were assigned to 25 neural clusters, from which left and right auditory clusters were identified. ICs from the “all” analysis were assigned to 66 possible clusters (both neural and non-neural), one of which contained peri-labial EMG activity.

Designation to auditory (STG) clusters for the “in head” STUDY was based primarily on the initial results of PCA, followed by inspection of all ICs in the auditory cluster and surrounding clusters based on spectra, dipoles, and scalp maps. Inclusion criteria for the auditory clusters were based on previously observed posterior dorsal stream activity in speech and, therefore, included components that were localized to the pSTG or pMTG regions, showed a characteristic α spectrum, and could be localized with RV <20%.

The majority of the 66 clusters generated in the “all” STUDY contained non-neural (myogenic) activity. The cluster containing peri-labial EMG activity was identified based on dipole location and verified by ERSP analysis, demonstrating activity during the overt speech conditions only.

Source Localization

Source localization for ECD clusters identified in the STUDY module is the mean of the Talairach coordinates (x , y , z) for each of the contributing dipole models (identified by the DIPFIT module). A further method of source localization is standardized low-resolution brain electromagnetic tomography (sLORETA), which addresses the inverse problem by using CSD from scalp recorded electrical signals to estimate source

location (Pascual-Marqui, 2002). Solutions are based on the Talairach cortical probability brain atlas, digitized at the Montreal Neurological Institute (MNI). Electrode locations are co-registered between both spherical models (BESA) and realistic head geometry (Towle et al., 1993). The 3-D brain space was divided into 6,239 voxels, yielding a spatial resolution of 5 mm. The inverse weight projections from the original EEG channels for each component contributing to the temporal α clusters were exported to sLORETA. Cross-spectra were computed and mapped to the Talairach atlas and cross-registered with MNI coordinates, resulting in CSD estimates for each contributing component. The analysis of statistical significance of CSD estimates across participants was performed in the sLORETA software package. The analysis was non-parametric, based on the estimation (via randomization) of the probability distribution of the t-statistic expected under the null hypothesis (Pascual-Marqui, 2002). This method corrects for multiple comparisons across all voxels and frequencies (3–34 Hz). Voxels that were significant at $p < 0.001$ were considered to be active across participants. Group level source localizations are based on the CSD source estimates computed via sLORETA, though the ECD localizations are also reported as they serve to demonstrate the inter-subject variability present in the data.

Analysis for Hypotheses 2 and 3

ERSP

ERSP analyses were used to measure fluctuations in spectral power (in normalized decibel units) across time in the frequency bands of interest (3–34 Hz). Time-frequency transformations were computed using a Morlet wavelet rising linearly from three cycles at 3 Hz to 34 cycles at 34 Hz. Trials were referenced to a 1000 ms pre-stimulus baseline selected from the inter-trial interval. A surrogate distribution was generated from 200 randomly sampled latency windows from this silent baseline (Makeig et al., 2004). Individual ERSP changes across time were calculated with a bootstrap resampling method ($p < 0.05$ uncorrected). Single trial data for all experimental conditions for frequencies between 4 and 30 Hz and ranging from −500 to 1500 ms were entered into the time-frequency analysis.

In the “in head” study, permutation statistics (2000 permutations) were used to assess inter-condition differences. The significance threshold was set at $p < 0.05$, and Type 1 error was controlled by false discovery rate (FDR) correction (Benjamini and Hochberg, 2000). Statistical analyses used a 1×5 repeated measures ANOVA design (PasN, Qdis, Ndis, SylP, WorP). Further *post hoc* analyses of differences in perception and production conditions used 1×3 and 1×2 repeated measures ANOVA designs, respectively.

Results

Discrimination Accuracy

All subjects that contributed to the temporal α clusters performed the discrimination tasks with a high degree of accuracy. As it has been shown that activity in sensorimotor regions is susceptible to the effects of response bias (Venezia et al., 2012), d' values also are reported to tease out the differential effects of sensitivity

and response bias on perceptual accuracy. The average number of useable trials (out of 80) for each condition was: PasN = 74.4 (SD 6.9), Qdis = 74 (SD 4.77), Ndis = 69.1 (SD 12.52), SylP = 73.73 (SD 5.19), and WorP = 73.14 (SD 6.39). Subjects performed the discrimination with similar high accuracy in both the Qdis [96.5%, SD 2.55; d' 3.38, SE 0.09] and Ndis [94.5, SD 8.69; d' 3.63, SE 0.19] conditions. The greater variability in the Ndis condition was due primarily to one participant, who performed the task with 65% accuracy. A paired *t*-test on d' values for each condition indicated that subject accuracy for these two conditions was not significantly different ($p > 0.05$). The mean reaction time for discrimination conditions was 506.3 ms in the Qdis conditions (SD 133.3) and 568.1 ms in the Ndis condition (SD 298.3). A paired *t*-test indicated that the mean response latency between conditions was not significantly different ($p > 0.05$). Taken together, these findings indicate that subjects performed both discrimination tasks with similar levels of accuracy and efficiency. A response contingent analysis was performed in which incorrectly discriminated trials were excluded from subsequent analysis, and thus the analysis of neural data was restricted only to those associated with correctly discriminated trials.

Results Pertaining to Hypothesis 1

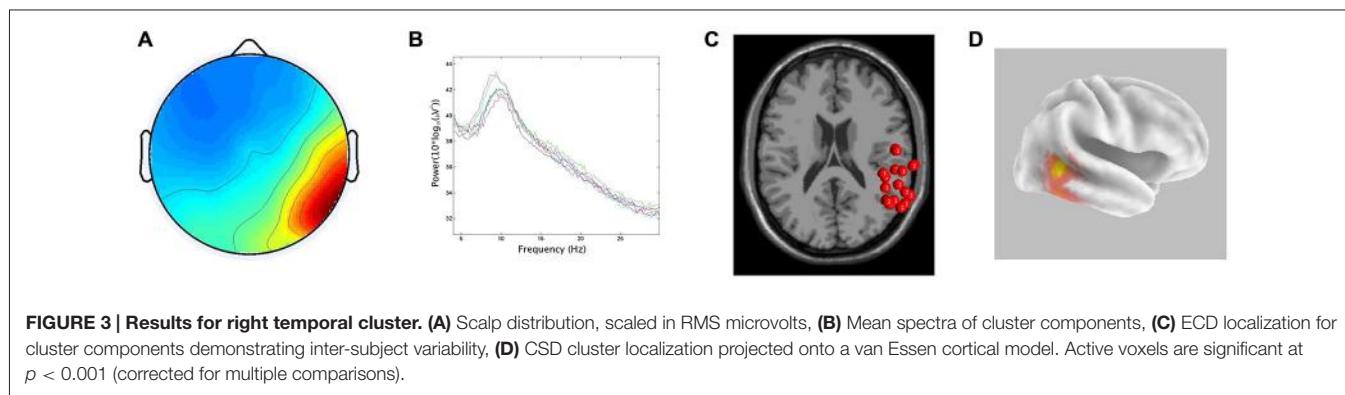
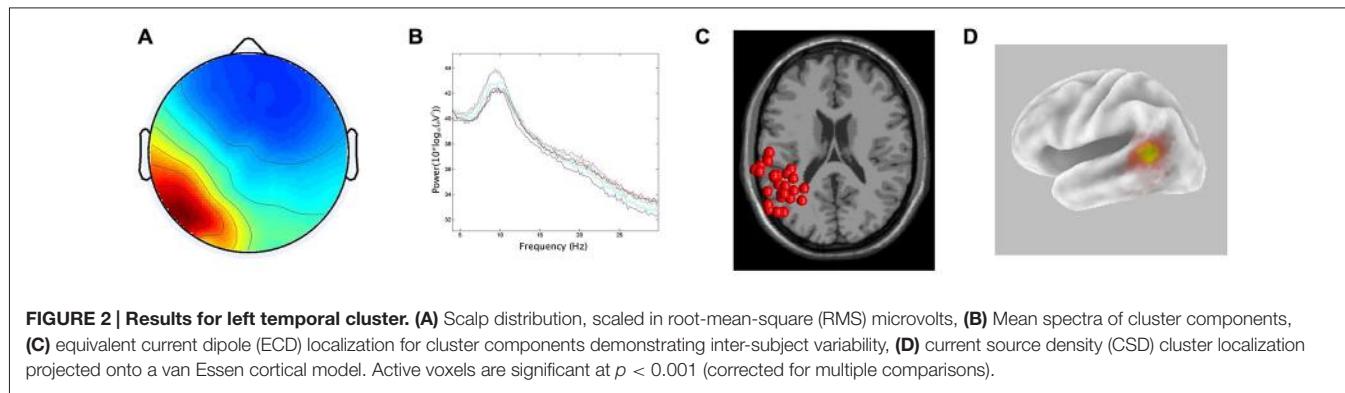
Temporal α Cluster Characteristics

In line with the hypothesis that ICA would identify bilateral α components localized to pSTG, 15/29 participants generated components with less than 20% RV contributing to both the left and right temporal α clusters. The clusters had peaks at 10 Hz on both the left and right. For the ECD dipole models, the average dipole localization was at Talairach [−48, −45, 15] for the left temporal cluster and Talairach [57, −42, 10] on the right. The percentage of unexplained variance for these two clusters was 11.7% and 11.9%, respectively. The CSD model computed with sLORETA showed active voxels ($p < 0.001$) localized to the pSTG on the left and the pMTG on the right. In both hemispheres, activation spread across the pSTG and pMTG. CSD source maxima were localized to MNI [−50, −55, 10] on the left and MNI [55, −45, 0] on the right. The Euclidean distance between ECD and CSD sources were 11.4 mm on the left and 10.6 mm on the right. The peri-labial EMG cluster, identified on the basis of dipole location and the time course of activity, consisted of non-neuronal components with an average of 20.07% RV. **Figures 2, 3** demonstrate: (A) the average scalp map; (B) the spectra; (C) the distribution of ECD dipoles; and (D) CSD source localization for the left and right temporal clusters, respectively. As the component activations were generated from data concatenated across conditions, the source localizations reported pertained to temporal lobe clusters from all experimental conditions.

Results Pertaining to Hypothesis 2

ERSP Analysis in Production (SylP, WorP)

The second hypothesis was that ERSP analysis of auditory α clusters would reveal α ERS in time periods coinciding with overt production. **Figure 4** shows van Essen maps (computed with sLORETA) demonstrating activated voxels at ($p < 0.001$)



in the (A) left and (C) right hemisphere temporal clusters. ERSP analyses show differential patterns of ERS/ERD measured against baseline across the two production conditions (SylP, WorP), within the 4–30 Hz bandwidth. The final column shows significant differences ($p_{FDR} < 0.05$) compared to PasN. **Figure 4B** shows the average ECD localization for the EMG cluster corresponding to peri-labial muscle activity, as well as the ERSP analysis of that component cluster.

In the left temporal cluster, α ERD in production conditions (SylP, WorP) began prior to stimulus onset, peaking after the cue to produce speech. Approximately 500 ms after the production cue, α ERD began to decrease accompanied by the emergence of α ERS, which extended into low β frequencies. As in perception conditions, the right temporal cluster showed identical patterns of activation, though with weaker spectral power.

Temporal Alignment Between Temporal α , Sensorimotor μ , and Peri-labial EMG Activity

EMG activity was found in the SylP and WorP conditions only. EMG ERS (corresponding to speech production) began at approximately 300 ms and peaked at about 500 ms post response cue. These response latencies are within the expected range for speech production tasks (Heinks-Maldonado et al., 2005). α ERS in left and right temporal clusters was aligned temporally with EMG ERS in the SylP and WorP conditions.

Jenson et al. (2014) analyzed data from the same subject pool in identical conditions and interpreted concurrent α and β ERD over the PMC as evidence of covert replay during perception and

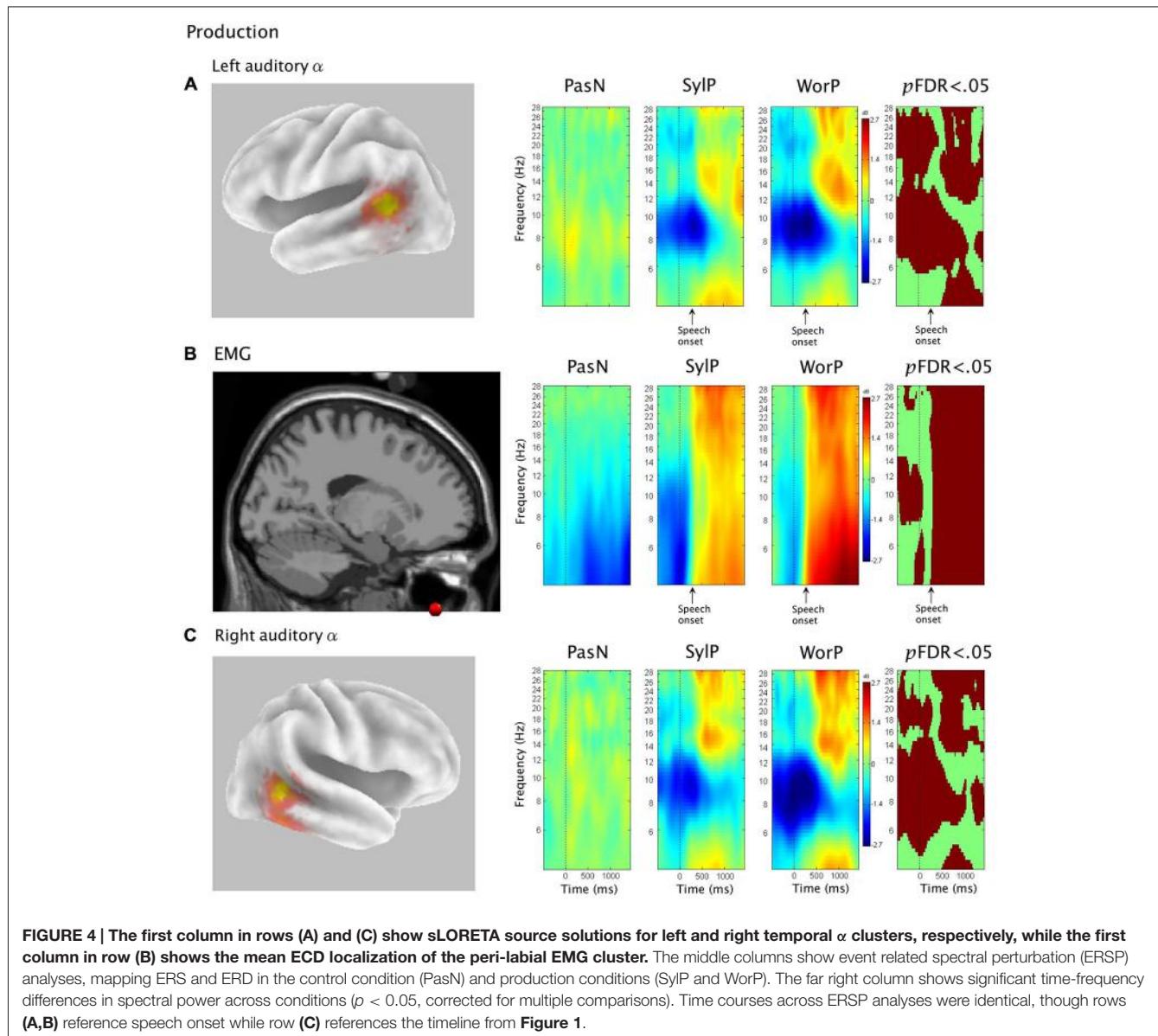
overt production of speech. In the current study, the emergence of α ERS in the temporal cluster also was aligned temporally with the peak α and β ERD in the sensorimotor μ rhythm reported by Jenson et al. (2014) in both perception and production. **Figure 5** demonstrates the temporal synchrony between α ERS in the temporal lobe, sensorimotor α/β ERD (representing SMI during production), and EMG ERS, as well as the alignment of temporal α ERS and sensorimotor α/β ERD (consistent with covert rehearsal) during discrimination tasks.

Results Pertaining to Hypothesis 3

ERSP Analysis in Perception (PasN, Qdis, Ndis)

The third hypothesis was that ERSP analysis of auditory α clusters in discrimination conditions would demonstrate α ERS during time periods of μ ERD, consistent with the interpretation of PMC activity during speech discrimination as evidence of covert replay. **Figure 6** shows van Essen maps (computed with sLORETA) showing activated voxels ($p < 0.001$) in the left (A) and right (B) hemisphere temporal clusters. ERSP analyses show differential patterns of ERS/ERD measured against baseline across the three perception conditions (PasN, Qdis, Ndis), within the 4–30 Hz bandwidth. The final column shows significant differences ($p_{FDR} < 0.05$) among the three conditions.

For the left temporal cluster, α ERD began subsequent to acoustic stimulation and persisted until approximately 500 ms post stimulus offset. At approximately 500 ms post stimulus offset, α ERD began to decrease, giving way to



α ERS in both discrimination conditions (Qdis and Ndis) that extended into low β frequencies. A *post hoc* comparison of the two discrimination conditions (Qdis and Ndis) revealed no significant differences between conditions. The right hemisphere temporal cluster showed the same patterns of α ERD fading to ERS as the left temporal cluster, though with weaker spectral power. *Post hoc* comparisons of Qdis and Ndis to the PasN condition produced identical results to those found in the left temporal cluster.

Discussion

The first hypothesis (that ICA would identify bilateral α clusters localized to auditory association regions) was well supported by the data obtained from the left hemisphere. This finding is consistent with the MEG/EEG findings of Weisz et al. (2011)

who also found evidence of an independent auditory α rhythm. Clusters of neural activity emanating from these regions with $<20\%$ unexplained RV were localized to the pSTG via both ECD and CSD localization methods (although their exact source averages varied by ~ 1 cm). This localization is in agreement with previous findings of auditory α oscillatory activity identified via EEG/ICA (Bowers et al., 2014) and MEG (Muller and Weisz, 2012) and is consistent with a left-hemisphere dominance for dorsal stream activity in speech-based tasks (Hickok and Poeppel, 2004). In the right hemisphere, the two localization techniques also produced source averages that were separated by ~ 1 cm. However, the average ECD source was in the pSTG, while the average CSD source was located slightly inferiorly in the pMTG, possibly highlighting the uncertainty of EEG source localization and a reduced role of the right hemisphere in speech processing.

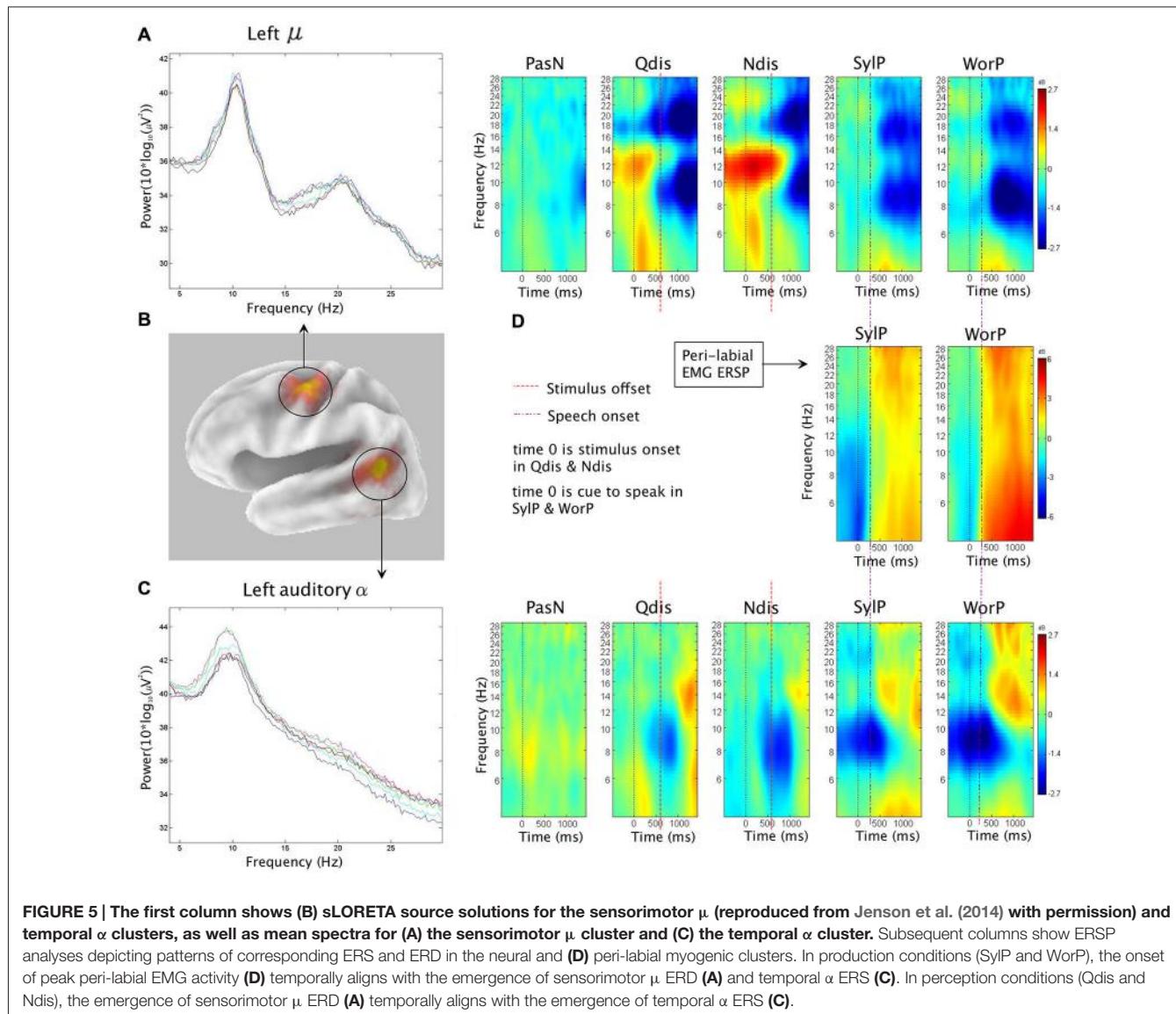


FIGURE 5 | The first column shows (B) sLORETA source solutions for the sensorimotor μ (reproduced from Jenson et al. (2014) with permission) and temporal α clusters, as well as mean spectra for (A) the sensorimotor μ cluster and (C) the temporal α cluster. Subsequent columns show ERSP analyses depicting patterns of corresponding ERS and ERD in the neural and (D) peri-labial myogenic clusters. In production conditions (SylP and WorP), the onset of peak peri-labial EMG activity (D) temporally aligns with the emergence of sensorimotor μ ERD (A) and temporal α ERS (C). In perception conditions (Qdis and Ndis), the emergence of sensorimotor μ ERD (A) temporally aligns with the emergence of temporal α ERS (C).

The finding that only 15/29 participants contributed to the clusters requires examination. Reasons for this include (1) the application of a standard head model reducing localization accuracy; (2) the location of auditory regions along the Sylvian fissure. As anatomic variability increases at greater distances from midline, the potential impact of a standard head model may have been maximal along the lateral and dorsal surfaces of the STG; (3) EEG's superior sensitivity to signals arising from cortical gyri rather than sulci; and (4) the fact that only components from pSTG and pMTG regions were included though all participants produced temporal lobe components. It should be noted that α activity has been observed across more anterior portions of the STG in addition to the pSTG (Weisz et al., 2011). However, due to the possibility that anterior and posterior regions of the STG perform functionally distinct tasks (Agnew et al., 2013; Chang et al., 2013; Houde and Chang, 2015) and that the current goal was to examine dorsal stream activity,

only components localized to posterior regions were included in this study. Therefore, it is likely that these inclusion criteria limited the number of contributors to the clusters. Even with the inherent limitations in source localization, the ERSP analyses produced significant changes in α spectral power across time in both speech perception and production conditions, providing evidence of auditory cortical deactivation that can be interpreted in light of current models.

Auditory α ERS/ERD in Speech Production

Auditory α activity in both production conditions (SylP, WorP) was characterized by ERD prior to production with no significant differences in activity between the two conditions. Before initiating speech, participants read the target and prepared to speak while attending to the visual cue to do so. Early activation of the auditory cortex (following stimulus presentation and prior to production) has previously been demonstrated during

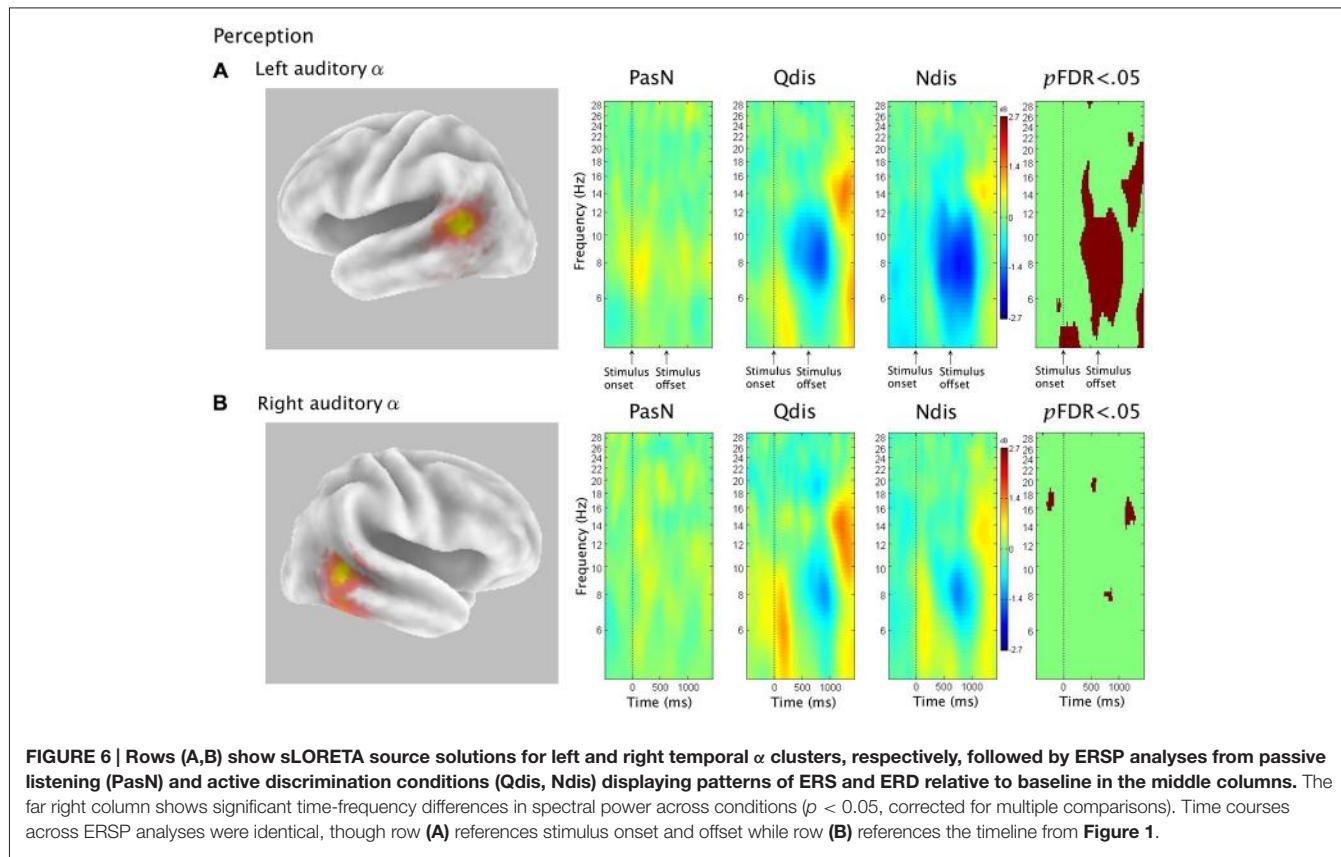


FIGURE 6 | Rows (A,B) show sLORETA source solutions for left and right temporal α clusters, respectively, followed by ERSP analyses from passive listening (PasN) and active discrimination conditions (Qdis, Ndis) displaying patterns of ERS and ERD relative to baseline in the middle columns. The far right column shows significant time-frequency differences in spectral power across conditions ($p < 0.05$, corrected for multiple comparisons). Time courses across ERSP analyses were identical, though row (A) references stimulus onset and offset while row (B) references the timeline from Figure 1.

silent and overt reading (Kell et al., 2011). Reduced levels of α activity in sensory regions prior to stimulus presentation have been shown to facilitate detection of near-threshold stimuli in the somatosensory domain (Weisz et al., 2014), and a similar mechanism in the auditory domain may facilitate monitoring of speech by an SMI loop. Additionally, α ERD also is known to result from simple increases in attention (Jensen and Mazaheri, 2010). Therefore, considering that the current speech production tasks required the coordination of multiple cognitive processes prior to speaking, all of which rely on attention to some extent, it is not possible to parse out the individual contributions of all cognitive processes to ERD prior to speech. Rather, it is likely that pre-speech α ERD resulted from contributions of attention, reading, and integration of auditory regions into an error-free SMI loop for speech.

The hypothesis that speech production would be characterized by increases in oscillatory α power was supported. A positive shift in auditory α power emerged concurrently with robust peri-labial EMG activity (i.e., muscle movement) that marked the initiation of speech. These patterns of neural and muscular activity were observed after ~ 300 ms (i.e., reaction time) from the cue to speak (time 0, Figure 4). It should be noted that these ERS/ERD changes were measured in reference to a “silent baseline” prior to each trial and that they were statistically significant when compared to a control passive listening condition in which little pSTG α activity was observed. As α ERS is associated with reduced activity (Laufs et al., 2003;

Brookes et al., 2005; Scheeringa et al., 2009; Mayhew et al., 2013), the current findings are consistent with speech-induced suppression (i.e., modulation of auditory cortical activity during speech production; Frith et al., 1991; Hirano et al., 1997; Curio et al., 2000; Houde et al., 2002; Heinks-Maldonado et al., 2005; Christoffels et al., 2007, 2011). However, it is also interesting to note that though auditory oscillatory activity was characterized by α spectra, the observed ERS spread into higher frequencies and may be somewhat consistent with ECoG and fMRI studies that have implicated modulations in auditory gamma frequencies during speech production (Greenlee et al., 2011; Agnew et al., 2013; Reznik et al., 2014).

It is also important to note that both myogenic and auditory activity during speech production is characterized by ERS, which may raise questions pertaining to the possibility of muscle activity contaminating neural activity. There are multiple reasons to refute this notion. First, if ICA was not able to adequately separate neural signals from myogenic artifact, muscle activity would have overwhelmed the α and β ERD recorded in the sensorimotor μ cluster (Jenson et al., 2014). Second, α ERS was noted in the perception conditions (coinciding with periods of covert rehearsal—see below), during which no overt response was required. Together, this evidence suggests that α ERS resulted from neural activity as opposed to myogenic artifact.

The larger picture of dorsal stream activity in speech production becomes apparent when the current data are viewed alongside those of Jenson et al. (2014). Data from the same

participants in the same conditions showed sensorimotor μ ERD (i.e., disinhibition) beginning with muscle movement in speech production. Thus, when viewed together, sensorimotor disinhibition and auditory inhibition coincided with speech production (as indicated by EMG activity; see **Figure 5**). Current models of SMI for speech indicate that the sensorimotor loop is initiated by the generation of a motor plan in PMC (Tian and Poeppel, 2010; Houde and Nagarajan, 2011; Jenson et al., 2014). Concurrent with the delivery of this motor plan to primary motor cortex (M1), an efference copy of the expected sensory consequences is sent to auditory regions for comparison with the goals and outcomes of the movements. Any deviations from expectations are detected and corrective feedback is sent to the PMC. As true auditory and somatosensory reafferent feedback is received, this information is also integrated into feedback to the PMC (Tian and Poeppel, 2010). During normal error-free production, articulatory predictions are matched to the available sensory (i.e., auditory) information, resulting in the observed net deactivation in auditory association regions. Based on this model, it was not surprising to see near perfect temporal concordance between $\alpha/\beta \mu$ ERD, auditory α ERS, and peri-labial EMG activity in normal, unperturbed speech production. Thus, the results of the current study demonstrate inhibition of auditory regions during overt speech, consistent with the suppression that would be expected based on the delivery of an efference copy from the PMC.

Auditory α ERS/ERD in Accurate Speech Discrimination

In both the quiet (Qdis) and noisy (Ndis) discrimination conditions, participants listened to pairs of syllables and then waited ~ 1400 ms to make an active same/different discrimination response. Considering that only correct responses were analyzed, the following interpretations of the oscillatory data are made in reference to accurate speech discrimination. Auditory activity prior to and during stimulus onset was characterized by α ERD. During this same time period, β ERD was observed within the μ components localized to anterior regions of the dorsal stream (e.g., PMC; Jenson et al., 2014). This pattern of PMC β ERD activity has previously been explained as early predictive coding (i.e., hypothesis generation via internal modeling) followed by hypothesis testing via auditory to motor integration (Alho et al., 2014), according to analysis by synthesis theories (Stevens and Halle, 1967). The current data from auditory regions which indicate increased auditory activity prior to and during stimulus presentation continue to support this interpretation, though it is necessary to consider how a predictive coding explanation might be favored over one of simple attention, which has also been known to modulate β activity in cognitive tasks (van Ede et al., 2014). Participants were briefed on the task prior to each discrimination condition and therefore knew what to expect. In addition, only four syllable pairs were possible. Therefore, across 80 trials per condition, it is likely that participants were able to formulate general internal models of the expected stimuli (i.e., syllables) to help constrain the upcoming sensory analysis. Further support for speech-related predictive coding comes from Bowers et al. (2013),

who reported early $\mu \beta$ ERD in similar syllable but not tone discrimination tasks.

The time period following stimulus offset and prior to the response was characterized by temporal α ERS, similar to that observed in the production conditions. Jenson et al. (2014) found sensorimotor μ ERD in the same time period. They interpreted this as evidence of covert rehearsal, during which the syllables were held in working memory to facilitate accurate discrimination and response. The current findings support this explanation. Clearly, it can be seen in **Figure 5** that in both Qdis and Ndis conditions, sensorimotor μ ERD is aligned temporally to auditory α ERS; a pattern similar to that observed in the production conditions. It should be noted that covert rehearsal has been used to explain motor activity sometimes observed in speech perception tasks (Burton et al., 2000; Baddeley, 2003; Jenson et al., 2014; Roa Romero et al., 2015). However, this assertion lacks support without temporal data showing when activity occurred relative to stimulus onset and offset. By demonstrating anterior sensorimotor disinhibition aligned with auditory inhibition following stimulus offset in the absence of peri-labial EMG activity, these data support the theory that covert rehearsal can account for some of the motor activity observed during accurate speech discrimination tasks such as these (Burton et al., 2000; Wilson et al., 2004; Callan et al., 2006, 2010; Bowers et al., 2013; Jenson et al., 2014). However, it should be noted, as indicated above, that covert rehearsal may not be the only explanation for this activity. Prior to and during syllable discrimination there is evidence of anterior sensorimotor activity characterizing internal modeling (Bowers et al., 2013, 2014; Jenson et al., 2014).

While the presence of anterior dorsal stream activity during covert production is relatively well established (Neuper et al., 2006; Gehrig et al., 2012; Jenson et al., 2014), it remains unclear how a subtractive mechanism in posterior dorsal stream regions could function in the absence of re-afferent feedback. It has been proposed that auditory inhibition is linked to the delivery of an efference copy. As no overt production took place during the discrimination conditions, it may seem surprising that the auditory cluster should demonstrate α ERS during covert rehearsal as the observed auditory suppression is thought to be contingent on efference copy delivery linked to motor plan execution. These findings are not without precedent, however, as lip-reading (Kauramäki et al., 2010; Balk et al., 2013) and covert speech (Tian and Poeppel, 2015) have been shown to reduce auditory cortical responses. There is also evidence that failure of this sensory suppression in covert productions may be associated with some of the positive symptoms (i.e., auditory hallucinations) of schizophrenia (Ford et al., 2001; Ford and Mathalon, 2004, 2005), indicating that auditory suppression during covert production is critical to normal function. One possible explanation for auditory suppression when efference copy delivery is questionable is that in the absence of an efference copy, auditory suppression may be based on higher order processes (Crapse and Sommer, 2008). In line with this explanation, sensory inhibition has recently been linked to motor intention prior to overt activity (Stenner et al., 2014). Additionally, it is possible that during covert production,

auditory suppression may be based on the comparison of sensory predictions to a higher-level sensory goal. This is consistent with a recent proposal by Skipper (2014), who suggested that the auditory hypotheses being evaluated are internal in nature based on context and prior experience rather than being dependent on available acoustic information. However, further investigation is warranted to better explain how auditory inhibition can occur during covert speech processing.

Summary

These results illustrate how fluctuations in oscillatory power in time characterize posterior dorsal stream activity across speech perception and production. Viewing these data alongside activity from anterior dorsal regions (Jenson et al., 2014; **Figure 5**) provides a window for understanding the temporal dynamics of dorsal stream activity in speech discrimination and production events. Prior to speech production, the pSTG is active (as evidenced by α ERD) and primed to receive speech. In this experiment, α ERD occurred as participants read the stimuli to be produced. As speech was initiated, anterior dorsal regions (e.g., PMC, μ ERD) became active as activity in the pSTG was attenuated (α ERS). The patterns of oscillatory activity across these cortical regions aligned temporally with muscle activity and are suggestive of auditory suppression arising from an efference copy driven sensorimotor loop that enables online monitoring during normal speech production (Guenther et al., 2006; Tian and Poeppel, 2010, 2015; Hickok et al., 2011; Houde and Nagarajan, 2011; Arnal and Giraud, 2012; Hickok, 2012). However, this interpretation is based solely on descriptions of the strength and timing of activity across regions and should be made with caution. Connectivity measures across these regions are necessary to provide more direct evidence of the efference copy mechanism.

Perhaps not surprisingly, dorsal stream activity in accurate speech discrimination is more complex. Prior to stimulus onset, anterior dorsal regions (e.g., PMC) are active, most likely reflecting the recruitment of motor/attentional mechanisms for internal modeling that help constrain the ensuing auditory analysis. Both anterior and posterior regions of the dorsal stream become active while stimuli are presented (evident by μ ERD), most likely indicative of hypothesis testing (analysis by synthesis). Finally, following stimulus offset and in the absence of reafferent feedback from overt production, activity in anterior dorsal regions is further enhanced (μ ERD), while activity in posterior regions (temporal α ERS) is suppressed. This pattern of late dorsal stream activity is similar to that observed during speech production and indicative of covert rehearsal following stimulus offset, potentially driven by efference copy. Based on these findings, it is feasible that the dorsal stream plays a variety of roles across the time course of stimuli expectancy, presentation, and rehearsal to facilitate accurate perception. However, because there were insufficient data from inaccurate discrimination trials for comparison, it is not currently possible to determine the extent to which each of these processes individually contributes to perceptual acuity. Oscillatory fluctuations reflecting activation changes across the time course of speech discrimination suggests a dynamic rather

than static role for the dorsal stream. Activity before, during, and after stimulus presentation may be explained as internal modeling, analysis by synthesis (or perhaps direct realism), and covert production, respectively. Taken together, the results converge on a dynamic constructivist perspective espousing the notion that speech discrimination is facilitated by embodied articulatory representations, attention, experience, and short-term memory (Callan et al., 2006, 2010, 2014; Galantucci et al., 2006; Bowers et al., 2013, 2014; Jenson et al., 2014).

Limitations and Future Directions

While the results of this study provide compelling evidence that the neural dynamics of the temporal α oscillator and the sensorimotor μ rhythm work in synchrony to accomplish online monitoring of speech in production and hypothesis testing in perception, certain limitations should be addressed. The source localization of auditory α clusters should be interpreted with caution based on the inherent uncertainty of source localization when performing EEG with 68 electrodes. In this study, the uncertainty was illustrated by the difference between ECD and CSD source localizations. The hypothesized communication between these two clusters of independent components in this study is based purely on temporally aligned concordant patterns of ERS/ERD. While it is clear that these regions are co-active in time periods that could support a sensorimotor feedback loop, direct transfer of information cannot be inferred solely on the basis of the data presented. Direct evidence for cortico-cortical communication between these two regions during speech perception and production requires further analysis with a measure that is able to assess coherence between cortical regions (e.g., any frequency-sensitive variant of Granger causality). However, such connectivity analyses are beyond the scope of this paper. In addition, analyses in the current study were restricted (according to our hypotheses) to activity in the α band, which characterized the spectrum of the components in the region of interest. However, there was also evidence of differential activity in other frequencies (e.g., theta-gamma nesting; Giraud and Poeppel, 2012), though their analysis was beyond the scope of the current study.

It should also be noted that the tasks used in this study (discrimination and production of syllable pairs in isolation) may not be representative of normal human communication (Hickok and Poeppel, 2000; Skipper, 2014). They lack normal audiovisual and semantic contextual cues, potentially requiring greater levels of processing than would be required in normal communicative situations. Despite these potential shortcomings, the temporal alignment of sensorimotor μ ERD, peri-labial EMG activity, and temporal α ERS strongly suggests the presence of a sensorimotor feedback loop for online monitoring and hypothesis testing, and warrants further investigation with methods able to establish cortico-cortical communication. Finally, deeper understanding of typical sensorimotor activity for speech enables the analysis and description of neural activity in clinical populations such as individuals who stutter, in whom auditory regions are found to show even greater deactivation than normal, possibly due to compromised SMI during speech production (Max et al., 2003; Brown et al., 2005; Watkins et al., 2008).

Conclusion

ICA identified components over auditory association cortices with expected characteristic α spectra. ERSP analysis of temporal α components demonstrated reduced activity concurrent with periods of overt and covert production. These findings demonstrate the utility of the ICA/ERSP analysis for localizing and temporally delineating neural activity in speech events. The temporal alignment of auditory α ERS, sensorimotor μ ERD, and peri-labial EMG activity in production tasks supports previous interpretations of temporal α ERS indexing a relative deactivation of auditory regions which may possibly

be attributed to an efference copy mechanism involved in the online monitoring of ongoing speech. In perception conditions, the synchrony of temporal α ERS and sensorimotor μ ERD likely represent a similar mechanism as subjects engaged in covert rehearsal of syllable pairs held in working memory. These observed phenomena reflecting the interactions of multiple dorsal stream regions may provide a framework for describing normal speech-related sensorimotor activity. The non-invasive and cost-effective nature of the technique supports its continued application to investigating neural network dynamics in normal and clinical populations of all ages.

References

- Agnew, Z. K., McGettigan, C., Banks, B., and Scott, S. K. (2013). Articulatory movements modulate auditory responses to speech. *Neuroimage* 73, 191–199. doi: 10.1016/j.neuroimage.2012.08.020
- Arnal, L. H., and Giraud, A. L. (2012). Cortical oscillations and sensory predictions. *Trends Cogn. Sci.* 16, 390–398. doi: 10.1016/j.tics.2012.05.003
- Alho, J., Lin, F. H., Sato, M., Tiitinen, H., Sams, M., and Jaaskelainen, I. P. (2014). Enhanced neural synchrony between left auditory and premotor cortex is associated with successful phonetic categorization. *Front. Psychol.* 5:394. doi: 10.3389/fpsyg.2014.00394
- Baddeley, A. (2003). Working memory and language: an overview. *J. Commun. Disord.* 36, 189–208. doi: 10.1016/s0021-9924(03)00019-4
- Balk, M. H., Kari, H., Kauramäki, J., Ahveninen, J., Sams, M., Autti, T., et al. (2013). Silent lipreading and covert speech production suppress processing of non-linguistic sounds in auditory cortex. *Open J. Neurosci.* 3, 1–17. doi: 10.13055/ojns_3_1_1.130206
- Behroozmand, R., and Larson, C. R. (2011). Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. *BMC Neurosci.* 12:54. doi: 10.1186/1471-2202-12-54
- Benjamini, Y., and Hochberg, Y. (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *J. Educ. Behav. Stat.* 25, 60–83. doi: 10.3102/10769986025001060
- Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., and Ward, B. D. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nat. Neurosci.* 7, 295–301. doi: 10.1038/nn1198
- Blakemore, S. J., Wolpert, D., and Frith, C. (2000). Why can't you tickle yourself? *Neuroreport* 11, R11–R16. doi: 10.1097/00001756-200008030-00002
- Blockcolskey, V., Frazer, J., and Frazer, D. (2008). *40,000 Selected Words*. San Antonio, TX: Communication Skill Builders.
- Bowers, A., Saltuklaroglu, T., Harkrider, A., and Cuellar, M. (2013). Suppression of the μ rhythm during speech and non-speech discrimination revealed by independent component analysis: implications for sensorimotor integration in speech processing. *PLoS One* 8:e72024. doi: 10.1371/journal.pone.0072024
- Bowers, A. L., Saltuklaroglu, T., Harkrider, A., Wilson, M., and Toner, M. A. (2014). Dynamic modulation of shared sensory and motor cortical rhythms mediates speech and non-speech discrimination performance. *Front. Psychol.* 5:366. doi: 10.3389/fpsyg.2014.00366
- Brookes, M. J., Gibson, A. M., Hall, S. D., Furlong, P. L., Barnes, G. R., Hillebrand, A., et al. (2005). GLM-beamformer method demonstrates stationary field, alpha ERD and gamma ERS co-localisation with fMRI BOLD response in visual cortex. *Neuroimage* 26, 302–308. doi: 10.1016/j.neuroimage.2005.01.050
- Brown, S., Ingham, R. J., Ingham, J. C., Laird, A. R., and Fox, P. T. (2005). Stuttered and fluent speech production: an ALE meta-analysis of functional neuroimaging studies. *Hum. Brain Mapp.* 25, 105–117. doi: 10.1002/hbm.20140
- Brunner, C., Delorme, A., and Makeig, S. (2013). Eeglab - an open source matlab toolbox for electrophysiological research. *Biomed. Tech. (Berl)* doi: 10.1515/bmt-2013-4182 [Epub ahead of print].
- Burton, M. W., Small, S. L., and Blumstein, S. E. (2000). The role of segmentation in phonological processing: an fMRI investigation. *J. Cogn. Neurosci.* 12, 679–690. doi: 10.1162/089892900562309
- Buszaki, G. (2006). *Rhythms of the Brain*. New York, NY: Oxford University Press.
- Callan, A. M., Callan, D. E., Tajima, K., and Akahane-Yamada, R. (2006). Neural processes involved with perception of non-native durational contrasts. *Neuroreport* 17, 1353–1357. doi: 10.1097/01.wnr.0000224774.66904.29
- Callan, D., Callan, A., Gamez, M., Sato, M. A., and Kawato, M. (2010). Premotor cortex mediates perceptual performance. *Neuroimage* 51, 844–858. doi: 10.1016/j.neuroimage.2010.02.027
- Callan, D. E., Callan, A. M., Honda, K., and Masaki, S. (2000). Single-sweep EEG analysis of neural processes underlying perception and production of vowels. *Brain Res. Cogn. Brain Res.* 10, 173–176. doi: 10.1016/s0926-6410(00)0025-2
- Callan, D. E., Jones, J. A., and Callan, A. (2014). Multisensory and modality specific processing of visual speech in different regions of the premotor cortex. *Front. Psychol.* 5:389. doi: 10.3389/fpsyg.2014.00389
- Chang, E. F., Niziolek, C. A., Knight, R. T., Nagarajan, S. S., and Houde, J. F. (2013). Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proc. Natl. Acad. Sci. U S A* 110, 2653–2658. doi: 10.1073/pnas.1216827110
- Christoffels, I. K., Formisano, E., and Schiller, N. O. (2007). Neural correlates of verbal feedback processing: an fMRI study employing overt speech. *Hum. Brain Mapp.* 28, 868–879. doi: 10.1002/hbm.20315
- Christoffels, I. K., van de Ven, V., Waldorp, L. J., Formisano, E., and Schiller, N. O. (2011). The sensory consequences of speaking: parametric neural cancellation during speech in auditory cortex. *PLoS One* 6:e18307. doi: 10.1371/journal.pone.0018307
- Crapse, T. B., and Sommer, M. A. (2008). Corollary discharge across the animal kingdom. *Nat. Rev. Neurosci.* 9, 587–600. doi: 10.1038/nrn2457
- Cuellar, M., Bowers, A., Harkrider, A. W., Wilson, M., and Saltuklaroglu, T. (2012). Mu suppression as an index of sensorimotor contributions to speech processing: evidence from continuous EEG signals. *Int. J. Psychophysiol.* 85, 242–248. doi: 10.1016/j.ijpsycho.2012.04.003
- Curio, G., Neuloh, G., Numminen, J., Jousmäki, V., and Hari, R. (2000). Speaking modifies voice-evoked activity in the human auditory cortex. *Hum. Brain Mapp.* 9, 183–191. doi: 10.1002/(sici)1097-0193(200004)9:4<183::aid-hbm1>3.0.co;2-z
- Delorme, A., Palmer, J., Onton, J., Oostenveld, R., and Makeig, S. (2012). Independent EEG sources are dipolar. *PLoS One* 7:e30135. doi: 10.1371/journal.pone.0030135
- Ford, J. M., and Mathalon, D. H. (2004). Electrophysiological evidence of corollary discharge dysfunction in schizophrenia during talking and thinking. *J. Psychiatr. Res.* 38, 37–46. doi: 10.1016/s0022-3956(03)00095-5

- Ford, J. M., and Mathalon, D. H. (2005). Corollary discharge dysfunction in schizophrenia: can it explain auditory hallucinations? *Int. J. Psychophysiol.* 58, 179–189. doi: 10.1016/j.ijpsycho.2005.01.014
- Ford, J. M., Mathalon, D. H., Kalba, S., Whitfield, S., Faustman, W. O., and Roth, W. T. (2001). Cortical responsiveness during talking and listening in schizophrenia: an event-related brain potential study. *Biol. Psychiatry* 50, 540–549. doi: 10.1016/s0006-3223(01)01166-0
- Frey, J. N., Mainy, N., Lachaux, J. P., Müller, N., Bertrand, O., and Weisz, N. (2014). Selective modulation of auditory cortical alpha activity in an audiovisual spatial attention task. *J. Neurosci.* 34, 6634–6639. doi: 10.1523/JNEUROSCI.4813-13.2014
- Frith, C. D., Friston, K. J., Liddle, P. F., and Frackowiak, R. S. (1991). A PET study of word finding. *Neuropsychologia* 29, 1137–1148. doi: 10.1016/0028-3932(91)90029-8
- Galantucci, B., Fowler, C. A., and Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychon. Bull. Rev.* 13, 361–377. doi: 10.3758/bf03193857
- Gehrig, J., Wibral, M., Arnold, C., and Kell, C. A. (2012). Setting up the speech production network: how oscillations contribute to lateralized information routing. *Front. Psychol.* 3:169. doi: 10.3389/fpsyg.2012.00169
- Giraud, A. L., and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* 15, 511–517. doi: 10.1038/nrn.3063
- Graimann, B., and Pfurtscheller, G. (2006). Quantification and visualization of event-related changes in oscillatory brain activity in the time-frequency domain. *Prog. Brain Res.* 159, 79–97. doi: 10.1016/s0079-6123(06)59006-5
- Greenlee, J. D., Jackson, A. W., Chen, F., Larson, C. R., Oya, H., Kawasaki, H., et al. (2011). Human auditory cortical activation during self-vocalization. *PLoS One* 6:e14744. doi: 10.1371/journal.pone.0014744
- Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *J. Commun. Disord.* 39, 350–365. doi: 10.1016/j.jcomdis.2006.06.013
- Guenther, F. H., and Hickok, G. (2015). Role of the auditory system in speech production. *Handb. Clin. Neurol.* 129, 161–175. doi: 10.1016/b978-0-444-62630-1.00009-3
- Guenther, F. H., Ghosh, S. S., and Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang.* 96, 280–301. doi: 10.1016/j.bandl.2005.06.001
- Guenther, F. H., and Vladusich, T. (2012). A neural theory of speech acquisition and production. *J. Neurolinguistics* 25, 408–422. doi: 10.1016/j.jneuroling.2009.08.006
- Hari, R. (2006). Action-perception connection and the cortical mu rhythm. *Prog. Brain Res.* 159, 253–260. doi: 10.1016/s0079-6123(06)59017-x
- Heinks-Maldonado, T. H., Mathalon, D. H., Gray, M., and Ford, J. M. (2005). Fine-tuning of auditory cortex during speech production. *Psychophysiology* 42, 180–190. doi: 10.1111/j.1469-8986.2005.00272.x
- Herman, A. B., Houde, J. F., Vinogradov, S., and Nagarajan, S. S. (2013). Parsing the phonological loop: activation timing in the dorsal speech stream determines accuracy in speech reproduction. *J. Neurosci.* 33, 5439–5453. doi: 10.1523/JNEUROSCI.1472-12.2013
- Hickok, G. (2012). The cortical organization of speech processing: feedback control and predictive coding the context of a dual-stream model. *J. Commun. Disord.* 45, 393–402. doi: 10.1016/j.jcomdis.2012.06.004
- Hickok, G., and Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends Cogn. Sci.* 4, 131–138. doi: 10.1016/s1364-6613(00)01463-7
- Hickok, G., and Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 92, 67–99. doi: 10.1016/j.cognition.2003.10.011
- Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402. doi: 10.1038/nrn2113
- Hickok, G., Houde, J., and Rong, F. (2011). Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron* 69, 407–422. doi: 10.1016/j.neuron.2011.01.019
- Hirano, S., Kojima, H., Naito, Y., Honjo, I., Kamoto, Y., Okazawa, H., et al. (1997). Cortical processing mechanism for vocalization with auditory verbal feedback. *Neuroreport* 8, 2379–2382. doi: 10.1097/00001756-199707070-00055
- Houde, J. F., and Chang, E. F. (2015). The cortical computations underlying feedback control in vocal production. *Curr. Opin. Neurobiol.* 33, 174–181. doi: 10.1016/j.conb.2015.04.006
- Houde, J. F., and Nagarajan, S. S. (2011). Speech production as state feedback control. *Front. Hum. Neurosci.* 5:82. doi: 10.3389/fnhum.2011.00082
- Houde, J. F., Nagarajan, S. S., Sekihara, K., and Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: an MEG study. *J. Cogn. Neurosci.* 14, 1125–1138. doi: 10.1162/089892902760807140
- Jasper, H. H. (1958). The ten twenty electrode system of the international federation. *Electroencephalogr. Clin. Neurophysiol.* 10, 371–375.
- Jensen, O., and Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Front. Hum. Neurosci.* 4:186. doi: 10.3389/fnhum.2010.00186
- Jenson, D., Bowers, A. L., Harkrider, A. W., Thornton, D., Cuellar, M., and Saltuklaroglu, T. (2014). Temporal dynamics of sensorimotor integration in speech perception and production: independent component analysis of EEG data. *Front. Psychol.* 5:656. doi: 10.3389/fpsyg.2014.00656
- Jokisch, D., and Jensen, O. (2007). Modulation of gamma and alpha activity during a working memory task engaging the dorsal or ventral stream. *J. Neurosci.* 27, 3244–3251. doi: 10.1523/jneurosci.5399-06.2007
- Kauramäki, J., Jääskeläinen, I. P., Hari, R., Möttönen, R., Rauschecker, J. P., and Sams, M. (2010). Lipreading and covert speech production similarly modulate human auditory-cortex responses to pure tones. *J. Neurosci.* 30, 1314–1321. doi: 10.1523/JNEUROSCI.1950-09.2010
- Kell, C. A., Morillon, B., Kouneiher, F., and Giraud, A. L. (2011). Lateralization of speech production starts in sensory cortices—a possible sensory origin of cerebral left dominance for speech. *Cereb. Cortex* 21, 932–937. doi: 10.1093/cercor/bhq167
- Klimesch, W., Sauseng, P., and Hanslmayr, S. (2007). EEG alpha oscillations: the inhibition-timing hypothesis. *Brain Res. Rev.* 53, 63–88. doi: 10.1016/j.brainres.2006.06.003
- Kort, N. S., Nagarajan, S. S., and Houde, J. F. (2014). A bilateral cortical network responds to pitch perturbations in speech feedback. *Neuroimage* 86, 525–535. doi: 10.1016/j.neuroimage.2013.09.042
- Laufs, H., Kleinschmidt, A., Beyerle, A., Eger, E., Salek-Haddadi, A., Preibisch, C., et al. (2003). EEG-correlated fMRI of human alpha activity. *Neuroimage* 19, 1463–1476. doi: 10.1016/s1053-8119(03)00286-6
- Lee, T. W., Girolami, M., and Sejnowski, T. J. (1999). Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural Comput.* 11, 417–441. doi: 10.1162/08997669930016719
- Lehtela, L., Salmelin, R., and Hari, R. (1997). Evidence for reactive magnetic 10-Hz rhythm in the human auditory cortex. *Neurosci. Lett.* 222, 111–114. doi: 10.1016/s0304-3940(97)13361-4
- Lin, C. T., Chuang, S. W., Chen, Y. C., Ko, L. W., Liang, S. F., and Jung, T. P. (2007). EEG effects of motion sickness induced in a dynamic virtual reality environment. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2007, 3872–3875. doi: 10.1109/emb.2007.4353178
- Lin, C. T., Chen, S. A., Chiu, T. T., Lin, H. Z., and Ko, L. W. (2011). Spatial and temporal EEG dynamics of dual-task driving performance. *J. Neuroeng. Rehabil.* 8:11. doi: 10.1186/1743-0003-8-11
- LoCastro, P. C., Krebs-Noble, D., Gullapalli, R. P., and Burton, M. W. (2004). An fMRI investigation of speech and tone segmentation. *J. Cogn. Neurosci.* 16, 1612–1624. doi: 10.1162/0898929042568433
- Makeig, S., Debener, S., Onton, J., and Delorme, A. (2004). Mining event-related brain dynamics. *Trends Cogn. Sci.* 8, 204–210. doi: 10.1016/j.tics.2004.03.008
- Max, L., Caruso, A. J., and Gracco, V. L. (2003). Kinematic analyses of speech, orofacial non-speech and finger movements in stuttering and non-stuttering adults. *J. Speech Lang. Hear. Res.* 46, 215–232. doi: 10.1044/1092-4388(2003/017)
- Mayhew, S. D., Ostwald, D., Porcaro, C., and Bagshaw, A. P. (2013). Spontaneous EEG alpha oscillation interacts with positive and negative BOLD responses

- in the visual-auditory cortices and default-mode network. *Neuroimage* 76, 362–372. doi: 10.1016/j.neuroimage.2013.02.070
- McGarry, L. M., Russo, F. A., Schalles, M. D., and Pineda, J. A. (2012). Audio-visual facilitation of the mu rhythm. *Exp. Brain Res.* 218, 527–538. doi: 10.1007/s00221-012-3046-3
- McMenamin, B. W., Shackman, A. J., Greischar, L. L., and Davidson, R. J. (2011). Electromyogenic artifacts and electroencephalographic inferences revisited. *Neuroimage* 54, 4–9. doi: 10.1016/j.neuroimage.2010.07.057
- McMenamin, B. W., Shackman, A. J., Maxwell, J. S., Bachhuber, D. R., Koppenhaver, A. M., Greischar, L. L., et al. (2010). Validation of ICA-based myogenic artifact correction for scalp and source-localized EEG. *Neuroimage* 49, 2416–2432. doi: 10.1016/j.neuroimage.2009.10.010
- Muller, N., and Weisz, N. (2012). Lateralized auditory cortical alpha band activity and interregional connectivity pattern reflect anticipation of target sounds. *Cereb. Cortex* 22, 1604–1613. doi: 10.1093/cercor/bhr232
- Neuper, C., Wörtz, M., and Pfurtscheller, G. (2006). ERD/ERS patterns reflecting sensorimotor activation and deactivation. *Prog. Brain Res.* 159, 211–222. doi: 10.1016/s0079-6123(06)59014-4
- Niziolek, C. A., Nagarajan, S. S., and Houde, J. F. (2013). What does motor efference copy represent? Evidence from speech production. *J. Neurosci.* 33, 16110–16116. doi: 10.1523/JNEUROSCI.2137-13.2013
- Numminen, J., and Curio, G. (1999). Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex. *Neurosci. Lett.* 272, 29–32. doi: 10.1016/s0304-3940(99)00573-x
- Oken, B. S. (1986). Filtering and aliasing of muscle activity in EEG frequency analysis. *Electroencephalogr. Clin. Neurophysiol.* 64, 77–80. doi: 10.1016/0013-4694(86)90045-3
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4
- Onton, J., Westerfield, M., Townsend, J., and Makeig, S. (2006). Imaging human EEG dynamics using independent component analysis. *Neurosci. Biobehav. Rev.* 30, 808–822. doi: 10.1016/j.neubiorev.2006.06.007
- Oostenveld, R., and Oostendorp, T. F. (2002). Validating the boundary element method for forward and inverse EEG computations in the presence of a hole in the skull. *Hum. Brain Mapp.* 17, 179–192. doi: 10.1002/hbm.10061
- Osnes, B., Hugdahl, K., and Specht, K. (2011). Effective connectivity analysis demonstrates involvement of premotor cortex during speech perception. *Neuroimage* 54, 2437–2445. doi: 10.1016/j.neuroimage.2010.09.078
- Pascual-Marqui, R. D. (2002). Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. *Methods Find. Exp. Clin. Pharmacol.* 24, 5–12.
- Rauschecker, J. P. (2012). Ventral and dorsal streams in the evolution of speech and language. *Front. Evol. Neurosci.* 4:7. doi: 10.3389/fnevo.2012.00007
- Reznik, D., Henkin, Y., Schadel, N., and Mukamel, R. (2014). Lateralized enhancement of auditory cortex activity and increased sensitivity to self-generated sounds. *Nat. Commun.* 5:4059. doi: 10.1038/ncomms5059
- Roa Romero, Y., Senkowski, D., and Keil, J. (2015). Early and late beta-band power reflect audiovisual perception in the McGurk illusion. *J. Neurophysiol.* 113, 2342–2350. doi: 10.1152/jn.00783.2014
- Rogalsky, C., Matchin, W., and Hickok, G. (2008). Broca's area, sentence comprehension and working memory: an fMRI Study. *Front. Hum. Neurosci.* 2:14. doi: 10.3389/neuro.09.014.2008
- Scheeringa, R., Petersson, K. M., Oostenveld, R., Norris, D. G., Hagoort, P., and Bastiaansen, M. C. (2009). Trial-by-trial coupling between EEG and BOLD identifies networks related to alpha and theta EEG power increases during working memory maintenance. *Neuroimage* 44, 1224–1238. doi: 10.1016/j.neuroimage.2008.08.041
- Schnitzler, A., and Gross, J. (2005). Normal and pathological oscillatory communication in the brain. *Nat. Rev. Neurosci.* 6, 285–296. doi: 10.1038/nrn1650
- Scott, S. K., McGetigan, C., and Eisner, F. (2009). A little more conversation, a little less action--candidate roles for the motor cortex in speech perception. *Nat. Rev. Neurosci.* 10, 295–302. doi: 10.1038/nrn2603
- Shackman, A. J., McMenamin, B. W., Slagter, H. A., Maxwell, J. S., Greischar, L. L., and Davidson, R. J. (2009). Electromyogenic artifacts and electroencephalographic inferences. *Brain Topogr.* 22, 7–12. doi: 10.1007/s10548-009-0079-4
- Sitek, K. R., Mathalon, D. H., Roach, B. J., Houde, J. F., Niziolek, C. A., and Ford, J. M. (2013). Auditory cortex processes variation in our own speech. *PLoS One* 8:e82925. doi: 10.1371/journal.pone.0082925
- Skipper, J. I. (2014). Echoes of the spoken past: how auditory cortex hears context during speech perception. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130297. doi: 10.1098/rstb.2013.0297
- Sohoglu, E., Peelle, J. E., Carlyon, R. P., and Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *J. Neurosci.* 32, 8443–8453. doi: 10.1523/JNEUROSCI.5069-11.2012
- Specht, K. (2014). Neuronal basis of speech comprehension. *Hear. Res.* 307, 121–135. doi: 10.1016/j.heares.2013.09.011
- Stenner, M. P., Bauer, M., Heinze, H. J., Haggard, P., and Dolan, R. J. (2014). Parallel processing streams for motor output and sensory prediction during action preparation. *J. Neurophysiol.* 113, 1752–1762. doi: 10.1152/jn.00616.2014
- Stevens, K. N., and Halle, M. (1967). *Remarks on Analysis by Synthesis and Distinctive Features*. Cambridge, MA: MIT press.
- Stone, J. V. (2004). *Independent Component Analysis*. Cambridge, MA: Wiley Online Library.
- Szenkovits, G., Peelle, J. E., Norris, D., and Davis, M. H. (2012). Individual differences in premotor and motor recruitment during speech perception. *Neuropsychologia* 50, 1380–1392. doi: 10.1016/j.neuropsychologia.2012.02.023
- Tian, X., and Poeppel, D. (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. *Front. Psychol.* 1:166. doi: 10.3389/fpsyg.2010.00166
- Tian, X., and Poeppel, D. (2015). Dynamics of self-monitoring and error detection in speech production: evidence from mental imagery and MEG. *J. Cogn. Neurosci.* 27, 352–364. doi: 10.1162/jocn_a_00692
- Tiihonen, J., Hari, R., Kajola, M., Karhu, J., Ahlfors, S., and Tissari, S. (1991). Magnetoencephalographic 10-Hz rhythm from the human auditory cortex. *Neurosci. Lett.* 129, 303–305. doi: 10.1016/0304-3940(91)90486-d
- Tourville, J. A., and Guenther, F. H. (2011). The DIVA model: a neural theory of speech acquisition and production. *Lang. Cogn. Process.* 26, 952–981. doi: 10.1080/01690960903498424
- Towle, V. L., Bolaños, J., Suarez, D., Tan, K., Grzeszczuk, R., Levin, D. N., et al. (1993). The spatial location of EEG electrodes: locating the best-fitting sphere relative to cortical anatomy. *Electroencephalogr. Clin. Neurophysiol.* 86, 1–6. doi: 10.1016/0013-4694(93)90061-y
- van Dijk, H., Nieuwenhuis, I. L., and Jensen, O. (2010). Left temporal alpha band activity increases during working memory retention of pitches. *Eur. J. Neurosci.* 31, 1701–1707. doi: 10.1111/j.1460-9568.2010.07227.x
- van Ede, F., Szűbenyi, S., and Maris, E. (2014). Attentional modulations of somatosensory alpha, beta and gamma oscillations dissociate between anticipation and stimulus processing. *Neuroimage* 97, 134–141. doi: 10.1016/j.neuroimage.2014.04.047
- Venezia, J. H., Saberi, K., Chubb, C., and Hickok, G. (2012). Response bias modulates the speech motor system during syllable discrimination. *Front. Psychol.* 3:157. doi: 10.3389/fpsyg.2012.00157
- von Holst, E. (1954). Relations between the central nervous system and the peripheral organs. *Br. J. Anim. Behav.* 2, 89–94. doi: 10.1016/s0950-5601(54)80044-x
- Watkins, K. E., Smith, S. M., Davis, S., and Howell, P. (2008). Structural and functional abnormalities of the motor system in developmental stuttering. *Brain* 131, 50–59. doi: 10.1093/brain/awm241
- Weisz, N., Hartmann, T., Muller, N., Lorenz, I., and Obleser, J. (2011). Alpha rhythms in audition: cognitive and clinical perspectives. *Front. Psychol.* 2:73. doi: 10.3389/fpsyg.2011.00073
- Weisz, N., Wuhle, A., Monittola, G., Demarchi, G., Frey, J., Popov, T., et al. (2014). Prestimulus oscillatory power and connectivity patterns predispose conscious somatosensory perception. *Proc. Natl. Acad. Sci. U.S.A.* 111, E417–E425. doi: 10.1073/pnas.1317267111

- Wilson, S. M., Saygin, A. P., Sereno, M. I., and Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nat. Neurosci.* 7, 701–702. doi: 10.1038/nn1263
- Wise, R. J., Greene, J., Buchel, C., and Scott, S. K. (1999). Brain regions involved in articulation. *Lancet* 353, 1057–1061. doi: 10.1016/s0140-6736(98)07491-1
- Wolpert, D. M., and Flanagan, J. R. (2001). Motor prediction. *Curr. Biol.* 11, R729–R732. doi: 10.1037/0894-4105.18.4.638.supp
- Ylinen, S., Nora, A., Leminen, A., Hakala, T., Huotilainen, M., Shtyrov, Y., et al. (2014). Two distinct auditory-motor circuits for monitoring speech production as revealed by content-specific suppression of auditory cortex. *Cereb. Cortex* 25, 576–586. doi: 10.1093/cercor/bht351

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Jenson, Harkrider, Thornton, Bowers and Saltuklaroglu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Withholding planned speech is reflected in synchronized beta-band oscillations

Vitória Piai^{1,2*}, Ardi Roelofs¹, Joost Rommers³, Kristoffer Dahlslätt⁴ and Eric Maris¹

¹ Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands, ² Knight Lab, Department of Psychology and Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, CA, USA,

³ Department of Psychology and The Beckman Institute for Advanced Science and Technology, University of Illinois, Urbana, IL, USA, ⁴ Independent Researcher, Berkeley, CA, USA

OPEN ACCESS

Edited by:

Johanna Maria Rimmele,
Max Planck Institute for Empirical
Aesthetics, Germany

Reviewed by:

Sundeep Teki,
École Normale Supérieure, France
Julian Keil,
Charité - Universitätsmedizin Berlin,
Germany

*Correspondence:

Vitória Piai,
Knight Lab, Department
of Psychology and Helen Wills
Neuroscience Institute, University of
California, 132 Barker Hall Berkeley,
CA 94720-3190, USA
v.piai.research@gmail.com

Received: 16 July 2015

Accepted: 18 September 2015

Published: 12 October 2015

Citation:

Piai V, Roelofs A, Rommers J,
Dahlslätt K and Maris E (2015)
Withholding planned speech is
reflected in synchronized beta-band
oscillations.
Front. Hum. Neurosci. 9:549.
doi: 10.3389/fnhum.2015.00549

When engaged in a conversation, speakers sometimes have to withhold a planned response, for example, before it is their turn to speak. In the present study, using magnetoencephalography (MEG) outside of a conversational setting, we investigate the oscillatory brain mechanisms involved in the process of withholding a planned verbal response until it is time to speak. Our participants viewed a sequence of four random consonant strings and one pseudoword, which they had to pronounce when the fifth string (the imperative stimulus) was presented. The pseudoword appeared either as the fourth or fifth stimulus in the sequence, creating two conditions. In the withhold condition, the pseudoword was the fourth string and the verbal response was withheld until the imperative stimulus was presented. In the control condition, the fifth string was the pseudoword, so no response was withheld. We compared oscillatory responses to the withhold relative to the control condition in the time period preceding speech. Alpha-beta power (8–30 Hz) decreased over occipital sensors in the withhold condition relative to the control condition. Source-level analysis indicated a posterior source (i.e., occipital cortex) associated with the alpha-beta power decreases. This occipital alpha-beta desynchronization likely reflects attentional allocation to the upcoming imperative stimulus. Moreover, beta (12–20 Hz) power increased over frontal sensors. Source-level analysis indicated a frontal source (i.e., middle and superior frontal gyri) associated with the beta-power increases. We interpret the frontal beta synchronization to reflect a mechanism aiding the maintenance of the current motor or cognitive state. Our results provide a window into a possible oscillatory mechanism implementing the ability of speakers to withhold a planned verbal response until they have to speak.

Keywords: beta oscillations, conversation, delayed naming, dorsolateral prefrontal cortex, go/no-go, magnetoencephalography, synchronization, turn-taking

Introduction

Conversation is marked by a well-coordinated taking of turns between the listener and the speaker. Although occasionally a speaker will fail to withhold her response until it is her turn to speak, the vast majority of transitions between a speaker's and a listener's turns is smooth with minimal gap in between (e.g., Sacks et al., 1974; Stivers et al., 2009). So the ability of a speaker to withhold a planned response seems to be an important component

of successful communication. Anecdotally, we are all familiar with the situation of feeling ready to say something, but doing so would cause our utterance to overlap with the speech of our conversation partner. Thus, we withhold our utterance until it is our turn to speak.

The fast and accurate transitions between speaker and listener roles in human conversation are an intriguing achievement considering that conversational turns are not fixed in length nor restricted to a particular phrase or syntactic construction (e.g., Sacks et al., 1974). The issue becomes even more interesting if one considers that a speaker may need about 600 ms to plan and begin to articulate a simple word (e.g., Indefrey and Levelt, 2004). Turn-taking studies have been conducted investigating what enables this precise timing of turn-taking in conversation (e.g., de Ruiter et al., 2006; Magyari and de Ruiter, 2012; Torreira et al., 2015). For example, Magyari et al. (2014) stated, "Given the latency of the speech production process, if speakers are going to come in on time, they must begin the production process well before the end of the other's turn" (p. 2536; see also Levinson and Torreira, 2015). This points to the potential need to withhold a planned response: If a speaker begins the production process long before it is her time to speak, she may plan an utterance that cannot yet be articulated, as producing it would create overlap with the utterance of her interlocutor. In the present study, we investigate the speaker in a non-conversational setting with the objective of identifying the neurophysiological correlate of her ability to withhold speech.

Researchers have only recently started examining oscillatory brain activity associated with planning and producing speech (e.g., Gehrig et al., 2012; Herman et al., 2013; Jenson et al., 2014). These studies have found that desynchronization in the alpha and beta bands (7–30 Hz), localized to left inferior frontal, motor, and premotor cortex (e.g., Herman et al., 2013), precedes speech onset. This desynchronization has been interpreted in relation to the well-known alpha-beta desynchronization characteristic of preparation and execution of (hand) movements (e.g., Pfurtscheller and Lopes da Silva, 1999; for review Cheyne, 2013). The finding that speech planning has a neurophysiological signature akin to that of general motor preparation (i.e., alpha-beta desynchronization) is exciting because it allows us to link the neural mechanisms supporting speech production to mechanisms supporting other brain functions, such as motor control (see Piai et al., 2015, for further discussion).

However, speakers must also be able to withhold a verbal response that has already been planned. Very little is known about the neurophysiology of this ability. In particular, no study has yet examined oscillatory activity related to withholding a planned verbal response. Note that robust beta power increases have been shown for withholding a planned manual response (i.e., "postural maintenance", see e.g., Engel and Fries, 2010; Kilavik et al., 2013; for reviews) Previous production studies employing electroencephalography and requiring participants to withhold a planned response have only reported event-related potentials (e.g., Jescheniak et al., 2003). Moreover, those studies were mainly concerned with addressing a psycholinguistic

question, rather than with neurophysiological mechanisms of speech.

In the present study, we focus on the neurophysiology of withholding planned speech, as measured through brain oscillations. Although our paradigm does not capture the real dynamics of naturalistic human conversation, it does allow us to investigate a core component of speaking in such a setting. Our participants engaged in planning a verbal response that sometimes had to be withheld. On half of the trials, they planned a response but withheld it until a cue was presented (withhold condition). On the other half of the trials, the response could only be planned after the cue was presented (control condition). Using magnetoencephalography (MEG), we compared pre-speech oscillatory brain responses in these two conditions.

Materials and Methods

Participants

Fifteen native speakers of Dutch (6 male, mean age = 23 years, $sd = 3.2$) voluntarily participated in the experiment for monetary compensation or for course credits after providing written informed consent. The datasets of four additional participants were not analyzed due to excessive blinking resulting in the loss of a large number of trials (less than 70% of the trials remaining). The present experiment was approved by the Ethics Committee for Behavioral Research of the Social Sciences Faculty at Radboud University Nijmegen in compliance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

Materials

A set of 204 pronounceable pseudowords was generated using WordGen (Duyck et al., 2004). All pseudowords had between two and ten orthographic neighbors and were of four, five, or six letters length (68 pseudowords of each length). Furthermore, a set of 204 random consonant strings was generated of four, five, and six characters length (68 strings of each) to serve as control items for the pseudowords. Finally, another set of 375 random consonant strings was generated of four, five, and six characters length (125 filler strings of each) to be presented in the first, second, and third position of the sequence.

Design

Each pseudoword was paired with a control consonant string of the same length. Half of the pairs appeared in the withhold condition and the other half in the control, counterbalanced across participants. The 204 pairs were pseudo-randomized using Mix (van Casteren and Davis, 2006) with at most five consecutive trials of the same condition. For each trial, three consonant strings were selected at random (without replacement) from the 375 filler strings. One unique randomized list per participant was used.

Behavioral Procedure

Participants were tested individually in an electrically, acoustically, and magnetically shielded room. The experimenter provided non-magnetic clothes to the participants. Participants

were instructed to keep fixation on the center of the screen, to minimize (head) movement during the experimental blocks, and to blink only during the blinking intervals (see below).

In every trial, participants were presented with five strings, and they had to respond to the fifth one (i.e., the imperative stimulus). One of the five strings was a pseudoword that had to be pronounced. This pseudoword was either on the fourth position, in which case the pronunciation had to be withheld and the fifth string served as a go cue (the withhold condition), or it was on the fifth position, in which case it had to be pronounced immediately (the control condition). In the following, we will denote the fifth string as the *imperative stimulus*, because it triggers the pronunciation of the pseudoword. Thus, in the withhold condition, the imperative stimulus is a consonant string that serves as a go cue for the pronunciation of the pseudoword that is presented as the fourth string. In the control condition, the imperative stimulus is a pseudoword that serves both as a go cue and provides the content of the pronunciation. Speed as well as accuracy were emphasized. Participants then practiced the task with 15 trials. After that, they were brought to the shielded room.

Stimuli were presented through an overhead projector on a screen placed 90 cm in front of the participants. The stimuli were in Arial font, size 20. A trial began with a fixation cross presented for 500 ms. Three consonant strings were then presented in green ink for 300 ms, interleaved with a black screen for 300 ms. The fourth and fifth stimuli were presented in white ink. The fourth stimulus was presented for 300 ms, followed by a black screen of 800 ms. The fifth stimulus (imperative stimulus) was then presented for 1.5 s, followed by *** for 2 s, which was the blinking interval. The use of two colors was intended to better guide participants in differentiating between the three initial consonant strings (presented in green), and the pre-speech and imperative stimuli (in white), discouraging them to count the stimuli. **Figure 1** presents an illustration of the trial structure. The 204 experimental trials were divided into four blocks with self-paced breaks in between.

MEG Procedure

The MEG system (CTF VSM MedTech) contained 274 axial gradiometers. Pairs of Ag/AgCl-electrodes were used to record

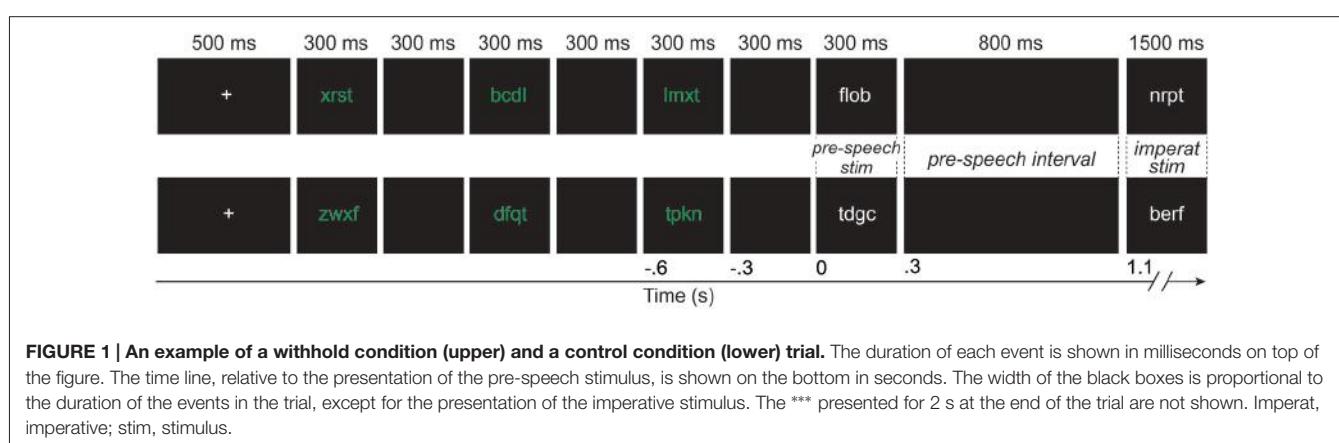
the surface electromyogram from the orbicularis oris muscle and the horizontal and vertical electro-oculogram (impedance <15 kΩ for all electrodes). Three localization coils were fixed to the nasion, right, and left ear canal to monitor the position of participants' heads relative to the gradiometers. Head localization was performed in real time, with the head position re-adjusted when it deviated more than 9 mm from the initial position (Stolk et al., 2013). The data were low-pass filtered by an anti-aliasing filter (300 Hz cutoff), digitized at 1200 Hz, and stored for offline analysis. A microphone in the magnetically shielded room was connected to a computer, which recorded the vocal responses and controlled stimulus presentation with the software package Presentation (Neurobehavioral Systems). Anatomical T1-weighted magnetic resonance images (MRI) of the participants' brains were acquired with a 1.5 T Siemens Magnetom Sonata system using a magnetization-prepared, rapid-acquisition gradient echo sequence.

Response-time Analysis

Verbal responses were evaluated in real time. Responses containing disfluencies were marked, as well as responses initiated before the cue stimulus was presented (3.2% of the trials). Their corresponding trials were subsequently excluded from all analyses. Response times (RTs) were calculated manually using the speech waveform editor Praat (Boersma and Weenink, 2013) before the trials were separated by condition. The statistical analysis was conducted using R (R Core Team, 2014). Participants' RTs were skewed. Given that the median is the best representative of central tendency with skewed data, participants' median RTs were computed for each condition. Paired-samples *t*-tests on participants' median RTs were used to evaluate the behavioral effect. Group RT distributions were also examined by rank-ordering the RTs for each participant, dividing them into 20% quantiles, and then computing quantile means.

MEG and EMG Data Analysis

The analyses were performed using FieldTrip version 20130515 (Oostenveld et al., 2011) in MatlabR2014b. The MEG data were down-sampled offline to 600 Hz and segmented into epochs time-locked to the pre-speech stimulus, from 0.3 s before the



pre-speech stimulus (corresponding to the beginning of the black screen) to 1.2 s post-stimulus (corresponding to 100 ms after the presentation of the imperative stimulus, see **Figure 1**). Since speaking causes artifacts that could potentially affect the MEG signal, all trials in which participants responded within 150 ms after cue onset were discarded from all MEG and EMG analyses.

MEG Preprocessing

All MEG epochs were inspected individually for artifacts. Excessively noisy channels were also removed. Artifact- and error-free data comprised on average 94 trials per condition.

Sensor-level analysis

Synthetic planar gradients were calculated (Bastiaansen and Knösche, 2000). Temporal smearing is an inherent property of time-resolved power estimation. Accordingly, signal components elicited by the imperative stimulus (and therefore also by the participants' initiation of speech) will affect power estimates for time intervals in the pre-speech interval. However, we do not suffer from smearing if we calculate the time-averaged power over the pre-speech interval (0.3–1.1 s after the pre-speech stimulus), and here we did this using multitaper-based spectral estimation. This method of spectral estimation allows for a precise control of the spectral smoothing. We estimated power between 5–30 Hz with 2 Hz spectral smoothing (i.e., 1 Hz above and 1 Hz below) over the pre-speech interval (0.3–1.1 s after the pre-speech stimulus). The data in the pre-speech interval was multiplied with discrete prolate spheroidal sequences as tapers and the Fourier transform was taken from the tapered signal. The power estimates were then averaged over trials for each condition and each participant.

Differences between the conditions were evaluated statistically using a non-parametric cluster-based permutation test (Maris and Oostenveld, 2007) applied to power as a function of frequency (8–30 Hz) and space (the MEG sensors). Given that the low-frequency range (5–7 Hz) was heavily contaminated by myogenic artifacts (see "Frontal Beta Power Increases are not due to Myogenic Artifacts" Section below), we restricted the statistical analysis to 8–30 Hz. For the statistical test, all parameters were the default settings of the Fieldtrip toolbox (version 20130515), except for the following parameters. Spatial clustering was performed on the basis of a neighborhood structure in which sensors had on average six neighbors. Only the sensors that were available for all participants were entered in the analyses (260 in total).

Source-level analysis

First, for each participant, the anatomical MRI was segmented using SPM8¹, which was then used for constructing a corrected-sphere model of the inside of the skull (the volume conduction model, Nolte, 2003). Next, the participant-specific MRI was first warped to a template MRI (Montreal Neurological

Institute (MNI), Montreal, QC, Canada) and then the inverse of that warp was applied to the dipole grid (a 3D grid with 1 cm resolution). This step yielded a grid in MNI coordinates for every participant, allowing us to directly compare grid points across participants in MNI space. The volume conduction model was then used to compute the lead field matrix for each grid point in the source model (Nolte, 2003).

Source-level power was estimated in the pre-speech interval (i.e., 0.3–1.1 s) using the dynamic imaging of coherent sources method (Gross et al., 2001). The sensor-level cross-spectral density matrix was computed from the data of the two conditions combined centered at 16 Hz (with 4 Hz spectral smoothing above and below) using discrete prolate spheroidal sequences as tapers. This frequency range was selected on the basis of the sensor-level results (see "Cortical Signatures of Planning and Withholding Planned Speech: Frontal Beta-power Increases and Posterior Alpha-Beta Power Decreases" Section below). The cross-spectral density matrix was then used together with the leadfields to compute the common spatial filters at each location of the 3-dimensional grid. The common spatial filters were applied to the Fourier transformed data from each condition separately to yield source-level spectral power estimates for each grid point in each condition. These power estimates were then averaged over the trials of each condition for each participant. Relative power change was calculated as the difference between the power in the two conditions divided by their average power. The differences in spectral power between conditions were evaluated using a non-parametric cluster-based permutation test (Maris and Oostenveld, 2007), resulting in a cluster of adjacent cortical locations exhibiting a similar difference across conditions (Fieldtrip toolbox, version 20130515, all parameters set to default).

EMG preprocessing and analysis

We analyzed the EMG to ensure that participants started planning their responses before the imperative stimulus. In that case, mouth muscles such as the orbicularis oris should show increased activity prior to the onset of the imperative stimulus. For one participant, EMG recordings failed so this analysis comprised 14 participants. For the EMG analysis, the EMG data were high-pass filtered offline at 15 Hz (Butterworth two-pass filter of 6th order, FieldTrip default settings) prior to segmentation (see van Boxtel, 2001, for a motivation of the cutoff frequency). The EMG was then Hilbert-transformed and rectified. The resulting signal was then segmented into epochs time-locked to the pre-speech stimulus, from 0.6 s before to 1.5 s after the pre-speech stimulus. Finally, the EMG was averaged over trials per participant for each condition separately. To test statistically whether EMG amplitude differed between conditions before the pre-speech stimulus, a non-parametric cluster-based permutation test was used (with 1,000 random permutations). On the basis of temporal adjacency, clusters exhibiting a similar difference between conditions were identified by means of dependent-samples *t*-tests thresholded at an alpha level of 0.05.

¹<http://www.fil.ion.ucl.ac.uk/spm/>

Results

Planned Responses are Articulated Earlier

Figure 2 (left panel) shows a bean plot of the participants' RTs (Kampstra, 2008), with the dashed line indicating the group mean and the two filled black lines indicating the mean of each condition. Each short white line represents the median RT of one participant for that condition. The cumulative distribution of participants' RTs as a function of condition is shown in the right panel. Verbal responses were on average 208 ms faster in the withhold than in the control condition, $t_{(14)} = 7.86$, $p < 0.001$, 95% CI (151, 264). The cumulative RT distribution shows that the effect is the result of a shift of the entire distribution as a function of whether participants could prepare their responses before the cue or not. Moreover, the larger difference between the two conditions for the 20% fastest responses is possibly due to the fact that the inter-stimulus interval preceding the pre-speech stimulus is fixed. As such, participants are more likely to predict exactly when to speak.

Planning and Withholding Speech Increases EMG Activity

Figure 3 shows the EMG from the orbicularis oris muscle for each condition. Shaded areas indicate the time intervals associated with the significant clusters. The EMG was increased in the withhold condition relative to control already during the pre-speech interval, and this difference increased further after the imperative stimulus was presented. These observations were confirmed by the cluster-based permutation test, which revealed two temporal clusters that exhibited a larger amplitude in the withhold than in the control condition ($p = 0.009$ and $p < 0.001$, respectively). These clusters were detected between 660 ms and 862 ms (left shaded area), and from 925 ms until the end of the segment (i.e., 1500 ms, right shaded area). Thus, we have a physiological indication that participants planned their responses in the withhold condition already prior to the imperative stimulus.

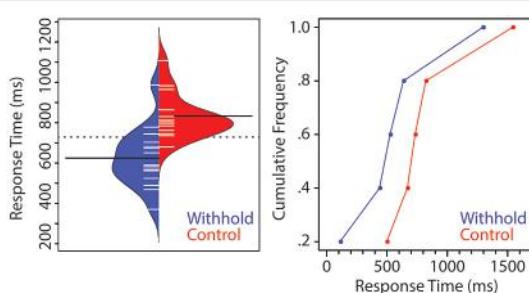


FIGURE 2 | Left. Beanplot of the group-average response times (RTs) for the withhold (blue) and control (red) conditions. The dashed line indicates the group mean whereas the two filled lines indicate the mean of each condition. Each short line represents the median RT of each participant. **Right.** Cumulative distribution of participants' RTs as a function of condition.

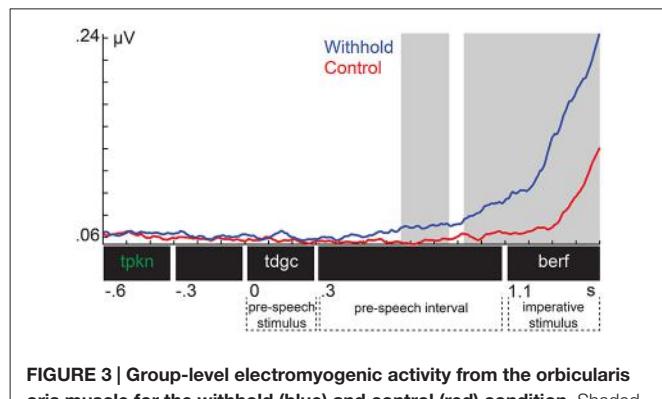


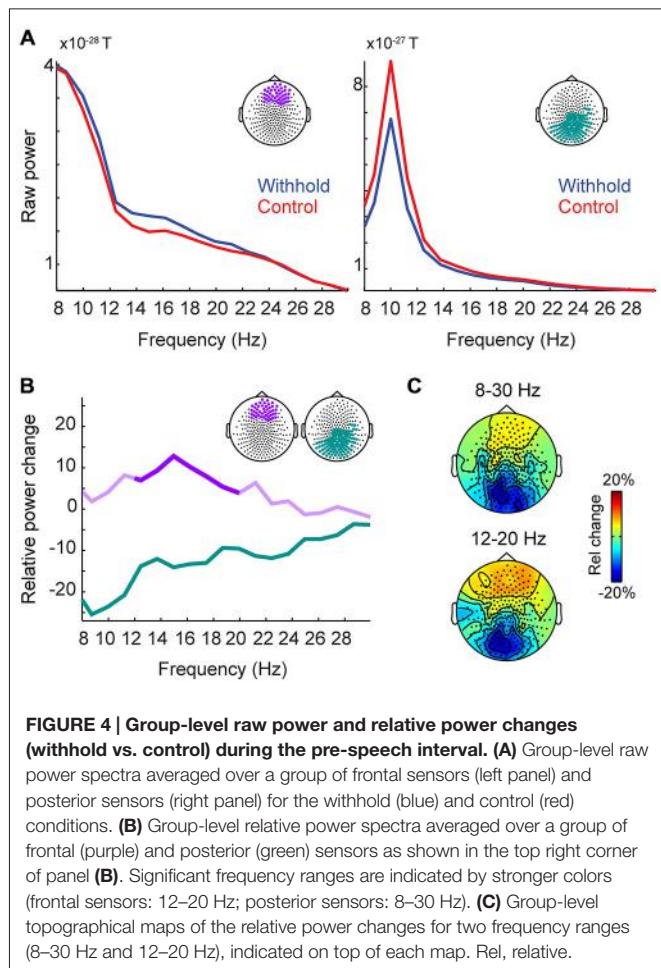
FIGURE 3 | Group-level electromyogenic activity from the orbicularis oris muscle for the withhold (blue) and control (red) condition. Shaded areas indicate the time intervals associated with the significant clusters.

Cortical Signatures of Planning and Withholding Planned Speech: Frontal Beta-Power Increases and Posterior Alpha-Beta Power Decreases

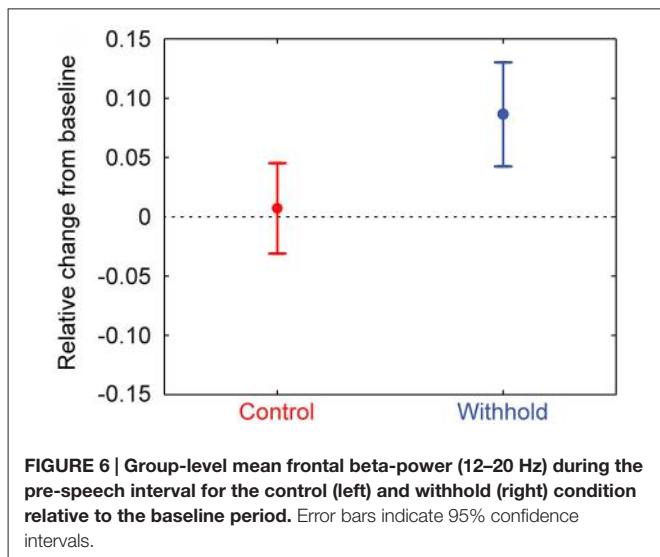
We statistically compared the withhold and the control condition with respect to time-averaged power (shown in **Figure 4A**) as a function of both frequency and space (sensor location). In a cluster-based permutation test, two clusters with significant p values were observed, one over frontal and one over posterior sensors. **Figure 4A** shows the power spectra for each condition during the pre-speech interval averaged over the significant frontal and posterior sensors shown on top of each spectrum. **Figure 4B** shows the relative power changes between the withhold and control conditions during the prespeech interval for the significant sensors. Over frontal sensors, power increased in the withhold relative to the control condition (withhold $>$ control, between 5–13%) in the 12–20 Hz range ($p = 0.012$). This range is indicated by the strong purple color in **Figure 4B**. Over posterior sensors, power decreased in the withhold relative to the control condition (withhold $<$ control, between 5–25%) in the 8–30 Hz range ($p < 0.001$). **Figure 4C** shows the topographical maps of the relative power changes for two frequency ranges, indicated on top of each map.

We source-localized both effects during the pre-speech interval using a frequency-domain beamformer analysis in the 12–20 Hz range, since this frequency range was optimal for the frontal power increases while capturing both the posterior power decreases and the frontal increases. We also statistically compared the withhold and the control condition at the source level using a cluster-level permutation test. **Figure 5** shows the results, masked by the statistically significant clusters, with the color scale indicating the percentage change in power. The results parallel those of the sensor-level analysis: one positive cluster (withhold $>$ control, $p = 0.005$) over bilateral frontal areas, localized to the superior and middle frontal gyri, and inferior frontal gyrus albeit less strong, and one negative cluster (withhold $<$ control, $p < 0.001$) over bilateral occipital cortex. When source localizing the effects in the 8–30 Hz range, very similar results were obtained to what we present here.

To assess whether the frontal beta power differences during the pre-speech interval are due to a pattern of synchronization



or desynchronization relative to baseline, frontal beta power (i.e., 12–20 Hz, statistically significant frontal sensors) was normalized to a baseline period ($-0.3\text{--}0$ s) for each participant. Participants' normalized mean frontal beta-power and 95% confidence intervals are shown in **Figure 6** for each condition separately. It is clear from the figure that frontal beta-power increases relative to the baseline period in the withhold condition, but remains similar to baseline levels in the control condition (the dashed lines indicate no (0) change from baseline). Participants' normalized frontal beta-power was assessed statistically by means of a *t*-test against zero in each condition at an alpha-level of

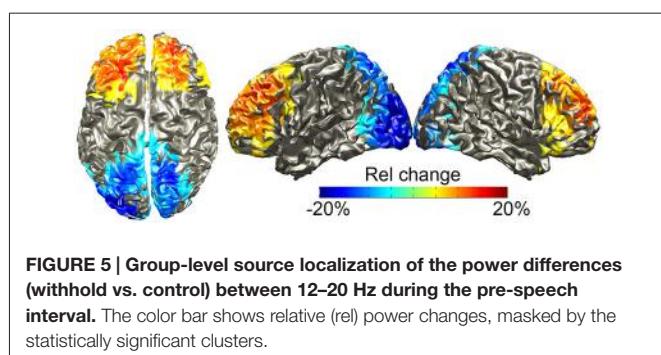


0.025 to correct for two comparisons. In the withhold condition, frontal beta-power increased during the pre-speech interval relative to baseline, $t_{(14)} = 4.22$, $p < 0.001$. In the control condition, frontal beta-power was not significantly different from baseline, $t_{(14)} = 0.40$, $p > 0.696$. Finally, a paired-sample *t*-test indicated that the baseline normalized frontal beta-power was larger in the withhold (mean: 0.086) than in the control (mean: 0.007) condition, $t_{(14)} = 3.50$, $p = 0.004$.

Frontal Beta Power Increases are not Due to Myogenic Artifacts

Given that the EMG was increased for the withhold relative to the control condition already during the pre-speech interval, it is important to assess whether the power increases below 8 Hz and between 12–20 Hz are caused by myogenic artifacts. Below, we evaluate this possibility for each of these frequency bands.

Firstly, if myogenic activity would explain the beta effect prior to speech, then beta power should be stronger during speech, when myogenic activity is greatest, than preceding speech. **Figure 7** shows the power spectrum for each condition pre-speech and during speech averaged over the frontal sensors indicated in light blue (i.e., the statistically significant frontal sensors). Contrary to the prediction, frontal beta power was lower during speech (orange and black lines) than in the pre-speech interval (blue and red lines) between 12–20 Hz. This observation was confirmed by a repeated measures analysis of variance on the frontal beta-band power as a function of condition (withhold vs. control) and interval (pre-speech vs. during speech) at an alpha level of 0.0125 to correct for four comparisons. Frontal beta power was lower during speech than in the pre-speech interval, as indicated by a main effect of interval, $F_{(1,15)} = 20.55$, $p < 0.001$. Condition and interval interacted, $F_{(1,15)} = 18.60$, $p < 0.001$, indicating that power was lower during speech than preceding speech for the withhold condition, $F_{(1,15)} = 36.76$, $p < 0.001$, but statistically similar for the control condition, $F_{(1,15)} = 5.49$, $p = 0.035$. With respect to the 5–7 Hz range, **Figure 7** suggests that frontal power is similar pre-speech and during speech for both conditions. If



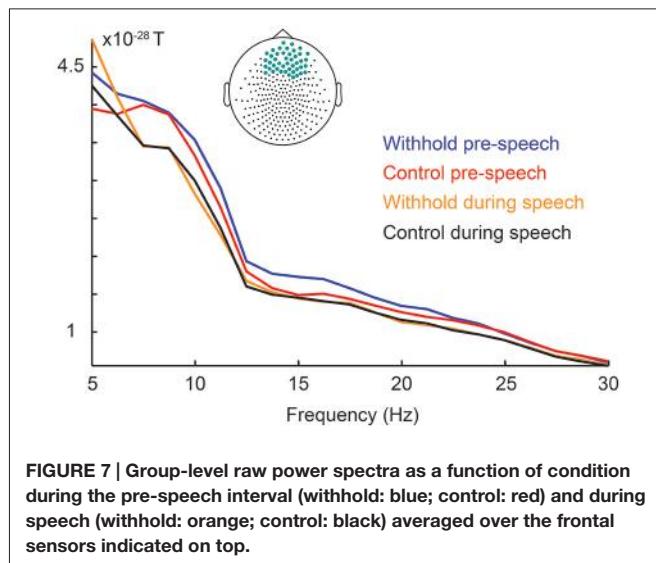


FIGURE 7 | Group-level raw power spectra as a function of condition during the pre-speech interval (withhold: blue; control: red) and during speech (withhold: orange; control: black) averaged over the frontal sensors indicated on top.

5–7 Hz frontal power during speech, when myogenic activity is greatest, is as high as preceding speech, it would indicate that the 5–7 Hz range is likely contaminated with myogenic activity. A repeated measures analysis of variance was conducted on the frontal power in the 5–7 Hz range as a function of condition (withhold vs. control) and interval (pre-speech vs. during speech) at an alpha level of 0.0125 to correct for four comparisons. Frontal power in the 5–7 Hz range was not statistically different during speech from preceding speech, $F_{(1,15)} < 1$, nor different between conditions, $F_{(1,15)} = 4.20$, $p = 0.063$. Condition and interval did not interact, $F_{(1,15)} < 1$. Thus, we can conclude that the frontal power increases between 12–20 Hz during the pre-speech interval are not caused by myogenic artifacts. The frontal power increases between 5–7 Hz, however, are likely caused by muscle activity.

Discussion

In the present study, we investigated the neurophysiology of withholding a planned verbal response through brain oscillations. It has been argued that neuronal oscillations may provide the key to understanding neuronal computations. Under this view, relating neurophysiological signatures in linguistic tasks to the signatures of other cognitive processes could help us understand language function in the context of more basic neurophysiological principles implemented in the brain (see for examples, Piai et al., 2014; Friederici and Singer, 2015).

Preceding the imperative stimulus (during the pre-speech interval), alpha-beta power (8–30 Hz) decreased 5–25% in the withhold relative to the control condition. This power decrease was restricted to occipital sensors and localized mainly to the occipital cortex bilaterally. By contrast, relative beta (12–20 Hz) power increased 5–13% over frontal sensors and was localized to a frontal source (middle and superior frontal gyri, and partly inferior frontal gyrus). Moreover, the EMG recorded from the orbicularis oris muscle was already increased for the withhold relative to the control condition during the pre-speech interval,

confirming that participants prepared their responses. Below, we discuss the oscillatory effects in more detail.

The most robust oscillatory signature of preparing to speak is alpha-beta desynchronization in speech motor areas such as left inferior frontal cortex and ventral motor and premotor cortex (e.g., Salmelin and Sams, 2002; Saarinen et al., 2006; Herman et al., 2013; Jenson et al., 2014; Piai et al., 2015). Our results of beta synchronization in superior and middle frontal gyri when withholding a planned verbal response are clearly different from the signature of speech preparation.

An interesting parallel with our beta synchronization effect can be found in instructed delay tasks, such as go/no-go. In a review of these tasks, it was noted that beta synchronization is commonly observed during an interval of stimulus processing while overt movement is withheld until the go signal (Kilavik et al., 2013). This interval in go/no-go tasks is equivalent to our pre-speech interval. In the literature beta synchronization has been found not only over sensorimotor cortex but also extending further into the entire frontal lobe (see Kilavik et al., 2013 for a review of these findings). This suggests that a similar spatial-spectral pattern underlies withholding speech and withholding other types of overt movement. The functional role of this beta synchronization while overt movement is withheld, however, has not been well specified (see for discussion Kilavik et al., 2013). A tentative functional explanation for this beta synchronization may be found in the proposal of Engel and Fries (2010). On their account, if the current sensorimotor or cognitive state has to be maintained, beta activity is increased. In fact, these authors explicitly predict that activity in the beta band should be increased “during delay-periods where the cognitive set has to be maintained following a cue” (p. 160). This prediction fits with our observation of beta-power increases in the withhold condition during the interval when participants prepare but do not execute their verbal responses. Presumably, in this period, the current motor or cognitive state has to be maintained to enable successful speech production.

The maintenance of the sensorimotor state modulates activity in sensorimotor brain regions (see for review, Engel and Fries, 2010). In our case, the sensorimotor areas associated with speech planning would be left inferior frontal cortex and ventral motor and premotor cortex (e.g., Salmelin and Sams, 2002; Saarinen et al., 2006; Herman et al., 2013; Jenson et al., 2014). Yet, the beta-power increases we observed were more prominent in bilateral superior and middle frontal gyri, less so in bilateral inferior frontal gyrus, and (statistically and descriptively) absent in ventral motor and premotor cortex. It could be the case that cognitive-set maintenance would be subserved by dorsolateral prefrontal cortex (e.g., Petrides, 2005; Buschman et al., 2012; Stoll et al., 2015), compatible with our source localization to superior and middle frontal gyri. Modulations of frontal beta-band oscillations have also been found in a study manipulating the demands for cognitive control (Stoll et al., 2015), which is possibly involved in our task. Future studies will hopefully clarify these issues.

Modulations of alpha and beta oscillatory power often reflect expectation and prediction (e.g., van Ede et al., 2010; Arnal and Giraud, 2012; Pomper et al., 2015). Most relevant to the

present study, alpha desynchronization in occipital cortex in expectation of a visual stimulus is a well-known finding (e.g., Foxe et al., 1998; Worden et al., 2000; Sauseng et al., 2005; Romei et al., 2010). Activity during our pre-speech interval is likely to encompass the anticipation of (or attention towards) the visually presented imperative stimulus, as well as preparation of the spoken response. The occipital source of our alpha-beta desynchronization speaks in favor of a visual attention interpretation, rather than preparation of the spoken response. Decreases in pre-stimulus posterior alpha-power have been often associated with improvements in visual perception (e.g., van Dijk et al., 2008; Jensen and Mazaheri, 2010; Jensen et al., 2012). There are at least two possibilities regarding attentional differences between the two conditions. One possibility is that in the control condition participants have to maximize visual processing to perceive the imperative stimulus and process its content, which is necessary for articulating the target. Another possibility is that in the withhold condition, the imperative stimulus is treated as a go signal and participants need to be maximally sensitive to it in order to respond as fast as possible upon its presentation. Under the assumption that alpha-band desynchronization reflects improved visual processing, or enhanced excitability of visual cortex (Lange et al., 2013), the first possibility would predict power decreases in the control relative to the withhold condition. However, this is the opposite of what we found. Thus, the power decreases are consistent with the second possibility, that is, the participants trying to be maximally sensitive to the imperative stimulus to respond as fast as possible after having prepared their responses.

It can be argued that the difference between the withhold and control conditions in the pre-speech interval is due to a different working-memory demand. Whereas in the control condition, participants were simply waiting for the imperative stimulus, in the withhold condition, they were maintaining the pseudoword in working memory. Although this hypothesis is compatible with our results, it is unclear whether it can fully account for our frontal beta synchronization effect. Firstly, the predominant oscillatory responses associated with working-memory maintenance are not within the frequency range in which we found power increases (12–20 Hz). In the working memory literature (Roux and Uhlhaas, 2014), oscillations between 4–13 Hz (theta and alpha) and above 30 Hz (gamma) have been associated with working-memory maintenance. Notably, the almost complete absence of beta-band effects in the working-memory literature has led some to question whether beta-band activity is even relevant for working memory (Roux and Uhlhaas, 2014). Moreover, if

found, beta-band activity during working-memory maintenance tends to localize to posterior, rather than frontal, brain areas (see for review, Roux and Uhlhaas, 2014). Some form of working memory retention is inevitably involved in the act of withholding speech. In the present study, we did not intend to distinguish between this specific form of working memory retention and other forms (such as those that do not involve motor programming).

Furthermore, it can be argued that the fixed timing of stimulus presentation is a confound in our study because participants learned the timing of stimulus presentation, increasing their expectations. Importantly, however, the fixed timing of presentation was the case for both conditions. Thus, although the expectation of when stimuli will be presented plays a role in our task, it cannot explain the observed spectral differences between the two conditions.

In summary, when participants planned a verbal response and withheld it until an imperative stimulus was presented, beta-power (12–20 Hz) increases were observed in frontal brain areas relative to a control condition not involving speech planning and retention. For the same comparison during the same interval, power decreased over a broad range of frequencies (8–30 Hz) in occipital cortex. Both posterior alpha- and beta-power decreases and frontal beta-power increases are comparable to findings in other cognitive tasks not employing language or verbal responses. In keeping with the extant literature, we interpret our beta-power increases in relation to the maintenance of a cognitive (and possibly motor) set during the pre-speech interval. Altogether, these results suggest that a speaker's ability to plan and withhold speech relies on similar neurophysiological computations as other cognitive functions outside of the language domain.

Author Contributions

Conceptualized and designed the experiment (VP, AR, JR, EM); acquired the data (VP, KD); analyzed the data (VP, KD); wrote the paper (VP, AR, JR, KD, EM). All authors have approved the final version of the paper and agree to be accountable for all aspects of this work.

Acknowledgments

This work has been funded by the Netherlands Organisation for Scientific Research, grant numbers 446-13-009 (to VP) and MaGW 400-09-138 (to AR). The authors thank Robert T. Knight, Nicki Swann, Ole Jensen, and the members of the Center for Aphasia and Related Disorders at VANCHCS, Martinez, CA, for their feedback on various aspects of this work.

References

- Arnal, L. H., and Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends Cogn. Sci.* 16, 390–398. doi: 10.1016/j.tics.2012.05.003
- Bastiaansen, M. C. M., and Knösche, T. R. (2000). Tangential derivative mapping of axial MEG applied to event-related desynchronization research. *Clin. Neurophysiol.* 111, 1300–1305. doi: 10.1016/s1388-2457(00)00272-8

Boersma, P., and Weenink, D. (2013). *Praat: Doing Phonetics by Computer (Version 5.3.42)*. Available online at: <http://www.praat.org> (accessed March 2, 2013).

Buschman, T. J., Denovellis, E. L., Diogo, C., Bullock, D., and Miller, E. K. (2012). Synchronous oscillatory neural ensembles for rules in the prefrontal cortex. *Neuron* 76, 838–846. doi: 10.1016/j.neuron.2012.09.029

- Cheyne, D. O. (2013). MEG studies of sensorimotor rhythms: a review. *Exp. Neurol.* 245, 27–39. doi: 10.1016/j.expneurol.2012.08.030
- de Ruiter, J. P., Mitterer, H., and Enfield, N. J. (2006). Projecting the end of a speaker's turn: a cognitive cornerstone of conversation. *Language* 82, 515–535. doi: 10.1353/lan.2006.0130
- Duyck, W., Desmet, T., Verbeke, L., and Brysbaert, M. (2004). WordGen: a tool for word selection and non-word generation in Dutch, German, English and French. *Behav. Res. Meth. Ins. C* 36, 488–499. doi: 10.3758/bf03195595
- Engel, A. K., and Fries, P. (2010). Beta-band oscillations – signalling the status quo? *Curr. Opin. Neurobiol.* 20, 156–165. doi: 10.1016/j.conb.2010.02.015
- Foxe, J. J., Simpson, G. V., and Ahlfors, S. P. (1998). Cued shifts of intermodal attention: parieto-occipital ~10 Hz activity reflects anticipatory state of visual attention mechanisms. *Neuroreport* 9, 3929–3933. doi: 10.1097/00001756-199812010-00030
- Friederici, A. D., and Singer, W. (2015). Grounding language processing on basic neurophysiological principles. *Trends Cogn. Sci.* 19, 329–338. doi: 10.1016/j.tics.2015.03.012
- Gehrig, J., Wibral, M., Arnold, C., and Kell, C. A. (2012). Setting up the speech production network: how oscillations contribute to lateralized information routing. *Front. Psychol.* 3:169. doi: 10.3389/fpsyg.2012.00169
- Gross, J., Kujala, J., Härmäläinen, M. S., Timmermann, L., Schnitzler, A., and Salmelin, R. (2001). Dynamic imaging of coherent sources: Studying neural interactions in the human brain. *Proc. Natl. Acad. Sci. U S A* 98, 694–699. doi: 10.1073/pnas.98.2.694
- Herman, A. B., Houde, J. F., Vinogradov, S., and Nagarajan, S. S. (2013). Parsing the phonological loop: activation timing in the dorsal speech stream determines accuracy in speech reproduction. *J. Neurosci.* 33, 5439–5453. doi: 10.1523/jneurosci.1472-12.2013
- Indefrey, P., and Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition* 92, 101–144. doi: 10.3389/fpsyg.2011.00255
- Jensen, O., Bonnefond, M., and VanRullen, R. (2012). An oscillatory mechanism for prioritizing salient unattended stimuli. *Trends Cogn. Sci.* 16, 200–206. doi: 10.1016/j.tics.2012.03.002
- Jensen, O., and Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Front. Hum. Neurosci.* 4:186. doi: 10.3389/fnhum.2010.00186
- Jenson, D., Bowers, A. L., Harkrider, A., Thornton, D., Cuellar, M., and Saltuklaroglu, T. (2014). Temporal dynamics of sensorimotor integration in speech perception and production: Independent component analysis of EEG data. *Front. Psychol.* 5:656. doi: 10.3389/fpsyg.2014.00656
- Jescheniak, J. D., Hahne, A., and Schriefers, H. (2003). Information flow in the mental lexicon during speech planning: evidence from event-related brain potentials. *Cogn. Brain Res.* 15, 261–276. doi: 10.1016/s0926-6410(02)00198-2
- Kampstra, P. (2008). Beanplot: A boxplot alternative for visual comparison of distributions. *J. Stat. Softw. Code Snippets* 28, 1–9. doi: 10.18637/jss.v028.c01
- Kilavik, B. E., Zaepffel, M., Brovelli, A., MacKay, W. A., and Riehle, A. (2013). The ups and downs of beta oscillations in sensorimotor cortex. *Exp. Neurol.* 245, 15–26. doi: 10.1016/j.expneurol.2012.09.014
- Lange, J., Oostenveld, R., and Fries, P. (2013). Reduced occipital alpha power indexes enhanced excitability rather than improved visual perception. *J. Neurosci.* 33, 3212–3220. doi: 10.1523/jneurosci.3755-12.2013
- Levinson, S. C., and Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Front. Psychol.* 6:731. doi: 10.3389/fpsyg.2015.00731
- Magyari, L., Bastiaansen, M. C. M., de Ruiter, J. P., and Levinson, S. C. (2014). Early anticipation lies behind the speed of response in conversation. *J. Cogn. Neurosci.* 26, 2530–2539. doi: 10.1162/jocn_a_00673
- Magyari, L., and de Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Front. Psychol.* 3:376. doi: 10.3389/fpsyg.2012.00376
- Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190. doi: 10.1016/j.jneumeth.2007.03.024
- Nolte, G. (2003). The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors. *Phys. Med. Biol.* 48, 3637–3652. doi: 10.1088/0031-9155/48/22/002
- Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG and invasive electrophysiological data. *Comp. Intell. Neurosci.* 2011:156869. doi: 10.1155/2011/156869
- Petrides, M. (2005). Lateral prefrontal cortex: architectonic and functional organization. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 781–795. doi: 10.1098/rstb.2005.1631
- Pfurtscheller, G., and Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin. Neurophysiol.* 110, 1842–1857. doi: 10.1016/s1388-2457(99)00141-8
- Piai, V., Roelofs, A., and Maris, E. (2014). Oscillatory brain responses in spoken word production reflect lexical frequency and sentential constraint. *Neuropsychologia* 53, 146–156. doi: 10.1016/j.neuropsychologia.2013.11.014
- Piai, V., Roelofs, A., Rommers, J., and Maris, E. (2015). Beta oscillations reflect memory and motor aspects of spoken word production. *Hum. Brain Mapp.* 36, 2767–2780. doi: 10.1002/hbm.22806
- Pomper, U., Keil, J., Foxe, J. J., and Senkowski, D. (2015). Intersensory selective attention and temporal orienting operate in parallel and are instantiated in spatially distinct sensory and motor cortices. *Hum. Brain Mapp.* 36, 3246–3259. doi: 10.1002/hbm.22845
- R Core Team. (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available online at: <http://www.R-project.org/>
- Romei, V., Gross, J., and Thut, G. (2010). On the role of prestimulus alpha rhythms over occipito-parietal areas in visual input regulation: correlation or causation? *J. Neurosci.* 30, 8692–8697. doi: 10.1523/jneurosci.0160-10.2010
- Roux, F., and Uhlhaas, P. J. (2014). Working memory and neural oscillations: alpha-gamma vs. theta-gamma codes for distinct WM-information. *Trends Cogn. Sci.* 18, 16–25. doi: 10.1016/j.tics.2013.10.010
- Saarinen, T., Laaksonen, H., Parviaainen, T., and Salmelin, R. (2006). Motor cortex dynamics in visuomotor production of speech and non-speech mouth movements. *Cereb. Cortex* 16, 212–222. doi: 10.1093/cercor/bhi099
- Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 696–735. doi: 10.1353/lan.1974.0010
- Salmelin, R., and Sams, M. (2002). Motor cortex involvement during verbal versus non-verbal lip and tongue movements. *Hum. Brain Mapp.* 16, 81–91. doi: 10.1002/hbm.10031
- Sauseng, P., Klimesch, W., Stadler, W., Schabus, M., Doppelmayr, M., Hanslmayr, S., et al. (2005). A shift of visual spatial attention is selectively associated with human EEG alpha activity. *Eur. J. Neurosci.* 22, 2917–2926. doi: 10.1111/j.1460-9568.2005.04482.x
- Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci. U S A* 106, 10587–10592. doi: 10.1073/pnas.0903616106
- Stolk, A., Todorovic, A., Schoffelen, J.-M., and Oostenveld, R. (2013). Online and offline tools for head movement compensation in MEG. *Neuroimage* 68, 39–48. doi: 10.1016/j.neuroimage.2012.11.047
- Stoll, F. M., Wilson, C. R. E., Faraut, M. C. M., Vezoli, J., Knoblauch, K., and Procyk, E. (2015). The effects of cognitive control and time on frontal beta oscillations. *Cereb. Cortex* doi: 10.1093/cercor/bhv006 [Epub ahead of print].
- Torreira, F., Bögels, S., and Levinson, S. C. (2015). Breathing for answering: The time course of response planning in conversation. *Front. Psychol.* 6:248. doi: 10.3389/fpsyg.2015.00284
- van Boxtel, A. (2001). Optimal signal bandwidth for the recording of surface EMG activity of facial, jaw, oral and neck muscles. *Psychophysiology* 38, 22–34. doi: 10.1111/1469-8986.3810022
- van Casteren, M., and Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behav. Res. Methods* 38, 584–589. doi: 10.3758/bf03193889

- van Dijk, H., Schoffelen, J. M., Oostenveld, R., and Jensen, O. (2008). Pre-stimulus oscillatory activity in the alpha band predicts visual discrimination ability. *J. Neurosci.* 28, 1816–1823. doi: 10.1523/jneurosci.1853-07.2008
- van Ede, F., Jensen, O., and Maris, E. (2010). Tactile expectation modulates pre-stimulus β -band oscillations in human sensorimotor cortex. *Neuroimage* 51, 867–876. doi: 10.1016/j.neuroimage.2010.02.053
- Worden, M. S., Foxe, J. J., Wang, N., and Simpson, G. V. (2000). Anticipatory biasing of visuospatial attention indexed by retinotopically specific α -band electroencephalography increases over occipital cortex. *J. Neurosci.* 20, 1–6.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Piai, Roelofs, Rommers, Dahlslätt and Maris. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Electrocorticographic Activation within Human Auditory Cortex during Dialog-Based Language and Cognitive Testing

Kirill V. Nourski^{1†}, Mitchell Steinschneider^{2*†} and Ariane E. Rhone¹

¹ Human Brain Research Laboratory, Department of Neurosurgery, The University of Iowa, Iowa City, IA, USA, ² Departments of Neurology and Neuroscience, Albert Einstein College of Medicine, Bronx, NY, USA

OPEN ACCESS

Edited by:

Johanna Maria Rimmele,
Max Planck Institute for Empirical
Aesthetics, Germany

Reviewed by:

Christian A. Kell,
Goethe University, Germany
Jordi Costa-Faidella,
University of Barcelona, Spain

*Correspondence:

Mitchell Steinschneider
mitchell.steinschneider@einstein.yu.edu

[†]These authors have contributed
equally to this work.

Received: 21 January 2016

Accepted: 20 April 2016

Published: 04 May 2016

Citation:

Nourski KV, Steinschneider M
and Rhone AE (2016)

Electrocorticographic Activation within
Human Auditory Cortex during
Dialog-Based Language
and Cognitive Testing.
Front. Hum. Neurosci. 10:202.
doi: 10.3389/fnhum.2016.00202

Current models of cortical speech and language processing include multiple regions within the temporal lobe of both hemispheres. Human communication, by necessity, involves complex interactions between regions subserving speech and language processing with those involved in more general cognitive functions. To assess these interactions, we utilized an ecologically salient conversation-based approach. This approach mandates that we first clarify activity patterns at the earliest stages of cortical speech processing. Therefore, we examined high gamma (70–150 Hz) responses within the electrocorticogram (ECoG) recorded simultaneously from Heschl's gyrus (HG) and lateral surface of the superior temporal gyrus (STG). Subjects were neurosurgical patients undergoing evaluation for treatment of medically intractable epilepsy. They performed an expanded version of the Mini-mental state examination (MMSE), which included additional spelling, naming, and memory-based tasks. ECoG was recorded from HG and the STG using multicontact depth and subdural electrode arrays, respectively. Differences in high gamma activity during listening to the interviewer and the subject's self-generated verbal responses were quantified for each recording site and across sites within HG and STG. The expanded MMSE produced widespread activation in auditory cortex of both hemispheres. No significant difference was found between activity during listening to the interviewer's questions and the subject's answers in posteromedial HG (auditory core cortex). A different pattern was observed throughout anterolateral HG and posterior and middle portions of lateral STG (non-core auditory cortical areas), where activity was significantly greater during listening compared to speaking. No systematic task-specific differences in the degree of suppression during speaking relative to listening were found in posterior and middle STG. Individual sites could, however, exhibit task-related variability in the degree of suppression during speaking compared to listening. The current study demonstrates that ECoG recordings can be acquired in time-efficient dialog-based paradigms, permitting examination of language and cognition in an ecologically salient manner. The results obtained from auditory cortex serve as a foundation for future studies addressing patterns of activity beyond auditory cortex that subserve human communication.

Keywords: Heschl's gyrus, high gamma, Mini-mental state examination, speech, superior temporal gyrus

INTRODUCTION

Intracranial recordings in humans have permitted evaluation of speech and language processing with unprecedented temporal and spatial resolution (e.g., Leonard and Chang, 2014; Nourski and Howard, 2015). Most of these intracranial studies have focused on neural activity on the lateral surface of the STG (e.g., Crone et al., 2001; Steinschneider et al., 2011; Mesgarani et al., 2014). For instance, Mesgarani et al. (2014) have demonstrated a role for the posterior lateral STG of the dominant hemisphere in acoustic-to-phonetic transformations of speech. Less explored are regions of auditory and auditory-related cortex envisioned to encode ever more complex features of speech and language. For instance, cortex within the superior temporal sulcus and middle temporal gyrus is critical for phonological and lexical-semantic processing, respectively (Binder et al., 2000; Obleser et al., 2008; Hickok, 2009; Leaver and Rauschecker, 2010). Furthermore, regions of the brain involved in cognitive processes such as attention, working memory, and declarative memory must by necessity interface with regions of the brain more directly involved in speech processing.

The opportunity to simultaneously explore multiple brain regions involved in speech and language is provided by the extensive electrode coverage in epilepsy patients undergoing chronic invasive monitoring. However, paradigms investigating complex speech and language functions must take into account that these studies are being carried out in patients in a hospital setting with the primary goal being remediation of their seizure disorders. These considerations mandate that these studies be time-efficient and performed with the recognition that prolonged experimental sessions often engender excessive patient fatigue and potentially lead to unwillingness to pursue further participation in research activities.

In this study, we initiated a conversation-based paradigm that incorporates multiple speech, language, and cognitive functions in a time-efficient manner. We hypothesized that such a paradigm would be a more ecologically salient means to study these complex functions than traditionally used trial-based protocols (e.g., Steinschneider et al., 2014; Nourski et al., 2015a). A conversation, by its very nature, will engage a wide array of auditory, speech, and language areas and interface with regions engaged in higher cognitive functions. This conversation-based approach has been shown to be an effective means for exploring the roles of human auditory and auditory-related cortex within the setting of clinically necessitated intracranial recordings (Creutzfeldt et al., 1989; Derix et al., 2012, 2014; see also Dastjerdi et al., 2013).

For these reasons, we utilized the MMSE, which is a commonly used tool to screen for language and cognitive impairments associated with dementia (Finney et al., 2016). It examines a range of functions, including orientation to time and place, immediate and delayed recall, attention, naming, repetition, and following multi-step commands (Folstein et al., 1975). However, it has been

recently noted that the MMSE lacks sufficient sensitivity and specificity in predicting dementia and thus should not be used as a standalone clinical test for screening of language and cognitive deficits (Arevalo-Rodriguez et al., 2015). Therefore, we have implemented additional tasks for a more comprehensive assay of the cortical regions involved in higher language and cognitive functions. These tasks included digit span, spelling, rhyming, abstract naming, verbal analogies, sentence comprehension, fund of knowledge, and identification of favorite items. The expanded paradigm is highly time-efficient and is typically completed within approximately 15 min.

Despite its potential utility, this conversation-based experimental paradigm presents several challenges when analyzing task-related cortical activity using ECoG (Nourski and Howard, 2015). Conventional trial-based paradigms typically rely on analyzing activity that is time-locked to particular events by averaging across multiple instances of these events. These analyses typically focus on low-frequency local field potentials or activity in the high gamma (70–150 Hz) ECoG frequency range (e.g., Crone et al., 2001; Nourski et al., 2015b). Studies examining high gamma ECoG often do so by referencing event-related activity to a pre-defined local baseline (ERBP). However, a conversation-based paradigm offers neither repetition of the same event, nor a stable local baseline. To deal with these issues in the present study, cortical high gamma activity was normalized relative to mean power over the entire duration of the recording, and then averaged across all utterances, done separately for the interviewer's and the subject's speech.

Due to the challenges of this new method, we initiated our investigation in lower auditory cortical areas with relatively well-described basic response properties (e.g., Brugge et al., 2008; Leonard and Chang, 2014; Mesgarani et al., 2014; Nourski et al., 2014a,b). Specifically, we focused our initial investigation on neural activity generated within the auditory cortex located in HG and on the lateral surface of the STG. These regions incorporate portions of auditory core, belt and parabelt cortex (e.g., Hackett et al., 2001; Brugge et al., 2008; Nourski et al., 2014a; Hackett, 2015). Analysis was restricted to activity in the high gamma frequency range, which has been shown to be useful in defining the basic physiological response properties of these cortical regions (e.g., Crone et al., 2001; Brugge et al., 2009; Steinschneider et al., 2011). Identification of high gamma response patterns within auditory cortex is a necessary prerequisite for clarifying patterns of activity at higher stages of cortical speech and language processing.

The posteromedial portion of HG has been consistently identified as part of core auditory cortex (e.g., Liegeois-Chauvel et al., 1991; Brugge et al., 2008; Nourski et al., 2014a). Electrophysiological studies have demonstrated that this brain region is strongly activated by a wide range of simple and complex sound stimuli. It is unclear, however, whether activity would be different for sounds generated by the interviewer versus sounds self-initiated by the subject. Suppression of activity during self-initiated speech has been demonstrated in both non-human primates (Müller-Preuss and Ploog, 1981; Eliades and Wang, 2003, 2005) and humans (Creutzfeldt et al., 1989; Houde et al., 2002; Greenlee et al., 2011). While suppression

Abbreviations: ECoG, electrocorticography; ERBP, event-related band power; FDR, false discovery rate; HG, Heschl's gyrus; MMSE, Mini-mental status exam; MNI, Montreal Neurological Institute; ROI, region of interest; SI, suppression index; STG, superior temporal gyrus; TTS, transverse temporal sulcus.

has been demonstrated within auditory core cortex in the non-human primate (Eliades and Wang, 2003, 2005), it has not been demonstrated in the human (Greenlee et al., 2014; Behroozmand et al., 2016). We therefore examined whether activity in posteromedial HG would be modulated by speaker during a conversation. A similar logic applies to whether suppression of activity elicited by self-initiated speech would occur within non-core cortex in anterolateral HG.

Auditory cortex on the lateral STG has been shown to be modulated by speech phonetic features, attention and task demands, and self-initiated vocalization (e.g., Chang et al., 2011; Greenlee et al., 2011; Mesgarani and Chang, 2012; Mesgarani et al., 2014; Steinschneider et al., 2014). While these studies have been performed in well-structured and controlled settings, it remains to be seen whether these effects can be reliably identified within the ecologically relevant context of a conversation-based paradigm.

Thus, in the present study, we examined modulation of activity elicited when listening and speaking during performance of the expanded MMSE within four ROIs: posteromedial HG, anterolateral HG, posterior STG, and middle STG. Decoding of complex and abstract features of speech occurs in more anterior regions of the temporal lobe (Hickok and Poeppel, 2007; Hickok, 2009). The TTS provides an anatomical landmark that may be useful for demarcating posterior from middle portions of STG. We therefore reasoned that modulation of activity due to self-vocalization might vary between these two regions of the STG. We further examined whether activity was modulated by the multiple tasks incorporated in our expanded version of the MMSE.

MATERIALS AND METHODS

Subjects

Experimental subjects were six neurosurgical patients (three female, three male, age 21–51 years old, median age 33 years old) diagnosed with medically refractory epilepsy undergoing chronic invasive ECoG monitoring to identify potentially resectable temporal lobe seizure foci. Demographic data for each subject are presented in **Table 1**. Research protocols were approved

TABLE 1 | Subject demographics.

Subject ¹	Age	Sex	Handedness	Language dominance	MMSE score ²
R288	21	M	R	L	26/27
L292	50	F	L	R	26/27
R294	35	M	R	L	26/27
L307	29	M	R	L	24/27
R316	31	F	R	L	24/27
R320	51	F	R	L	24/27

¹Letter prefix of the subject code denotes the side of electrode implantation (L, left; R, right). ²Three of the standard MMSE tasks (reading, writing, and copying) were not included in the current protocol, therefore the maximum MMSE score was 27 rather than 30.

by the University of Iowa Institutional Review Board and the National Institutes of Health. Written informed consent was obtained from all subjects. Research participation did not interfere with acquisition of clinically required data, and subjects could rescind consent at any time without interrupting their clinical evaluation.

All subjects underwent audiometric evaluation before the study, and none was found to have hearing deficits that should impact the findings presented in this study. All subjects had pure-tone thresholds within 25 dB HL between 250 Hz and 4 kHz, with the exception of subject L307, who had a mild (40 dB HL) notch at 4 kHz in the right ear only. All subjects were native English speakers. Intracranial recordings revealed that auditory cortical areas within the four ROI in HG and on STG were not epileptic foci in any subject.

Procedure

Experiments were carried out in a dedicated electrically shielded suite in The University of Iowa Clinical Research Unit. The subjects were comfortably reclining in a hospital bed or an armchair while performing the MMSE (Folstein et al., 1975). In subjects L307, R316, and R320, testing was expanded beyond the MMSE to include other tasks (digit span, spelling, rhyming, naming, verbal analogies, sentence comprehension, and fund of knowledge). These subjects were also asked to identify favorite items (e.g., favorite food or movie; Supplementary Table 1).

All subjects had comparable performance in aspects of the MMSE, with “Delayed Verbal Recall” being the only section where all subjects had difficulty (see **Table 1**). Three subjects failed to recall one out of three words, while three others could not recall any of the three words. It should be noted that the interviewer did not specifically emphasize that the subjects would be asked to recall the three words later in the test. Overall, the subjects’ successful performance on the exam indicated that neural activity was not biased by cognitive deficits specifically revealed by the MMSE.

Recordings

Electrocorticography recordings were simultaneously made from HG and the lateral cortical surface using multicontact depth and subdural grid electrodes, respectively. Details of electrode implantation, recording, and analysis of high gamma cortical activity have been previously described in depth (Howard et al., 1996, 2000; Reddy et al., 2010; Nourski et al., 2013; Nourski and Howard, 2015). All electrode arrays were placed solely on the basis of clinical requirements, and were part of a more extensive set of recording arrays meant to identify seizure foci. Electrodes remained in place under the direction of the patients’ treating neurologists.

Depth electrode arrays (eight macro contacts, spaced 5 mm apart) were implanted in each subject stereotactically into HG, along its anterolateral-to-posteromedial axis. The approach used at The University of Iowa is modeled in part after the well-established stereo-EEG techniques developed and used widely in epilepsy centers in Europe. The technique involves implantation of electrodes within the superior temporal plane in order to

provide broad coverage of the suspected seizure focus. With this strategy, electrodes are implanted in the superior temporal plane regardless of whether a patient with suspected temporal lobe seizures describes auditory auras (Munari, 1987; Bartolomei et al., 1999, 2008; Maillard et al., 2004; Gavaret et al., 2006; McGonigal et al., 2007). Review of all patients who had been implanted with depth electrodes in the superior temporal plane within the last 3 years revealed the strong clinical utility of the ECoG data provided by these electrodes in clinical decision making with regard to the extent of surgical resections (data available upon request).

Subdural grid arrays were implanted over the lateral surface of the cerebral hemisphere, including the auditory cortex on the lateral STG. The grid arrays consisted of platinum–iridium disk electrodes (2.3 mm exposed diameter) embedded in a silicon membrane. In subjects R288, L307, and R320 high density (5 mm center-to-center inter-electrode distance) research grids were used, with electrodes arranged in an 8×12 grid, yielding a 3.5 cm \times 5.5 cm array of 96 contacts. In subject R316, a 32-contact clinical grid (4 \times 8 array with a 10 mm inter-electrode distance) was used. In subjects L292 and R294, 16-contact clinical grids (2 \times 8 array, 10 mm inter-electrode distance) were placed over the lateral surface of the STG. In all subjects, a subgaleal contact was used as a reference.

As with the depth electrodes, decisions regarding what surface regions and to what extent should be covered, are driven exclusively by clinical considerations. High resolution research grids do not increase the risks of surgery or alter the area of cortex from which records are obtained. Also, the materials used to fabricate the arrays that are in contact with the brain surface are the same for research and clinical electrodes. Information about electrodes modified for research purposes was conveyed to each patient prior to surgery.

Subjects underwent whole-brain high-resolution T1-weighted structural MRI scans (resolution 0.78 mm \times 0.78 mm, slice thickness 1.0 mm) before electrode implantation. Two volumes were averaged to improve the signal-to-noise ratio of the MRI data sets and minimize the effects of movement artifact on image quality. After electrode implantation, subjects underwent thin-sliced volumetric computerized tomography scans (resolution 0.51 mm \times 0.51 mm, slice thickness 1.0 mm).

Locations of recording sites were determined by co-registering pre- and post-implantation structural imaging data using a linear algorithm with six degrees of freedom (Jenkinson et al., 2002), aided by intraoperative photographs.

Data acquisition was controlled by a TDT RZ2 real-time processor (Tucker-Davis Technologies, Alachua, FL, USA). Collected ECoG data were amplified, filtered (0.7–800 Hz bandpass, 12 dB/octave rolloff), digitized at a sampling rate of 2034.5 Hz, and stored for subsequent offline analysis. The conversation between the interviewer and subject was recorded simultaneously using an in-room Behringer ECM 8000 microphone (Behringer, Willich, Germany) and digitized at a sampling rate of 12207 Hz.

Data Analysis

Utterances spoken by the interviewer and the subject were parsed using Praat software based upon specific phrases and natural breaks in the conversation, generally following a question-answer format. This method was chosen in order to compare activity elicited during listening versus speaking across ROIs. Average durations of utterances by the interviewer and the subjects parsed using this method were not significantly different (Wilcoxon rank sum test, **Table 2**). Voice fundamental frequency (F_0) was estimated for each utterance using YIN fundamental frequency estimator (de Chevigné and Kawahara, 2002). Two of the subjects (L292, R316) had median F_0 s significantly higher than the interviewer, one subject (L307) had significantly lower F_0 , while the other three subjects did not exhibit significant differences in F_0 from the interviewer (Wilcoxon rank sum test; see **Table 2**).

Electrocorticography data obtained from each recording site were downsampled to 1000 Hz. To minimize contamination from power line noise, ECoG waveforms were de-noised using an adaptive notch filtering procedure (Nourski et al., 2013). Data analysis was performed using custom software written in the MATLAB Version 7.14 programming environment (MathWorks, Natick, MA, USA).

Analysis of cortical activity focused on the high gamma ECoG frequency band. High gamma power envelope was calculated for each recording site. ECoG waveforms were bandpass filtered between 70 and 150 Hz (300th order finite impulse response

TABLE 2 | Acoustic properties of dialogues between the interviewer and each subject.

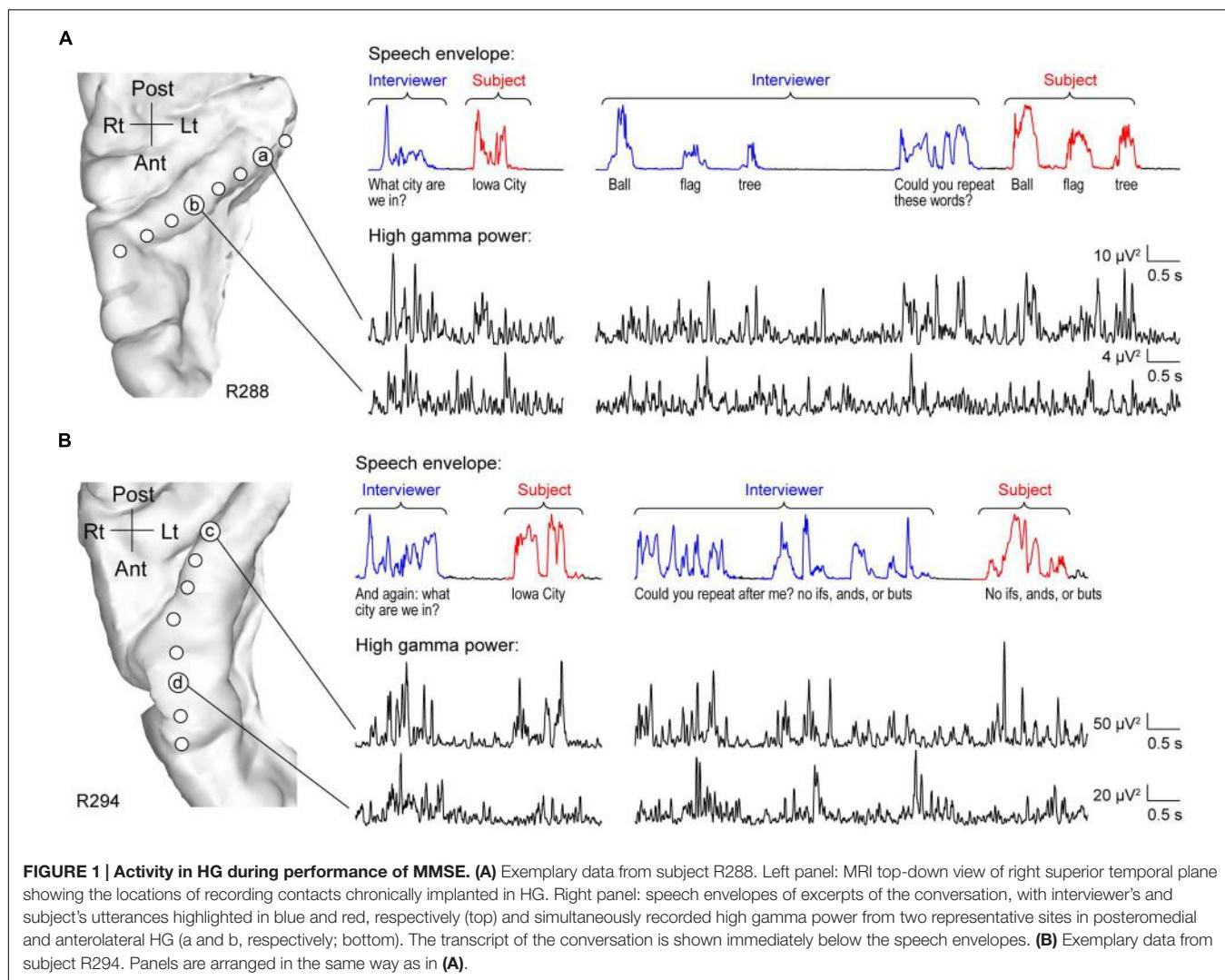
Subject	Utterance duration			F_0		
	Median utterance duration (interviewer) (s)	Median utterance duration (subject) (s)	Interviewer vs. subject comparison p	Median F_0 (interviewer) (Hz)	Median F_0 (subject) (Hz)	Interviewer vs. subject comparison p
R288	1.934	1.457	0.205	127.7	142.9	0.305
L292	1.738	1.464	0.271	155.0	204.4	<0.001
R294	1.441	1.435	0.918	137.9	141.4	0.593
L307	1.709	1.440	0.815	131.2	120.7	<0.001
R316	1.622	1.139	0.142	139.8	205.2	<0.001
R320	1.309	1.168	0.477	163.7	181.2	0.711
All subjects	1.519	1.441				

filter), followed by Hilbert envelope extraction and smoothing using a moving average filter with a span of 25 ms.

For quantitative analysis, high gamma ERBP was computed in all subjects as follows: power envelope waveforms were log-transformed, high-pass filtered (fourth order Butterworth filter, 0.1 Hz cutoff) to eliminate long-term baseline changes, and normalized to the mean power over the entire duration of the recording. ERBP was then averaged within time windows corresponding to each utterance (between 50 ms after the onset and 200 ms after the offset of each utterance), and averaged separately across all utterances spoken by the interviewer and the subject. This time window has been shown to capture the excitatory responses to speech, as well as suppression in high gamma activity during self-vocalization (see Greenlee et al., 2011). Supplementary Figure 1 demonstrates this window for high gamma activity elicited by all utterances in subjects L307 and R320. The analysis to establish the time window of interest was carried out in these two subjects because they had extensive coverage of the STG and were presented with the expanded MMSE questionnaire. On average,

onset of activity began approximately 50 ms after the onset of the utterance, and persisted for approximately 200 ms following the offset of the utterance. It must be acknowledged that this approach limits the ability to assess the neural dynamics underlying the processing fine-grain spectrotemporal attributes within speech stimuli (cf. Mesgarani et al., 2014). However, the purpose of this paradigm is to characterize brain regions processing the utterances as a whole, thus promoting identification of neural dynamics related to specific language and cognitive tasks. Finally, activity during silent intervals between the interviewer's questions and the subject's verbal responses was averaged within time windows between 250 ms after the interviewer's utterance offset and the onset of the next utterance. These time windows were then used for quantitative analysis of high gamma activity elicited during listening, speaking, and the intervening silence in all six subjects.

Previous studies have demonstrated that acoustically responsive cortex in HG and on STG comprises multiple fields, with posteromedial HG consistently interpreted as core



auditory cortex. To approximate this complex multi-field functional organization, both HG and STG in each subject were subdivided into ROIs for quantitative analysis of high gamma activity recorded during the MMSE. Recording sites within HG were subdivided into posteromedial and anterolateral ROIs based on physiological criteria (Brugge et al., 2008, 2009). Specifically, recording sites were assigned to the posteromedial HG ROI if they exhibited phase-locked ECoG responses to 100 Hz click trains and averaged evoked potentials to these stimuli featured short-latency (<20 ms) components. Such response features are not present within anterolateral HG. Recording sites on the lateral surface of STG were subdivided into posterior and middle STG ROIs based on their location relative to the TTS, which is a continuation of Heschl's sulcus onto the lateral surface of the STG. This anatomical demarcation is supported by previous work demonstrating that phonological processing primarily engaged areas of the STG posterior to the TTS (Hickok and Poeppel, 2007; Hickok, 2009).

Following the approach of Eliades and Wang (2003) and Greenlee et al. (2011), differences in high gamma activity between listening and speaking were first evaluated for each recording site using the SI metric:

$$SI = \frac{\gamma_{\text{listening}} - \gamma_{\text{speaking}}}{\gamma_{\text{listening}} + \gamma_{\text{speaking}}}$$

where $\gamma_{\text{listening}}$ and γ_{speaking} are median high gamma power within the time windows corresponding to listening and speaking, respectively. For each ROI, SI values were compared to zero using Wilcoxon signed-rank tests.

The use of SI in this study differs from previous studies that compared auditory responses to self-initiated vocalizations with responses elicited by playback of the same utterances (e.g., Eliades and Wang, 2003; Greenlee et al., 2013). In contrast, the present study defined SI based on different speech material, specifically, comparisons between auditory responses elicited during listening to the interviewer and during verbally responding. The SI was used in a manner similar to a study that examined suppression of auditory activity on lateral STG during a repetition task (Flinker et al., 2010). Our study is novel in that it extends the findings of previous studies that used the same speech material to a conversational scenario.

Non-parametric statistical analysis was used for comparisons of high gamma ERBP between speaker conditions (interviewer vs. subject) and ROIs (posteromedial vs. anterolateral HG and posterior vs. middle STG). Wilcoxon rank sum test was used to compare average high gamma ERBP during listening to instructions of the interviewer and to the subject's own verbal responses. Wilcoxon signed-rank test was used for ROI comparisons. Correction for multiple comparisons was done by controlling FDR (Benjamini and Hochberg, 1995) using the linear step-up procedure, as implemented in MATLAB Version 7.14 Bioinformatics Toolbox. Previous work has demonstrated the utility of this statistical approach when examining ECoG recorded during a conversation-based paradigm (Derix et al., 2012).

RESULTS

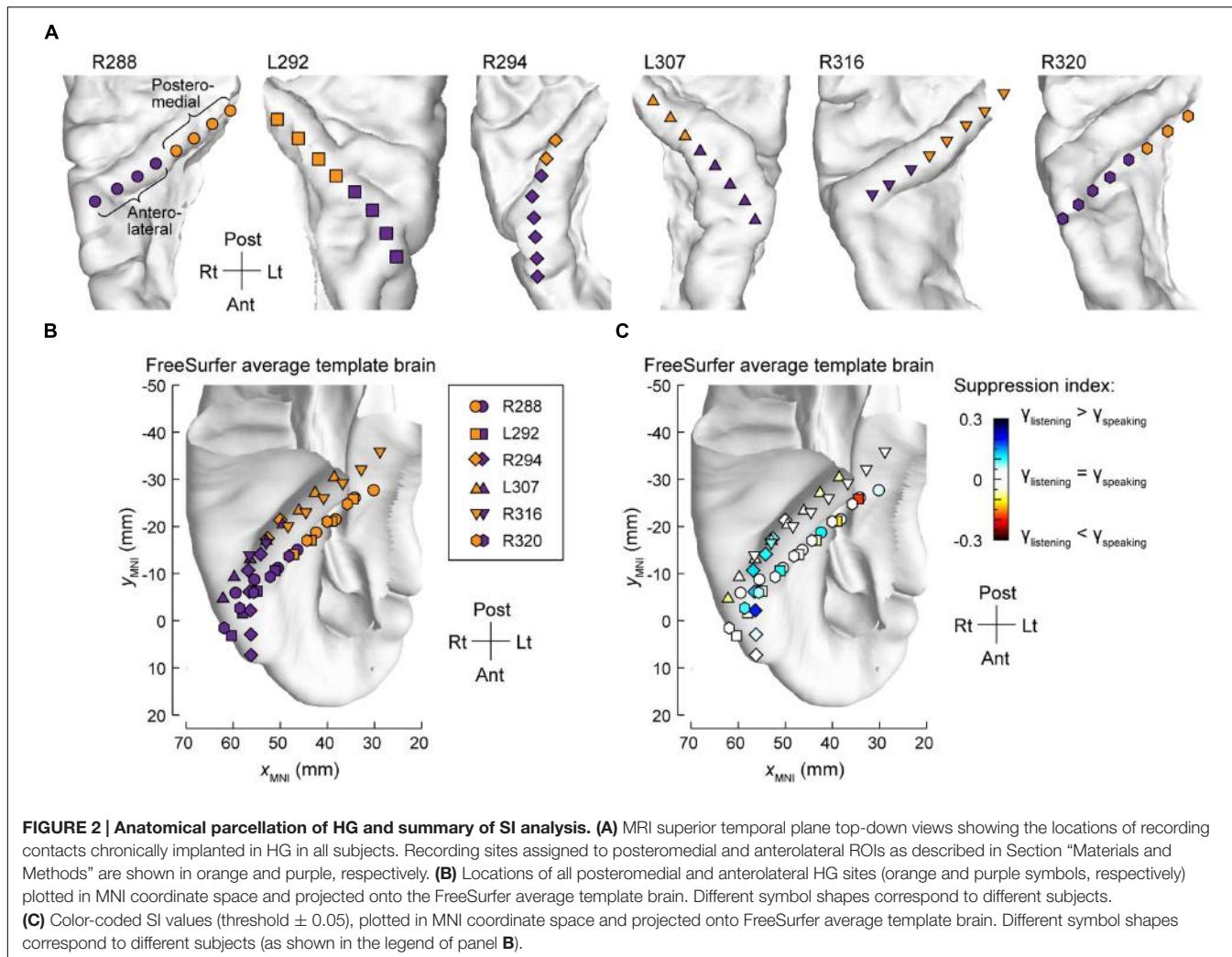
Heschl's Gyrus

As expected, HG was strongly activated by speech. However, activity was not uniform across its length. Two principal patterns of neural activity were identified that related to whether the utterances were the interviewer's questions, or were self-generated by the subject in response to these questions. These two patterns were anatomically segregated along HG. Specifically, activity recorded from sites within posteromedial HG was characterized by robust increases in high gamma power when the subject was both listening and speaking. This pattern is exemplified by data from two subjects (R288 and R294) in **Figure 1** (sites 'a' and 'c'). Increases in high gamma power were time-locked to the utterances of both the interviewer and subject. The second pattern was observed in anterolateral HG (sites 'b' and 'd' in **Figure 1**), wherein high gamma activity was generally of lower amplitude in response to self-initiated speech compared to listening.

The differences between high gamma activity in posteromedial and anterolateral HG were quantified for all subjects on an utterance-by-utterance basis by comparing activity elicited during listening and self-vocalizations (**Figure 2**). Locations of the recording sites along HG in all six subjects are shown in **Figure 2A**. Recording sites are color-coded according to whether they were in posteromedial or anterolateral portions of HG as determined physiologically by responses to simple non-speech stimuli (see Materials and Methods). These locations, pooled across all subjects and transferred onto the right HG, are plotted in MNI coordinate space over the FreeSurfer average template brain in **Figure 2B**. Pooling anatomical data across subjects demonstrated that ROI demarcation based on physiological response properties in individual subjects translated into anatomically distinct regions within HG at the population level. This finding supports the reliability of the physiology-based operational definitions of posteromedial (core) and anterolateral HG (non-core) cortex as implemented in the present study.

Changes in high gamma activity during listening vs. speaking were quantified as SIs for each recording site across the entire conversation (see Materials and Methods). Recording sites in posteromedial HG were characterized by SIs that were not significantly different than zero (Wilcoxon signed-rank test $p = 0.57$), indicating a comparable degree of activation during listening and speaking (**Figure 2C**). In contrast, sites localized to the anterolateral portion of HG did exhibit positive SIs (Wilcoxon signed-rank test $p < 0.005$), corresponding to a greater degree of activation during listening versus speaking.

Site-by-site analysis of SIs was effective in identifying differential patterns of speech-elicited activity along HG based on whether or not it was self-generated. This finding was confirmed by quantifying the differences between normalized high gamma activity (ERBP) measured during listening and speaking within the two HG ROIs (**Figure 3**). Utterance-by-utterance average high gamma power elicited during listening and self-initiated speech was calculated for each ROI in each subject. In posteromedial HG, activity elicited during listening



and self-vocalization was of similar magnitude (Wilcoxon rank sum test, FDR-corrected, $p > 0.05$) in five out of six subjects. In the sixth subject (L307) activity was greater during self-vocalization ($p < 0.05$). In contrast, activity in anterolateral HG was greater while listening in five out of six subjects ($p < 0.05$). In the sixth subject (L307), responses were not significantly different.

In summary, there was a significant change in high gamma activity patterns along HG, wherein its posteromedial portion exhibited robust responses to conversational speech regardless of the speaker, while its anterolateral aspect responded more strongly during listening.

Superior Temporal Gyrus

Similar to anterolateral HG, there was significant suppression of high gamma activity in response to self-initiated speech relative to listening on most sites along STG, as exemplified in Figure 4. In the language-dominant hemisphere of subject L307, site 'a' exhibited marked suppression of high gamma activity when the subject was speaking regardless of the task (Figure 4A). On a more anterior site 'b', this suppression was more nuanced,

with greater suppression occurring during the Verbal Analogies task compared to the Repetition task. The latter finding was comparable in the Immediate Recall task of the MMSE. Similar response patterns were observed in the non-language dominant hemisphere, exemplified by sites 'c' and 'd' in subject R316 (Figure 4B). In this subject, site 'c' again showed a more nuanced pattern of activity. In contrast to site 'b', responses to the subject's own speech were comparable to those when listening during the Verbal Analogies task, whereas suppression during speaking was evident during the Repetition task. A more anterior site 'd' showed a uniform pattern of marked suppression of activity when speaking, similar to site 'a' of subject L307.

It is likely that lateral STG contains multiple functional fields along its posterior-to-anterior axis (e.g., Hickok, 2009; Rauschecker and Scott, 2009). Accordingly, the distribution of electrodes along STG was examined to determine whether there were differences in suppression in posterior vs. middle portions of the STG. As physiological criteria currently do not provide a reliable means of identifying spatially distinct functional fields along the STG, anatomical criteria were used instead, based on the location of electrodes relative to the TTS (Figure 5A).

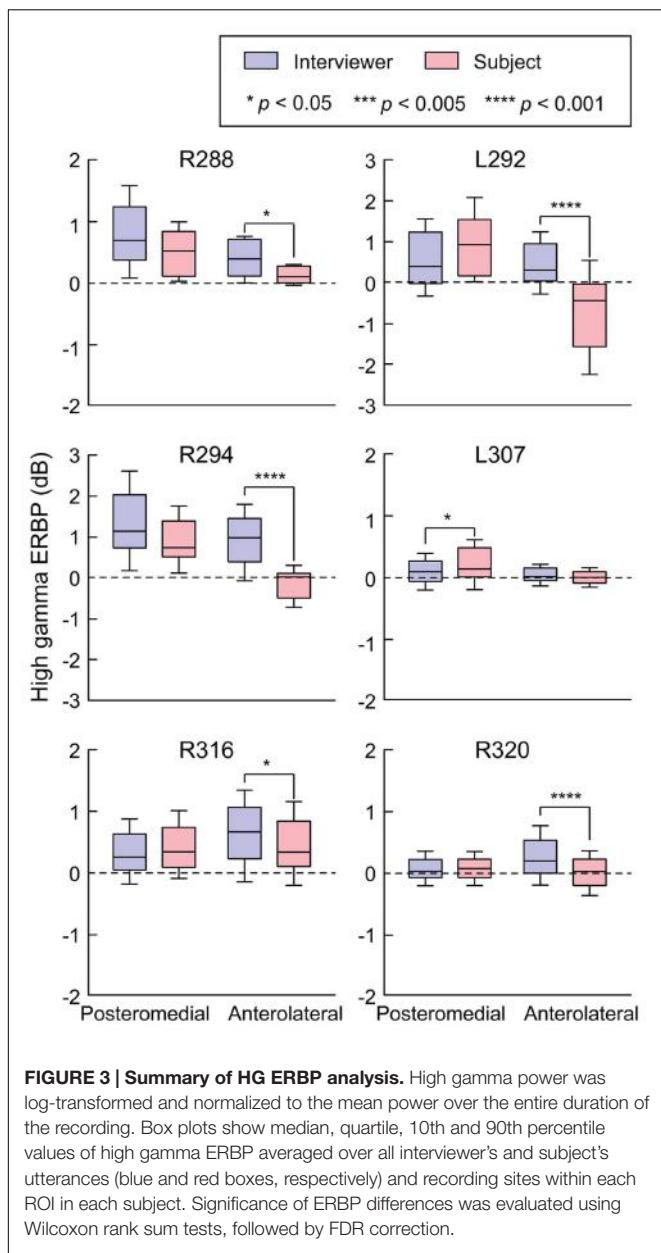


FIGURE 3 | Summary of HG ERBP analysis. High gamma power was log-transformed and normalized to the mean power over the entire duration of the recording. Box plots show median, quartile, 10th and 90th percentile values of high gamma ERBP averaged over all interviewer's and subject's utterances (blue and red boxes, respectively) and recording sites within each ROI in each subject. Significance of ERBP differences was evaluated using Wilcoxon rank sum tests, followed by FDR correction.

Superior temporal gyrus recording sites were then pooled across all six subjects and plotted in MNI coordinate space over the right hemisphere of the FreeSurfer average template brain (**Figure 5B**). In parallel with the evaluation of HG parcellation (cf. **Figure 2B**), there was concordance between STG ROI demarcation in each subject, and clustering of the recording sites into two ROIs in the MNI coordinate space with little overlap. The TTS thus provided a reliable gross anatomical criterion for STG ROI parcellation.

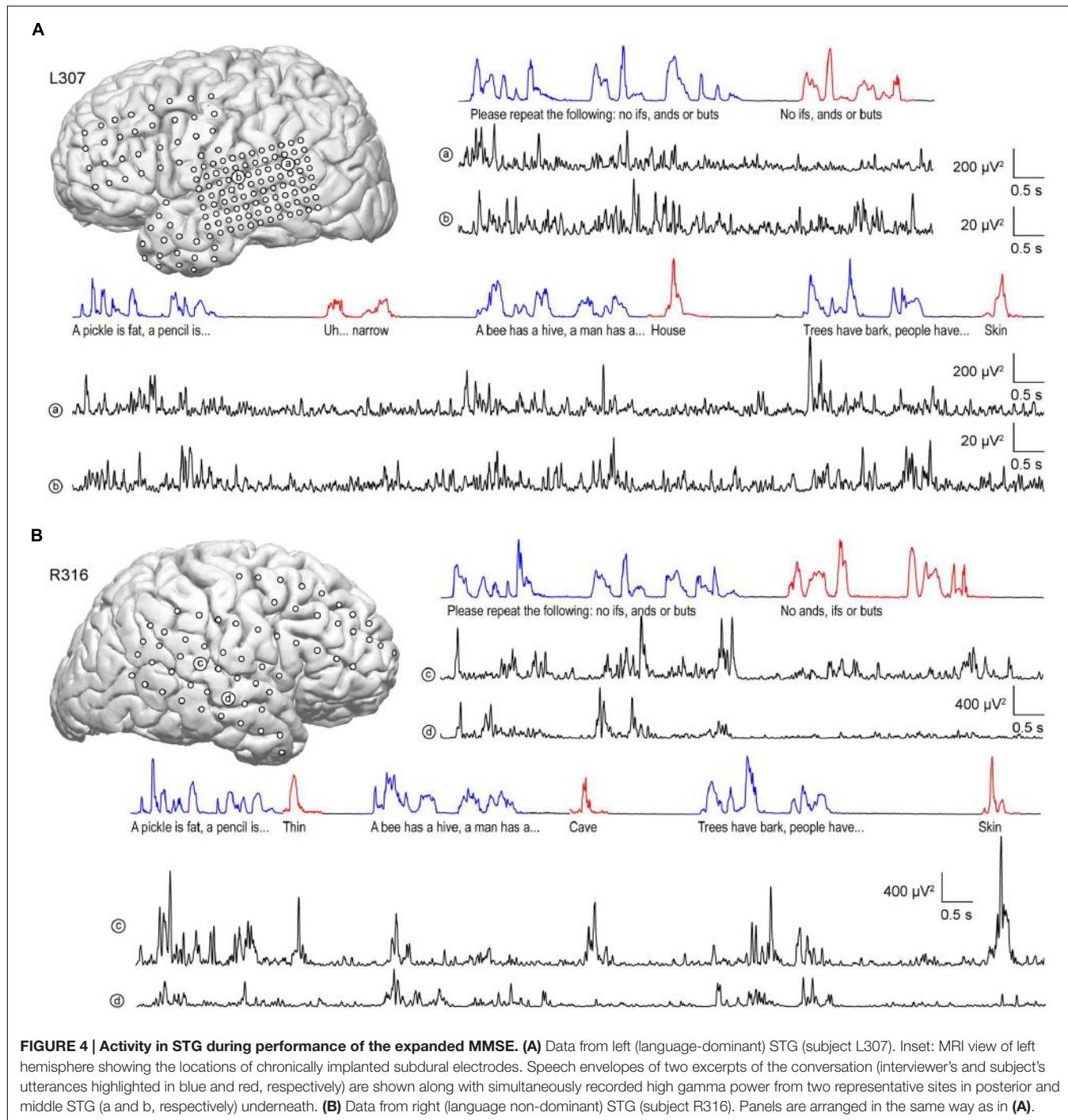
Differences between high gamma activity elicited during listening and speaking were quantified as SIs at each STG recording site (**Figure 5C**). On the population level, significant suppression ($p < 0.001$, Wilcoxon signed rank tests) was observed in both STG ROIs, with no significant difference identified

between the two ROIs ($p = 0.63$, Wilcoxon rank sum test). Instead, regions of suppression were interspersed with those exhibiting little-to-no suppression (cf. **Figure 4**). There appeared to be an overall lack of suppression between -20 and -40 mm on the y_{mni} axis when the data were pooled across subjects (white symbols, corresponding to $-0.05 < \text{SI} < 0.05$). However, most of those data points were contributed by the most posterior STG recording sites of subject R288 (hexagons). Therefore, the data should not be interpreted as suggesting that there is an orderly distribution of SIs along the long axis of the STG. This conclusion can only be made following a formal assessment of spatial distribution in the MNI coordinate space, which would require a larger number of subjects (see Nourski et al., 2014a) and is outside the scope of the current study.

As with examination of HG (see **Figure 3**), STG ROIs were further characterized using comparisons of high gamma activity normalized to the mean over the entire recording epoch (**Figure 6**). Significant suppression of high gamma activity during speaking was found in both posterior and middle STG in each subject. This suppression was further examined on a site-by-site basis in the three subjects with comprehensive lateral STG electrode coverage (L307, R288, and R320). In subject L307, 23 out of 26 STG sites (88.4%) exhibited significantly greater high gamma activity elicited during listening compared to speaking (Wilcoxon rank sum test, FDR-corrected, $p < 0.05$). No sites showed preference for self-vocalization. In subject R288, 12 out of 32 STG sites (37.5%) exhibited a significantly greater response when listening ($p < 0.05$), while two sites (6.25%) showed a reverse pattern, and 18 sites (56.25%) showed no difference. In subject R320, 15 out of 23 STG sites (65.2%) exhibited a significantly ($p < 0.05$) greater response when listening, while two sites (8.7%) showed a reverse pattern, and six sites (26.1%) showed no difference. Finally, there was no reliable difference between posterior and middle portions of lateral STG when comparing either responses elicited during listening or during speaking for all six subjects ($p > 0.05$).

Modulation by Task

Modulation of high gamma activity on STG as a function of task can occur at a single site level, as exemplified by site 'c' in **Figure 4B**. At this site, activity during the Repetition task was suppressed when speaking relative to listening, yet was not suppressed during the Rapid Naming task. We further examined this property at a population level by exploring whether there were any systematic differences while listening and speaking as a function of specific tasks in the expanded MMSE. For this exploration, we included periods of silence between listening to questions and responding in order to account for activity related to either processing of the former or planning the latter. This analysis is illustrated in **Figure 7**. Although the low number of exemplars for each task within the dialog precluded a formal statistical assessment, it can be observed that no systematic task effects were apparent at the population level of STG. Periods of silence between questions and answers were typically associated with negative ERBP values, and, in general, responses while speaking were less than while listening. These findings indicate that the comparisons of high



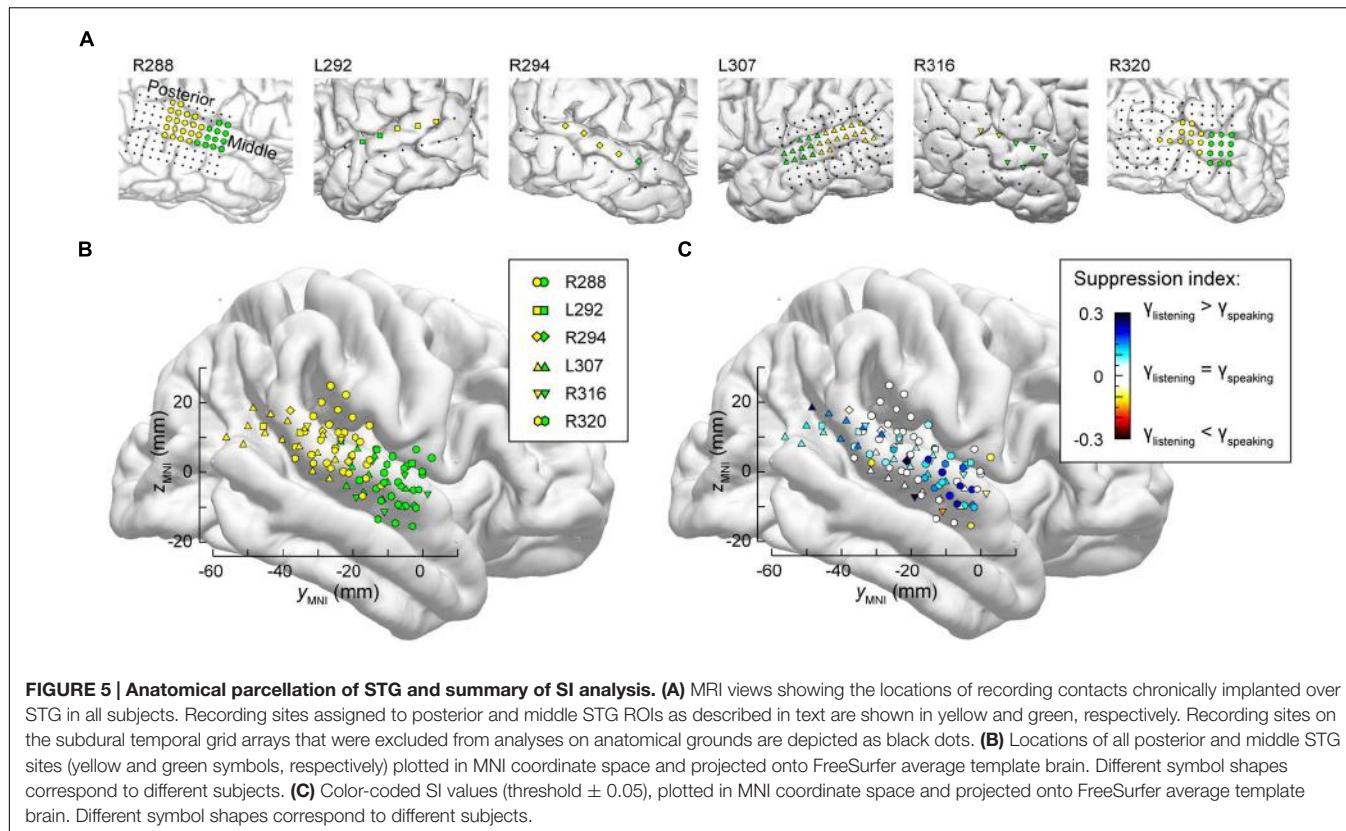
gamma activity while listening versus speaking, as depicted in Figures 5C and 6, were not affected by systematic task-specific biases on the group (ROI) level. Given that individual sites on the STG can be modulated by task, these results may represent a “fine-grain” property that would not be seen at the ROI level. Acquisition of additional data would be required to systematically evaluate this property of the auditory cortex of the STG. At the ROI level, current observations provide a comparison point when examining higher cortical areas likely involved in the

comprehension of questions, and the planning and execution of answers.

DISCUSSION

Summary of Findings

Using a conversation-based paradigm modeled after a commonly used neurological screening tool for dementia (the MMSE),



we examined high gamma ERBP at three stages of auditory cortical processing with regard to modulation when listening versus speaking. In posteromedial HG (core auditory cortex), no significant difference was found between activity during listening to the interviewer's questions and the subject's answers. This non-discriminative pattern changed within both anterolateral HG and lateral STG (non-core auditory cortical areas), where responses were significantly greater during listening compared to speaking. These observations are consistent with the idea that suppression of cortical activity to self-initiated speech is an emerging property of human non-core auditory cortex.

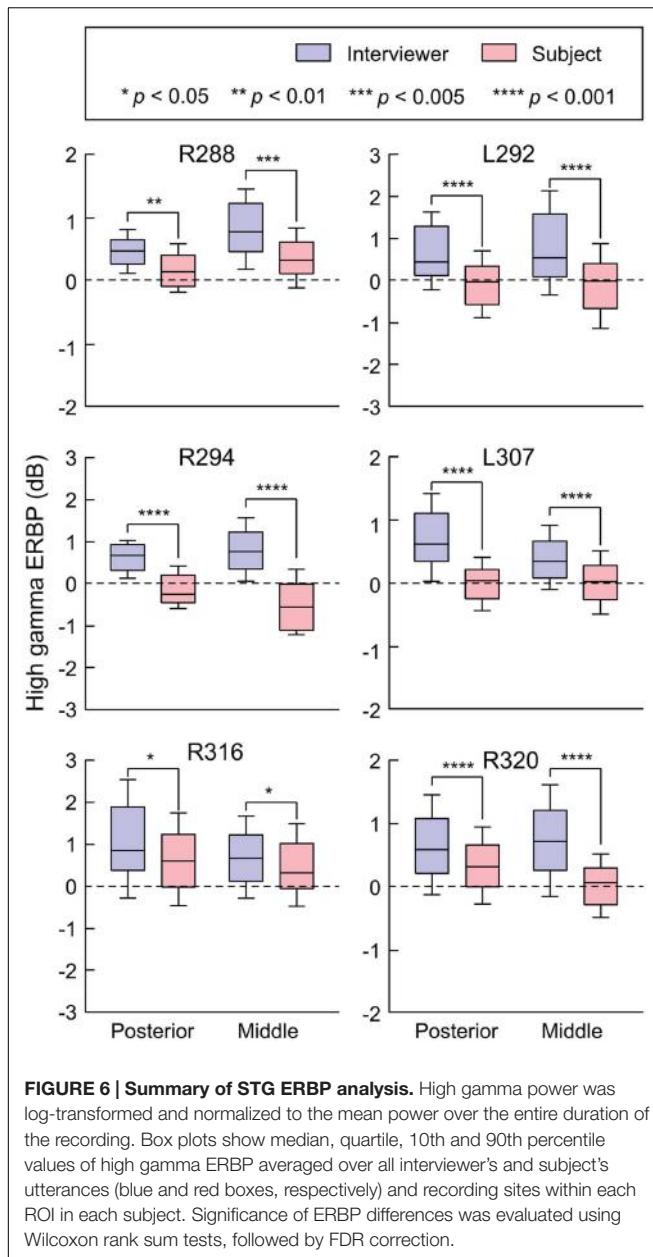
Heschl's Gyrus

This is the first detailed report to compare neural activity in human core auditory cortex during listening and speaking in a dialog-based paradigm. High gamma activity in posteromedial HG was not significantly modulated by speaker during the performance of the expanded MMSE. This observation is consistent with previous reports examining cortical high gamma activity in posteromedial HG, showing that this area responds indiscriminately to a wide array of simple and complex sounds, including intelligible and unintelligible speech (e.g., Brugge et al., 2009; Nourski et al., 2009a; Steinschneider et al., 2013) as well as while speaking or listening to playback of one's own speech (Greenlee et al., 2014; Behroozmand et al., 2016). Further, high gamma activity in posteromedial HG is not strongly modulated by experimental context or specific task requirements (Steinschneider et al., 2014). Preliminary observations also

demonstrate that early high gamma activity in posteromedial HG is even preserved under general anesthesia (Nourski et al., 2009b). In the setting of the current study, high gamma responses elicited by self-initiated vocalizations provide a further example of the breadth of acoustic inputs that activate core auditory cortex.

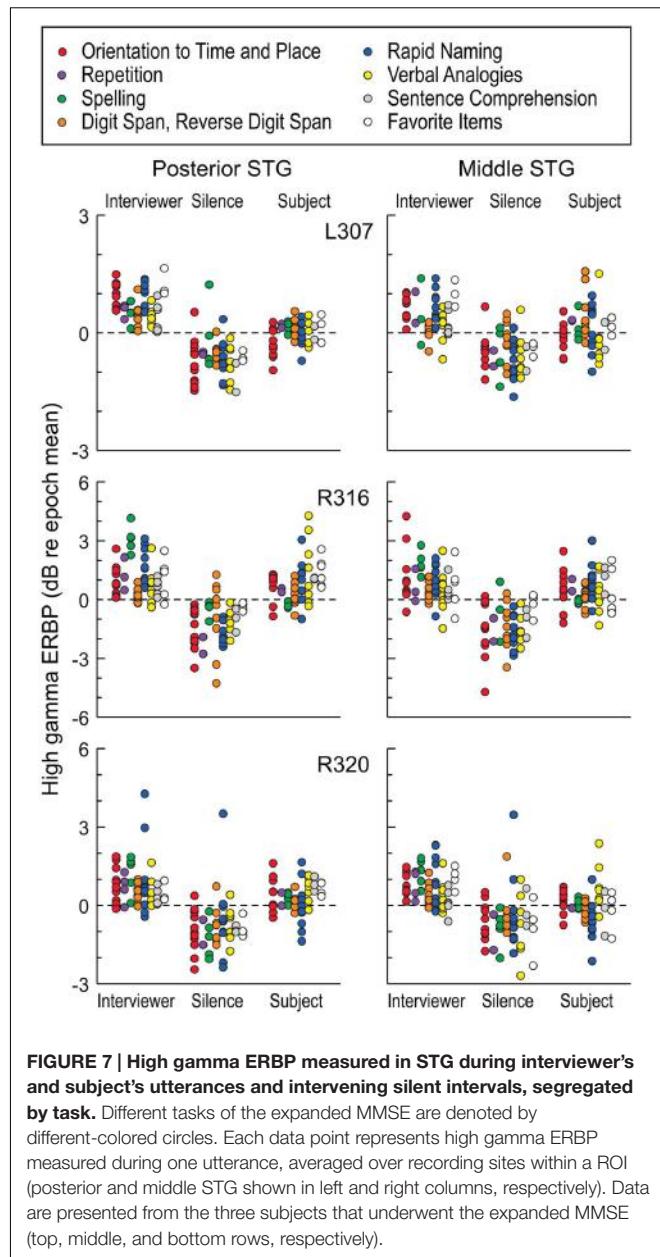
Auditory cortex in posteromedial HG exhibits phase locking to voice F_0 , particularly for male talkers whose speech is typically characterized by lower F_0 values (e.g., Nourski and Brugge, 2011; Steinschneider et al., 2013; Behroozmand et al., 2016). These phase-locked responses would contribute to high gamma ERBP measured in posteromedial HG, and thus introduce a potential confound for comparisons between responses to utterances of different talkers with different F_0 s. Three out of six subjects in the present study (L292, R316, and R320) were female, and two of them (L292 and R316) had average F_0 values higher than that of the male interviewer (see Table 2). Activity in posteromedial HG was not greater when listening to the interviewer compared to speaking in these subjects (see Figures 2 and 3). Further, the average voice F_0 of the interviewer during these conversations (155 and 139.8 Hz) was at frequencies that were borderline with regard to the ability to elicit phase-locked responses (see Steinschneider et al., 2013; Behroozmand et al., 2016), again minimizing their potential contribution to our results.

It should be noted that the only subject where high gamma activity was significantly greater during speaking (L307) had the lowest voice F_0 (120.7 Hz), and it was significantly lower than the interviewer's voice F_0 . Even though phase-locked activity may have contributed to the observed significant difference in high



gamma ERBP in this subject, it does not alter the conclusion that there is no systematic suppression of high gamma activity during self-generated speech at the level of posteromedial HG when compared to listening.

Utterances phrased as questions are often characterized by higher F_0 values than utterances phrased as statements (e.g., Eady and Cooper, 1986). It's not likely, however, that higher F_0 s associated with the interviewer's questions would affect the results reported in the present study, as many of the interviewer's utterances were phrased as statements (see Supplementary Table 1). Also, upward inflections in the F_0 are often seen toward the end of a question, and do not substantially contribute to the overall high gamma response profiles when averaged over the entire utterance.



Given that responses when listening were greater than during self-generated speech in anterolateral HG and lateral STG, it is conceivable that these results could be skewed by the differences in voice F_0 s between the interviewer and the subjects. However, multiple studies have shown that these ROIs do not phase-lock to speech with voice F_0 s within the range occurring in the current study (e.g., Nourski and Brugge, 2011; Steinschneider et al., 2011; Steinschneider, 2013). This indicates that results represent genuine suppression of activity to self-initiated speech in these ROIs.

The finding that high gamma activity within posteromedial HG was not suppressed during self-vocalizations apparently contradicts human non-invasive studies. Neuromagnetic studies have revealed a decrement in the M100 component during

speaking compared to listening (Houde et al., 2002; see also Numminen et al., 1999). However, the M100 is the sum of multiple generators with greater contributions from non-primary cortex on the superior temporal plane than HG (Scherg et al., 1989; Liégeois-Chauvel et al., 1994). Thus, the decrements seen while speaking could be a property of those non-primary areas rather than posteromedial HG.

In the marmoset, a New World monkey, two types of single-cell activity within primary and surrounding secondary auditory cortical areas have been described to occur during self-vocalization (Eliades and Wang, 2003). Vocalization-induced suppression of activity was seen in the majority of cells, but a significant minority showed increased discharges during self-vocalizations. Overall, summation of net activity generated by these cell populations was excitatory (Eliades and Wang, 2005). Our failure to find significant differences between responses during listening and speaking at the level of posteromedial HG may reflect limitations inherent to population responses (such as high gamma activity) in differentiating the fine-grain excitatory and inhibitory patterns associated with these two sources of acoustic inputs. On the other hand, mechanisms that preserve responses to self-vocalizations as seen in the current study at the level of core auditory cortex may be a necessary component of cortical pathways involved in self-monitoring of one's own speech (Eliades and Wang, 2003, 2008; Rauschecker and Scott, 2009).

In contrast to posteromedial HG, high gamma activity within anterolateral portions of HG was both generally lower in magnitude and exhibited suppression during speaking. The decrement in response magnitude along HG has been a consistent finding in previous studies that examined high gamma activity using multiple sound stimuli in more controlled trial-based paradigms (e.g., Brugge et al., 2009; Nourski et al., 2009a; Nourski and Brugge, 2011). The change in magnitude of response along HG has been interpreted as reflecting a change from a core to a non-core field, and is consistent with anatomic parcellations of HG (e.g., Hackett et al., 2001). This interpretation is further supported by the transformation that occurs between posteromedial and anterolateral HG in terms of sensitivity to self-vocalization vs. listening as seen in the present study.

It is premature to draw conclusions regarding comparisons between the results obtained from HG in the only language-dominant hemisphere examined (subject L307) with those obtained from the five other subjects. Comparisons regarding response properties in HG (see **Figure 3**) require special caution because of the limited sampling in each subject. Thus, enhanced activity during speaking in posteromedial HG of subject L307 does not necessarily reflect a consistent difference in auditory processing between language dominant and non-dominant hemispheres at the level of auditory core cortex. What is consistent across all subjects, and which is a main finding of the present study, is that there is a lack of suppression of activity within auditory core cortex during speaking compared to listening regardless of the language dominance. Inclusion of many more subjects who clinically require placement of depth electrodes in the superior temporal plane of the language-dominant hemisphere would be required to reveal systematic

differences across the hemispheres. It should also be noted that many models of speech perception posit that such differences emerge at later stages of auditory cortical processing (e.g., superior temporal sulcus; Leaver and Rauschecker, 2010).

Superior Temporal Gyrus

The STG was strongly activated during our conversation-based paradigm in all subjects, including the five subjects in which the non-language dominant hemisphere was studied, as well as in the single subject (L307) with language dominant hemisphere electrode coverage. As previously reported by Greenlee et al. (2011), high gamma activity during speaking was generally attenuated when compared to listening to the playback of one's own vocalizations. Suppressed activity during speaking occurred at sites in both posterior and middle portions of STG, which were intermingled with sites that exhibited no such suppression. This patchy distribution has been described in both humans and non-human primate models (Eliades and Wang, 2003; Greenlee et al., 2011). Interestingly, suppression of neural activity during self-vocalizations in the monkey was primarily seen in upper cortical laminae (Eliades and Wang, 2005). Activity generated within upper laminae would provide a major contribution to the population responses (high gamma) as captured by subdural electrodes immediately over lamina 1.

It is tempting to compare the overall magnitude of responses and the degree of self-vocalization suppression between anterolateral HG and STG. However, the extent of sampling was less for anterolateral HG and lateral STG responses were obtained from the pial surface as opposed to the brain parenchyma. For these reasons, we refrain from making conclusions regarding the relative degree of suppression of activity to self-vocalizations between anterolateral HG and STG.

Phonetic Feature Representation

The lateral STG has been shown to encode phonetic features at both the single-electrode and population level (Mesgarani et al., 2014). The role of phonetic modulation in the neural activity within STG was not currently studied due to several technical restraints. First, the density of coverage over the posterior and middle STG in our subject cohort (between 5 and 32 recording sites) was considerably smaller than that in the study of Mesgarani et al. (2014), where the number of STG sites in each subject was generally greater than 80 and reached a maximum of 102. Next, the number of spoken sentences that was drawn upon for analysis of phonemic representation by Mesgarani et al. (2014) came from a well-designed acoustic-phonetic speech corpus (TIMIT; Garofolo, 1993) and greatly exceeded those in our data sets. Further, the conversational nature of the experimental paradigm in our study precluded the use of a local prestimulus baseline as utilized by Mesgarani et al. (2014). Finally, our study required participants to perform multiple verbal tasks while listening to the interviewer as opposed to passive listening to continuously presented sentences. It is possible that task demands might greatly increase the overall complexity of neural response patterns and thus partially mask effects based on phonetic representation. It should be stressed that our findings do not contradict the results of

Mesgarani et al. (2014), but instead shed light on complementary organizational properties of the STG in an active conversation-based paradigm.

Task Modulation

While at the population level of the STG, there was no systematic variation of high gamma activity according to task, activity at individual recording sites could show task-specific modulation during the subject's verbal responses (see **Figure 4**). Modulation of high gamma activity at the level of the STG was not observed during the listening phase of the dialog. It is unclear what mechanisms drive this effect, and further work is clearly needed to categorize the functional specialization underlying task modulation observed at the level of single electrodes, and whether these effects occur in specific regions of posterior and middle STG.

CONCLUSION

The utility of this conversation-based paradigm is supported by its ability to reliably reproduce findings such as speaker modulation on the lateral STG, and transformation of patterns of activity across regions of auditory cortex. It follows in the footsteps of previous intracranial studies demonstrating the ability to study social interactions, "cognitive ideas" and numerical processing in non-experimental settings (Derix et al., 2012, 2014; Dastjerdi et al., 2013). As such, this study lays the groundwork for analysis of this paradigm's ability to rapidly evaluate task-specific activity related to language processing at higher levels of auditory-related cortex and its interface with regions of the brain involved in cognitive and affective functions. The expanded MMSE permits these examinations in a rapid and efficient manner, taking into account factors such as fatigue that commonly occur in patients being evaluated for their medically intractable epilepsy. While this study was limited to

REFERENCES

- Arevalo-Rodriguez, I., Smailagic, N., Roqué, I., Figuls, M., Ciapponi, A., Sanchez-Perez, E., et al. (2015). Mini-mental state examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI). *Cochrane Database Syst. Rev.* 3:CD010783. doi: 10.1002/14651858.CD010783.pub2
- Bartolomei, F., Chauvel, P., and Wendling, F. (2008). Epileptogenicity of brain structures in human temporal lobe epilepsy: a quantified study from intracerebral EEG. *Brain* 131, 1818–1830. doi: 10.1093/brain/awn111
- Bartolomei, F., Wendling, F., Vignal, J. P., Kochen, S., Bellanger, J. J., Badier, J. M., et al. (1999). Seizures of temporal lobe epilepsy: identification of subtypes by coherence analysis using stereo-electro-encephalography. *Clin. Neurophysiol.* 110, 1741–1754. doi: 10.1016/S1388-2457(99)00107-8
- Behroozmand, R., Oya, H., Nourski, K. V., Kawasaki, H., Larson, C., Brugge, J. F., et al. (2016). Neural correlates of vocal production and motor control in human Heschl's gyrus. *J. Neurosci.* 36, 2302–2315. doi: 10.1523/JNEUROSCI.3305-14.2016
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57, 289–300.
- Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Springer, J. A., Kaufman, J. N., et al. (2000). Human temporal lobe activation by high gamma activity, it is recognized that future studies must also incorporate examination of lower frequency bands and coherence across sensory, cognitive, and affective areas. Finally, the results obtained from the expanded MMSE should permit formulation of novel hypotheses that can be tested using more formal, controlled experimental designs.

AUTHOR CONTRIBUTIONS

MS conceived the study; KN and MS designed the study; KN and AR collected the data; KN and MS analyzed and interpreted the data. All authors wrote the manuscript, approved its final version, and agreed to be accountable for all aspects of the work.

FUNDING

This study was supported by grants NIH R01-DC04290, UL1RR024979, NSF CRCNS-1515678 and the Hoover Fund.

ACKNOWLEDGMENTS

We thank Jeremy Greenlee and Matthew Howard for helpful comments on the manuscript, and Timothy Ando, Haiming Chen, Phillip Gander, Hiroto Kawasaki, Christopher Kovach, Hiroyuki Oya and Xiayi Wang for help with data acquisition and analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fnhum.2016.00202>

- speech and nonspeech sounds. *Cereb. Cortex* 10, 512–528. doi: 10.1093/cercor/10.5.512
- Brugge, J. F., Nourski, K. V., Oya, H., Reale, R. A., Kawasaki, H., Steinschneider, M., et al. (2009). Coding of repetitive transients by auditory cortex on Heschl's gyrus. *J. Neurophysiol.* 102, 2358–2374. doi: 10.1152/jn.91346.2008
- Brugge, J. F., Volkov, I. O., Oya, H., Kawasaki, H., Reale, R. A., Fenoy, A., et al. (2008). Functional localization of auditory cortical fields of human: click-train stimulation. *Hear. Res.* 238, 12–24. doi: 10.1016/j.heares.2007.11.012
- Chang, E. F., Edwards, E., Nagarajan, S. S., Fogelson, N., Dalal, S. S., Canolty, R. T., et al. (2011). Cortical spatio-temporal dynamics underlying phonological target detection in humans. *J. Cogn. Neurosci.* 23, 1437–1446. doi: 10.1162/jocn.2010.21466
- Creutzfeldt, O., Ojemann, G., and Lettich, E. (1989). Neuronal activity in the human lateral temporal lobe. II. Responses to the subjects own voice. *Exp. Brain Res.* 77, 476–489. doi: 10.1007/BF00249600
- Crone, N. E., Boatman, D., Gordon, B., and Hao, L. (2001). Induced electrocorticographic gamma activity during auditory perception. Brazier Award-winning article. *Clin. Neurophysiol.* 112, 565–582. doi: 10.1016/S1388-2457(00)00545-9
- Dastjerdi, M., Ozker, M., Foster, B. L., Rangarajan, V., and Parvizi, J. (2013). Numerical processing in the human parietal cortex during experimental and natural conditions. *Nat. Commun.* 4:2528. doi: 10.1038/ncomms3528

- de Cheveigné, A., and Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *J. Acoust. Soc. Am.* 111, 1917–1930. doi: 10.1121/1.1458024
- Derix, J., Iljina, O., Schulze-Bonhage, A., Aertsen, A., and Ball, T. (2012). “Doctor” or “darling”? Decoding the communication partner from ECoG of the anterior temporal lobe during non-experimental, real-life social interaction. *Front. Hum. Neurosci.* 6:251. doi: 10.3389/fnhum.2012.00251
- Derix, J., Iljina, O., Weiske, J., Schulze-Bonhage, A., Aertsen, A., and Ball, T. (2014). From speech to thought: the neuronal basis of cognitive units in non-experimental, real-life communication investigated using ECoG. *Front. Hum. Neurosci.* 8:383. doi: 10.3389/fnhum.2014.00383
- Eady, S. J., and Cooper, W. E. (1986). Speech intonation and focus location in matched statements and questions. *J. Acoust. Soc. Am.* 80, 402–415. doi: 10.1121/1.394091
- Eliades, S. J., and Wang, X. (2003). Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. *J. Neurophysiol.* 89, 2194–2207. doi: 10.1152/jn.00627.2002
- Eliades, S. J., and Wang, X. (2005). Dynamics of auditory-vocal interaction in monkey auditory cortex. *Cereb. Cortex* 15, 1510–1523. doi: 10.1093/cercor/bhi030
- Eliades, S. J., and Wang, X. (2008). Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature* 453, 1102–1106. doi: 10.1038/nature06910
- Finney, G. R., Minagar, A., and Heilman, K. M. (2016). Assessment of mental status. *Neurol. Clin.* 34, 1–16. doi: 10.1016/j.ncl.2015.08.001
- Flinker, A., Chang, E. F., Kirsch, H. E., Barbaro, N. M., Crone, N. E., and Knight, R. T. (2010). Single-trial speech suppression of auditory cortex activity in humans. *J. Neurosci.* 30, 16643–16650. doi: 10.1523/JNEUROSCI.1809-10.2010
- Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 12, 189–198. doi: 10.1016/0022-3956(75)90026-6
- Garfolo, J. S. (1993). *TIMIT: Acoustic-Phonetic Continuous Speech Corpus*. Philadelphia: Linguistic Data Consortium.
- Gavaret, M., Badier, J. M., Marquis, P., McGonigal, A., Bartolomei, F., Regis, J., et al. (2006). Electric source imaging in frontal lobe epilepsy. *J. Clin. Neurophysiol.* 23, 358–370. doi: 10.1097/01.wnp.0000214588.94843.c2
- Greenlee, J. D., Behroozmand, R., Larson, C. R., Jackson, A. W., Chen, F., Hansen, D. R., et al. (2013). Sensory-motor interactions for vocal pitch monitoring in non-primary human auditory cortex. *PLoS ONE* 8:e60783. doi: 10.1371/journal.pone.0060783
- Greenlee, J. D., Behroozmand, R., Nourski, K. V., Oya, H., Kawasaki, H., and Howard, M. A. (2014). Using speech and electrocorticography to map human auditory cortex. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2014, 6798–6801. doi: 10.1109/EMBC.2014.6945189
- Greenlee, J. D., Jackson, A. W., Chen, F., Larson, C. R., Oya, H., Kawasaki, H., et al. (2011). Human auditory cortical activation during self-vocalization. *PLoS ONE* 6:e14744. doi: 10.1371/journal.pone.0014744
- Hackett, T. A. (2015). Anatomic organization of the auditory cortex. *Handb. Clin. Neurol.* 129, 27–53. doi: 10.1016/B978-0-444-62630-1.00002-0
- Hackett, T. A., Preuss, T. M., and Kaas, J. H. (2001). Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *J. Comp. Neurol.* 441, 197–222. doi: 10.1002/cne.1407
- Hickok, G. (2009). The functional neuroanatomy of language. *Phys. Life Rev.* 6, 121–143. doi: 10.1016/j.plrev.2009.06.001
- Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402. doi: 10.1038/nrn2113
- Houde, J. F., Nagarajan, S. S., Sekihara, K., and Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: an MEG study. *J. Cogn. Neurosci.* 14, 1125–1138. doi: 10.1162/089892902760807140
- Howard, M. A. III, Volkov, I. O., Granner, M. A., Damasio, H. M., Ollendieck, M. C., and Bakken, H. E. (1996). A hybrid clinical-research depth electrode for acute and chronic *in vivo* microelectrode recording of human brain neurons. *Technical J. Neurosurg.* 84, 129–132. doi: 10.3171/jns.1996.84.1.0129
- Howard, M. A., Volkov, I. O., Mirsky, R., Garell, P. C., Noh, M. D., Granner, M., et al. (2000). Auditory cortex on the human posterior superior temporal gyrus. *J. Comp. Neurol.* 416, 79–92. doi: 10.1002/(SICI)1096-9861(20000103)416:1.1.1
- Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17, 825–841. doi: 10.1006/nimg.2002.1132
- Leaver, A. M., and Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *J. Neurosci.* 30, 7604–7612. doi: 10.1523/JNEUROSCI.0296-10.2010
- Leonard, M. K., and Chang, E. F. (2014). Dynamic speech representations in the human temporal lobe. *Trends Cogn. Sci.* 18, 472–479. doi: 10.1016/j.tics.2014.05.001
- Liégeois-Chauvel, C., Musolino, A., Badier, J. M., Marquis, P., and Chauvel, P. (1994). Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components. *Electroencephalogr. Clin. Neurophysiol.* 92, 204–214. doi: 10.1016/0168-5597(94)90064-7
- Liégeois-Chauvel, C., Musolino, A., and Chauvel, P. (1991). Localization of the primary auditory area in man. *Brain* 114, 139–151.
- Maillard, L., Vignal, J. P., Gavaret, M., Guye, M., Biraben, A., McGonigal, A., et al. (2004). Semiotic and electrophysiologic correlations in temporal lobe seizure subtypes. *Epilepsia* 45, 1590–1599. doi: 10.1111/j.0013-9580.2004.09704.x
- McGonigal, A., Bartolomei, F., Régis, J., Guye, M., Gavaret, M., Trébuchon-Da Fonseca, A., et al. (2007). Stereoelectroencephalography in presurgical assessment of MRI-negative epilepsy. *Brain* 130, 3169–3183. doi: 10.1093/brain/awm218
- Mesgarani, N., and Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* 485, 233–236. doi: 10.1038/nature11020
- Mesgarani, N., Cheung, C., Johnson, K., and Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science* 343, 1006–1010. doi: 10.1126/science.1245994
- Müller-Preuss, P., and Ploog, D. (1981). Inhibition of auditory cortical neurons during phonation. *Brain Res.* 215, 61–76. doi: 10.1016/0006-8993(81)90491-1
- Munari, C. (1987). “Depth electrode implantation at Hôpital Sainte Anne, Paris,” in *Surgical Treatment of the Epilepsies*, ed. J. Engel Jr. (New York, NY: Raven Press), 583–588.
- Nourski, K. V., and Brugge, J. F. (2011). Representation of temporal sound features in the human auditory cortex. *Rev. Neurosci.* 22, 187–203. doi: 10.1515/RNS.2011.016
- Nourski, K. V., Brugge, J. F., Reale, R. A., Kovach, C. K., Oya, H., Kawasaki, H., et al. (2013). Coding of repetitive transients by auditory cortex on posterolateral superior temporal gyrus in humans: an intracranial electrophysiology study. *J. Neurophysiol.* 109, 1283–1295. doi: 10.1152/jn.00712.2012
- Nourski, K. V., and Howard, M. A. III. (2015). Invasive recordings in the human auditory cortex. *Handb. Clin. Neurol.* 129, 225–244. doi: 10.1016/B978-0-444-62630-1.00013-5
- Nourski, K. V., Reale, R. A., Oya, H., Kawasaki, H., Kovach, C. K., Chen, H., et al. (2009a). Temporal envelope of time-compressed speech represented in the human auditory cortex. *J. Neurosci.* 29, 15564–15574. doi: 10.1523/JNEUROSCI.3065-09.2009
- Nourski, K. V., Reddy, C., Oya, H., Kawasaki, H., and Brugge, J. (2009b). “General anesthesia differentially affects responses from auditory cortical fields in humans,” in *Proceeding of the 3rd International Conference on Auditory Cortex*, Magdeburg, 116.
- Nourski, K. V., Steinschneider, M., McMurray, B., Kovach, C. K., Oya, H., Kawasaki, H., et al. (2014a). Functional organization of human auditory cortex: investigation of response latencies through direct recordings. *Neuroimage* 101, 598–609. doi: 10.1016/j.neuroimage.2014.07.004
- Nourski, K. V., Steinschneider, M., Oya, H., Kawasaki, H., Jones, R. D., and Howard, M. A. (2014b). Spectral organization of the human lateral superior temporal gyrus revealed by intracranial recordings. *Cereb. Cortex* 24, 340–352. doi: 10.1093/cercor/bhs314
- Nourski, K. V., Steinschneider, M., Oya, H., Kawasaki, H., and Howard, M. A. III. (2015a). Modulation of response patterns in human auditory cortex during a target detection task: an intracranial electrophysiology study. *Int. J. Psychophysiol.* 95, 191–201. doi: 10.1016/j.jippsycho.2014.03.006

- Nourski, K. V., Steinschneider, M., Rhone, A. E., Oya, H., Kawasaki, H., Howard, M. A. III, et al. (2015b). Sound identification in human auditory cortex: differential contribution of local field potentials and high gamma power as revealed by direct intracranial recordings. *Brain Lang.* 148, 37–50. doi: 10.1016/j.bandl.2015.03.003
- Numminen, J., Salmelin, R., and Hari, R. (1999). Subject's own speech reduces reactivity of the human auditory cortex. *Neurosci. Lett.* 265, 119–122. doi: 10.1016/S0304-3940(99)00218-9
- Obleser, J., Eisner, F., and Kotz, S. A. (2008). Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *J. Neurosci.* 28, 8116–8124. doi: 10.1523/JNEUROSCI.1290-08.2008
- Rauschecker, J. P., and Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* 12, 718–724. doi: 10.1038/nn.2331
- Reddy, C. G., Dahdaleh, N. S., Albert, G., Chen, F., Hansen, D., Nourski, K., et al. (2010). A method for placing Heschl gyrus depth electrodes. *J. Neurosurg.* 112, 1301–1307. doi: 10.3171/2009.7.JNS09404
- Scherg, M., Vajsar, J., and Picton, T. W. (1989). A source analysis of the late human auditory evoked potentials. *J. Cogn. Neurosci.* 1, 336–355. doi: 10.1162/jocn.1989.1.4.336
- Steinschneider, M. (2013). "Phonemic representations and categories," in *Neural Correlates of Auditory Cognition, Springer Handbook of Auditory Research*, Vol. 45, eds Y. E. Cohen, A. N. Popper, and R. R. Fay (New York, NY: Springer Science+Business Media), 151–191.
- Steinschneider, M., Nourski, K. V., Kawasaki, H., Oya, H., Brugge, J. F., and Howard, M. A. III (2011). Intracranial study of speech-elicited activity on the human posterolateral superior temporal gyrus. *Cereb. Cortex* 21, 2332–2347. doi: 10.1093/cercor/bhr014
- Steinschneider, M., Nourski, K. V., and Fishman, Y. I. (2013). Representation of speech in human auditory cortex: is it special? *Hear. Res.* 305, 57–73. doi: 10.1016/j.heares.2013.05.013
- Steinschneider, M., Nourski, K. V., Rhone, A. E., Kawasaki, H., Oya, H., and Howard, M. A. III (2014). Differential activation of human core, non-core and auditory-related cortex during speech categorization tasks as revealed by intracranial recordings. *Front. Neurosci.* 8:240. doi: 10.3389/fnins.2014.00240

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Nourski, Steinschneider and Rhone. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The coordination dynamics of social neuromarkers

Emmanuelle Tognoli^{1*} and J. A. Scott Kelso^{1,2}

¹ Human Brain and Behavior Laboratory, Center for Complex Systems and Brain Sciences, Florida Atlantic University, Boca Raton, FL, USA, ² Intelligent System Research Centre, Ulster University, Derry ~ Londonderry, UK

Social behavior is a complex integrative function that entails many aspects of the brain's sensory, cognitive, emotional and movement capacities. Its neural processes are seldom simultaneous but occur according to precise spatiotemporal choreographies, manifested by the coordination of their oscillations within and between brains. Methods with good temporal resolution can help to identify so-called "neuromarkers" of social function and aid in disentangling the dynamical architecture of social brains. In our ongoing research, we have used dual-electroencephalography (EEG) to study neuromarker dynamics during synchronic interactions in which pairs of subjects coordinate behavior spontaneously and intentionally (social coordination) and during diachronic transactions that require subjects to perceive or behave in turn (action observation, delayed imitation). In this paper, after outlining our dynamical approach to the neurophysiological basis of social behavior, we examine commonalities and differences in the neuromarkers that are recruited for both kinds of tasks. We find the neuromarker landscape to be task-specific: synchronic paradigms of social coordination reveal medial mu, alpha and the phi complex as contributing neuromarkers. Diachronic tasks recruit alpha as well, in addition to lateral mu rhythms and the newly discovered nu and kappa rhythms whose functional significance is still unclear. Social coordination, observation, and delayed imitation share commonality of context: in each of our experiments, subjects exchanged information through visual perception and moved in similar ways. Nonetheless, there was little overlap between their neuromarkers, a result that hints strongly of task-specific neural mechanisms for social behavior. The only neuromarker that transcended both synchronic and diachronic social behaviors was the ubiquitous alpha rhythm, which appears to be a key signature of visually-mediated social behaviors. The present paper is both an entry point and a challenge: much work remains to determine the nature and scope of recruitment of other neuromarkers, and to create theoretical models of their within- and between-brain dynamics during social interaction.

OPEN ACCESS

Edited by:

Joachim Gross,
University of Glasgow, UK

Reviewed by:

Anne Keitel,
University of Glasgow, UK

Anne Hauswald,
University of Trento, Italy

***Correspondence:**

Emmanuelle Tognoli
tognoli@ccs.fau.edu

Received: 27 January 2015

Accepted: 25 September 2015

Published: 20 October 2015

Citation:

Tognoli E and Kelso JAS (2015)
The coordination dynamics of social
neuromarkers.
Front. Hum. Neurosci. 9:563.
doi: 10.3389/fnhum.2015.00563

Keywords: social coordination, alpha, mu, phi complex, brain rhythms, coordination dynamics, complexity

INTRODUCTION

Social neuroscience has garnered tremendous interest over the past decade, as readily appreciated from the large number of dedicated reviews (e.g., Frith and Frith, 2001; Ochsner and Lieberman, 2001; Cacioppo, 2002; Blakemore et al., 2004; Gallese et al., 2004; Insel and Fernald, 2004; Saxe, 2006; Cacioppo et al., 2007;

Adolphs, 2009; Behrens et al., 2009; Hari and Kujala, 2009; Schilbach, 2010; Farah, 2011). The entire armamentarium of non-invasive brain imaging methods has been harnessed toward the goal of discovering neural mechanisms of human social behavior, for instance electroencephalography (EEG; Babiloni et al., 2002; Sebanz et al., 2006b; Tognoli et al., 2007a; Lindenberger et al., 2009; De Vico Fallani et al., 2010; Dumas et al., 2010; Thiroux et al., 2010), magnetoencephalography (MEG; Hari et al., 1998), PET (Decety et al., 2002), functional magnetic resonance imaging or functional MRI (fMRI; Iacoboni et al., 1999; Montague et al., 2002; Beauchamp et al., 2003; Olson and Phelps, 2007; Izuma et al., 2008; Saito et al., 2010; Schilbach et al., 2010; Guionnet et al., 2012) and optical imaging (Suda et al., 2010; Funane et al., 2011). However, knowledge of the brain mechanisms involved in social behaviors has tended to lag far behind knowledge of the individual brain. The stakes are high: social behaviors show intricate symptomatic and etiologic ties with a vast number of brain disorders as well as with their treatments (see **Table 1**). The perspective offered in this review is that many neurophysiological biomarkers (neuromarkers) exist to support distinct aspects of social behavior. We may therefore envision in the future a matrix with each of the conditions of **Table 1** having a specific profile of neuromarkers: a trans-nosographic approach. Such a neuromarker profile might help both for diagnosis and for monitoring potential and actual therapies. However, basic discoveries and understanding are much needed before this translational goal can be achieved.

Neuromarkers are important tools to describe the transient and sustained activity of the brain's functional networks during social behavior. They may appear as oscillatory patterns in electrophysiological measurements due to electrical activity that reverberates in specific brain circuits (Kelso, 1995; Buzsáki, 2006; Kelso and Tognoli, 2007; Tognoli and Kelso, 2009, 2014). Or they may appear as spatial activity patterns in imaging approaches such as fMRI. In the following (See Sections entitled: "The neuromarker framework: finding local oscillations," "The neuromarker framework: brain coordination dynamics," and "The neuromarker framework: functional inferences"), we review methodological advances developed in our laboratory and findings that followed from them (See Sections entitled: "Neuromarker commonalities and differences" and "Toward dynamical models of social brains") within the context of experimental paradigms from social coordination dynamics. The dynamical approach is geared toward the analysis and understanding of network-specific oscillatory patterns that are engaged and disengaged during social behavior. The present research aims to elucidate the mapping between dynamic brain patterns and two categorically distinct social functions, namely, synchronic behaviors during which individuals coordinate simultaneously occurring actions; and diachronic behaviors, during which individuals alternate in the perception and production of social behavior (See Sections entitled: "Synchronic social behaviors" and "Diachronic social behaviors"). The emphasis of our approach is on continuous brain recordings rather than the more typical average evoked potentials or average spectra and related measures. Similar efforts are growing quickly

TABLE 1 | Brain conditions affecting social behavior.

Reference	Domain of investigation
Compton et al. (2005)	Alcohol/drug use disorders
Bennett et al. (2006)	Alzheimer's disease
Crooks et al. (2008) and Mendez et al. (2005)	Dementia
Segrin (2000), Clark et al. (2003) and Inoue et al. (2006)	Depression
Bassuk et al. (1999), Gleit et al. (2005) and Béland et al. (2005)	Ageing
Monetta et al. (2009) and Poletti et al. (2011)	Parkinson's disease
Mundy et al. (1986), Dawson et al. (2004), Baron-Cohen and Belmonte (2005), Hadjikhani et al. (2006) and Volkmar (2011)	Autism
Giovagnoli et al. (2011)	Epilepsy
Kirsch (2006)	Epilepsy surgery
Genizi et al. (2011)	Benign childhood epilepsy
Yeates et al. (2007)	Children brain disorder
Greenham et al. (2010)	Children brain insults
Lezak and O'Brien (1988) and Gomez-Hernandez et al. (1997)	TBI
Anderson et al. (1999)	Prefrontal lesions
Willis et al. (2010)	Orbitofrontal cortex lesions
Farrant et al. (2005)	Frontal lobe epilepsy
Green and Phillips (2004)	Paranoia
Russell et al. (2000), Couture et al. (2006), Brunet-Gouet and Decety (2006), Schimansky et al. (2010) and Varlet et al. (2012)	Schizophrenia
Jones et al. (2000) and Meyer-Lindenberg et al. (2005)	Williams syndrome

The long list of clinical conditions in which social behavior is altered suggests that basic discovery of neuromarkers and their functional organization could ultimately have large translational benefits.

in the field of brain-machine interfaces (Guger et al., 2000; Townsend et al., 2004; Kübler et al., 2005; Birbaumer, 2006; Blankertz et al., 2010; Hsu, 2011; Veluvolu et al., 2012), but have yet to be deployed to interpret the dynamics of social behavior. Given the complexity of most social functions, it is likely that multiple routes are available for the realization of particular tasks. This means that to explain social behavior we need to embrace such "degeneracy"—which is what the dynamical neuromarker approach aims to do.

One of the original quests of social neuroscience was toward discovery of "the" neuromarker of social behavior, that is, brain activity emanating from a functional network that transcends social interaction contexts—perhaps in the form of a system of mirror neurons (Gallese et al., 2004; Uddin et al., 2007). However, from many studies it has come to pass that more neuromarkers are recruited and modulated over the course of social behavior than initially presumed. Using EEG to investigate social interactions, our findings reveal that social neuromarkers have oftentimes taken the form of oscillations in the 10 Hz frequency band, a dominant frequency in the cerebral cortex and in cortico-thalamic loops (e.g., Bollimunta et al., 2011). In addition to portraying neuromarkers from this very active region of the EEG spectrum, we will briefly discuss the meaning and relevance of the 10 Hz time scale for social behavior (Note that we use "10 Hz frequency range" as opposed to "alpha range" to describe the band spanning from about 7.5–13 Hz,

in order to emphasize that this band contains a variety of potential neuromarkers besides the prominent and well-known alpha rhythm, and to disambiguate ranges and rhythms; see also Bazanova and Vernon, 2014).

Neuromarker multiplicity has led to a number of basic questions about the functional and dynamical architecture of social brains: which major functional system do such neuromarkers support; how do neuromarkers differ from one another; and how do they arise and interact over the course of ordinary social interaction? Questions like these motivate us to propose the methodological framework outlined in Sections “The neuromarker framework: finding local oscillations,” “The neuromarker framework: brain coordination dynamics” and “The neuromarker framework: functional inferences”. Our hope is that revealing the dynamics of neural oscillations will lead to a deeper understanding of the mechanisms underlying social behavior.

An enduring challenge in behavioral, cognitive, affective and social neuroscience is to develop a theory of tasks (Saltzman and Kelso, 1987). This development is especially critical when dealing with dynamical models of the brain, as it may help to infer covert mental processes and determine the timing of their recruitment and dissolution. Today, it seems, we are at a crossroads—having explored a sufficient task repertoire (the *behavior* side of the story) and identified a number of neuromarkers (the *brain* side)—it becomes possible to enquire about the integration of results and their modeling. These are early days in such an enterprise: many elements are still missing and others not yet in definitive place. The present paper is contributed in this spirit. Through methodological advances, systematic experimentation and neurobehavioral theorizing, we attempt to chart a path toward understanding social brains. We end the present review with some ideas on how to cross this frontier in social neuroscience.

SYNCHRONIC SOCIAL BEHAVIORS

Synchronous social behaviors engage simultaneous action and perception processes. Tango dancing, choir singing, driving in traffic, executing shared-tasks such as lifting heavy furniture in tandem or performing surgery are examples of synchronous behaviors, with varying degrees of symmetry between the actions performed and the varying effector and sensory pathways involved in action~perception couplings. In such interactions, information flows continuously and reciprocally between people through perceptual channels (**Figure 1A**, blue arrows), creating linkages at both brain and behavioral levels.

A unique characteristic of synchronous behavior is that the actions of one individual (e.g., **Figure 1A**, annotation 2) are readily able to modulate a partner’s behavior (**Figure 1A**, annotation 4) with information flowing in a reciprocal, bidirectional fashion. Information about self-produced movement is returned back to oneself and is updated based on the actions of one’s partner (**Figure 1A**, f_{4-1}). With both partners simultaneously engaged in such informational exchange—each continuously perturbing the other—a system is

formed that enters a kind of self-organization that exhibits rich dynamical behavior (Kelso, 1995; Sebanz et al., 2006a; Tognoli et al., 2007a; Oullier et al., 2008; Tognoli, 2008; Oullier and Kelso, 2009; Konvalinka et al., 2010; Riley et al., 2011; Duran and Dale, 2014). To probe this dynamics, the social coordination paradigm assesses the behavioral and neural organization of subjects as they continuously perform simple rhythmic index finger movements (extensions/flexions). Differences between behaviors produced in individual and social contexts are assessed by manipulating the perceptual flow between people, switching vision of each others’ action on and off with the help of an optical barrier (see supplementary materials S1–S2). The advantage of this very basic, canonical situation is that it provides explicit and continuous measures of social coupling through the dynamics of a collective variable, the relative phase (Tognoli et al., 2007a; Oullier et al., 2008; Tognoli, 2008; Oullier and Kelso, 2009; Tognoli et al., 2011), akin to studies of bimanual (e.g., Kelso, 1984; Haken et al., 1985; Swinnen and Wenderoth, 2004; Banerjee et al., 2012), sensorimotor (Kelso et al., 1990; Schmidt et al., 1990; Wimmers et al., 1992) and postural coordination (Varlet et al., 2011). To study the potency of perceptual coupling and corresponding neural correlates during this spontaneous form of social coordination, we further distinguish trials during which subjects entered a state of phase-locked collective behavior from those that did not (Tognoli et al., 2007a).

DIACHRONIC SOCIAL BEHAVIORS

In contrast to synchronic behavior, in diachronic social transactions only one participant acts at a given time. Examples of such behavioral transactions include conversation with well-defined turn-taking, imitating a person’s facial expression or accent, and learning a surgical gesture by observing a demonstrator in medical school. Coupling in the system is ensured by perceptual flows to the observer’s brain (**Figure 1B**, blue arrows), but there information flow reaches an endpoint—at least momentarily until role settings are eventually modified. As a result, information flows do not circulate continuously in the system. If all relevant influences stopped in the immediacy of perceptual exchange, this type of social transaction would seem less useful than its ubiquity suggests. However, it appears that such exchanges rely upon delayed influences—possibly buffered in the observer’s brain through memory processes—and mutual social influences are therefore allowed to resume at slower time scales (see Tognoli and Kelso, 2013 for a theoretical discussion on time scales and causality in complex systems). Experimental tasks that probe such diachronic behaviors include action observation and delayed imitation. In our implementation (Sutari et al., 2010), we instructed pairs of participants to first observe then imitate index finger movements in turn, during two periods of continuous behavior (8 s long) separated by retention, pause and control intervals for individual behaviors (see supplementary materials S3). We studied social neuromarkers and their dynamics when subjects observed their partner’s action, performed an action alone or under the observation of their

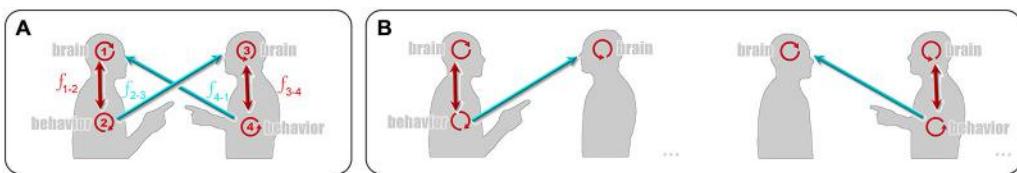


FIGURE 1 | Task settings. The flow of information during synchronic (**A**), and diachronic (**B**) social interactions in a dyadic setting. Circular red arrows describe intrinsic dynamics in neural and behavioral subsystems respectively. Straight red arrows describe movement and perceptual information flows that are circumscribed to an individual; blue arrows represent information flows that cross to the other individual (social coupling). During synchronic social behaviors (**A**), information flows bidirectionally between all parts of the system. In contrast, during diachronic social behavior (**B**), only one person acts at a given time and one behavioral subsystem is disengaged. The two vignettes in (**B**) illustrate turns of behavior between the two individuals. See details in text.

partner, imitated the action they observed earlier, and during rest.

THE NEUROMARKER FRAMEWORK: FINDING LOCAL OSCILLATIONS

From dual EEG recordings, we examined the repertoire of brainwaves (brain rhythms, periodic and aperiodic oscillations) recruited for social tasks. Brainwaves carry a 3-sided signature of underlying neurophysiological processes: (1) spatial organization (how energy is distributed over the scalp—an indirect manifestation of the originating neural structures); (2) spectral properties (the frequencies at which brainwaves operate—a

manifestation of their temporal extent and affordance for interaction with other neuromarkers); and (3) functional dependency (i.e., which behavioral/mental/affective processes modulate them). In other words, analysis of brainwaves addresses the structure, dynamics and function of the brain (e.g., Kelso, 1995; Freeman, 2000; Basar, 2004; Bressler and Tognoli, 2006; Buzsáki, 2006; Kelso and Tognoli, 2007).

Importantly, from the theory of coupled oscillators, it ensues that neural oscillations meant to work together need to operate on similar time scales, or equivalently, frequencies. If the binding/coordination mechanism at play is phase- and frequency-locking or a more subtle metastability (Kelso, 1995, 2012; Tognoli and Kelso, 2009, 2014) this constraint translates

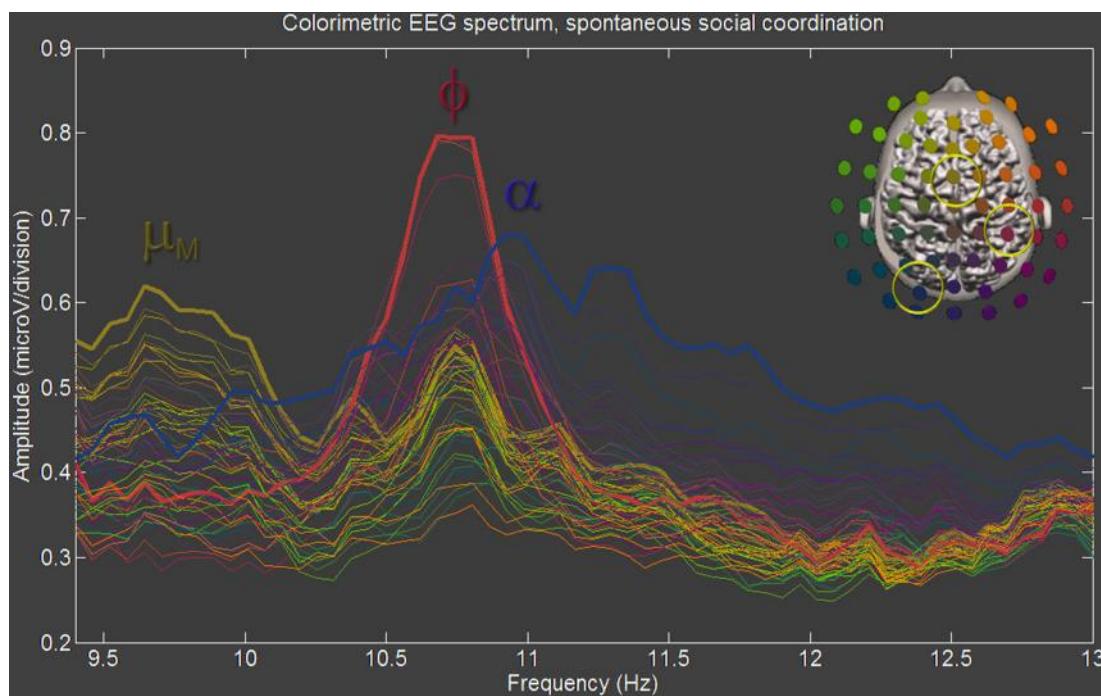


FIGURE 2 | Parsing neuromarkers. Neuromarkers can be parsed using multi-electrode spectra with high spectral resolution (here bin size is 0.06 Hz) and colorimetric encoding of spatial organization (following colorimetric legend shown in upper right corner). In this figure adapted from Tognoli et al. (2007a), sampled from a subject performing spontaneous social coordination—a synchronic task—3 neuromarkers are observed that include mu medial (appearing in brown color as a result of its fronto-central topography), left alpha (blue, left occipital region) and phi (burgundy, right centro-parietal region). Note spectral proximity, especially for phi and alpha. Neuromarkers are quantified by identifying the boundaries of spectral peaks, when power departs from and returns to background power, and by integrating power over all the bins included in this interval (see supplementary materials S4).

into neural ensembles' operating with similar (or near integer-related) frequencies (see, e.g., deGuzman and Kelso, 1991; Bressler and Kelso, 2001; Bressler and Tognoli, 2006; Palva and Palva, 2007, 2012; Tognoli and Kelso, 2009, 2014; Tass et al., 1999). As a result, spectral overlap is often present, a feature that is poorly accounted for in traditional EEG studies. For example, when examining the 10 Hz band at the usual spectral resolution of ~ 1 Hz, overlap translates into a blurred spectral and spatial differentiation of neural oscillations. More specifically, one sees an irregular-shaped peak in the spectrum, with its power distribution changing from place to place over the surface of the scalp. This amorphous view conceals a number of discrete peaks each with their own frequency and topography (such as the three peaks shown in **Figure 2**), but so close that they may merge spatially and spectrally at low resolution. Our framework of brain coordination dynamics rests on high-resolution spectral analysis of EEG with colorimetric encoding of topography- a set of techniques that performs well at distinguishing oscillations with spectral and spatial proximity (Tognoli and Kelso, 2009). When sufficient spatial and spectral resolution are achieved (increasing sensor density to augment spatial resolution and either increasing the amount of continuous time in Fourier analysis or lengthening the time interval artificially using zero padding to augment spectral resolution), crisp regional distributions of power do appear. Using such techniques, it is possible to measure the functional specificity of brain rhythms without the corrupting effect of other oscillations located nearby.

Although the general architecture of human brains may be the same, on a fine grain level every brain circuit is different. Hence, a neuromarker may shift slightly in frequency and topography from one subject to another. Critically, identification of neuromarkers needs to be conducted on a subject by subject basis (see also Veluvolu et al., 2012, for a related account). At this stage, interindividual comparisons are performed on parsed neuromarkers (their conditional power or/and their time course), not on the less refined picture of power distribution that is obtained from grand-average spectra (mean of all individual spectra, which again causes blurring due to spatial and spectral variations between subjects).

THE NEUROMARKER FRAMEWORK: BRAIN COORDINATION DYNAMICS

Oscillations may be studied in average spectra (as in **Figure 2**) and continuous time. We hypothesize that such oscillations reveal the transient activation of unique functional networks in the brain. Under such an hypothesis, it is possible to establish a time-course describing the engagement and disengagement of brain networks. The latter coexist with another timeline of descriptors, namely one that refers to the brain's functional organization at the level of behavior, perception, cognition and volition (See Section below entitled: "*The neuromarker framework: functional inferences*"). The challenge for social neuroscience (and for neuroscience in general) is to recognize that both neural and behavioral/cognitive levels may be characterized in terms of their dynamics and that dynamics offers

a means by which to relate them (Kelso, 1995, 2012; Buzsáki, 2006).

Neuromarker dynamics can be probabilistically approached using wavelet analysis (see, e.g., Tognoli et al., 2007a; Suutari et al., 2010) within the spatio-spectral domain identified from a "static" neuromarker approach (**Figure 2**). This provides a picture of the brain in which macroscopic ensembles fluctuate smoothly in amplitude over time, an imperfect but heuristic means to explore macroscale neural dynamics. The wavelet approach is heuristic in the sense that following selection of the right electrode and frequency band for a neuromarker of interest, it tends to maximize the correspondence between signal power fluctuations and the genuine time course of a functional process. Fundamentally, the inverse problem prevents one from identifying source dynamics solely on the basis of information from scalp recordings. As a result, electrode-based wavelet approaches (and related methods) are far from perfect. Since a number of distant neural ensembles contribute to the scalp signal in the same scalp neighborhood, there is no guarantee that a unique neural ensemble is tracked continuously by monitoring power at selected electrodes. Rather, electrode power is determined by a number of neural ensembles in turn. A much more precise approach includes segmentation and classification of transient spatiotemporal patterns and analysis of their coordination dynamics (Tognoli and Kelso, 2009; Benites et al., 2010; Fuchs et al., 2010; Tognoli et al., 2011; and **Figure 3**; see also Lehmann et al., 2006), to be followed by reconstruction of their source dynamics (Pascual-Marqui et al., 2002; Murzin et al., 2011). Such methods provide a picture in which sources are intermittently on and off. As discussed in Tognoli and Kelso (2009), we are less interested in power/amplitude quantifications (which are inappropriate measures of neural source strength in the first place, Tognoli and Kelso, 2009), than with the lifespan of large scale patterns (duration and recurrence) and their dynamical interaction with other neural ensembles (e.g., phase relationships within patterns; vicinity of other patterns that entertain causal precedence and consequence). In our approach, all such dynamical attributes are scrutinized in terms of their possible functional significance.

THE NEUROMARKER FRAMEWORK: FUNCTIONAL INFERENCES

Inferences about brain~behavior correspondences (a temporal puzzle, see **Figure 4**) represent a key challenge that must be overcome in order to achieve adequate explanatory models of social brains. The rich phenomenological language of human behavior and cognition has been developed over centuries of scholarly enquiry, accelerated in recent decades due to the thrust of cognitive (neuro)science. We postulate that the functional language of human behavior (e.g., sociocognitive and affective processes) maps onto discrete neural patterns, i.e., those that can be captured from segmentation of continuous EEG (See above Section entitled "*The neuromarker framework: brain coordination dynamics*"). Due to the convergent~divergent connectivity of the brain, the mapping is likely to be degenerate: the same output pattern may be produced by

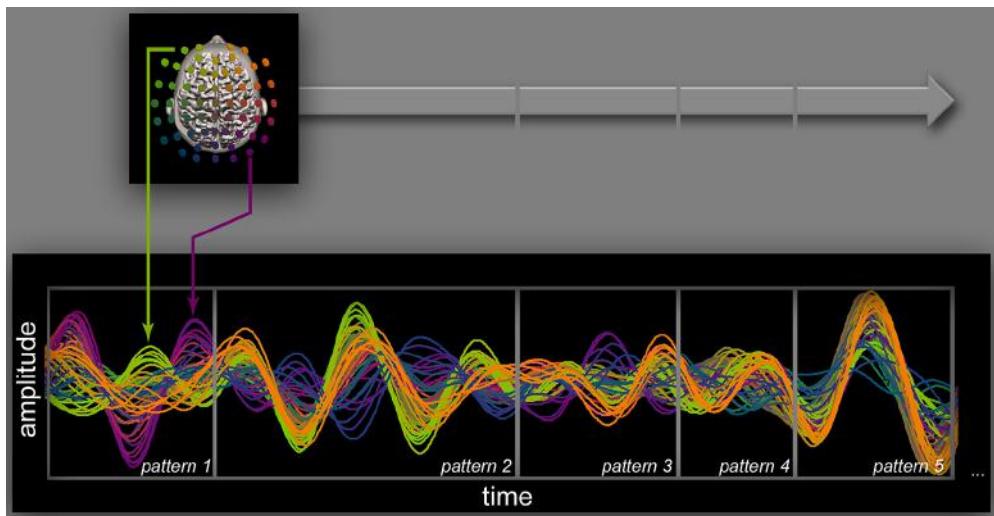


FIGURE 3 | Brain dynamics. A sequence of five oscillatory patterns segmented from continuous, band-pass filtered EEG in a synchronic task of intentional social coordination. Filters (7–13.5 Hz) were set to retain activity in the 10 Hz range, a prominent feature of human waking EEG. Patterns were segmented manually by two trained examiners who analyzed the spatiotemporal evolution of phase aggregates (Benites et al., 2010). Results were later confirmed using an automatic segmentation algorithm. Each pattern inside the gray frames is best explained by the transient organization of a few macroscopic ensembles that interact through phase-locking or metastability. For instance, the first pattern shows phase aggregates that are suggestive of one gyral and one sulcal source (green and magenta arrows respectively; source estimation provides some indication on their cortical origin). Short-lived configurations tend to succeed one another (e.g., magenta phase aggregate ends with the edge of the first box, giving way to new phase aggregates in the second gray box). Putatively, this organization provides support for ongoing functional processes. Note that such neural organization in the 10 Hz frequency band sustains transient patterns with a typical duration of 1–200 ms, a crucial time-scale for human behavior, both individual and social.

a number of different interacting brain structures, and alternative pathways between neural structures are capable of producing functionally equivalent cortical patterns (Edelman and Gally, 2001; Tononi, 2010; Kelso, 2012)—the key signature

of self-organized synergies or coordinative structures (Kelso et al., 1984; Kelso, 1995). The empirical challenge then becomes one of matching temporally inferred functional processes and observed brain patterns (Figure 4, left). Such

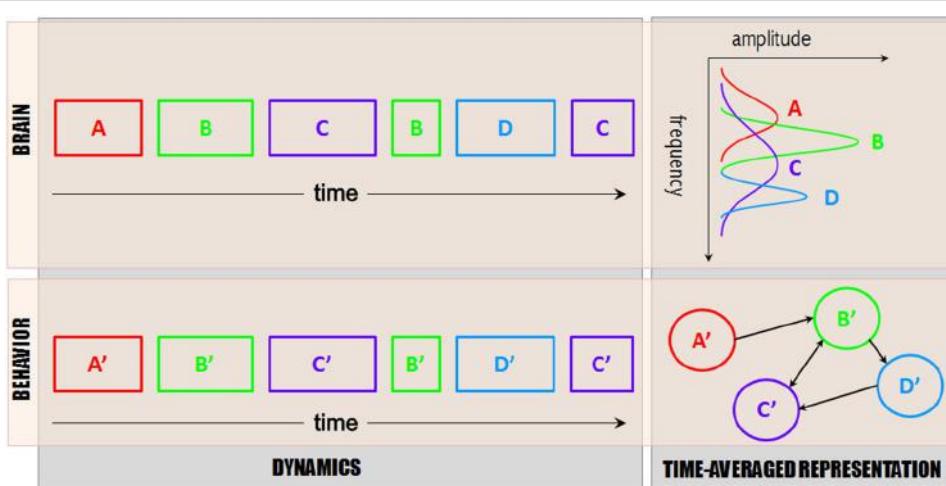


FIGURE 4 | Brain~behavior scheme. Dynamical descriptions of brain functional networks (top left) and inferred functional processes (bottom left), along with their time-averaged representation (functional graph on lower right and power spectrum on the upper right (note rotated axes to reflect the fact that amplitude is largely inherited from the cumulative duration of the patterns, along with their frequency consistency over time). For simplicity, only one frequency band is represented (say, 10 Hz), and only one process at a time (i.e., no network interaction). In reality, multiple frequency bands (and associated functional processes) occur at the same time. Typically, networks are co-activated and exhibit transient interactions, e.g., via phase locking and metastability. The goal of functional inference is to identify the functional processes (bottom rectangles) that match spatiotemporal patterns of brain activity (top rectangles) and their temporal footprints, so that correspondences between brain and behavior can be uncovered. Though simplistic, a translational language along these lines would propel our understanding of social brain functions and lead the way toward explanatory models.

inference is guided by the study of neuromarkers (as in **Figure 2**), and neuromarker dynamics (as in **Figure 3**). A sound strategy consists of meta-analyses: after a neuromarker has been revealed through the study of multiple tasks and experimental manipulations, it becomes possible to narrow down its functional significance more precisely, thereby separating its true functional meaning from sporadically co-varying effects.

Difficulties lie in the fact that (1) theories of tasks are seldom based on explicit, observable quantities and (2) such descriptions, despite their ready reduction into serial models, are not grounded in a dynamical framework that allows one to establish unambiguous time addresses for the engagement and disengagement of functional processes. A place to begin such an endeavor is with functional processes that have explicit temporal footprints, as in our social coordination paradigms. Time-averaged neuromarkers (obtained from the methodology spelled out in Section “Synchronous social behaviors”) and their reactivity also provide tractable material that may lead to establishing neuro-functional relationships (see **Table 2** below).

Descriptions of behavior and cognition are especially fruitful for slower and more global functional processes, the timescale of which was amenable to observational and experimental tools of earlier times. In contrast, faster processes (timescales of

tens of milliseconds and less) have not systematically received distinct names and descriptions. Short-lived patterns that are uniquely tracked with dynamic brain imaging techniques such as EEG and MEG may hold keys to advancing understanding of social behavior (for instance, irrespective of their functional brevity, they may be keys to certain deficits). Identifying causal chains of neuro-functional processes at faster time-scales—not typically available in social cognition/behavior settings—may be one of the most valued advances that social neuroscience can make.

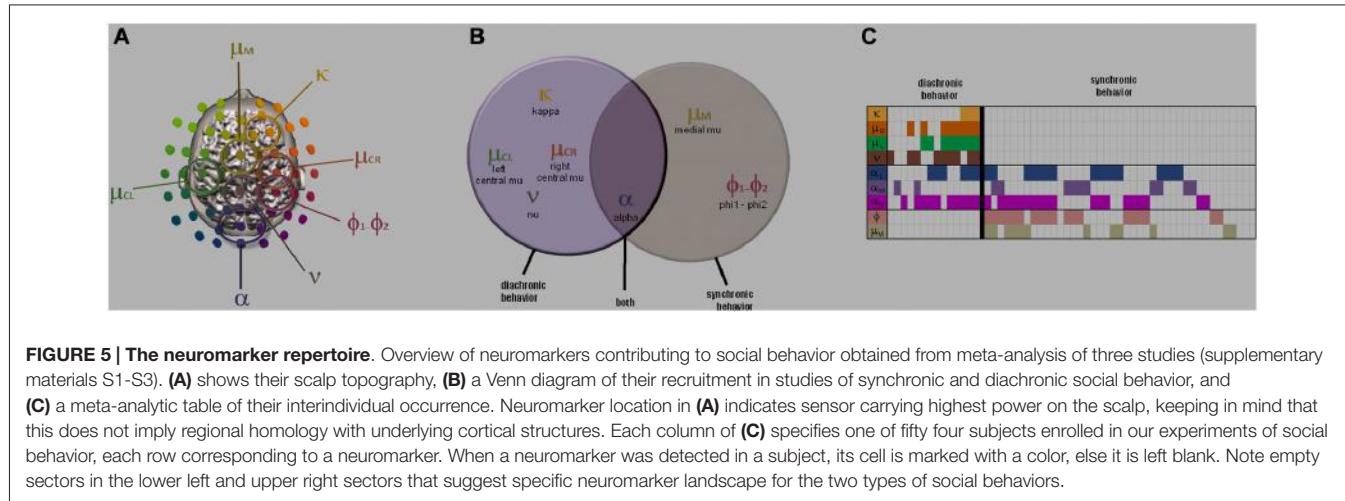
NEUROMARKER COMMONALITIES AND DIFFERENCES

The repertoire of neuromarkers observed during our social tasks (synchronous social behaviors of spontaneous and intentional social coordination; diachronic social behaviors of action observation and delayed imitation; Tognoli et al. (2007a,b, 2011); Tognoli (2008) and Suutari et al. (2010); (see also supplementary materials) is summarized in **Figure 5**. During synchronous social behaviors, a set of neuromarkers was recruited that included the alpha rhythm, the phi complex and especially when interaction was spontaneous, a medial mu rhythm (Tognoli et al., 2007a,b). During diachronic social behaviors, alpha was also observed, but

TABLE 2 | Neuromarker meta-analysis.

	Neuromarker	Peak frequency	Topography	Task dependence
Only in synchronous tasks (Tognoli et al., 2007a,b, 2011; Tognoli, 2008)	Mu medial	9.1 Hz (1.1)	FCz	- recruited during spontaneous coordination - suppressed during social interaction - rebounds during intrinsic, self-produced movement - phi 1 increases in spontaneous coordination when subjects do not synchronize with each other - phi 2 increases when subjects synchronize with each other - phi complex recruitment and modulation is strongest during intentional social coordination
	Phi complex	10.9 Hz (0.9)	CP4	- increased by drowsiness - decreased by visual input - decreased further by vision of the partner's movement (larger decrease than with non-social stimuli) - further decreased if partner's movement is more variable
Both task types	Alpha	9.9 Hz (0.9)	PO7 (left), PO8 (right), POz (midline aggregate)	- left and right mu depressed during self movement; rebound at self-movement arrest - no systematical decrease during social observation - exhibit dynamic aftereffects (suppression, rebound) at cessation of observed movement - right mu vanishes tonically when people memorize observed behavior (right hand movement)
Only in diachronic tasks (Suutari et al., 2010; Tognoli et al., 2011)	Left and right central mu	Left: 10.6 Hz (0.9) Right: 11.4 Hz (0.8)	C3 (left), and C4 (right)	- decreases during self-movement - increases when self-movement is performed in view of another person - tendency to decrease any time either partner performs a movement in view of the other
	Nu	10.1 Hz (1.1)	CPz	
	Kappa	11.2 Hz (0.7)	FC2	

Summary of spatial, spectral and functional properties of neuromarkers involved in synchronous and diachronic social behavior (see also **Figure 6** for topographical maps and colorimetric spectra). The data presented are group results obtained from the samples of subjects that have participated in our studies. Peak frequency (measured from high-resolution spectra) describes the arithmetic mean of the samples with standard deviation in parenthesis. The electrode reported in column topography refers to the mode (electrodes most frequently observed across subjects that bear largest spectral energy, named according to the 10 percent system, Chatrian et al., 1985). All recordings were performed with linked-mastoid reference. “Task dependence” refers to conditions in which power is modulated, a precursor to inferences about function.



mu medial and the phi complex were not detectable. In addition, left and right central mu appeared as did two newly described nu and kappa rhythms (Sutari et al., 2010). The spatial, spectral and functional properties observed for these rhythms in our samples of subjects are reported in **Table 2** and **Figure 6**. Keeping in mind the high-resolution spectral analysis implemented here, accuracy of estimation is aligned with the spectral resolution of the coarsest dataset, i.e., 0.1 Hz. The data presented in **Table 2**

are group results obtained from the samples of subjects that have participated in our studies (peak frequency describes the arithmetic mean of the samples; electrode location refers to the mode). Of course, large populations would be helpful to establish robust normative properties of neuromarkers (something that at this time, we forgo in favor of smaller, discovery-based studies). **Table 2** summarizes spatial, spectral and functional properties as a starting point to identifying new neuromarkers

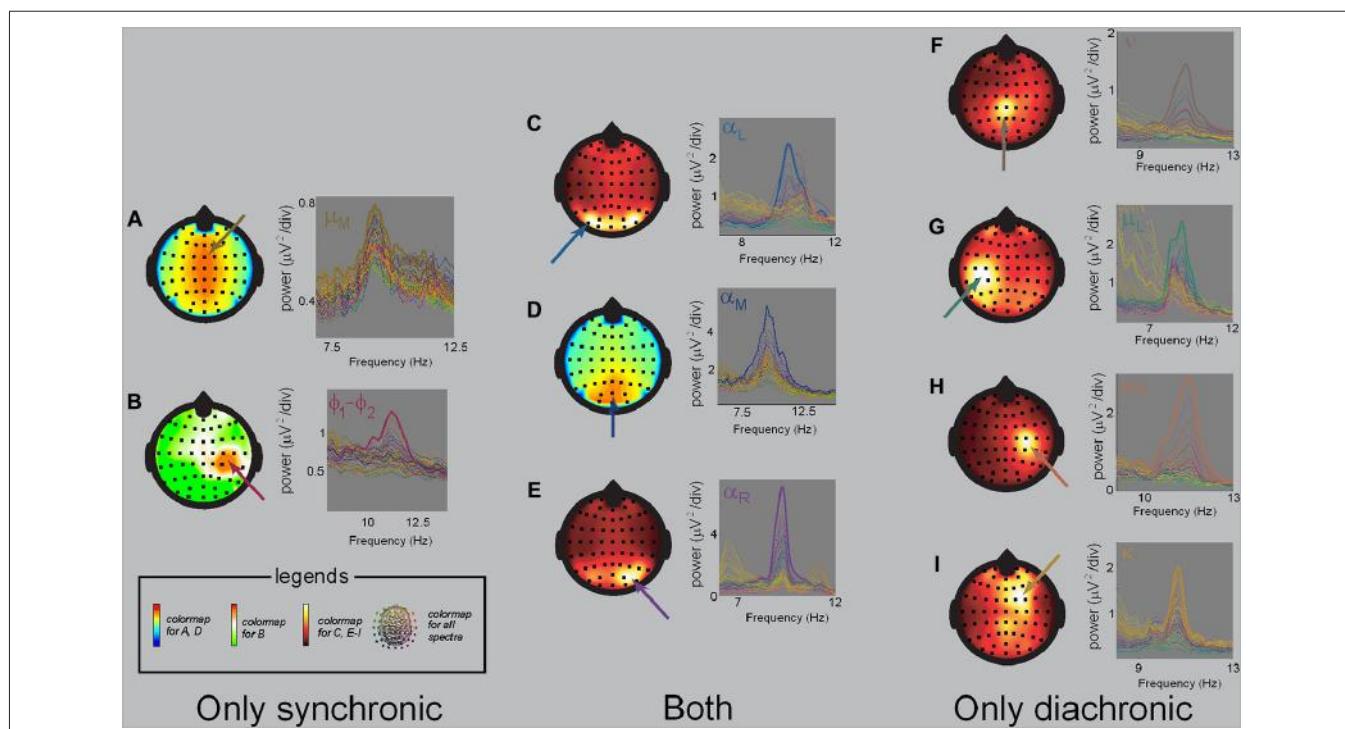


FIGURE 6 | Task-specific neuromarkers. Spatial and spectral properties of the neuromarkers referenced in **Table 2** (adapted from Tognoli et al., 2007a; Sutari et al., 2010). On the left (**A–B**) are neuromarkers mu medial (**A**) and phi (**B**), that were only observed in synchronic tasks. On the right are neuromarkers nu (**F**), left and right mu (**G,H**) and kappa (**I**), that were only observed in diachronic tasks. (**C–E**) has alpha neuromarkers that were observed in both types of task, and shows the medial form (**D**), and its lateralized variants, emphasizing the left (**C**) and right hemisphere (**E**). Neuromarker discovery is aided by the color of their spectra, which are not chosen arbitrarily but inherited from neuromarker's spatial organization.

and with the aim of helping others in the field who share similar goals.

The only neuromarker that transcended both synchronic and diachronic social behavior was the alpha rhythm, a neuromarker associated with visual attention (Mulholland, 1972; Klimesch et al., 1998; Palva and Palva, 2011). All of our studies revealed that vision of the partner substantially reduces alpha power. With its separation of social and self behaviors in distinct experimental phases, our study of action observation further allowed us to show that alpha fluctuated with the complexity of behavioral information acquired about the partner. In Suutari et al. (2010), single trial alpha power was low when observers were exposed to finger movements with high cycle to cycle variance. By contrast, alpha increased with more regular movements. Put another way, the individual brain's alpha rhythm appears to be a pertinent measuring instrument of the complexity embedded in interpersonal information flows (see also Müller et al., 2003 for related account in non-social visual perception).

A social interaction exists only if social partners acquire information about each other (see blue arrows in **Figure 1**). Our results suggest that the alpha rhythm is a key neuromarker of visually-mediated social behavior (putatively, social transactions mediated by other sensory channels would have their own signatures, see, e.g., Pineda, 2005 for candidates). Alpha modulation is often overlooked in EEG/MEG studies of social interaction in favor of mu rhythms. We suggest however that alpha's sensitivity to informational exchange between partners, its large amplitude in human EEG and robust presence in most subjects makes it an important neuromarker of social behavior (see The neuromarker framework: "finding local oscillations" Section for strategies to disambiguate alpha, mu and other spectrally similar neuromarkers). Furthermore, in visual detection tasks, it has been shown that alpha suppression is spatially informative, with attention to the right hemifield depressing specifically left alpha rhythm and vice-versa (Worden et al., 2000; Sauseng et al., 2005). Such lateralization could be useful to disentangle self and social attention in experimental designs that carefully manipulate the spatial arrangement of self and other—with the potential outcome that roles in social interactions could be quantified as a function of the spatial deployment of attentional resources. Moreover, interindividual variation in alpha suppression could reveal the extent of social engagement and task-related social affinities, with consequent applications to a variety of domains relevant to human social behavior.

TOWARD DYNAMICAL MODELS OF SOCIAL BRAINS

As we observe many neuromarkers and their intermittent dynamics in dual-EEG recordings (see "The neuromarker framework: brain coordination dynamics" Section), we are led to question their spatiotemporal organization—how the functional processes that participate in social behavior are orchestrated. Until now, at the largest scale of complete dual-EEG

experiments, we have achieved either a static neuromarker description (as in "The neuromarker framework: finding local oscillations" Section), or a probabilistic description of their dynamics using wavelet analysis on selected frequency bands and spatial sites (e.g., Tognoli et al., 2007a; Suutari et al., 2010; see discussion in "The neuromarker framework: brain coordination dynamics" Section). Based on theoretical and methodological work (Tognoli and Kelso, 2009), we have also started to study the dynamic patterns of dual-EEG (see **Figure 3** and text thereafter) on particularly interesting aspects of social behavior such as the loss or establishment of coordinated action. The first stage of this analysis is a segmentation of continuous (band-selected) EEG. We have implemented either a manual analysis of the oscillations' phase, frequency and topography (Benites et al., 2010), or an automatic segmentation method examining the eigenvalue tradeoff between two principal modes of the EEG power envelope derived from a rotating wave approximation (Fuchs et al., 2010). The result of both approaches is to parse each participant's EEG into a sequence of dynamic patterns (see **Figure 3**). This sequence is then matched to an estimation of the time course of inferred functional processes (**Figure 4**), with the goal of connecting their dynamics. This framework extends our earlier efforts that found a tight connection between behavioral and neural dynamics once an appropriate space of collective variables was identified. Spatiotemporal measures of brain activity tracked kinematic measures of sensorimotor coordination both empirically (Kelso et al., 1998) and in a theoretical model of the underlying neural field dynamics (Fuchs et al., 2000).

As more and more insights into the function of neuromarkers becomes available, it should become possible to solve the temporal puzzle of brain~behavior as presented in **Figure 4**. When that point is reached, we will be able to draft dynamical models of social processes at the combined levels of brain and behavior and to study their variation in different situations (e.g., social skill development, disease, effects of pharmacological treatment, etc.).

In the preceding, we have examined collective behavior and its relation to brain activity, but only a single brain at a time. With social neuroscience born from cognitive neuroscience, there is a temptation to segregate the neural activity of participants to fit the existing framework of single-brain neuroscience. A true social neuroscience, however, will only realize itself when it fully integrates neural activity of every participant in a common analysis scheme. Efforts to do so have been undertaken by collecting synced records of brain activity from multiple people (e.g., dual-EEG: Tognoli et al., 2007a; or fMRI hyperscanning: Montague et al., 2002) and by formulating novel analysis frameworks that combine the neural dynamics from multiple subjects (Lindenberger et al., 2009; Dumas et al., 2010; Dodel et al., 2011; Tognoli et al., 2011). With brains chock full of oscillations that are coupled between people through inter-personal perceptual flows, a straightforward hypothesis is that oscillations enter collective states of phase-locking and frequency coupling between the

brains of interacting partners—a hypothesis that has been pursued by ourselves and others (e.g., Lindenberger et al., 2009; Dumas et al., 2010; see also Funane et al., 2011; for related hemodynamic account). Our research has yet to uncover unambiguous evidence of phase-locking between the brains of people as they engage in social behavior. Moreover, our longstanding theoretical inclination is toward metastable coordination dynamics, where tendencies for integration coexist with tendencies for segregation (e.g., Kelso, 1995; Kelso and Tognoli, 2007; Tognoli and Kelso, 2009, 2014). The reason we suspect that phase synchrony is seldom observed is that at the level of dynamic patterns (and in the frequency bands examined, especially around 10 Hz), limited symmetry exists between the instantaneous networks formed in each person's brain (see example **Figure 7**). However, in applying the aforementioned segmentation methods to social coordination tasks, we encountered evidence of another, less expected mechanism of coupling between brains (Benites et al., 2010; Fuchs et al., 2010). On the one hand, each subject's neuro-functional activity was distinct (compare upper and lower white frames in **Figure 7A**, and note patterns' lack of correspondence in topography, frequency and phase), yet on the other hand, the moment at which those patterns changed in each partner coincided (note temporal coincidence of white frames' edges marked with asterisks in **Figure 7**). In other words, it was not the oscillatory neural activity proper that was synchronized between people but rather the underlying temporal structure of their recruitment and dissolution. An analogy to such inter-brain coordination is a group of musicians, each playing different notes yet achieving a harmonious outcome by following the same tempo—without, of course, a conductor (see Kelso and Engstrom, 2006, p.93). We hypothesize that this mechanism of inter-brain coordination springs from the very weak coupling engendered by perceptual flows (i.e., weaker than connectivity-based information flows within brains). We further speculate that this weak coupling promotes the emergence of complexity in social interaction (Tognoli et al., 2011).

RELATION TO OTHER WORK

A vast literature has emerged in the previous decade regarding neural oscillations involved in social behaviors (reviews in Hari, 2006; Perry et al., 2010; Konvalinka and Roepstorff, 2012; Keller et al., 2014). This literature grew—in the wake of the discovery of the mirror neuron system—with much emphasis on mu rhythm's suppression during action observation and related social activities (e.g., Cochin et al., 1999; Babiloni et al., 2002; Muthukumaraswamy et al., 2004; Oberman et al., 2007; Cheng et al., 2008; Arnstein et al., 2011; Perry et al., 2011; Woodruff et al., 2011; Derix et al., 2012; Dumas et al., 2012; Lachat et al., 2012; Liao et al., 2012; Moore et al., 2012; Naeem et al., 2012; Vanderwert et al., 2013; Hogeveen et al., 2014; Sebastiani et al., 2014; Fitzpatrick et al., 2015; Moreno et al., 2015, to cite a few). The multiple designations given by different scientists to identical rhythms (e.g., central alpha and mu) and the identical name given to distinct neural activity (e.g., alpha used to

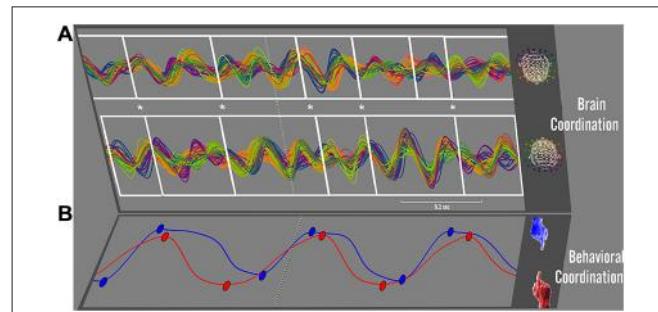


FIGURE 7 | Brain~behavior coordination. Synchronized patterns between brains, in a synchronous behavior of intentional social coordination (after Tognoli et al., 2007b). Continuous dual-EEG is shown in the 10 Hz frequency band for a pair of interacting subjects in **(A)**, with electrode signals encoded using the colorimetric legend shown on the right (EEG from one subject on top, the other on the bottom). Changes in spatiotemporal organization of brainwaves were determined by two trained examiners who were blind to the associated behavioral variables (Benites et al., 2010). A manual segmentation was performed separately on each subject's EEG. Transitions are marked by successive white frames, following the method outlined in Section "The neuromarker framework: brain coordination dynamics" and **Figure 3**. In this sample trial, subjects were instructed to coordinate finger movements inphase (see red and blue movement trajectories of right index fingers in **B**). The dashed line in **(B)** indicates the moment at which they successfully coordinated their behavior (with the movements' relative phase exhibiting a sudden phase transition to inphase, not shown). The entire temporal window displayed is about 1 s long and relates to the intentional transition process from independent to coordinated behavior. In this window, the transition between subjects' brain patterns reveals strong tendencies for coincidence (see series of asterisks in **(A)**, cueing temporal proximity of each participant's brain pattern transitions). Note that the dynamic patterns of each participant's brain activity have distinct spatial, spectral, and phase organization. Neural transitions are coupled, but not the spatiotemporal neural patterns located between them.

designate the parieto-occipital rhythm as well as many other oscillatory activities) are obstacles to advances in the field. Our view is that progress toward understanding the relationship between neural oscillations and (social) function will emerge after a standardized taxonomy of EEG rhythms is in place to facilitate inter-study comparisons; the names that we give to neuromarkers represent an effort to organize our own findings with this goal in mind.

We can classify the foregoing literature depending on the methodology and its ability to resolve spatially and functionally specific neural oscillations (**Table 3**). Many of the earlier studies (type I), and still some today, used power in predetermined frequency bands at electrodes of interest. For instance, mu can be analyzed at electrodes C3 and C4 in the alpha band or one of its subdivisions. This approach incurs a substantial risk that the results are driven by another rhythmic activity than the one that is assumed (for instance, some unpublished analyses in our laboratory suggest that during social tasks such as action observation with college students as subjects, the specific contribution of mu to power at electrodes C3 and C4 in the complete "alpha" band varies from 13–23%, and is commonly dwarfed by parieto-occipital alpha whose large amplitude attenuates slowly across space. This heterogeneity is in line

TABLE 3 | Neural epistemology.

	a. Establish existence of a neuromarker, oscillation, rhythm	b. Assess modulation of power in task and control conditions	c. Analyze unbiased coordination dynamics	d. Estimate source location
Type I: Power in standard frequency bands, at electrodes of interest, sans spatial context				
Type II: Power in standard frequency bands, at peak electrodes, with spatial context	✓	✓		
Type III: Power in adjusted frequency bands, at peak electrodes, with spatial context	✓	✓	✓	
Type IV: Power in standard frequency bands, in source space	✓	✓	✓	✓

Summary of approaches to assess oscillatory power in studies of social functions, and their appropriateness to different scientific questions. See details in text.

with others' findings (Braadbaart et al., 2013) that power in the mu band at electrode C3 negatively modulates the BOLD signal from a constellation of brain areas within and beyond the mirror neuron system. Other studies (type II) use scalp signal and canonical frequency bands, but provide contextual information about the power's spatial distribution exhibiting peaks at the expected location for a rhythm of interest. Due to suboptimal frequency boundaries, the risk of contamination in these studies lies in the aggregation of power from multiple rhythms—though it may be identified somewhat from the complexity of the rhythmic activity's spatial patterns (with simpler patterns suggestive of lesser bias). Our own approach (type III) also starts with the scalp signal but adapts the frequency band to each rhythm and each subject, in order to further enhance functional specificity. Finally (type IV), efforts to eliminate extraneous variance take the form of source estimations: provided good head models, electrode density and adequate algorithms, such studies attempt to provide information about the involvement of specific brain areas.

To date, we are not aware of source estimation studies that tuned frequency boundaries (as here) in order to further eliminate extraneous variance. Some of our work strongly suggests that at the macroscale of EEG signals, the brain's spatiotemporal patterns are intermittent rather than continuously modulated in amplitude (Tognoli and Kelso, 2009, 2014; see also **Figures 3, 7**), although it is highly probable that continuous activity underlies the smaller scales (Figure 5 in Tognoli and Kelso, 2014). Under this hypothesis, the common finding of type IV studies that brain dynamics is continuously modulated (as opposed to a discrete succession of onsets and offsets) appears unlikely. A further possibility is that when a main spatio-temporal pattern recedes, other sources fill-in and contaminate the former source dynamics. With the typically complex brain activities involved in social behavior this problem is aggravated because of the enhanced likelihood that task-related neuromarkers overlap. In our view, an ideal approach,

yet to be realized, combines type III and type IV studies in that order.

With the above considerations in mind, and with due caution regarding direct comparisons between topographies obtained using different EEG montages and methods, in the following we attempt to map some of our neuromarker findings with the literature (question a in **Table 3**), for those studies in which we found sufficient spatial and spectral information to do so. Resolution of questions b and c (power modulation, brain patterns' spatiotemporal dynamics) would require replications or reanalysis of the respective studies due to the unforeseen effects of extraneous variance—an important issue but well beyond the scope of this work.

Our finding of alpha as an important neuromarker of social function echoes other studies that suggested its importance for the integration of sensory information into social perception, social behavior and (joint) attention (Babiloni et al., 2002; Perry et al., 2010, 2011; Lachat et al., 2012; and with MEG: Sebastiani et al., 2014). The latter work is of both a synchronic and diachronic nature and is in agreement with our findings. We also observed a medial mu in our synchronic studies of social coordination (Tognoli et al., 2007a, 2011). This rhythm distributed its power broadly in frequency and in space, with a mellow peak in the low part of the 10 Hz range over the midline at the level of electrode FCz; power was attenuated during social interactions irrespective of how people coordinated. This rhythm's frequency and topography might relate to the finding by Moreno et al. (2015), of a central mu that is suppressed during reading of action language (as opposed to abstract language)—although it is difficult to classify this study with respect to synchronic or diachronic behavior since it is a study of single subjects.

In diachronic behaviors such as action observation and delayed imitation, we observed the occurrence of two other mu rhythms with a clear lateralization and a slightly faster frequency than mu medial. The mu rhythms we found perhaps reflect their

historical definition since they were located above the Rolandic fissure. Our findings seem to map in a congenial way with a large number of studies of action observation, execution, imagination and imitation (Babiloni et al., 2002; Muthukumaraswamy et al., 2004; Cheng et al., 2008; Perry et al., 2010, 2011; Arnstein et al., 2011; Avanzini et al., 2012; Lachat et al., 2012; Moore et al., 2012; Braadbaart et al., 2013; Sebastiani et al., 2014).

A further finding in our diachronic studies, a parietal rhythm, nu, appeared to be suppressed during action execution, but comparatively less so when the action was being observed. It is possible that this rhythm concurs with findings of parietal mu modulation (Babiloni et al., 2002; Avanzini et al., 2012). Though less obvious because of its smaller spectral footprint and amplitude, we hypothesize that the nu rhythm may well be present in other studies, yet elude detection due to methodological factors. In the same manner, the other neuromarkers that were discovered in our synchronic and diachronic studies (phi and kappa respectively) were of modest size as compared to alpha and mu, and may not make themselves apparent unless specifically parsed as described in Section “The neuromarker framework: finding local oscillations”.

SUMMARY AND CONCLUSIONS

Social neuroscience is a young discipline. Accordingly in this review we have focused more on finding the right questions than providing definitive answers about the functional and dynamic architecture of social brains. Our aim was to establish a comprehensive framework to study the dynamics of brains as they evolve through successive phases of social interaction. Such a dynamical framework seems necessary if we are to understand normal and pathological social function. Using a novel set of techniques, a number of neuro-functional signatures of social behavior were uncovered, each with a specific topography and frequency, and each based on continuous brain dynamics requiring high temporal precision. We have drafted some tentative directions for functional inference on newly discovered and lesser known neuromarkers, keeping in mind that more information is needed to converge upon solid interpretations.

Social behavior is grounded in perception~action coupling, a fundamental organizing principle of intentional living beings (see also Prinz, 1997): in the absence of action from an individual, there is no information flow to another’s brain. Without sensitivity to this information by the receiver’s perceptual system, there can be no effective social interaction. We have stressed the primacy of information flows across individuals, and we have shown their fundamental importance for attention—an aspect, perhaps, that has received insufficient scrutiny in social neuroscience.

We examined interpersonal perception-action coupling from the standpoint of the relative phase between individuals (simultaneous or diachronic action~perception). Of course, what we describe as synchronic and diachronic behaviors are limit-cases of a continuum of social circumstances that varies systematically with the phase of each participant’s action. Yet, heuristically, this taxonomy proved useful in revealing little

overlap between respective neuromarker landscapes. At several levels of temporal precision (e.g., across tasks, through average activity over trials, and through instantaneous activity), we emphasized the complex reorganization of endogenous brain networks leading to different phases and facets of social behavior.

From the multiplicity of functional processes, and from our findings that the underlying neuromarkers tend not to arise simultaneously, we have begun to enquire about their engagement and disengagement over the course of social interaction, a step that we hope will help refine functional (dynamical) modeling. In our opinion, much work remains to unravel the neural choreography of the cognitive, affective and behavioral processes that participate in social behavior and to embed them in theoretical/computational models of social brain function. Keys to future progress lie with studies of neuromarker coordination in social settings, which, as in other systems such as bimanual and sensorimotor coordination, will lead to modeling the neuro-functional architecture of the social brain.

Already, the present dynamical approach to social brains has revealed some unique coordinative mechanisms that truly relate to social neuroscience (as opposed to a generalization of cognitive neuroscience to social tasks). That is, with the help of the dynamical framework presented in Section “The neuromarker framework: brain coordination dynamics”, we have encountered preliminary evidence that spatiotemporal patterns of brain activity tend to switch in synchrony in pairs of subjects that establish or dissolve behavioral coordination (Benites et al., 2010; Fuchs et al., 2010). These synchronized transitions happened even as one subject’s neural activity differed from that of the other. This finding reveals once more that the interplay of integrative and segregative tendencies within (and now between) brains is a powerful mechanism of nature to enhance system complexity (Kelso, 1995; Edelman, 1999; Sporns, 2003; Kelso and Tognoli, 2007). It is at the level of multiple brains and multiple behaviors, within a complex systems framework, that dynamical models of social function are likely to be ultimately formulated.

ACKNOWLEDGMENTS

We acknowledge the contribution of HBBL group members who took part in the work that led to this review, especially Gonzalo de Guzman, Julien Lagarde, Daniela Benites, Benjamin Suutari, Seth Weisberg, William McLean and Armin Fuchs. We are grateful to the agencies that supported the theoretical, methodological and empirical work of our Social Neuroscience research program, and especially, NIMH (MH080838), NSF (BCS026897), the US ONR (N00014-09-1-0527), and the Davimos Family Endowment for Excellence in Science. JASK was also supported by the Chaire d’excellence Pierre de Fermat.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fnhum.2015.00563/abstract>

REFERENCES

- Adolphs, R. (2009). The social brain: neural basis of social knowledge. *Annu. Rev. Psychol.* 60, 693–716. doi: 10.1146/annurev.psych.60.110707.163514
- Anderson, S. W., Bechara, A., Damasio, H., Tranel, D., and Damasio, A. R. (1999). Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Nat. Neurosci.* 2, 1032–1037. doi: 10.1038/14833
- Arnstein, D., Cui, F., Keysers, C., Maurits, N. M., and Gazzola, V. (2011). μ suppression during action observation and execution correlates with BOLD in dorsal premotor, inferior parietal and SI cortices. *J. Neurosci.* 31, 14243–14249. doi: 10.1523/JNEUROSCI.0963-11.2011
- Avanzini, P., Fabbri-Destro, M., Dalla Volta, R., Daprati, E., Rizzolatti, G., and Cantalupo, G. (2012). The dynamics of sensorimotor cortical oscillations during the observation of hand movements: an EEG study. *PLoS One* 7:e37534. doi: 10.1371/journal.pone.0037534
- Babiloni, C., Babiloni, F., Carducci, F., Cincotti, F., Cocozza, G., Del Percio, C., et al. (2002). Human Cortical Electroencephalography (EEG) rhythms during the observation of simple aimless movements: a high-resolution EEG study. *NeuroImage* 17, 559–572. doi: 10.1006/nimg.2002.1192
- Banerjee, A., Tognoli, E., Kelso, J. A. S., and Jirsa, V. K. (2012). Spatiotemporal re-organization of large-scale neural assemblies underlies bimanual coordination. *Neuroimage* 62, 1582–1592. doi: 10.1016/j.neuroimage.2012.05.046
- Baron-Cohen, S., and Belmonte, M. K. (2005). Autism: a window onto the development of the social and the analytic brain. *Annu. Rev. Neurosci.* 28, 109–126. doi: 10.1146/annurev.neuro.27.070203.144137
- Başar, E. (2004). *Memory and Brain Dynamics: Oscillations Integrating Attention, Perception, Learning and Memory*. Boca Raton, Fla: CRC Press.
- Bassuk, S. S., Glass, T. A., and Berkman, L. F. (1999). Social disengagement and incident cognitive decline in community-dwelling elderly persons. *Ann. Intern. Med.* 131, 165–173. doi: 10.7326/0003-4819-131-3-199908030-00002
- Bazanova, O. M., and Vernon, D. (2014). Interpreting EEG alpha activity. *Neurosci. Biobehav. Rev.* 44, 94–110. doi: 10.1016/j.neubiorev.2013.05.007
- Beauchamp, M. S., Lee, K. E., Haxby, J. V., and Martin, A. (2003). fMRI responses to video and point-light displays of moving humans and manipulable objects. *J. Cogn. Neurosci.* 15, 991–1001. doi: 10.1162/089892903770007380
- Behrens, T. E. J., Hunt, L. T., and Rushworth, M. F. (2009). The computation of social behavior. *Science* 324, 1160–1164. doi: 10.1126/science.1169694
- Béland, F., Zunzunegui, M.-V., Alvarado, B., Otero, A., and Del Ser, T. (2005). Trajectories of cognitive decline and social relations. *J. Gerontol. B Psychol. Sci. Soc. Sci.* 60, P320–P330. doi: 10.1093/geronb/60.6.p320
- Benites, D., Tognoli, E., de Guzman, G. C., and Kelso, J. A. S. (2010). Brain coordination dynamics: continuous EEG tracking of the neural functional organization in a social task. *Psychophysiology* 47:S75. doi: 10.1111/j.1469-8986.2010.01111.x
- Bennett, D. A., Schneider, J. A., Tang, Y., Arnold, S. E., and Wilson, R. S. (2006). The effect of social networks on the relation between Alzheimer's disease pathology and level of cognitive function in old people: a longitudinal cohort study. *Lancet Neurol.* 5, 406–412. doi: 10.1016/s1474-4422(06)70417-3
- Birbaumer, N. (2006). Breaking the silence: brain-computer interfaces (BCI) for communication and motor control. *Psychophysiology* 43, 517–532. doi: 10.1111/j.1469-8986.2006.00456.x
- Blakemore, S.-J., Winston, J., and Frith, U. (2004). Social cognitive neuroscience: where are we heading? *Trends Cogn. Sci.* 8, 216–222. doi: 10.1016/j.tics.2004.03.012
- Blankertz, B., Tangermann, M., Vidaurre, C., Fazli, S., Sannelli, C., Haufe, S., et al. (2010). The Berlin brain-computer interface: non-medical uses of BCI technology. *Front. Neurosci.* 4:198. doi: 10.3389/fnins.2010.00198
- Bollimunta, A., Mo, J., Schroeder, C. E., and Ding, M. (2011). Neuronal mechanisms and alpha modulation of corticothalamic alpha oscillations. *J. Neurosci.* 31, 4935–4943. doi: 10.1523/JNEUROSCI.5580-10.2011
- Braadbaart, L., Williams, J. H. G., and Waiter, G. D. (2013). Do mirror neuron areas mediate mu rhythm suppression during imitation and action observation? *Int. J. Psychophysiol.* 89, 99–105. doi: 10.1016/j.ijpsycho.2013.05.019
- Bressler, S. L., and Kelso, J. A. S. (2001). Cortical coordination dynamics and cognition. *Trends Cogn. Sci.* 5, 26–36. doi: 10.1016/s1364-6613(00)01564-3
- Bressler, S. L., and Tognoli, E. (2006). Operational principles of neurocognitive networks. *Int. J. Psychophysiol.* 60, 139–148. doi: 10.1016/j.ijpsycho.2005.12.008
- Brunet-Gouet, E., and Decety, J. (2006). Social brain dysfunctions in schizophrenia: a review of neuroimaging studies. *Psychiatry Res.* 148, 75–92. doi: 10.1016/j.psychresns.2006.05.001
- Buzsáki, G. (2006). *Rhythms of the Brain*. Oxford: Oxford University Press.
- Cacioppo, J. T. (2002). Social neuroscience: understanding the pieces fosters understanding the whole and vice versa. *Am. Psychol.* 57, 819–831. doi: 10.1037/0003-066x.57.11.819
- Cacioppo, J. T., Amaral, D. G., Blanchard, J. J., Cameron, J. L., Carter, C. S., Crews, D., et al. (2007). Social neuroscience: progress and implications for mental health. *Perspect. Psychol. Sci.* 2, 99–123. doi: 10.1111/j.1745-6916.2007.00032.x
- Chatrian, G. E., Lettich, E., and Nelson, P. L. (1985). Ten percent electrode system for topographic studies of spontaneous and evoked EEG activities. *Am. J. EEG Techno.* 25, 83–92. doi: 10.1080/00029238.1985.11080163
- Cheng, Y., Lee, P.-L., Yang, C.-Y., Lin, C.-P., Hung, D., and Decety, J. (2008). Gender differences in the mu rhythm of the human mirror-neuron system. *PLoS One* 3:e2113. doi: 10.1371/journal.pone.0002113
- Clark, L. A., Vittengl, J. R., Kraft, D., and Jarrett, R. B. (2003). Shared, not unique, components of personality and psychosocial functioning predict depression severity after acute-phase cognitive therapy. *J. Pers. Disord.* 17, 406–430. doi: 10.1521/pedi.17.5.406.22975
- Cochin, S., Barthelemy, C., Roux, S., and Martineau, J. (1999). Observation and execution of movement: similarities demonstrated by quantified electroencephalography. *Eur. J. Neurosci.* 11, 1839–1842. doi: 10.1046/j.1460-9568.1999.00598.x
- Compton, M. T., Weiss, P. S., West, J. C., and Kaslow, N. J. (2005). The associations between substance use disorders, schizophrenia-spectrum disorders, and Axis IV psychosocial problems. *Soc. Psychiatry Psychiatr. Epidemiol.* 40, 939–946. doi: 10.1007/s00127-005-0964-4
- Couture, S. M., Penn, D. L., and Roberts, D. L. (2006). The functional significance of social cognition in schizophrenia: a review. *Schizophr. Bull.* 32(Suppl. 1), S44–S63. doi: 10.1093/schbul/sbl029
- Crooks, V. C., Lubben, J., Petitti, D. B., Little, D., and Chiu, V. (2008). Social network, cognitive function and dementia incidence among elderly women. *Am. J. Public Health* 98, 1221–1227. doi: 10.2105/ajph.2007.115923
- Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., et al. (2004). Early social attention impairments in autism: social orienting, joint attention and attention to distress. *Dev. Psychol.* 40, 271–283. doi: 10.1037/0012-1649.40.2.271
- deGuzman, G. C., and Kelso, J. A. S. (1991). Multifrequency behavioral patterns and the phase attractive circle map. *Biol. Cybern.* 64, 485–495. doi: 10.1007/bf00202613
- De Vico Fallani, F., Nicosia, V., Sinatra, R., Astolfi, L., Cincotti, F., Mattia, D., et al. (2010). Defecting or not defecting: how to 'read' human behavior during cooperative games by EEG measurements. *PLoS One* 5:e14187. doi: 10.1371/journal.pone.0014187
- Decety, J., Chaminade, T., Grèzes, J., and Meltzoff, A. N. (2002). A PET exploration of the neural mechanisms involved in reciprocal imitation. *Neuroimage* 15, 265–272. doi: 10.1006/nimg.2001.0938
- Derix, J., Iljina, O., Schulze-Bonhage, A., Aertsen, A., and Ball, T. (2012). "Doctor" or "darling"? decoding the communication partner from ECoG of the anterior temporal lobe during non-experimental, real-life social interaction. *Front. Hum. Neurosci.* 6:251. doi: 10.3389/fnhum.2012.00251
- Dodel, S., Cohn, J., Mersmann, J., Luu, P., Forsythe, C., and Jirsa, V. K. (2011). "Brain signatures of team performance," in *Foundations of Augmented Cognition: Directing the Future of Adaptive Systems*, eds D. D. Schmorow and C. M. Fidopiastis (Berlin, Heidelberg: Springer Berlin Heidelberg), 288–297.
- Dumas, G., Chavez, M., Nadel, J., and Martinierie, J. (2012). Anatomical connectivity influences both intra-and inter-brain synchronizations. *PLoS One* 7:e36414. doi: 10.1371/journal.pone.0036414
- Dumas, G., Nadel, J., Soussignan, R., Martinierie, J., and Garner, L. (2010). Inter-brain synchronization during social interaction. *PLoS One* 5:e12166. doi: 10.1371/journal.pone.0012166

- Duran, N. D., and Dale, R. (2014). Perspective-taking in dialogue as self-organization under social constraints. *New Ideas Psychol.* 32, 131–146. doi: 10.1016/j.newideapsych.2013.03.004
- Edelman, G. M. (1999). Building a picture of the brain. *Ann. N. Y. Acad. Sci.* 882, 68–89. doi: 10.1111/j.1749-6632.1999.tb08535.x
- Edelman, G. M., and Gally, J. (2001). Degeneracy and complexity in biological systems. *Proc. Natl. Acad. Sci. U S A* 98, 13763–13768. doi: 10.1073/pnas.231499798
- Farah, M. J. (2011). “Ethical, legal and societal issues in social neuroscience,” in *The Oxford Handbook of Social Neuroscience*, eds J. Decety and J. T. Cacioppo (New York: Oxford University Press), 1015–1024.
- Farrant, A., Morris, R. G., Russell, T., Elwes, R., Akanuma, N., Alarcón, G., et al. (2005). Social cognition in frontal lobe epilepsy. *Epilepsy Behav.* 7, 506–516. doi: 10.1016/j.yebeh.2005.07.018
- Fitzpatrick, P., Mitchell, T., Frazier, J., Cochran, D., Coleman, C., and Schmidt, R. C. (2015). “Exploring the behavioral and neural processes underlying social synchronization of individuals with and without social deficits”, in *Studies in Perception and Action XIII: eighteenth International Conference on Perception and Action*, 41. Psychology Press.
- Freeman, W. J. (2000). *Neurodynamics: an Exploration in Mesoscopic Brain Dynamics*. London, New York: Springer.
- Frith, U., and Frith, C. (2001). The biological basis of social interaction. *Curr. Dir. Psychol. Sci.* 10, 151–155. doi: 10.1111/1467-8721.00137
- Fuchs, A., Jirsa, V. K., and Kelso, J. A. S. (2000). Theory of the relation between human brain activity (MEG) and hand movements. *Neuroimage* 11, 359–369. doi: 10.1006/nimg.1999.0532
- Fuchs, A., Tognoli, E., Benites, D., and Kelso, J. A. S. (2010). “Neural correlates of social coordination: spatiotemporal analysis of brain and behavioral measures”, in *Society for Neuroscience Itinerary Planner Program*. San Diego, CA: Society for Neuroscience.
- Funane, T., Kiguchi, M., Atsumori, H., Sato, H., Kubota, K., and Koizumi, H. (2011). Synchronous activity of two people’s prefrontal cortices during a cooperative task measured by simultaneous near-infrared spectroscopy. *J. Biomed. Opt.* 16:077011. doi: 10.1117/1.3602853
- Gallese, V., Keysers, C., and Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends Cogn. Sci.* 8, 396–403. doi: 10.1016/j.tics.2004.07.002
- Genizi, J., Shamay-Tsoory, S. G., Shahar, E., Yaniv, S., and Aharon-Perez, J. (2011). Impaired social behavior in children with benign childhood epilepsy with centrotemporal spikes. *J. Child Neurol.* 27, 156–161. doi: 10.1177/0883073811414420
- Giovagnoli, A. R., Franceschetti, S., Reati, F., Parente, A., Maccagnano, C., Villani, F., et al. (2011). Theory of mind in frontal and temporal lobe epilepsy: cognitive and neural aspects. *Epilepsia* 52, 1995–2002. doi: 10.1111/j.1528-1167.2011.03215.x
- Glei, D. A., Landau, D. A., Goldman, N., Chuang, Y.-L., Rodríguez, G., and Weinstein, M. (2005). Participating in social activities helps preserve cognitive function: an analysis of a longitudinal, population-based study of the elderly. *Int. J. Epidemiol.* 34, 864–871. doi: 10.1093/ije/dyi049
- Gomez-Hernandez, R., Max, J. E., Kosier, T., Paradiso, S., and Robinson, R. G. (1997). Social impairment and depression after traumatic brain injury. *Arch. Phys. Med. Rehabil.* 78, 1321–1326. doi: 10.1016/s0003-9993(97)90304-x
- Green, M. J., and Phillips, M. L. (2004). Social threat perception and the evolution of paranoia. *Neurosci. Biobehav. Rev.* 28, 333–342. doi: 10.1016/j.neubiorev.2004.03.006
- Greenham, M., Spencer-Smith, M. M., Anderson, P. J., Coleman, L., and Anderson, V. A. (2010). Social functioning in children with brain insult. *Front. Hum. Neurosci.* 4:22. doi: 10.3389/fnhum.2010.00022
- Guger, C., Ramoser, H., and Pfurtscheller, G. (2000). Real-time EEG analysis with subject-specific spatial patterns for a brain-computer interface (BCI). *IEEE Trans. Rehabil. Eng.* 8, 447–456. doi: 10.1109/86.895947
- Guionnet, S., Nadel, J., Bertasi, E., Sperduti, M., Delaveau, P., and Fossati, P. (2012). Reciprocal imitation: toward a neural basis of social interaction. *Cereb. Cortex* 22, 971–978. doi: 10.1093/cercor/bhr177
- Hadjikhani, N., Joseph, R. M., Snyder, J., and Tager-Flusberg, H. (2006). Anatomical differences in the mirror neuron system and social cognition network in autism. *Cereb. Cortex* 16, 1276–1282. doi: 10.1093/cercor/bhj069
- Haken, H., Kelso, J. A. S., and Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biol. Cybern.* 51, 347–356. doi: 10.1007/bf00336922
- Hari, R. (2006). Action-perception connection and the cortical mu rhythm. *Prog. Brain Res.* 159, 253–260. doi: 10.1016/s0079-6123(06)59017-x
- Hari, R., and Kujala, M. V. (2009). Brain basis of human social interaction: from concepts to brain imaging. *Physiol. Rev.* 89, 453–479. doi: 10.1152/physrev.00041.2007
- Hari, R., Forss, N., Avikainen, S., Kirveskari, E., Salenius, S., and Rizzolatti, G. (1998). Activation of human primary motor cortex during action observation: a neuromagnetic study. *Proc. Natl. Acad. Sci. U S A* 95, 15061–15065. doi: 10.1073/pnas.95.25.15061
- Hogeveen, J., Chartrand, T. L., and Obhi, S. S. (2014). Social mimicry enhances mu-suppression during action observation. *Cereb. Cortex* 25, 2076–2082. doi: 10.1093/cercor/bhu016
- Hsu, W. Y. (2011). Continuous EEG signal analysis for asynchronous BCI application. *Int. J. Neural Syst.* 21, 335–350. doi: 10.1142/s0129065711002870
- Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazzoni, J. C., and Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science* 286, 2526–2528. doi: 10.1126/science.286.5449.2526
- Inoue, Y., Yamada, K., and Kanba, S. (2006). Deficit in theory of mind is a risk for relapse of major depression. *J. Affect. Disord.* 95, 125–127. doi: 10.1016/j.jad.2006.04.018
- Insel, T. R., and Fernald, R. D. (2004). How the brain processes social information: searching for the social brain. *Annu. Rev. Neurosci.* 27, 697–722. doi: 10.1146/annurev.neuro.27.070203.144148
- Izuma, K., Saito, D. N., and Sadato, N. (2008). Processing of social and monetary rewards in the human striatum. *Neuron* 58, 284–294. doi: 10.1016/j.neuron.2008.03.020
- Jones, W., Bellugi, U., Lai, Z., Chiles, M., Reilly, J., Lincoln, A., et al. (2000). II. Hypersociability in williams syndrome. *J. Cogn. Neurosci.* 12(Suppl. 1), 30–46. doi: 10.1162/089892900561968
- Keller, P. E., Novembre, G., and Hove, M. J. (2014). Rhythm in joint action: psychological and neurophysiological mechanisms for real-time interpersonal coordination. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130394. doi: 10.1098/rstb.2013.0394
- Kelso, J. A. S. (1984). Phase transitions and critical behavior in human bimanual coordination. *Am. J. Physiol.* 246, R1000–R1004.
- Kelso, J. A. S. (1995). *Dynamic Patterns: The Self-Organization of Brain and Behavior*. Cambridge, MA: MIT Press.
- Kelso, J. A. S., DelColle, J. D., and Schöner, G. (1990). “Action-perception as a pattern formation process,” in *Attention and Performance XIII* ed. M. Jeannerod (Hillsdale, NJ: Erlbaum), 139–169.
- Kelso, J. A. S., Fuchs, A., Lancaster, R., Holroyd, T., Cheyne, D., and Weinberg, H. (1998). Dynamic cortical activity in the human brain reveals motor equivalence. *Nature* 392, 814–818. doi: 10.1038/33922
- Kelso, J. A. S., and Tognoli, E. (2007). “Toward a complementary neuroscience: metastable coordination dynamics of the brain,” in *Neurodynamics of Cognition and Consciousness*, eds L. I. Perlovsky and R. Kozma, (Berlin, Heidelberg: Springer Berlin Heidelberg), 39–59.
- Kelso, J. A. S. (2012). Multistability and metastability: understanding dynamic coordination in the brain. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 36, 906–918. doi: 10.1098/rstb.2011.0351
- Kelso, J. A. S., and Engstrom, D. A. (2006). *The Complementary Nature*. Cambridge, MA: The MIT Press.
- Kelso, J. A. S., Tuller, B., Vatikiotis-Bateson, E., and Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: evidence for coordinative structures. *J. Exp. Psychol. Hum. Percept. Perform.* 10, 812–832. doi: 10.1037/0096-1523.10.6.812
- Kirsch, H. E. (2006). Social cognition and epilepsy surgery. *Epilepsy Behav.* 8, 71–80. doi: 10.1016/j.yebeh.2005.09.002
- Klimesch, W., Doppelmayr, M., Russegger, H., Pachinger, T., and Schwaiger, J. (1998). Induced alpha band power changes in the human EEG and attention. *Neurosci. Lett.* 244, 73–76. doi: 10.1016/s0304-3940(98)00122-0
- Konvalinka, I., and Roepstorff, A. (2012). The two-brain approach: how can mutually interacting brains teach us something about social interaction? *Front. Hum. Neurosci.* 6:215. doi: 10.3389/fnhum.2012.00215

- Konvalinka, I., Vuust, P., Roepstorff, A., and Frith, C. D. (2010). Follow you, follow me: continuous mutual prediction and adaptation in joint tapping. *Q. J. Exp. Psychol.* 63, 2220–2230 doi: 10.1080/17470218.2010.497843
- Kübler, A., Nijboer, F., Mellinger, J., Vaughan, T. M., Pawelzik, H., Schalk, G., et al. (2005). Patients with ALS can use sensorimotor rhythms to operate a brain-computer interface. *Neurology* 64, 1775–1777. doi: 10.1212/01.wnl.0000158616.43002.6d
- Lachat, F., Hugueville, L., Lemaréchal, J.-D., Conty, L., and George, N. (2012). Oscillatory brain correlates of live joint attention: a dual-EEG study. *Front. Hum. Neurosci.* 6:156. doi: 10.3389/fnhum.2012.00156
- Lehmann, D., Faber, P. L., Gianotti, L., Kochi, K., and Pascual-Marqui, D. R. (2006). Coherence and phase locking in the scalp EEG and between LORETA model sources and microstates as putative mechanisms of brain temporo-spatial functional organization. *J. Physiol. Paris* 99, 29–36. doi: 10.1016/j.jphsparis.2005.06.005
- Lezak, M. D., and O'Brien, P. K. (1988). Longitudinal study of emotional, social and physical changes after traumatic brain injury. *J. Learn. Disabil.* 21, 456–463. doi: 10.1177/002221948802100802
- Liao, Y., Makeig, S., Acar, Z., and Deak, G. (2012). EEG imaging of toddlers during live dyadic turn-taking: mu-rhythm modulation and source clusters in natural action observation and execution *Human Brain Mapping*. Beijing: Organization for Human Brain Mapping.
- Lindenberger, U., Li, S.-C., Gruber, W., and Müller, V. (2009). Brains swinging in concert: cortical phase synchronization while playing guitar. *BMC Neurosci.* 10:22. doi: 10.1186/1471-2202-10-22
- Mendez, M. F., Chen, A. K., Shapira, J. S., and Miller, B. L. (2005). Acquired sociopathy and frontotemporal dementia. *Dement. Geriatr. Cogn. Disord.* 20, 99–104. doi: 10.1159/000086474
- Meyer-Lindenberg, A., Hariri, A. R., Munoz, K. E., Mervis, C. B., Mattay, V. S., Morris, C. A., et al. (2005). Neural correlates of genetically abnormal social cognition in williams syndrome. *Nat. Neurosci.* 8, 991–993. doi: 10.1038/nn1494
- Monetta, L., Grindrod, C. M., and Pell, M. D. (2009). Irony comprehension and theory of mind deficits in patients with parkinson's disease. *Cortex* 45, 972–981. doi: 10.1016/j.cortex.2009.02.021
- Montague, P. R., Berns, G. S., Cohen, J. D., McClure, S. M., Pagnoni, G., Dhamala, M., et al. (2002). Hyperscanning: simultaneous fMRI during linked social interactions. *Neuroimage* 16, 1159–1164. doi: 10.1006/nimg.2002.1150
- Moore, A., Gorodnitsky, I., and Pineda, J. (2012). EEG mu component responses to viewing emotional faces. *Behav. Brain Res.* 226, 309–316. doi: 10.1016/j.bbr.2011.07.048
- Moreno, I., de Vega, M., León, I., Bastiaansen, M., Glen Lewis, A., and Magyari, L. (2015). Brain dynamics in the comprehension of action-related language. A time-frequency analysis of mu rhythms. *Neuroimage* 109, 50–62. doi: 10.1016/j.neuroimage.2015.01.018
- Mulholland, T. B. (1972). Occipital alpha revisited. *Psychol. Bull.* 78, 176–182. doi: 10.1037/h0032962
- Müller, V., Lutzenberger, W., Preissl, H., Pulvermüller, F., and Birbaumer, N. (2003). Complexity of visual stimuli and non-linear EEG dynamics in humans. *Brain Res. Cogn. Brain Res.* 16, 104–110. doi: 10.1016/s0926-6410(02)00225-2
- Mundy, P., Sigman, M., Ungerer, J., and Sherman, T. (1986). Defining the social deficits of autism: the contribution of non-verbal communication measures. *J. Child Psychol. Psychiatry* 27, 657–669. doi: 10.1111/j.1469-7610.1986.tb00190.x
- Murzin, V., Fuchs, A., and Kelso, J. A. S. (2011). Anatomically constrained minimum variance beamforming applied to EEG. *Exp. Brain Res.* 214, 515–528. doi: 10.1007/s00221-011-2850-5
- Muthukumaraswamy, S. D., Johnson, B. X., and McNair, N. A. (2004). Mu rhythm modulation during observation of an object-directed grasp. *Brain Res. Cogn. Brain Res.* 19, 195–201. doi: 10.1016/j.cogbrainres.2003.12.001
- Naem, M., Prasad, G., Watson, D. R., and Kelso, J. A. S. (2012). Electrophysiological signatures of intentional social coordination in the 10–12Hz range. *Neuroimage* 59, 1795–1803. doi: 10.1016/j.neuroimage.2011.08.010
- Oberman, L. M., Pineda, J. A., and Ramachandran, V. S. (2007). The human mirror neuron system: a link between action observation and social skills. *Soc. Cogn. Affect. Neurosci.* 2, 62–66. doi: 10.1093/scan/nsl022
- Ochsner, K. N., and Lieberman, M. D. (2001). The emergence of social cognitive neuroscience. *Am. Psychol.* 56, 717–734. doi: 10.1037/0003-066x.56.9.717
- Olson, A., Phelps, E. A. (2007). “Understanding social evaluations: what we can (and cannot) learn from neuroimaging,” in *C Implicit Measures of Attitudes*, eds B. Wittenbrink and N. Schwarz (New York, NY: Guilford Press), 159–175.
- Oullier, O., de Guzman, G. C., Jantzen, K. J., Lagarde, J., and Kelso, J. A. S. (2008). Social coordination dynamics: measuring human bonding. *Soc. Neurosci.* 3, 178–192. doi: 10.1080/17470910701563392
- Oullier, O., and Kelso, J. A. S. (2009). “Social coordination from the perspective of coordination dynamics,” in *Encyclopedia of Complexity and Systems Science*, ed. R. A. Meyers (New York, NY: Springer New York), 8198–8213.
- Palva, S., and Palva, J. M. (2007). New vistas for alpha-frequency band oscillations. *Trends Neurosci.* 30, 150–158. doi: 10.1016/j.tins.2007.02.001
- Palva, S., and Palva, J. M. (2011). Functional roles of alpha-band phase synchronization in local and large-scale cortical networks. *Front. Psychol.* 2:204. doi: 10.3389/fpsyg.2011.00204
- Palva, S., and Palva, J. M. (2012). Discovering oscillatory interaction networks with M/EEG: challenges and breakthroughs. *Trends Cogn. Sci.* 16, 219–230. doi: 10.1016/j.tics.2012.02.004
- Pascual-Marqui, R. D., Esslen, M., Kochi, K., and Lehmann, D. (2002). Functional imaging with low-resolution brain electromagnetic tomography (LORETA): a review. *Methods Find. Exp. Clin. Pharmacol.* 24, 91–95.
- Perry, A., Stein, L., and Bentin, S. (2011). Motor and attentional mechanisms involved in social interaction—evidence from mu and alpha EEG suppression. *Neuroimage* 58, 895–904. doi: 10.1016/j.neuroimage.2011.06.060
- Perry, A., Troje, N. F., and Bentin, S. (2010). Exploring motor system contributions to the perception of social information: evidence from EEG activity in the mu/alpha frequency range. *Soc. Neurosci.* 5, 272–284. doi: 10.1080/17470910903395767
- Pineda, J. A. (2005). The functional significance of mu rhythms: translating ‘seeing’ and ‘hearing’ into ‘doing’. *Brain Res. Brain Res. Rev.* 50, 57–68. doi: 10.1016/j.brainresrev.2005.04.005
- Poletti, M., Enrici, I., Bonuccelli, U., and Adenzato, M. (2011). Theory of mind in parkinson's disease. *Behav. Brain Res.* 219, 342–350. doi: 10.1016/j.bbr.2011.01.010
- Prinz, W. (1997). Perception and action planning. *Eur. J. Cogn. Psychol.* 9, 129–154. doi: 10.1080/713752551
- Riley, M. A., Richardson, M. J., Shockley, K., and Ramenzoni, V. C. (2011). Interpersonal synergies. *Front. Psychol.* 2:38. doi: 10.3389/fpsyg.2011.00038
- Russell, T. A., Rubia, K., Bullmore, E. T., Soni, W., Suckling, J., Brammer, M. J., et al. (2000). Exploring the social brain in schizophrenia: left prefrontal underactivation during mental state attribution. *Am. J. Psychiatry* 157, 2040–2042. doi: 10.1176/appi.ajp.157.12.2040
- Saito, D. N., Tanabe, H. C., Izuma, K., Hayashi, M. J., Morito, Y., Komeda, H., et al. (2010). Stay tuned: inter-individual neural synchronization during mutual gaze and joint attention. *Front. Integr. Neurosci.* 4:127. doi: 10.3389/fnint.2010.00127
- Saltzman, E., and Kelso, J. A. S. (1987). Skilled actions: a task-dynamic approach. *Psychol. Rev.* 94, 84–106. doi: 10.1037/0033-295x.94.1.84
- Sauseng, P., Klimesch, W., Stadler, W., Schabus, M., Doppelmayr, M., Hanslmayr, S., et al. (2005). A shift of visual spatial attention is selectively associated with human EEG alpha activity. *Eur. J. Neurosci.* 22, 2917–2926. doi: 10.1111/j.1460-9568.2005.04482.x
- Saxe, R. (2006). Uniquely human social cognition. *Curr. Opin. Neurobiol.* 16, 235–239. doi: 10.1016/j.conb.2006.03.001
- Schilbach, L. (2010). A second-person approach to other minds. *Nat. Rev. Neurosci.* 11:449. doi: 10.1038/nrn2805-c1
- Schilbach, L., Wilms, M., Eickhoff, S. B., Romanzetti, S., Tepest, R., Bente, G., et al. (2010). Minds made for sharing: initiating joint attention recruits reward-related neurocircuitry. *J. Cogn. Neurosci.* 22, 2702–2715. doi: 10.1162/jocn.2009.21401
- Schimansky, J., David, N., Rössler, W., and Haker, H. (2010). Sense of agency and mentalizing: dissociation of subdomains of social cognition in patients with schizophrenia. *Psychiatry Res.* 178, 39–45. doi: 10.1016/j.psychres.2010.04.002

- Schmidt, R. C., Carello, C., and Turvey, M. T. (1990). Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people. *J. Exp. Psychol. Hum. Percept. Perform.* 16, 227–247. doi: 10.1037/0096-1523.16.2.227
- Sebanz, N., Knoblich, G., Prinz, W., and Wascher, E. (2006a). Twin peaks: an ERP study of action planning and control in co-acting individuals. *J. Cogn. Neurosci.* 18, 859–870. doi: 10.1162/jocn.2006.18.5.859
- Sebanz, N., Bekkering, H., and Knoblich, G. (2006b). Joint action: bodies and minds moving together. *Trends Cogn. Sci.* 10, 70–76. doi: 10.1016/j.tics.2005.12.009
- Sebastiani, V., de Pasquale, F., Costantini, M., Mantini, D., Pizzella, V., Romani, G. L., et al. (2014). Being an agent or an observer: different spectral dynamics revealed by MEG. *Neuroimage* 102, 717–728. doi: 10.1016/j.neuroimage.2014.08.031
- Segrin, C. (2000). Social skills deficits associated with depression. *Clin. Psychol. Rev.* 20, 379–403. doi: 10.1016/S0272-7358(98)00104-4
- Sporns, O. (2003). Network analysis, complexity and brain function. *Complexity* 8, 56–60. doi: 10.1002/cplx.10047
- Suda, M., Takei, Y., Aoyama, Y., Narita, K., Sato, T., Fukuda, M., et al (2010). Frontopolar activation during face-to-face conversation: an in situ study using near-infrared spectroscopy. *Neuropsychologia* 48, 441–447. doi: 10.1016/j.neuropsychologia.2009.09.036
- Sutari, B. S., Weisberg, S. A., Tognoli, E., and Kelso, J. A. S. (2010). A fresh look on mu rhythms: individual and social behaviors. *Society for Neuroscience Itinerary Planner*, (San Diego, CA: Society for Neuroscience), 12.
- Swinnen, S. P., and Wenderoth, N. (2004). Two hands, one brain: cognitive neuroscience of bimanual skill. *Trends Cogn. Sci.* 8, 18–25. doi: 10.1016/j.tics.2003.10.017
- Tass, P., Kurths, J., Rosenblum, M., Weule, J., Pikovsky, A., Volkmann, J., et al. (1999). “Complex phase synchronization in neurophysiological data,” in *Analysis of Neurophysiological Brain Functioning*, ed. C. Uhl (Berlin: Springer-Verlag), 252–273.
- Thirion, B., Mercier, M. R., Jorland, G., Berthoz, A., and Blanke, O. (2010). Mental imagery of self-location during spontaneous and active self-other interactions: an electrical neuroimaging study. *J. Neurosci.* 30, 7202–7214. doi: 10.1523/jneurosci.3403-09.2010
- Tognoli, E., and Kelso, J. A. S. (2009). Brain coordination dynamics: true and false faces of phase synchrony and metastability. *Prog. Neurobiol.* 87, 31–40. doi: 10.1016/j.pneurobio.2008.09.014
- Tognoli, E., and Kelso, J. A. S. (2013). “On the brain’s dynamical complexity: coupling and causal influences across spatiotemporal scales,” in *Advances in Cognitive Neurodynamics (III)*, ed. Y. Yamaguchi (Dordrecht: Springer), 259–265.
- Tognoli, E., and Kelso, J. A. S. (2014). The metastable brain. *Neuron* 81, 35–48. doi: 10.1016/j.neuron.2013.12.022
- Tognoli, E. (2008). “EEG coordination dynamics: neuromarkers of social coordination,” in *Coordination: Neural, Behavioral and Social Dynamics*, eds A. Fuchs and V. K. Jirsa (Berlin, Heidelberg: Springer Berlin Heidelberg), 309–323.
- Tognoli, E., de Guzman, G. C. and Kelso, J. A. S. (2011). “Interacting humans and the dynamics of their social brains,” in *Advances in Cognitive Neurodynamics (II)*, eds R. Wang and F. Gu (Dordrecht: Springer), 139–143.
- Tognoli, E., Lagarde, J., DeGuzman, G. C., and Kelso, J. A. S. (2007a). The phi complex as a neuromarker of human social coordination. *Proc. Natl. Acad. Sci. U.S.A.* 104, 8190–8195. doi: 10.1073/pnas.0611453104
- Tognoli, E., Magne, C., de Guzman, G. C., Tuller, B., and Kelso, J. A. S. (2007b). Brain rhythms underlying intentional social coordination. *Society for Neuroscience Itinerary Planner* (San Diego, CA: Society for Neuroscience).
- Tononi, G. (2010). Information integration: its relevance to brain function and consciousness. *Arch. Ital. Biol.* 148, 299–322.
- Townsend, G., Graimann, B., and Pfurtscheller, G. (2004). Continuous EEG classification during motor imagery - simulation of an asynchronous BCI. *IEEE Trans. Neural Syst. Rehabil. Eng.* 12, 258–265. doi: 10.1109/tnsre.2004.827220
- Uddin, L. Q., Iacoboni, M., Lange, C., and Keenan, J. P. (2007). The self and social cognition: the role of cortical midline structures and mirror neurons. *Trends Cogn. Sci.* 11, 153–157. doi: 10.1016/j.tics.2007.01.001
- Vanderwert, R. E., Fox, N. A., and Ferrari, P. F. (2013). The mirror mechanism and mu rhythm in social development. *Neurosci. Lett.* 540, 15–20. doi: 10.1016/j.neulet.2012.10.006
- Varlet, M., Marin, L., Lagarde, J., and Bardy, B. G. (2011). Social postural coordination. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 473–483. doi: 10.1037/a0020552
- Varlet, M., Marin, L., Raffard, S., Schmidt, R. C., Capdevielle, D., Boulenger, J.-P., et al. (2012). Impairments of social motor coordination in schizophrenia. *PLoS One* 7:e29772. doi: 10.1371/journal.pone.0029772
- Veluvolu, K. C., Wang, Y., and Kavuri, S. S. (2012). Adaptive estimation of EEG-rhythms for optimal band identification in BCI. *J. Neurosci. Methods* 203, 163–172. doi: 10.1016/j.jneumeth.2011.08.035
- Volkmar, F. R. (2011). Understanding the social brain in autism. *Dev. Psychobiol.* 53, 428–434. doi: 10.1002/dev.20556
- Willis, M. L., Palermo, R., Burke, D., McGrillen, K., and Miller, L. (2010). Orbitofrontal cortex lesions result in abnormal social judgements to emotional faces. *Neuropsychologia* 48, 2182–2187. doi: 10.1016/j.neuropsychologia.2010.04.010
- Wimmers, R. H., Beek, P. J., and van Wieringen, P. C. (1992). Phase transitions in rhythmic tracking movements: a case of unilateral coupling. *Human Movement Science* 11, 217–226. doi: 10.1016/0167-9457(92)90062-G
- Woodruff, C. C., Martin, T., and Bilyk, N. (2011). Differences in self-and other-induced mu suppression are correlated with empathic abilities. *Brain Res.* 1405, 69–76. doi: 10.1016/j.brainres.2011.05.046
- Worden, M. S., Foxe, J. J., Wang, N., and Simpson, G. V. (2000). Anticipatory biasing of visuospatial attention indexed by retinotopically specific alpha-band electroencephalography increases over occipital cortex. *J. Neurosci.* 20:RC63.
- Yeates, K. O., Bigler, E. D., Dennis, M., Gerhardt, C. A., Rubin, K. H., Stancin, T., et al. (2007). Social outcomes in childhood brain disorder: a heuristic integration of social neuroscience and developmental psychology. *Psychol. Bull.* 133, 535–556. doi: 10.1037/0033-2909.133.3.535

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Tognoli and Kelso. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Functional Role of Neural Oscillations in Non-Verbal Emotional Communication

Ashley E. Symons^{1*}, Wael El-Deredy^{1,2}, Michael Schwartze^{3,4} and Sonja A. Kotz^{1,3,4}

¹ School of Psychological Sciences, University of Manchester, Manchester, UK, ² School of Biomedical Engineering, Universidad de Valparaíso, Valparaíso, Chile, ³ Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, ⁴ Faculty of Psychology and Neuroscience, Department of Neuropsychology and Psychopharmacology, Maastricht University, Maastricht, Netherlands

Effective interpersonal communication depends on the ability to perceive and interpret nonverbal emotional expressions from multiple sensory modalities. Current theoretical models propose that visual and auditory emotion perception involves a network of brain regions including the primary sensory cortices, the superior temporal sulcus (STS), and orbitofrontal cortex (OFC). However, relatively little is known about how the dynamic interplay between these regions gives rise to the perception of emotions. In recent years, there has been increasing recognition of the importance of neural oscillations in mediating neural communication within and between functional neural networks. Here we review studies investigating changes in oscillatory activity during the perception of visual, auditory, and audiovisual emotional expressions, and aim to characterize the functional role of neural oscillations in nonverbal emotion perception. Findings from the reviewed literature suggest that theta band oscillations most consistently differentiate between emotional and neutral expressions. While early theta synchronization appears to reflect the initial encoding of emotionally salient sensory information, later fronto-central theta synchronization may reflect the further integration of sensory information with internal representations. Additionally, gamma synchronization reflects facilitated sensory binding of emotional expressions within regions such as the OFC, STS, and, potentially, the amygdala. However, the evidence is more ambiguous when it comes to the role of oscillations within the alpha and beta frequencies, which vary as a function of modality (or modalities), presence or absence of predictive information, and attentional or task demands. Thus, the synchronization of neural oscillations within specific frequency bands mediates the rapid detection, integration, and evaluation of emotional expressions. Moreover, the functional coupling of oscillatory activity across multiples frequency bands supports a predictive coding model of multisensory emotion perception in which emotional facial and body expressions facilitate the processing of emotional vocalizations.

OPEN ACCESS

Edited by:

Anne Keitel,
University of Glasgow, UK

Reviewed by:

Didier Grandjean,
University of Geneva, Switzerland
Bahar Güntekin,
Istanbul Kultur University, Turkey

*Correspondence:

Ashley E. Symons
ashley.symons@
manchester.ac.uk

Received: 11 January 2016

Accepted: 09 May 2016

Published: 25 May 2016

Citation:

Symons AE, El-Deredy W, Schwartze M and Kotz SA (2016) The Functional Role of Neural Oscillations in Non-Verbal Emotional Communication. *Front. Hum. Neurosci.* 10:239.
doi: 10.3389/fnhum.2016.00239

Keywords: emotion, nonverbal communication, multisensory, cross-modal prediction, neural oscillations

INTRODUCTION

Effective communication is crucial for the formation and maintenance of social relationships in complex societies. Emotional communication is a complex process where the expression and perception of emotional signals exchanges information about internal affective states. While some of these signals can be expressed through verbal means, much of our emotional communication occurs nonverbally through changes in facial, body, and vocal expressions. Therefore, our ability to perceive and interpret nonverbal expressions of emotion can have a profound impact on the quality of our social interactions, affecting our mental health and wellbeing. To this end, deficits in emotion perception have been observed in a number of neurological and psychiatric conditions (Phillips et al., 2003; Garrido-Vásquez et al., 2011) and may negatively correlate with subjective quality of life in a number of these conditions (i.e., Phillips et al., 2010, 2011; Fulford et al., 2014). Despite this importance, the neural mechanisms and dynamics underpinning the perception of emotional cues within and between sensory modalities is poorly understood. This review explores the functional role of neural oscillations in mediating neural communication within and between sensory modalities in order to facilitate the detection, integration, and evaluation of emotional expressions.

Emotions are commonly defined as brief, coordinated neural, physiological, and behavioral responses to relevant events (Scherer, 2000). These responses can manifest behaviorally as changes in facial expression, body language, tone of voice (prosody), or any combination thereof. Thus, emotion perception can be described as the process of detecting salient signals, integrating those signals with prior knowledge of emotional meaning, and evaluating the integrated representation within the context of the current environment. According to current models, emotion perception of visual (Adolphs, 2002a; De Gelder, 2006), auditory (Schirmer and Kotz, 2006; Wildgruber et al., 2009; Kotz and Paulmann, 2011), and audiovisual (Brück et al., 2011) signals unfolds in three fast yet distinct stages: detection, integration, and evaluation.

The first stage consists of early perceptual processing in what are traditionally considered modality-specific cortices. For visual expressions of emotion, this includes regions of the occipito-temporal cortex, most notably the fusiform gyrus (Adolphs, 2002a,b; Vuilleumier and Pourtois, 2007) with distinct subregions of the fusiform gyrus responding preferentially to facial and body expressions (Schwarzlose et al., 2005), and extratectal body area (Grèzes et al., 2007; Kret et al., 2011; Meeren et al., 2013). Early detection of complex acoustic cues occurs in the belt region of the primary auditory cortex (Woods and Alain, 2009) and later in multiple voice-sensitive areas in the temporal lobe (Belin et al., 2004; Wiethoff et al., 2008; Ehofer et al., 2009, 2011; Pernet et al., 2015).

Following the extraction of low-level visual and acoustic features, a more detailed representation of the emotional expression is generated in the superior temporal sulcus (STS). Evidence from neuroimaging research suggests a functional

subdivision within the STS with face-sensitive regions in the posterior terminal ascending branch and voice-sensitive regions in the trunk section (Kreifels et al., 2009). Functional differentiation between middle and anterior regions of the STS has also been noted during the perception of emotional vocal expressions (Kotz and Paulmann, 2011). Receiving input from both visual and auditory cortices, the STS also plays a key role in audiovisual integration (i.e., Calvert et al., 2000, 2001; Beauchamp et al., 2004a,b; Stevenson et al., 2007; Stevenson and James, 2009). To this end, facial and vocal expressions activate overlapping face- and voice-sensitive regions within the STS, suggesting that the STS is essential for the integration of audiovisual emotional information (Robins et al., 2009; Watson et al., 2013, 2014).

In the final stage, the behavioral and motivational significance of the expression is interpreted and evaluated within the inferior frontal gyrus (IFG; Frühholz and Grandjean, 2013a) and orbitofrontal cortex (OFC; Adolphs, 2002b; Kotz et al., 2012). Involved in the processing of reward and punishment (Kringelbach and Rolls, 2004; Rolls, 2004), the OFC is thought to be involved in the representation of stimulus value across sensory modalities. Thus, during emotion perception, the OFC may be responsible for evaluating the emotional value of the expression within the context of the current environment.

In addition to these cortical regions, many studies also support a key role for subcortical structures such as the amygdala and basal ganglia in the perception of emotions. For example, the amygdala has been implicated in the processing of facial (i.e., Phillips et al., 1997; Blair et al., 1999; Whalen et al., 2001, 2013; Williams et al., 2004), body (Hadjikhani and de Gelder, 2003; Grèzes et al., 2007), and vocal (Fecteau et al., 2007; Frühholz and Grandjean, 2013b) expressions in both early and late stages of emotion perception. Studies also support a key role for the basal ganglia in the processing of facial (Adolphs, 2002b) and vocal (Kotz et al., 2003) expressions. Furthermore, deep brain stimulation of the basal ganglia, specifically the subthalamic nucleus, can impair emotion perception from facial and vocal expressions (Péron et al., 2010a,b). Given the importance of the basal ganglia in other aspects of emotion processing (i.e., subjective feeling and production of emotional expressions), it has been proposed that the basal ganglia coordinates the synchronization of different components of emotion processing (Péron et al., 2013). While the amygdala appears to be involved in both the early and late stages of emotion perception, consistent with dual-pathway models of emotion processing (LeDoux, 1996), basal ganglia activity is more often observed in the later stages as a function of attention (Kotz et al., 2012).

In sum, the perception of emotion from facial, body and vocal expressions involves a distributed neural network of cortical and subcortical structures. The question then becomes, how does the brain selectively attend to and integrate these signals across space and time in order to give rise to a unified representation of an emotional expression?

The investigation of such rapid online processing of dynamic changes in sensory input requires adequate methods

to capture neural information processing in real time. Electroencephalography (EEG) and magnetoencephalography (MEG) are particularly well suited for the study of emotion perception due to their millisecond temporal resolution.

Results from event-related potential (ERPs) suggest differentiation between emotional and neutral facial expressions within 120 ms of stimulus onset (Eimer and Holmes, 2002; Eimer et al., 2003). The time course of emotion-related effects on evoked responses to vocal expressions depends on the stimulus type with earlier effects for affective bursts such as laughs and screams (Liu et al., 2012) compared to changes in emotion prosody (Paulmann and Kotz, 2008; Paulmann et al., 2012, 2013; Pell et al., 2015). These early ERP effects are thought to reflect rapid detection of salience. Visual (Stefanics et al., 2012) and auditory (Schirmer et al., 2005) deviance detection, in the form of the mismatch negativity (MMN) is observed at approximately 200 ms, supporting the idea that integration of emotional signals occurs at this stage. Across domains, emotional expressions also elicit a sustained positivity (the late positive component or LPC) beginning between 300–400 ms post-stimulus, reflecting the interpretation and evaluation of emotional significance (visual, Eimer and Holmes, 2007; auditory, Paulmann et al., 2013).

Taken together, findings from ERP studies largely support the staged models of visual and auditory emotion perception and further establish a time course for the stages of detection, integration, and evaluation of emotional expressions. While studies of ERPs have undoubtedly advanced our understanding of the time course and neural bases of emotion perception, they can only provide limited insight into the dynamic interaction within and between nodes of functional neural networks. That is, we know substantially more about when and where certain processes may occur, than about *how* these processes arise and unfold within the human brain.

Neural oscillations, which reflect rhythmic fluctuations in the synchronization of neuronal populations, provide a measure of the dynamic interactions within and between regions involved in the different stages of emotion perception. Changes in oscillatory activity are commonly analyzed by treating the on-going EEG (or MEG) signal as the sum of pure sinusoids, which are separated into characteristic frequency bands each associated with distinct cognitive and computational operations. Decomposing the EEG/MEG signal into its constituent sinusoids, allows for the measurement of changes in power (amplitude) and phase within and between each frequency band, at different time points and in different brain regions. While increases or decreases in power—referred to as event-related synchronization (ERS) or desynchronization (ERD) respectively—indicate changes in neural synchronization within a specific node or region, phase coherence across brain regions reflects synchrony between brain regions that make up a functional neural network (Bastiaansen et al., 2012). According to one hypothesis, phase coherence enables the effective communication between neuronal populations (Fries, 2005). Moreover, cross-frequency coupling may facilitate the integration of information across different spatial and temporal

scales (Canolty and Knight, 2010). Thus, neural oscillations can provide an index of the dynamic interaction between brain regions involved in emotion perception as well as a plausible mechanism by which the brain can integrate rapidly changing emotional information from facial, body, and vocal expressions. To date, the majority of studies investigating emotion perception have focused solely on power changes. Thus, this review will primarily focus on ERS and ERD, while noting the critical importance of phase coherence and cross-frequency coupling in elucidating the functional dynamics of emotion perception within and between sensory modalities.

PERCEPTION OF EMOTION FROM FACIAL EXPRESSIONS

Facial expressions are by far the most commonly studied means of emotional communication. In a typical study, participants are presented with images of facial expressions and asked to respond to the emotion (explicit) or identity/gender (implicit) of the face. Using this type of paradigm, studies have found changes in oscillatory activity across multiple frequency bands during the perception of emotion from facial expressions.

Delta

Delta oscillations have been implicated in a wide range of processes including the perception of faces and facial expressions (Knyazev, 2012; Güntekin and Başar, 2015). While frontal delta synchronization is characteristic of many more “cognitive” tasks, face processing is associated with delta synchronization over more posterior regions (Güntekin and Başar, 2009). Moreover, emotional expressions appear to induce stronger delta synchronization than neutral expressions over occipito-parietal regions, which is suggested to reflect stimulus updating (Balconi and Lucchiari, 2006; Balconi and Pozzoli, 2007, 2008, 2009). Effects of emotion on delta oscillations have also been observed over fronto-central regions, correlating with behavioral measures emotional involvement (Knyazev et al., 2009b). Of note here is that the studies observing occipito-parietal delta synchronization have typically used passive viewing paradigms while Knyazev et al. (2009a) used both implicit (gender identification) and explicit (emotion categorization) tasks. Thus, emotion may differentially affect delta responses to facial expressions depending on task demands. Together these findings suggest a role for delta oscillations in the perception of emotional facial expressions; yet the functional significance of delta synchronization in this context remains unclear. Further research is needed in order to determine more precisely the functional role of delta oscillations within the context of emotion perception.

Theta (4–7 Hz)

Most commonly associated with memory encoding and retrieval (Klimesch, 1999), theta band oscillations are thought to play a key role in the processing of emotion (Knyazev, 2007). To this end, recent studies have shown enhanced theta synchronization for emotional compared to neutral facial

expressions, suggesting that theta oscillations may facilitate the rapid encoding of emotionally salient sensory information. For instance, Balconi and colleagues have observed enhanced theta synchronization over predominantly right frontal regions of the scalp between 150–250 ms extending into the later time window of 250–350 ms which they suggest reflects the orienting of attention toward the emotional significance of the stimulus during the early stages of conceptual processing (Balconi and Lucchiari, 2006; Balconi and Pozzoli, 2009). Similar results were reported by Knyazev et al. (2009b, 2010), who found increased early theta ERS over right frontal regions during the implicit processing of emotional facial expressions, that is, when participants performed a gender categorization task in which attention was directed away from the emotional content of the stimulus. Furthermore, these authors observed a second distinct theta ERS between 230–350 ms that was greater when the emotional content of the stimulus was processed explicitly during an emotion categorization task. Source localization revealed differential activation in the right parietal cortex (angry) and insula (happy) in the early time window and left temporal lobe (angry) and bilateral PFC (happy) in the later time window. Interestingly, some studies have also observed theta synchronization over more posterior (occipital and occipito-parietal) regions within a similar (early) time window (i.e., Başar et al., 2006; Balconi and Pozzoli, 2007, 2008), an effect that increases as a function of visual awareness (Zhang et al., 2012). However, the extent to which this theta synchronization is emotion-specific may be called into question on the basis of a study by González-Roldán et al. (2011) showing no effect of emotion on theta synchronization during an explicit task. Instead, the authors observed a main effect of intensity on theta ERS between 200–400 ms over frontal, central, and parietal regions. This may suggest that theta synchronization in response to emotional facial expressions may reflect facilitated encoding of the biological or motivational significance rather than the emotional quality of the expression *per se*. That is, emotional expressions (relative to neutral expressions) contain more behaviorally relevant sensory information, which reduces uncertainty, resulting in stronger neural synchronization in the theta frequency. This enhanced theta synchronization facilitates the dynamic between brain regions involved in the early detection and integration of static emotional facial expressions.

Alpha (8–12 Hz)

As first noted by Berger (1929), neural oscillations in the alpha frequency band show strong synchronization over occipital regions in the absence of visual stimulation (i.e., with eyes closed). Based on further evidence showing alpha ERS over cortical regions not necessary for a given task, alpha synchronization was initially taken as an indicator of cortical idling (Pfurtscheller et al., 1996). However, more recent hypotheses suggest that alpha synchronization serves an active role in the inhibition of task-irrelevant brain regions (Klimesch et al., 2007; Jensen and Mazaheri, 2010). The rhythmic fluctuation of alpha oscillations thus produces temporal windows in which neurons are more

or less likely to fire. Larger amplitudes (reflecting stronger inhibition) result in smaller temporal windows and thus more precise timing of neuronal firing. Smaller amplitudes, associated with release of inhibition, result in greater cortical excitability over longer temporal intervals. Within the context of emotion perception, alpha oscillations may be involved in the selective attention to emotionally salient social cues through active inhibition of task-irrelevant regions and pathways. It is notable, however, that many studies using static faces have found no difference in alpha synchronization between emotional and neutral expressions (Balconi and Lucchiari, 2006; Balconi and Pozzoli, 2007, 2008, 2009). Differences in alpha power emerge more reliably when comparing expressions of positive and negative valence. While perception of negative emotional expressions was associated with right-lateralized alpha ERD, perception of positive emotional expressions was associated with left-lateralized alpha ERD (Balconi and Ferrari, 2012). Although greater when facial expressions were presented supraluminally, these valence-specific differences were also observed when expressions were presented subliminally. Further support for these findings comes from a study by Del Zotto et al. (2013) showing valence-specific lateralization of frontal alpha power in a patient with cortical blindness. Results from this study showed that alpha ERD was greatest for fear compared to happy expressions over right frontal regions even though the patient could not report seeing the stimuli. Other evidence suggests that alpha synchronization over posterior regions may also differentiate between stimuli of negative and positive valence (Başar et al., 2006), though this effect was only observed when selecting the stimuli with the most extreme valence ratings for analysis.

While studies using static stimuli highlight the roles of valence in alpha responses to facial expressions, they have two important limitations. Firstly, in naturalistic human communication, facial expressions are inherently dynamic and therefore the extent to which these findings would be valid in naturalistic settings is unclear. Secondly, although providing a rough estimate as to the topographical distribution of alpha ERD, these studies only provide limited insight into the patterns of functional connectivity underpinning the perception of emotion from facial expressions. Addressing these issues, a recent MEG study used dynamic facial expressions to explore changes in spatial connectivity during emotion perception (Popov et al., 2013). Findings from this study provide evidence for two stages of upper alpha desynchronization during facial emotion perception: a pre-recognition stage associated with increased alpha power over frontal and sensorimotor regions and decreased alpha power over occipital regions followed by a post-recognition stage associated with the reversed pattern. Moreover, these power changes were associated with inverse patterns of functional connectivity, suggesting that alpha synchronization and desynchronization may regulate the exchange of information between visual and sensorimotor. That these effects were stronger in response to emotional compared to neutral expressions implies that emotion may enhance the functional

coupling, facilitating recognition of facial expressions of emotion.

Beta (13–30 Hz)

Oscillatory activity in the beta frequency is typically associated with sensorimotor processing (Brovelli et al., 2004). However, recent evidence suggests a broader role for beta synchronization in the maintenance of current sensory, motor, and cognitive sets (Engel and Fries, 2010). Beta band oscillations have also been implicated in the perception of emotion from facial expressions. However, the direction, time course, and topography of beta modulation vary considerably between studies. For example, Güntekin and Başar (2007a) found increased beta power for angry compared to happy expressions over frontal and central regions. In a similar study including occipital electrodes, however, the authors found no main effect of emotion on beta band activity (Güntekin and Başar, 2007b). Thus, it seems that only fronto-central beta synchronization reflects differentiation between emotional and neutral facial expressions. Other evidence suggests that such differences in beta synchronization may also be modulated by attention. To this end, asymmetry in resting-state parietal beta band activity has been negatively correlated with attentional bias towards angry facial expressions (Schutter et al., 2001).

Given the importance of beta oscillations in the perception of biological motion, which is thought to be critical for social cognition in naturalistic environments (Pavlova, 2012). Jabbi et al. (2015) used MEG to compare evoked beta band activity in response to dynamic and static facial expressions. Perhaps unsurprisingly, greater beta power was observed for dynamic compared to static facial expressions in occipital, superior temporal and sensorimotor cortices. When comparing dynamic emotional to neutral expressions, the authors found stronger beta power in regions such as the amygdala, STS, and OFC. Furthermore, beta power in the left STS was negatively correlated with the time course of fearful facial expressions but positively correlated with the time course of happy facial expressions. These emotion-specific differences suggest that the observed changes in beta power were not solely due to the processing of biological motion. Although this study investigates evoked rather than induced oscillatory activity, its findings support a putative role for beta oscillations—particularly within the STS—in tracking the temporal dynamics of facial expressions of emotion.

Gamma (>30 Hz)

Reflecting neuronal communication on a more local scale, gamma oscillations have been implicated in a number of cognitive processes including feature integration (Singer and Gray, 1995) and sensory selection (Fries et al., 2002). Within the context of emotion perception, event-related gamma synchronization has been commonly used to explore the functional dynamics underpinning the conscious and unconscious processing of emotional facial expressions. Studies investigating the spatial and temporal dynamics of emotion perception support a dual-pathway model of emotion perception consisting of a cortical and subcortical pathway (i.e., LeDoux,

1996). Accordingly, in an MEG study, Luo et al. (2007, 2009) have reported that fearful expressions elicit early gamma band activity in the amygdala followed by later responses in the occipital, parietal, and prefrontal cortices. These authors have also observed a later attention-dependent gamma response localized to the amygdala, presumably due to feedback from prefrontal regions (Luo et al., 2010). However, these studies, as with any EEG or MEG study reporting activation from deep, subcortical structures, should be considered respect to the current limitations in source analysis techniques. Although greater for supraluminally-presented facial expressions, gamma synchronization is also observed in response to facial expressions processed subliminally (Balconi and Lucchiari, 2008; Luo et al., 2009), suggesting that gamma synchronization can be influenced by emotion even in the absence of visual awareness.

These findings are supported by intracranial studies showing localized gamma synchronization in brain regions implicated in emotion processing—most notably, the amygdala and OFC. Recording intracranial field potentials from the amygdala of pre-surgical epileptic patients, Sato et al. (2011) found increased gamma synchronization in the amygdala for fearful compared to neutral facial expressions. The early time course of gamma synchronization (50–150 ms) supports the presence of a subcortical pathway involved in the rapid detection of emotionally salient facial features. Gamma synchronization has also been observed over prefrontal cortices during the later stage of emotional face perception. Consistent with findings from functional neuroimaging studies demonstrating functional subdivisions between medial and lateral regions of the OFC (i.e., Kringelbach and Rolls, 2004), gamma responses in the lateral OFC are greater in response to negative emotions (Jung et al., 2011). However, this effect only occurred when attention was explicitly directed to the emotional quality of the expressions. Thus, during an implicit processing task, no responses in the lateral OFC were observed. Moreover, Jung et al. (2011) observed increased gamma band activity in the medial OFC only in response to target stimuli, regardless of emotional valence. These findings suggest that the medial-lateral distinction between sub-regions of the OFC cannot be explained simply in terms of valence but may instead reflect the processing of relative value within the context of the current environment. Recent studies have also observed differential effects of attention on gamma band activity in a network of brain regions the amygdala and OFC during the perception of emotional facial expressions (Müschat et al., 2014). Thus, gamma synchronization in the OFC may reflect the attention-dependent binding of emotionally salient stimuli with internal representations of their motivational significance.

Summary

Taken together, the current evidence supports the idea that the perception of emotional facial expressions is mediated by the synchronization of neural oscillations across multiple frequency bands (Güntekin and Başar, 2014). Overall, it appears that lower frequency bands may coordinate patterns of long-range connectivity necessary for the encoding and selection of emotionally salient facial features

while higher frequency bands may be associated with the integration of these features at multiple stages of emotion processing.

PERCEPTION OF EMOTION FROM VOCAL EXPRESSIONS

Within the auditory domain, emotion can be communicated via affective bursts (laughs, screams, cries, etc.) or more subtle changes in tone of voice, or emotion prosody. While both convey important affective information, perception of emotion from these two types of vocal expressions occurs along different time scales and may rely on different patterns of neural activity and connectivity. Although very few studies have investigated the role of neural oscillations in perception of emotion from either type of vocal expression, current evidence suggests that theta synchronization may play a particularly important role in facilitating the detection of emotionally salient vocal cues.

Detection of Prosodic Change

A considerable body of research suggests that theta band oscillations drive the processing of slow acoustic changes in speech perception (Peelle and Davis, 2012). To explore the role of oscillatory activity in the detection of emotional prosodic change, Chen et al. (2012) used a cross-splicing procedure to artificially combine vocalizations spoken in angry and neutral prosodies. Thus, vocalizations could change from neutral to angry, angry to neutral, or remain constant. In this paradigm, detection of prosodic change was associated with an increase in fronto-central theta synchronization between 100–600 ms. Furthermore, for angry prosodies only, theta synchronization was modulated by intensity with greater power for high compared to low intensity vocalizations. Subsequent research by the same group has extended these findings, showing increased theta synchronization for neutral to angry change compared to no change for both implicit and explicit tasks suggesting that the emotional content of the stimulus may facilitate the detection of acoustic change (Chen et al., 2014). In this study, significant beta desynchronization was also observed between 400–750 ms, but only when the task required explicit processing of emotional change, which the authors interpret as re-integration of the cross-spliced portion of the sentence with its preceding context. Although these findings provide preliminary support for the role of theta synchronization and beta desynchronization in the detection of emotion prosody, the precise temporal and spatial dynamics of these effects needs to be addressed in order to provide a better characterization of the function of these frequency bands in vocal emotion perception.

Oscillatory Response to Affective Bursts

With regards to affective bursts, what little evidence there is suggests that gender differences may also influence theta band activity. In a study by Bekkedaal et al. (2011), the authors found no main effect of emotion on frontal theta synchronization. Instead, they found an interaction between

emotion and gender such that women showed increased theta synchronization for angry expressions over bilateral anterior regions while men showed increased theta synchronization for expressions of pleasure over right anterior regions. As noted by the authors, this gender difference in theta synchronization may be due to differences in arousal, although behavioral measures would certainly be needed to support this claim. Moreover, the wide time intervals used for analysis (500 ms) make the functional interpretation of these gender differences in theta synchronization difficult and may partially account for the absence of any statistically significant differences in other frequency bands.

Summary

Though few in number, the existing studies suggest that theta synchronization may facilitate the perception of emotion from vocal expressions. Consistent with findings from the speech literature, theta synchronization appears to mediate the detection of acoustic change, an effect which is modulated by emotion. Additionally, beta desynchronization may also play a role in vocal emotion perception, but only when explicitly attending to the change in prosody. Thus, theta synchronization may be involved in the detection of emotionally significant acoustic features during vocal emotion perception while beta desynchronization may facilitate the integration of these features with contextual information.

INTEGRATION OF FACIAL, BODY, AND VOCAL EXPRESSIONS OF EMOTION

In natural environments, emotion perception requires the integration of emotional cues from both visual and auditory modalities. Based on current models of visual and auditory emotion perception, it could be hypothesized that multisensory emotional expressions are integrated in a convergent manner such that visual and auditory cues are processed separately in modality-specific cortices, integrated into a coherent multisensory percept the STS, and evaluated in the PFC. However, it is important to note that facial and vocal expressions occur along different temporal scales with changes in facial expression often preceding changes in vocal expressions. Therefore, based on dynamic changes in facial and body expressions, the brain can generate predictions about the timing and content of forthcoming vocal expressions. Evidence from ERP studies suggests that emotional facial expressions elicit stronger (i.e., more reliable) predictions than neutral expressions (Jessen and Kotz, 2011; Jessen et al., 2012; Ho et al., 2015; Kokinous et al., 2015), resulting in facilitated processing of predicted emotional vocalizations. Together with recent proposals suggesting that neural oscillations play an important role in multisensory processing (Schroeder et al., 2008; Senkowski et al., 2008; Arnal and Giraud, 2012), this suggests that neural synchronization may facilitate the processing of multisensory emotional expressions through: (i) the selective binding of emotionally-salient sensory input from different modalities; and (ii) the formation and modification of sensory predictions.

Multisensory Integration of Facial and Vocal Expressions

Many earlier studies of multisensory emotion perception relied on the use of static facial expressions paired with words or phrases spoken in emotional or neutral prosody. In one such study, Chen et al. (2010) sought to determine whether multisensory integration effects could be observed in the primary sensory cortices during emotional face-voice processing. Using MEG, the authors recorded changes in oscillatory activity during visual, auditory, and audiovisual processing of emotional expressions. However, no integration effects were observed in either visual or auditory cortices. While this finding is interpreted as absence of audiovisual integration in primary sensory cortices, it could also be explained by the absence of predictive visual information since visual and auditory cues were presented simultaneously (see Vroomen and Stekelenburg, 2010). Interestingly, however, the authors observed alpha synchronization over superior frontal and cingulate cortices, which may suggest that increasing the amount of information available to the sensory systems via multiple modalities reduces the cognitive demand on prefrontal regions (Schelenz et al., 2013). Other studies using static facial expressions have found cross-modal interactions in other frequency bands and brain regions. For instance, by presenting participants with static fearful and neutral facial expressions paired with congruent vocal expressions, Hagan et al. (2009) demonstrated supra-additive increases in oscillatory activity in the STS, with theta and gamma bands contributing most to the increase in broadband activity. Subsequent research by the same group showed that supra-additive increases in the STS occurred in both congruent and incongruent conditions (albeit later in the incongruent condition), suggesting automatic integration of emotional facial and vocal expressions (Hagan et al., 2013). Consistent with these findings, other studies have observed theta synchronization during the integration of facial and prosodic change (Chen et al., 2015). Together, these findings suggest that oscillatory activity in the alpha and theta frequency bands drive the integration of facial and vocal expressions. Thus, without predictive visual information, theta synchronization in the STS may facilitate the feedforward integration of visual and auditory input into a coherent percept, reducing the processing demands on prefrontal regions involved in the interpretation and evaluation of the expression.

Cross-Modal Predictive Coding of Emotional Expressions

Although these studies using static facial expressions have undoubtedly contributed to our understanding of audiovisual integration of emotional expression, their findings could be challenged on the grounds of ecological validity. Therefore, more recent studies have moved towards the use of dynamic facial, body, and vocal expression in order to explore the oscillatory correlates of emotion perception in more naturalistic environments. In among the first to do so, Jessen and Kotz (2011) presented participants with video clips of dynamic

facial, body, and vocal expressions. Using EEG, the authors found significant decreases in both alpha and beta power for audiovisual compared to the sum of auditory- and visual-only conditions with additional suppression for emotional compared to neutral expressions. These findings were replicated in a subsequent study, which also showed that while beta suppression for the contrast between multimodal and unimodal conditions was localized to the premotor cortex, suppression for the contrast between emotional and neutral conditions was localized to the posterior parietal cortex (Jessen et al., 2012). Since previous studies have demonstrated beta suppression in these regions during the processing of biological motion (Muthukumaraswamy et al., 2006; Muthukumaraswamy and Singh, 2008), it could be argued that the observed differences in beta power are due to differences in the motion content between emotional and neutral expressions. However, for the stimuli used in these studies, there was no difference in the motion content before the onset of the vocal expression (see Jessen and Kotz, 2011) making it unlikely that beta suppression was an artifact of differences in motion content. Instead, beta oscillations may play a broader role in the predictive coding of audiovisual information (i.e., Arnal and Giraud, 2012). Furthermore, the observed differences in beta ERD between emotional and neutral expressions provide support for the hypothesis that emotional expressions generate stronger cross-modal predictions compared to neutral expressions (Jessen and Kotz, 2013).

Summary

Taken together, these studies support previous research suggesting that neural oscillations play an important role in multisensory processing. Furthermore, these findings show that the emotional content of the stimulus may facilitate flexible integration of facial, body, and vocal expressions. The simultaneous presentation of visual and auditory expressions results in synchronization of theta oscillations in the STS (i.e., the STS) and alpha oscillations over prefrontal regions, suggesting that theta synchronization mediates the integration of audiovisual emotional expressions. Previous evidence suggests that multimodal expressions generally are more easily recognizable than unimodal expressions (Collignon et al., 2008; Tanaka et al., 2010; Föcker et al., 2011), frontal alpha synchronization may reflect relative inhibition of regions needed to resolve any remaining uncertainty with regards to the emotional content of the stimulus. Since this effect was observed in emotion categorization tasks, it is possible that different task demands will induce different spatial and temporal patterns of alpha synchronization. In contrast, the natural temporal delay between visual and auditory expressions enables the brain use changes in facial and body expression to generate predictions about the timing and content of forthcoming vocal expressions. Thus, cross-modal prediction results in ERD, particularly in the alpha and beta frequencies. These findings support the idea that multisensory integration and cross-modal prediction are distinct yet interactive mechanisms underpinning the multisensory emotion perception (Jessen and Kotz, 2013, 2015).

DISCUSSION

Nonverbal emotion perception is driven by dynamic, context-dependent interactions within and between brain regions involved in the detection, integration, and evaluation of emotional expressions. Where and when such interactions occur depends on the sensory modality (or modalities) through which the emotion is expressed as well as the emotional quality of the stimulus itself. However, emotional expressions are dynamic events that continuously evolve over time. Therefore, the neural system(s) supporting emotion perception must be able to flexibly adapt to and integrate rapidly changing sensory input from multiple modalities. Based on the reviewed evidence, we propose that neural synchronization underpins the selective attention to and the flexible binding of emotionally salient sensory input across different spatial and temporal scales. Furthermore, neural oscillations provide a mechanism through which emotional facial and body expressions can predictively modulate the processing of subsequent vocal expressions.

The recognition of an expression as “emotional” requires the selective binding of emotionally relevant sensory information. However, individual features of an emotional expression can occur at different points in time and are processed in spatially distinct regions of the brain. Thus, the brain is challenged with the task of binding only those features belonging to the same event across both space and time. One mechanism through which this may occur is the synchronization of neural oscillations, which creates temporal windows in which information belonging to the same event can be selected and integrated (Singer and Gray, 1995). Moreover, coherence between distinct neuronal populations may enable the flexible neuronal communication across different regions of the brain (Fries, 2005). Consistent with this idea, current evidence suggests that the synchronization of neural oscillations supports the selection and integration of sensory information within and between modalities (Senkowski et al., 2008; van Atteveldt et al., 2014). Gamma band oscillations, in particular, are thought to be important for sensory binding and feature integration on a local scale (Tallon-Baudry and Bertrand, 1999). As previously discussed, perception of emotion from facial expressions results in increased gamma band synchronization, suggesting that gamma band oscillations may mediate the rapid integration of emotionally salient sensory input. However, gamma band synchronization may be modulated by lower-frequency oscillations. Since lower frequency bands represent the activity of larger neuronal populations and longer temporal windows, such cross-frequency coupling between low and high frequency oscillations may enable the integration of information across different spatial and temporal scales (Canolty and Knight, 2010).

Natural communicative signals exhibit strong regularities that enable the brain to generate predictions about forthcoming sensory information within and between sensory modalities. This process may be mediated by the functional coupling of neural oscillations, which can facilitate the efficient allocation of processing resources to the predicted sensory input. For instance, synchronization of low-frequency oscillations may

coordinate the allocation of processing resources, via high-frequency oscillations, at the phase in which the predicted sensory input occurs (Hyafil et al., 2015). As an example, the natural temporal delay between visual and acoustic speech signals provides a means through which the visual signal can alter the phase of ongoing neural oscillations such that the expected acoustic signal occurs at the phase of optimal neuronal excitability (Schroeder et al., 2008). While the phase of low-frequency oscillations may create temporal windows for the selection of relevant sensory information, higher-frequency beta and gamma oscillations may be involved in the transmission of top-down predictions (both formal and temporal) and bottom-up prediction errors, respectively (Arnal et al., 2011; Arnal and Giraud, 2012). If this is indeed the case, then it follows that neural oscillations, particularly within these frequency bands, may facilitate the predictive coding of nonverbal communicative signals such as dynamic facial, body, and vocal expressions. In this respect, emotion perception is similar to other forms of perception, with emotion acting as a highly salient source of relevant information that must be encoded and integrated with other sources of sensory information.

FUTURE DIRECTIONS

Effect of Modality

Although early on, Charles Darwin recognized the equal importance of facial, body, and vocal expressions in emotional communication, research over the past 50 years has focused predominantly on the perception of emotion from facial expressions. Thus, the role of neural oscillations in emotion perception has primarily been studied by presenting participants with images of static facial expressions. While this approach has yielded some valuable results, it does not necessarily reflect how emotions are expressed and perceived in natural human communication.

In everyday life, emotional expressions are dynamic, characterized by changes in facial expression, body language, and prosody unfolding over time. To this end, previous functional neuroimaging research has shown distinct neural pathways involved in the perception of emotion from static and dynamic facial expressions (i.e., Kilts et al., 2003). Consistent with these findings, results from Jabbi et al. (2015) suggest that oscillatory activity in the beta frequency band may track dynamic changes in sensory input facilitating the differentiation of emotional expressions. Although the use of dynamic facial expressions adds an additional level of stimulus complexity, it also affords greater ecological validity, which can improve our understanding of the neural dynamics underpinning naturalistic emotion perception. Moreover, the dynamic nature of emotional expressions enables the brain to use incoming sensory input to generate predictions about future events. Future studies using methods such as dynamic causal modeling (DCM) can be used to compare convergent and predictive coding models of multisensory emotion perception.

A second issue relates to the fact that facial expressions are also not the only means of emotional communication. Changes in emotional body language (De Gelder, 2006) and prosody (Schirmer and Kotz, 2006) also provide important information about one's emotional state. Compared to facial expressions, however, little is known about the oscillatory dynamics underpinning the perception of emotion from body and vocal expressions. Therefore, further research into: (i) the perception of emotion from dynamic body and vocal expressions; and (ii) the integration of emotional expressions from multiple modalities is needed if we are to understand the neural bases of emotion perception in human social interactions.

Emotional Differentiation

Each emotion is associated with a unique physiological, cognitive, and behavioral profile that serves an adaptive and, in social species, a communicative function. Therefore, it is likely that distinct patterns of neural activity and connectivity drive the expression and perception of different emotions.

One of the broadest distinctions between emotions is that of valence, which categorizes emotions as positive (pleasant) or negative (unpleasant). Within the brain, some have proposed that the right hemisphere is dominant for the processing of negative emotions while the left hemisphere is dominant for positive emotions (Ahern and Schwartz, 1979, 1985; Silberman and Weingartner, 1986). Although valence-specific asymmetry has primarily been discussed within the context of emotional experience, studies in healthy individuals and in patients with unilateral brain damage suggest that there may also be hemispheric asymmetry in the perception of emotion (i.e., Jansari et al., 2000; Adolphs et al., 2001), though this may be influenced by task demands (Kotz et al., 2003, 2006). Consistent with this hypothesis, there is preliminary support for valence-specific hemispheric asymmetry of alpha desynchronization during the emotion perception (i.e., Balconi and Ferrari, 2012). However, given support for alternative hypotheses such as the approach-withdrawal model of hemispheric lateralization (Davidson, 1992), future studies examining patterns of coherence across brain regions during the perception of positive and negative emotions are needed in order to elucidate the functional dynamics underpinning the differentiation of emotional valence.

Since each emotion serves a distinct function, it has been hypothesized that there may be different, yet partially overlapping, neural pathways specialized for the processing of different emotions (i.e., LeDoux, 2000). Thus, we may expect specific patterns of neural synchronization during the perception of different emotions. In support of this idea, distinct spatial and temporal patterns of theta (Knyazev et al., 2009b) and gamma (Luo et al., 2007) band activity have been observed in response to different emotions. So although perception of different emotions may rely partially on overlapping networks, further investigations into patterns of neural synchronization and coherence may reveal subtle changes in functional dynamics that enable us to differentiate between emotions.

Individual Differences

Due to the interaction between neurophysiological and environmental factors, individual differences can have a profound effect on how we perceive and interpret nonverbal expressions of emotion. Underlying these individual differences are changes in functional coupling that can be investigated by examining patterns of neural synchrony. To this end, gender differences are reflected in beta (Güntekin and Başar, 2007b) and theta (Knyazev et al., 2010) synchronization in response to emotional facial expressions. Furthermore, alpha desynchronization has been negatively associated with extraversion (Fink, 2005) and hostility (Knyazev et al., 2009b) and positively associated with anxiety (Knyazev et al., 2008) and depression (Knyazev et al., 2015). Individual differences have also been observed in the theta band, with reduced frontal theta synchronization in individuals with high levels of anxiety (Knyazev et al., 2008) and depression (Knyazev et al., 2015) and increased theta synchronization in those scoring high on measures of emotional intelligence (Knyazev et al., 2013). Additionally, hostility has been associated with gender differences in alpha and theta synchronization over posterior regions (Knyazev et al., 2009b) while dominance motivation is associated with delta/beta asymmetry (Hofman et al., 2013). Taken together, these findings suggest that changes in patterns of neural synchronization may mediate individual differences in the perception of emotional expressions.

Clinical Implications

Deficits in the ability to accurately perceive and interpret emotions have been observed in a number of neurological and psychiatric conditions, the neural bases of which remain poorly understood. By enabling us to look beyond the activity of specific brain regions into the dynamics of functional neural networks, investigations into changes in neural synchronization and coherence can advance our understanding of the specific impairments associated with different clinical conditions. Work in this area has already begun with studies showing reduced theta synchronization during perception of emotional facial expressions in individuals with schizophrenia (Ramos-Loyo et al., 2009; Csukly et al., 2014). Schizophrenia has also been associated with abnormal patterns of alpha synchronization (Ramos-Loyo et al., 2009; Popov et al., 2014), though this may be improved through targeted training in facial affect recognition (Popova et al., 2014). Other studies have found that oscillatory responses to facial expressions in the gamma band differentiate between unipolar and bipolar depression; while individuals with unipolar depression show reduced gamma power in response to sad facial expressions, those with bipolar show increased gamma band activity in response to highly arousing emotions (Liu et al., 2014). Finally, adolescents with Autism Spectrum Disorder show reduced interregional beta synchronization in response to angry facial expressions, suggesting that impairments in functional connectivity within networks involved in emotion processing may contribute to the deficits in facial emotion perception observed in autism (Leung et al., 2014). Thus, a better characterization of oscillatory responses to emotional

expressions may aid in the diagnosis and treatment of a number of clinical conditions.

CONCLUSION

From the reviewed studies, it is clear that the perception of facial, body, and vocal expressions of emotion is mediated by oscillatory activity in multiple frequency bands. Although research on delta synchronization has been primarily restricted to the visual domain, the important delta oscillations and their functional coupling with higher (beta/gamma) frequency bands, in basic biological, cognitive, and emotional processes highlights the need for further research into the functional role of delta oscillations in emotion perception within and between sensory modalities. Across modalities, theta synchronization most consistently differentiates between emotional and neutral expressions and may reflect the initial encoding and derivation of emotional significance. Changes in alpha power have been primarily observed in studies with a visual component, with some evidence of valence-specific lateralization over frontal regions. Based on the hypothesis that alpha synchronization reflects active inhibition of task irrelevant brain regions (Klimesch et al., 2007), modulation of alpha power may reflect sensory selection and inhibition of behaviorally relevant sensory information. Although evidence is still inconclusive as to the role of beta

oscillations in emotion perception, changes in beta power are more likely to be observed in studies using dynamic stimuli or in those involving shifts in attention, consistent with the idea that beta band activity reflects the maintenance of the current cognitive or sensorimotor set (Engel and Fries, 2010). Gamma synchronization has been observed in emotion processing regions such as the amygdala, STS, and OFC, suggesting that oscillatory activity in this frequency band is associated with the binding of emotionally salient sensory input. The modulation of specific frequency bands by emotion enables the selective detection, integration, and evaluation of emotional signals through coordinated changes in effective connectivity. From a predictive coding perspective, the emotional quality of the expression may act as a particularly salient source of information, strengthening the precision of sensory predictions through enhanced neural synchronization. However, further research, particularly in the auditory and audiovisual domains, is clearly necessary to gain a deeper understanding of the neural dynamics underpinning the perception of emotion within and between sensory modalities.

AUTHOR CONTRIBUTIONS

Main contribution by first author (AES). All the other authors contributed equally to this work.

REFERENCES

- Adolphs, R. (2002a). Neural systems for recognizing emotion. *Curr. Opin. Neurobiol.* 12, 169–177. doi: 10.1016/s0959-4388(02)00301-x
- Adolphs, R. (2002b). Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behav. Cogn. Neurosci. Rev.* 1, 21–62. doi: 10.1177/15345823020010003
- Adolphs, R., Jansari, A., and Tranel, D. (2001). Hemispheric perception of emotional valence from facial expressions. *Neuropsychology* 15, 516–524. doi: 10.1037/0894-4105.15.4.516
- Ahern, G. L., and Schwartz, G. E. (1979). Differential lateralization for positive versus negative emotion. *Neuropsychologia* 17, 693–698. doi: 10.1016/0028-3932(79)90045-9
- Ahern, G. L., and Schwartz, G. E. (1985). Differential lateralization for positive and negative emotion in the human brain: EEG spectral analysis. *Neuropsychologia* 23, 745–755. doi: 10.1016/0028-3932(85)90081-8
- Arnal, L. H., and Giraud, A. L. (2012). Cortical oscillations and sensory predictions. *Trends Cogn. Sci.* 16, 390–398. doi: 10.1016/j.tics.2012.05.003
- Arnal, L. H., Wyart, V., and Giraud, A. L. (2011). Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nat. Neurosci.* 14, 797–801. doi: 10.1038/nn.2810
- Balconi, M., and Ferrari, C. (2012). Subliminal and supraliminal processing of facial expression of emotions: brain oscillation in the left/right frontal area. *Brain Sci.* 2, 85–100. doi: 10.3390/brainsci2020085
- Balconi, M., and Lucchiari, C. (2006). EEG correlates (event-related desynchronization) of emotional face elaboration: a temporal analysis. *Neurosci. Lett.* 392, 118–123. doi: 10.1016/j.neulet.2005.09.004
- Balconi, M., and Lucchiari, C. (2008). Consciousness and arousal effects on emotional face processing as revealed by brain oscillations. A γ band analysis. *Int. J. Psychophysiol.* 67, 41–46. doi: 10.1016/j.ijpsycho.2007.10.002
- Balconi, M., and Pozzoli, U. (2007). Event-related oscillations (EROs) and event-related potentials (ERPs) comparison in facial expression recognition. *J. Neuropsychol.* 1, 283–294. doi: 10.1348/174866407x184789
- Balconi, M., and Pozzoli, U. (2008). Event-related oscillations (ERO) and event-related potentials (ERP) in emotional face recognition. *Int. J. Neurosci.* 118, 1412–1424. doi: 10.1080/00207450601047119
- Balconi, M., and Pozzoli, U. (2009). Arousal effect on emotional face comprehension. Frequency band changes in different time intervals. *Physiol. Behav.* 97, 455–462. doi: 10.1016/j.physbeh.2009.03.023
- Başar, E., Güntekin, B., and Öñiz, A. (2006). Principles of oscillatory brain dynamics and a treatise of recognition of faces and facial expressions. *Prog. Brain Res.* 159, 43–62. doi: 10.1016/s0079-6123(06)59004-1
- Bastiaansen, M., Mazaheri, A., and Jensen, O. (2012). “Beyond ERPs: oscillatory neuronal dynamics,” in *The Oxford Handbook of Event-Related Potential Components*, eds S. J. Luck, and E. S. Kappenman (New York, NY: Oxford University Press), 31–50.
- Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., and Martin, A. (2004a). Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nat. Neurosci.* 7, 1190–1192. doi: 10.1038/nn1333
- Beauchamp, M. S., Lee, K. E., Argall, B. D., and Martin, A. (2004b). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron* 41, 809–823. doi: 10.1016/s0896-6273(04)00070-4
- Bekkedaal, M. Y. V., Rossi, J. III, and Panksepp, J. (2011). Human brain EEG indices of emotions: delineating responses to affective vocalizations by measuring frontal theta event-related synchronization. *Neurosci. Biobehav. Rev.* 35, 1959–1970. doi: 10.1016/j.neubiorev.2011.05.001
- Belin, P., Fecteau, S., and Bédard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends Cogn. Sci.* 8, 129–135. doi: 10.1016/j.tics.2004.01.008
- Berger, H. (1929). Über das elektrenkephalogramm des menschen. *Eur. Arch. Psychiatry Clin. Neurosci.* 87, 527–570. doi: 10.1007/bf01797193
- Blair, R. J. R., Morris, J. S., Frith, C. D., Perrett, D. I., and Dolan, R. J. (1999). Dissociable neural responses to facial expressions of sadness and anger. *Brain* 122, 883–893. doi: 10.1093/brain/122.5.883
- Brovelli, A., Ding, M., Ledberg, A., Chen, Y., Nakamura, R., and Bressler, S. L. (2004). α oscillations in a large-scale sensorimotor cortical network: directional

- influences revealed by Granger causality. *Proc. Natl. Acad. Sci. U S A* 101, 9849–9854. doi: 10.1073/pnas.0308538101
- Brück, C., Kreifels, B., and Wildgruber, D. (2011). Emotional voices in context: a neurobiological model of multimodal affective information processing. *Phys. Life Rev.* 8, 383–403. doi: 10.1016/j.plrev.2011.10.002
- Calvert, G. A., Campbell, R., and Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr. Biol.* 10, 649–657. doi: 10.1016/s0960-9822(00)00513-3
- Calvert, G. A., Hansen, P. C., Iversen, S. D., and Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage* 14, 427–438. doi: 10.1006/nimg.2001.0812
- Canolty, R. T., and Knight, R. T. (2010). The functional role of cross-frequency coupling. *Trends Cogn. Sci.* 14, 506–515. doi: 10.1016/j.tics.2010.09.001
- Chen, Y. H., Edgar, J. C., Holroyd, T., Dammers, J., Thönenßen, H., Roberts, T. P. L., et al. (2010). Neuromagnetic oscillations to emotional faces and prosody. *Eur. J. Neurosci.* 31, 1818–1827. doi: 10.1111/j.1460-9568.2010.07203.x
- Chen, X., Pan, Z., Wang, P., Yang, X., Liu, P., You, X., et al. (2015). The integration of facial and vocal cues during emotional change perception: EEG markers. *Soc. Cogn. Affect. Neurosci.* doi: 10.1093/scan/nsv083 [Epub ahead of print].
- Chen, X., Pan, Z., Wang, P., Zhang, L., and Yuan, J. (2014). EEG oscillations reflect task effects for the change detection in vocal emotion. *Cogn. Neurodyn.* 9, 351–358. doi: 10.1007/s11571-014-9326-9
- Chen, X., Yang, J., Gan, S., and Yang, Y. (2012). The contribution of sound intensity in vocal emotion perception: behavioral and electrophysiological evidence. *PLoS One* 7:e30278. doi: 10.1371/journal.pone.0030278
- Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., et al. (2008). Audio-visual integration of emotion expression. *Brain Res.* 1242, 126–135. doi: 10.1016/j.brainres.2008.04.023
- Csukly, G., Stefanics, G., Komlósi, S., Czigler, I., and Czobor, P. (2014). Event-related theta synchronization predicts deficit in facial affect recognition in schizophrenia. *J. Abnorm. Psychol.* 123, 178–189. doi: 10.1037/a0035793
- Davidson, R. J. (1992). Anterior cerebral asymmetry and the nature of emotion. *Brain Cogn.* 20, 125–151. doi: 10.1016/0278-2626(92)90065-t
- De Gelder, B. (2006). Towards the neurobiology of emotional body language. *Nat. Rev. Neurosci.* 7, 242–249. doi: 10.1038/nrn1872
- Del Zotto, M., Deiber, M. P., Legrand, L. B., De Gelder, B., and Pegna, A. J. (2013). Emotional expressions modulate low alpha and α oscillations in a cortically blind patient. *Int. J. Psychophysiol.* 90, 358–362. doi: 10.1016/j.ijpsycho.2013.10.007
- Eimer, M., and Holmes, A. (2002). An ERP study on the time course of emotional face processing. *Neuroreport* 13, 427–431. doi: 10.1097/00001756-200203250-00013
- Eimer, M., and Holmes, A. (2007). Event-related brain potential correlates of emotional face processing. *Neuropsychologia* 45, 15–31. doi: 10.1016/j.neuropsychologia.2006.04.022
- Eimer, M., Holmes, A., and McGlone, F. P. (2003). The role of spatial attention in the processing of facial expression: an ERP study of rapid brain responses to six basic emotions. *Cogn. Affect. Behav. Neurosci.* 3, 97–110. doi: 10.3758/cabn.3.2.97
- Engel, A. K., and Fries, P. (2010). β -band oscillations-signalling the status quo? *Curr. Opin. Neurobiol.* 20, 156–165. doi: 10.1016/j.conb.2010.02.015
- Ethofer, T., Bretscher, J., Gschwind, M., Kreifels, B., Wildgruber, D., and Vuilleumier, P. (2011). Emotional voice areas: anatomic location, functional properties and structural connections revealed by combined fMRI/DTI. *Cereb. Cortex* 22, 191–200. doi: 10.1093/cercor/bhr113
- Ethofer, T., Van De Ville, D., Scherer, K., and Vuilleumier, P. (2009). Decoding of emotional information in voice-sensitive cortices. *Curr. Biol.* 19, 1028–1033. doi: 10.1016/j.cub.2009.04.054
- Fecteau, S., Belin, P., Joanette, Y., and Armony, J. L. (2007). Amygdala responses to nonlinguistic emotional vocalizations. *Neuroimage* 36, 480–487. doi: 10.1016/j.neuroimage.2007.02.043
- Fink, A. (2005). Event-related desynchronization in the EEG during emotional and cognitive information processing: differential effects of extraversion. *Biol. Psychol.* 70, 152–160. doi: 10.1016/j.biopsych.2005.01.013
- Föcker, J., Gondan, M., and Röder, B. (2011). Preattentive processing of audio-visual emotional signals. *Acta Psychol. (Amst)* 137, 36–47. doi: 10.1016/j.actpsy.2011.02.004
- Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn. Sci.* 9, 474–480. doi: 10.1016/j.tics.2005.08.011
- Fries, P., Schröder, J.-H., Roelfsema, P. R., Singer, W., and Engel, A. K. (2002). Oscillatory neuronal synchronization in primary visual cortex as a correlate of stimulus selection. *J. Neurosci.* 22, 3739–3754.
- Frühholz, S., and Grandjean, D. (2013a). Processing of emotional vocalizations in bilateral inferior frontal cortex. *Neurosci. Biobehav. Rev.* 37, 2847–2855. doi: 10.1016/j.neubiorev.2013.10.007
- Frühholz, S., and Grandjean, D. (2013b). Amygdala subregions differentially respond and rapidly adapt to threatening voices. *Cortex* 49, 1394–1403. doi: 10.1016/j.cortex.2012.08.003
- Fulford, D., Peckham, A. D., Johnson, K., and Johnson, S. L. (2014). Emotion perception and quality of life in bipolar I disorder. *J. Affect. Disord.* 152–154, 491–497. doi: 10.1016/j.jad.2013.08.034
- Garrido-Vásquez, P., Jessen, S., and Kotz, S. A. (2011). Perception of emotion in psychiatric disorders: on the possible role of task, dynamics and multimodality. *Soc. Neurosci.* 6, 515–536. doi: 10.1080/17470919.2011.620771
- González-Roldán, A. M., Martínez-Jauand, M., Muñoz-García, M. A., Sitges, C., Cifre, I., and Montoya, P. (2011). Temporal dissociation in the brain processing of pain and anger faces with different intensities of emotional expression. *Pain* 152, 853–859. doi: 10.1016/j.pain.2010.12.037
- Grèzes, J., Pichon, S., and de Gelder, B. (2007). Perceiving fear in dynamic body expressions. *Neuroimage* 35, 959–967. doi: 10.1016/j.neuroimage.2006.11.030
- Güntekin, B., and Başar, E. (2007a). Emotional face expressions are differentiated with brain oscillations. *Int. J. Psychophysiol.* 64, 91–100. doi: 10.1016/j.ijpsycho.2006.07.003
- Güntekin, B., and Başar, E. (2007b). Gender differences influence brain's α oscillatory responses in recognition of facial expressions. *Neurosci. Lett.* 424, 94–99. doi: 10.1016/j.neulet.2007.07.052
- Güntekin, B., and Başar, E. (2009). Facial affect manifested by multiple oscillations. *Int. J. Psychophysiol.* 71, 31–36. doi: 10.1016/j.ijpsycho.2008.07.019
- Güntekin, B., and Başar, E. (2014). A review of brain oscillations in perception of faces and emotional pictures. *Neuropsychologia* 58, 33–51. doi: 10.1016/j.neuropsychologia.2014.03.014
- Güntekin, B., and Başar, E. (2015). Review of evoked and event-related delta responses in the human brain. *Int. J. Psychophysiol.* doi: 10.1016/j.ijpsycho.2015.02.001 [Epub ahead of print].
- Hadjikhani, N., and de Gelder, B. (2003). Seeing fearful body expressions activates the fusiform cortex and amygdala. *Curr. Biol.* 13, 2201–2205. doi: 10.1016/j.cub.2003.11.049
- Hagan, C. C., Woods, W., Johnson, S., Calder, A. J., Green, G. G. R., and Young, A. W. (2009). MEG demonstrates a supra-additive response to facial and vocal emotion in the right superior temporal sulcus. *Proc. Natl. Acad. Sci. U S A* 106, 20010–20015. doi: 10.1073/pnas.0905792106
- Hagan, C. C., Woods, W., Johnson, S., Green, G. G. R., and Young, A. W. (2013). Involvement of right STS in audio-visual integration for affective speech demonstrated using MEG. *PLoS One* 8:e70648. doi: 10.1371/journal.pone.0070648
- Ho, H. T., Schröger, E., and Kotz, S. A. (2015). Selective attention modulates early human evoked potentials during emotional face-voice processing. *J. Cogn. Neurosci.* 27, 798–818. doi: 10.1162/jocn_a_00734
- Hofman, D., Terburg, D., van Wielink, L., and Schutter, D. J. L. G. (2013). Coalescence of dominance motivation and responses to facial anger in resting-state and event-related electrophysiology. *Neuroimage* 79, 138–144. doi: 10.1016/j.neuroimage.2013.04.088
- Hyafil, A., Giraud, A. L., Fontolan, L., and Gutkin, B. (2015). Neural cross-frequency coupling: connecting architectures, mechanisms and functions. *Trends Neurosci.* 38, 725–740. doi: 10.1016/j.tins.2015.09.001
- Jabbi, M., Kohn, P. D., Nash, T., Ianni, A., Coutlee, C., Holroyd, T., et al. (2015). Convergent BOLD and α -band activity in superior temporal sulcus and frontolimbic circuitry underpins human emotion cognition. *Cereb. Cortex* 25, 1878–1888. doi: 10.1093/cercor/bht427

- Jansari, A., Tranel, D., and Adolphs, R. (2000). A valence-specific lateral bias for discriminating emotional facial expressions in free field. *Cogn. Emot.* 14, 341–353. doi: 10.1080/026999300378860
- Jensen, O., and Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Front. Hum. Neurosci.* 4:186. doi: 10.3389/fnhum.2010.00186
- Jessen, S., and Kotz, S. A. (2011). The temporal dynamics of processing emotions from vocal, facial and bodily expressions. *Neuroimage* 58, 665–674. doi: 10.1016/j.neuroimage.2011.06.035
- Jessen, S., and Kotz, S. A. (2013). On the role of crossmodal prediction in audiovisual emotion perception. *Front. Hum. Neurosci.* 7:369. doi: 10.3389/fnhum.2013.00369
- Jessen, S., and Kotz, S. A. (2015). Affect differentially modulates brain activation in uni- and multisensory body-voice perception. *Neuropsychologia* 66, 134–143. doi: 10.1016/j.neuropsychologia.2014.10.038
- Jessen, S., Obleter, J., and Kotz, S. A. (2012). How bodies and voices interact in early emotion perception. *PLoS One* 7:e36070. doi: 10.1371/journal.pone.0036070
- Jung, J., Bayle, D., Jerbi, K., Vidal, J. R., Hénaff, M. A., Ossandon, T., et al. (2011). Intracerebral γ modulations reveal interaction between emotional processing and action outcome evaluation in the human orbitofrontal cortex. *Int. J. Psychophysiol.* 79, 64–72. doi: 10.1016/j.ijpsycho.2010.09.014
- Kilts, C. D., Egan, G., Gideon, D. A., Ely, T. D., and Hoffman, J. M. (2003). Dissociable neural pathways are involved in the recognition of emotion in static and dynamic facial expressions. *Neuroimage* 18, 156–168. doi: 10.1006/nimg.2002.1323
- Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Res. Rev. Res. Rev.* 29, 169–195. doi: 10.1016/s0165-0173(98)00056-3
- Klimesch, W., Sauseng, P., and Hanslmayr, S. (2007). EEG alpha oscillations: the inhibition-timing hypothesis. *Brain Res. Rev.* 53, 63–88. doi: 10.1016/j.brainres.2006.06.003
- Knyazev, G. G. (2007). Motivation, emotion and their inhibitory control mirrored in brain oscillations. *Neurosci. Biobehav. Rev.* 31, 377–395. doi: 10.1016/j.neubiorev.2006.10.004
- Knyazev, G. G. (2012). EEG delta oscillations as a correlate of basic homeostatic and motivational processes. *Neurosci. Biobehav. Rev.* 36, 677–695. doi: 10.1016/j.neubiorev.2011.10.002
- Knyazev, G. G., Bocharov, A. V., Levin, E. A., Savostyanov, A. N., and Slobodskoy-Plusnin, J. Y. (2008). Anxiety and oscillatory responses to emotional facial expressions. *Brain Res.* 1227, 174–188. doi: 10.1016/j.brainres.2008.06.108
- Knyazev, G. G., Bocharov, A. V., Savostyanov, A. N., and Slobodskoy-Plusnin, J. (2015). Predisposition to depression and implicit emotion processing. *J. Clin. Exp. Neuropsychol.* 37, 701–709. doi: 10.1080/13803395.2015.1061483
- Knyazev, G. G., Bocharov, A. V., and Slobodskoy-Plusnin, J. Y. (2009a). Hostility- and gender-related differences in oscillatory responses to emotional facial expressions. *Aggress. Behav.* 35, 502–513. doi: 10.1002/ab.20318
- Knyazev, G. G., Mitrofanova, L. G., and Bocharov, A. V. (2013). Emotional intelligence and oscillatory responses to emotional facial expressions. *Hum. Physiol.* 39, 371–377. doi: 10.1134/s0362119713030110
- Knyazev, G. G., Slobodskoy-Plusnin, J. Y., and Bocharov, A. V. (2009b). Event-related delta and theta synchronization during explicit and implicit emotion processing. *Neuroscience* 164, 1588–1600. doi: 10.1016/j.neuroscience.2009.09.057
- Knyazev, G. G., Slobodskoy-Plusnin, J. Y., and Bocharov, A. V. (2010). Gender differences in implicit and explicit processing of emotional facial expressions as revealed by event-related theta synchronization. *Emotion* 10, 678–687. doi: 10.1037/a0019175
- Kokinous, J., Kotz, S. A., Tavano, A., and Schröger, E. (2015). The role of emotion in dynamic audiovisual integration of faces and voices. *Soc. Cogn. Affect. Neurosci.* 10, 713–720. doi: 10.1093/scan/nsu105
- Kotz, S. A., Hasting, A. S., and Paulmann, S. (2012). “On the orbito-striatal interface in (acoustic) emotional processing,” in *The Evolution of Emotional Communication: From Sounds in Nonhuman Mammals to Speech and Music in Man*, eds E. Altenmüller, S. Schmidt and E. Zimmermann (New York, NY: Oxford University Press), 229–240.
- Kotz, S. A., Meyer, M., Alter, K., Besson, M., von Cramon, D. Y., and Friederici, A. D. (2003). On the lateralization of emotional prosody: an event-related functional MR investigation. *Brain Lang.* 86, 366–376. doi: 10.1016/s0093-934x(02)00532-1
- Kotz, S. A., Meyer, M., and Paulmann, S. (2006). Lateralization of emotional prosody in the brain: an overview and synopsis on the impact of study design. *Prog. Brain Res.* 156, 285–294. doi: 10.1016/s0079-6123(06)56015-7
- Kotz, S. A., and Paulmann, S. (2011). Emotion, language and the brain. *Lang. Linguist. Compass* 5, 108–125. doi: 10.1111/j.1749-818x.2010.00267.x
- Kreifels, B., Ethofer, T., Shiozawa, T., Grodd, W., and Wildgruber, D. (2009). Cerebral representation of non-verbal emotional perception: fMRI reveals audiovisual integration area between voice- and face-sensitive regions in the superior temporal sulcus. *Neuropsychologia* 47, 3059–3066. doi: 10.1016/j.neuropsychologia.2009.07.001
- Kret, M. E., Pichon, S., Grèzes, J., and de Gelder, B. (2011). Similarities and differences in perceiving threat from dynamic faces and bodies. An fMRI study. *Neuroimage* 54, 1755–1762. doi: 10.1016/j.neuroimage.2010.08.012
- Kringelbach, M. L., and Rolls, E. T. (2004). The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. *Prog. Neurobiol.* 72, 341–372. doi: 10.1016/j.pneurobio.2004.03.006
- LeDoux, J. (1996). *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York, NY: Simon and Schuster.
- LeDoux, J. E. (2000). Emotion circuits in the brain. *Annu. Rev. Neurosci.* 23, 155–184. doi: 10.1146/annurev.neuro.23.1.155
- Leung, R. C., Ye, A. X., Wong, S. M., Taylor, M. J., and Doesburg, S. M. (2014). Reduced α connectivity during emotional face processing in adolescents with autism. *Mol. Autism* 5:51. doi: 10.1186/2040-2392-5-51
- Liu, T. Y., Chen, Y. S., Su, T. P., Hsieh, J. C., and Chen, L. F. (2014). Abnormal early γ responses to emotional faces differentiate unipolar from bipolar disorder patients. *Biomed Res. Int.* 2014:906104. doi: 10.1155/2014/906104
- Liu, T., Pinheiro, A. P., Deng, G., Nestor, P. G., McCarley, R. W., and Niznikiewicz, M. A. (2012). Electrophysiological insights into processing nonverbal emotional vocalizations. *Neuroreport* 23, 108–112. doi: 10.1097/WNR.0b013e32834ea757
- Luo, Q., Holroyd, T., Jones, M., Hendler, T., and Blair, J. (2007). Neural dynamics for facial threat processing as revealed by γ band synchronization using MEG. *Neuroimage* 34, 839–847. doi: 10.1016/j.neuroimage.2006.09.023
- Luo, Q., Holroyd, T., Majestic, C., Cheng, X., Schechter, J., and Blair, R. J. (2010). Emotional automaticity is a matter of timing. *J. Neurosci.* 30, 5825–5829. doi: 10.1523/JNEUROSCI.5668-09.2010
- Luo, Q., Mitchell, D., Cheng, X., Mondillo, K., McCaffrey, D., Holroyd, T., et al. (2009). Visual awareness, emotion and γ band synchronization. *Cereb. Cortex* 19, 1896–1904. doi: 10.1093/cercor/bhn216
- Meeren, H. K., de Gelder, B., Ahlfors, S. P., Hämäläinen, M. S., and Hadjikhani, N. (2013). Different cortical dynamics in face and body perception: an MEG study. *PLoS One* 8:e71408. doi: 10.1371/journal.pone.0071408
- Müsch, K., Hamamé, C. M., Perrone-Bertolotti, M., Minotti, L., Kahane, P., Engel, A. K., et al. (2014). Selective attention modulates high-frequency activity in the face-processing network. *Cortex* 60, 34–51. doi: 10.1016/j.cortex.2014.06.006
- Muthukumaraswamy, S. D., Johnson, B. W., Gaetz, W. C., and Cheyne, D. O. (2006). Neural processing of observed oro-facial movements reflects multiple action encoding strategies in the human brain. *Brain Res.* 1071, 105–112. doi: 10.1016/j.brainres.2005.11.053
- Muthukumaraswamy, S. D., and Singh, K. D. (2008). Modulation of the human mirror neuron system during cognitive activity. *Psychophysiology* 45, 896–905. doi: 10.1111/j.1469-8986.2008.00711.x
- Paulmann, S., Bleichner, M., and Kotz, S. A. (2013). Valence, arousal and task effects in emotional prosody processing. *Front. Psychol.* 4:345. doi: 10.3389/fpsyg.2013.00345
- Paulmann, S., and Kotz, S. A. (2008). Early emotional prosody perception based on different speaker voices. *Neuroreport* 19, 209–213. doi: 10.1097/WNR.0b013e3282f454db
- Paulmann, S., Jessen, S., and Kotz, S. A. (2012). It's special the way you say it: an ERP investigation on the temporal dynamics of two types of prosody. *Neuropsychologia* 50, 1609–1620. doi: 10.1016/j.neuropsychologia.2012.03.014

- Pavlova, M. A. (2012). Biological motion processing as a hallmark of social cognition. *Cereb. Cortex* 22, 981–995. doi: 10.1093/cercor/bhr156
- Peelle, J. E., and Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Front. Psychol.* 3:320. doi: 10.3389/fpsyg.2012.00320
- Pell, M. D., Rothermich, K., Liu, P., Paulmann, S., Sethi, S., and Rigoulot, S. (2015). Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. *Biol. Psychol.* 111, 14–25. doi: 10.1016/j.biopsych.2015.08.008
- Pernet, C. R., McAleer, P., Latinus, M., Gorgolewski, K. J., Charest, I., Bestelmeyer, P. E. G., et al. (2015). The human voice areas: spatial organisation and inter-individual variability in temporal and extra-temporal cortices. *Neuroimage* 119, 164–174. doi: 10.1016/j.neuroimage.2015.06.050
- Péron, J., Biseul, I., Leray, E., Vicente, S., Le Jeune, F., Drapier, S., et al. (2010a). Subthalamic nucleus stimulation affects fear and sadness recognition in Parkinson's disease. *Neuropsychology* 24, 1–8. doi: 10.1037/a017433
- Péron, J., Fröhholz, S., Vérin, M., and Grandjean, D. (2013). Subthalamic nucleus: a key structure for emotional component synchronization in humans. *Neurosci. Biobehav. Rev.* 37, 358–373. doi: 10.1016/j.neubiorev.2013.01.001
- Péron, J., Grandjean, D., Le Jeune, F., Sauleau, P., Haegelen, C., Drapier, D., et al. (2010b). Recognition of emotional prosody is altered after subthalamic nucleus deep brain stimulation in Parkinson's disease. *Neuropsychologia* 48, 1053–1062. doi: 10.1016/j.neuropsychologia.2009.12.003
- Pfurtscheller, G., Stancák, A., and Neuper, C. (1996). Event-related synchronization (ERS) in the alpha band - an electrophysiological correlate of cortical idling: a review. *Int. J. Psychophysiol.* 24, 39–46. doi: 10.1016/s0167-8760(96)00066-9
- Phillips, M. L., Drevets, W. C., Rauch, S. L., and Lane, R. (2003). Neurobiology of emotion perception II: implications for major psychiatric disorders. *Biol. Psychiatry* 54, 515–528. doi: 10.1016/s0006-3223(03)00171-9
- Phillips, L. H., Henry, J. D., Scott, C., Summers, F., Whyte, M., and Cook, M. (2011). Specific impairments of emotion perception in multiple sclerosis. *Neuropsychology* 25, 131–136. doi: 10.1037/a0020752
- Phillips, L. H., Scott, C., Henry, J. D., Mowat, D., and Bell, J. S. (2010). Emotion perception in Alzheimer's disease and mood disorder in old age. *Psychol. Aging* 25, 38–47. doi: 10.1037/a0017369
- Phillips, M. L., Young, A. W., Senior, C., Brammer, M., Andrew, C., Calder, A. J., et al. (1997). A specific neural substrate for perceiving facial expressions of disgust. *Nature* 389, 495–498. doi: 10.1038/39051
- Popov, T., Miller, G. A., Rockstroh, B., and Weisz, N. (2013). Modulation of α power and functional connectivity during facial affect recognition. *J. Neurosci.* 33, 6018–6026. doi: 10.1523/JNEUROSCI.2763-12.2013
- Popova, P., Popov, T. G., Wienbruch, C., Carolus, A. M., Miller, G. A., and Rockstroh, B. S. (2014). Changing facial affect recognition in schizophrenia: effects of training on brain dynamics. *Neuroimage Clin.* 6, 156–165. doi: 10.1016/j.nicl.2014.08.026
- Popov, T. G., Rockstroh, B. S., Popova, P., Carolus, A. M., and Miller, G. A. (2014). Dynamics of alpha oscillations elucidate facial affect recognition in schizophrenia. *Cogn. Affect. Behav. Neurosci.* 14, 364–377. doi: 10.3758/s13415-013-0194-2
- Ramos-Loyo, J., González-Garrido, A. A., Sánchez-Loyo, L. M., Medina, V., and Basar-Eroglu, C. (2009). Event-related potentials and event-related oscillations during identity and facial emotional processing in schizophrenia. *Int. J. Psychophysiol.* 71, 84–90. doi: 10.1016/j.ijpsycho.2008.07.008
- Robins, D. L., Hunyadi, E., and Schultz, R. T. (2009). Superior temporal activation in response to dynamic audio-visual emotional cues. *Brain Cogn.* 69, 269–278. doi: 10.1016/j.bandc.2008.08.007
- Rolls, E. T. (2004). The functions of the orbitofrontal cortex. *Brain Cogn.* 55, 11–29. doi: 10.1016/S0278-2626(03)00277-X
- Sato, W., Kochiyama, T., Uono, S., Matsuda, K., Usui, K., Inoue, Y., et al. (2011). Rapid amygdala γ oscillations in response to fearful facial expressions. *Neuropsychologia* 49, 612–617. doi: 10.1016/j.neuropsychologia.2010.12.025
- Schellenz, P. D., Klasen, M., Reese, B., Regenbogen, C., Wolf, D., Kato, Y., et al. (2013). Multisensory integration of dynamic emotional faces and voices: method for simultaneous EEG-fMRI measurements. *Front. Hum. Neurosci.* 7:729. doi: 10.3389/fnhum.2013.00729
- Scherer, K. R. (2000). "Psychological models of emotion," in *The Neuropsychology of Emotion*, ed J. Board (Oxford, NY: Oxford University Press), 137–162.
- Schirmer, A., and Kotz, S. A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends Cogn. Sci.* 10, 24–30. doi: 10.1016/j.tics.2005.11.009
- Schirmer, A., Striano, T., and Friederici, A. D. (2005). Sex differences in the preattentive processing of vocal emotional expressions. *Neuroreport* 16, 635–639. doi: 10.1097/00001756-200504250-00024
- Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., and Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends Cogn. Sci.* 12, 106–113. doi: 10.1016/j.tics.2008.01.002
- Schutter, D. J. L. G., Putman, P., Hermans, E., and van Honk, J. (2001). Parietal electroencephalogram or asymmetry and selective attention to angry facial expressions in healthy human subjects. *Neurosci. Lett.* 314, 13–16. doi: 10.1016/s0304-3940(01)02246-7
- Schwarzlose, R. F., Baker, C. I., and Kanwisher, N. (2005). Separate face and body selectivity on the fusiform gyrus. *J. Neurosci.* 25, 11055–11059. doi: 10.1523/jneurosci.2621-05.2005
- Senkowski, D., Schneider, T. R., Foxe, J. J., and Engel, A. K. (2008). Crossmodal binding through neural coherence: implications for multisensory processing. *Trends Neurosci.* 31, 401–409. doi: 10.1016/j.tins.2008.05.002
- Silberman, E. K., and Weingartner, H. (1986). Hemispheric lateralization of functions related to emotion. *Brain Cogn.* 5, 322–353. doi: 10.1016/0278-2626(86)90035-7
- Singer, W., and Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annu. Rev. Neurosci.* 18, 555–586. doi: 10.1146/annurev.neuro.18.1.555
- Stefanics, G., Csukly, G., Komlósi, S., Czobor, P., and Czigler, I. (2012). Processing of unattended facial emotions: a visual mismatch negativity study. *Neuroimage* 59, 3042–3049. doi: 10.1016/j.neuroimage.2011.10.041
- Stevenson, R. A., Geoghegan, M. L., and James, T. W. (2007). Superadditive BOLD activation in superior temporal sulcus with threshold non-speech objects. *Exp. Brain Res.* 179, 85–95. doi: 10.1007/s00221-006-0770-6
- Stevenson, R. A., and James, T. W. (2009). Audiovisual integration in human superior temporal sulcus: inverse effectiveness and the neural processing of speech and object recognition. *Neuroimage* 44, 1210–1223. doi: 10.1016/j.neuroimage.2008.09.034
- Tallon-Baudry, C., and Bertrand, O. (1999). Oscillatory γ activity and its role in object representation. *Trends Cogn. Sci.* 3, 151–162. doi: 10.1016/s1364-6613(99)01299-1
- Tanaka, A., Koizumi, A., Imai, H., Hiramatsu, S., Hiramoto, E., and de Gelder, B. (2010). I feel your voice. Cultural differences in the multisensory perception of emotion. *Psychol. Sci.* 21, 1259–1262. doi: 10.1177/0956797610308698
- van Atteveldt, N., Murray, M. M., Thut, G., and Schroeder, C. E. (2014). Multisensory integration: flexible use of general operations. *Neuron* 81, 1240–1253. doi: 10.1016/j.neuron.2014.02.044
- Vroomen, J., and Stekelenburg, J. J. (2010). Visual anticipatory information modulates multisensory interactions of artificial audiovisual stimuli. *J. Cogn. Neurosci.* 22, 1583–1596. doi: 10.1162/jocn.2009.21308
- Vuilleumier, P., and Pourtois, G. (2007). Distributed and interactive brain mechanisms during emotion face perception: evidence from functional neuroimaging. *Neuropsychologia* 45, 174–194. doi: 10.1016/j.neuropsychologia.2006.06.003
- Watson, R., Latinus, M., Noguchi, T., Garrod, O., Crabbe, F., and Belin, P. (2013). Dissociating task difficulty from incongruence in face-voice emotion integration. *Front. Hum. Neurosci.* 7:744. doi: 10.3389/fnhum.2013.00744
- Watson, R., Latinus, M., Noguchi, T., Garrod, O., Crabbe, F., and Belin, P. (2014). Crossmodal adaptation in right posterior superior temporal sulcus during face-voice emotional integration. *J. Neurosci.* 34, 6813–6821. doi: 10.1523/JNEUROSCI.4478-13.2014

- Whalen, P. J., Raila, H., Bennett, R., Mattek, A., Brown, A., Taylor, J., et al. (2013). Neuroscience and facial expressions of emotion: the role of amygdala-prefrontal interactions. *Emot. Rev.* 5, 78–83. doi: 10.1177/1754073912457231
- Whalen, P. J., Shin, L. M., McInerney, S. C., Fischer, H., Wright, C. I., and Rauch, S. L. (2001). A functional MRI study of human amygdala responses to facial expressions of fear versus anger. *Emotion* 1, 70–83. doi: 10.1037/1528-3542.1.1.70
- Wiethoff, S., Wildgruber, D., Kreifelts, B., Becker, H., Herbert, C., Grodd, W., et al. (2008). Cerebral processing of emotional prosody—fluence of acoustic parameters and arousal. *Neuroimage* 39, 885–893. doi: 10.1016/j.neuroimage.2007.09.028
- Wildgruber, D., Ethofer, T., Grandjean, D., and Kreifelts, B. (2009). A cerebral network model of speech prosody comprehension. *Int. J. Speech Lang. Pathol.* 11, 277–281. doi: 10.1080/17549500902943043
- Williams, M. A., Morris, A. P., McGlone, F., Abbott, D. F., and Mattingley, J. B. (2004). Amygdala responses to fearful and happy facial expressions under conditions of binocular suppression. *J. Neurosci.* 24, 2898–2904. doi: 10.1523/jneurosci.4977-03.2004
- Woods, D. L., and Alain, C. (2009). Functional imaging of human auditory cortex. *Curr. Opin. Otolaryngol. Head Neck Surg.* 17, 407–411. doi: 10.1097/MOO.0b013e3283303330
- Zhang, D., Wang, L., Luo, Y., and Luo, Y. (2012). Individual differences in detecting rapidly presented fearful faces. *PLoS One* 7:e49517. doi: 10.1371/journal.pone.0049517

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Symons, El-Deredy, Schwartze and Kotz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Synchronization by the hand: the sight of gestures modulates low-frequency activity in brain responses to continuous speech

Emmanuel Biau^{1*} and Salvador Soto-Faraco^{1,2}

¹ Multisensory Research Group, Center for Brain and Cognition, Universitat Pompeu Fabra, Barcelona, Spain, ² Institutó Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

OPEN ACCESS

Edited by:

Anne Keitel,
University of Glasgow, UK

Reviewed by:

Daniel Callan,

Center for Information and Neural Networks (CiNet), National Institute of Information and Communications

Technology (NICT), Japan
Spencer Kelly,
Colgate University, USA

***Correspondence:**

Emmanuel Biau,

Multisensory Research Group, Center for Brain and Cognition, Universitat Pompeu Fabra, Roc Boronat, 138,
08018 Barcelona, Spain
emmanuel.biau@free.fr

Received: 30 July 2015

Accepted: 10 September 2015

Published: 24 September 2015

Citation:

Biau E and Soto-Faraco S (2015) Synchronization by the hand: the sight of gestures modulates low-frequency activity in brain responses to continuous speech. *Front. Hum. Neurosci.* 9:527.
doi: 10.3389/fnhum.2015.00527

During social interactions, speakers often produce spontaneous gestures to accompany their speech. These coordinated body movements convey communicative intentions, and modulate how listeners perceive the message in a subtle, but important way. In the present perspective, we put the focus on the role that congruent non-verbal information from beat gestures may play in the neural responses to speech. Whilst delta-theta oscillatory brain responses reflect the time-frequency structure of the speech signal, we argue that beat gestures promote phase resetting at relevant word onsets. This mechanism may facilitate the anticipation of associated acoustic cues relevant for prosodic/syllabic-based segmentation in speech perception. We report recently published data supporting this hypothesis, and discuss the potential of beats (and gestures in general) for further studies investigating continuous AV speech processing through low-frequency oscillations.

Keywords: audiovisual speech, gestures, beats, low-frequency oscillations, EEG

Speakers spontaneously gesture to accompany their speech, and listeners definitely seem to take advantage of this source of complementary information from the visual modality (Goldin-Meadow, 1999). The aim of the present perspective is to bring attention to the relevance of this visual concomitant information when investigating continuous speech. Here, we argue that part of this explanation may have to do with the modulations that speaker's gestures impose on low-frequency oscillatory activity related to speech segmentation in the listener's brain. The speaker modulates the amplitude envelope of the utterance (i.e., the summed acoustic power across all frequency ranges for each time point of the signal) in a regular manner, providing quasi-rhythmic acoustic cues in at least two low-frequency ranges. First, speech syllables are produced rhythmically at frequency of 4–7 Hz, corresponding to a theta rate imposed by voicing after breath taking and jaw aperture (Peelle and Davis, 2012). Second, the speaker modulates pitch accents in her/his vocalization to convey particular speech acts (e.g., declarative or ironic), and emphasize relevant information to convey communicative intentions. These pitch peaks also occur with a quasi-rhythmic rate of 1–3 Hz corresponding to a delta frequency and constituting part of prosody (Munhall et al., 2004; Park et al., 2015). Recently, Electroencephalography (EEG) and Magnetoencephalography (MEG) studies investigated auditory speech segmentation mechanisms, taking advantage of time-frequency analyses to look at brain activities that are not time-locked to stimuli onsets, and measure the amount of activity in frequency bands of interest [typically missing in the classic Event-Related Potential (ERPs) averages]. These studies reported that spontaneous delta-theta

activities in the auditory cortex reset their phase to organize in structured patterns, highly similar to the spectro-temporal architecture of the auditory speech envelope, reflecting entrainment mechanism (Ahissar et al., 2001; Luo and Poeppel, 2007; Abrams et al., 2008; Nourski et al., 2009; Giraud and Poeppel, 2012; Gross et al., 2013; Park et al., 2015; Zoefel and VanRullen, 2015). Then, delta-theta periodicity seems to constitute a fundamental window of compatibility between brain's activity and speech segmentation (Ghitza and Greenberg, 2009; Peele and Davis, 2012). Thus, when the natural delta-theta periodicity in the auditory signal is affected by time compression, speech comprehension worsens significantly. But more interestingly, the degradation of the delta-theta rhythms also decreases the spectro-temporal similarity between the speech envelope and the low-frequency activities in the auditory cortex (Ahissar et al., 2001). These important spectro-temporal features of the acoustic signal seem to be, therefore, important in determining brain responses to speech.

Yet, the acoustic signal is not the only communicative cue between speaker and listener. Coherent face and body movements often accompany verbalization. Before placing the focus on the speaker's hand gestures, it is important to note that the relevance of non-verbal information has been first established regarding the speaker's face (van Wassenhove et al., 2005). Corresponding lip movements have been long shown to facilitate comprehension in noisy conditions (Sumbay and Pollack, 1954), or in contrast, affect speech processing when incongruent with utterance, e.g., the famous McGurk illusion (McGurk and MacDonald, 1976). More recently, visual speech information has been proposed to play a role in the extraction of the aforementioned rhythmic aspects of the speech signal (van Wassenhove et al., 2005). Due to the natural precedence of visual speech cues over their auditory counterparts in natural situations (i.e., the sight of articulation often precedes its auditory consequence; see Sánchez-García et al., 2011), it has been hypothesized that visual information conveys predictive information about the timing and contents of corresponding auditory information, facilitating its anticipation (van Wassenhove et al., 2005; Stekelenburg and Vroomen, 2007; Vroomen and Stekelenburg, 2010). For example, van Wassenhove et al. (2005) presented isolated consonant-vowel syllables in audio, visual or audiovisual modalities. They showed that the N1-P2 component in the auditory evoked responses time-locked to the phoneme onset were significantly reduced in amplitude and speeded up in time in the AV modality, compared to the responses to auditory syllables. In the time-frequency dimension, delta-theta entrainment has been proposed to underlie predictive coding mechanism based on the temporal correlation between audio-visual speech cues (Lakatos et al., 2008; Schroeder et al., 2008; Schroeder and Lakatos, 2009; Arnal and Giraud, 2012). Thus, Arnal and Giraud (2012) hypothesized that visual information provided by lip movements increases delta-theta phase resetting at relevant associated acoustic cue onsets (word onsets), reflecting predictive coding mechanisms that minimize the uncertainty about when regular event are likely to occur, and a better speech segmentation.

Along these lines, one could ask whether other speech-related visible body movements of the speaker may also bear predictive information and have an impact on low-frequency neural activity in the listeners' brain. In continuous speech production, which movements may be correlated with delta-theta acoustic cues in the auditory signal? Head movements for example, were shown to be highly correlated with pitch peaks and facilitate comprehension of speech perception in noisy conditions (Munhall et al., 2004). Looking at public addressees, and in particular political discourses, we observed that speakers almost all the time accompany their speech with spontaneous hand gestures called "beats" (McNeill, 1992). Beats are simple and biphasic arm/hand movements that often bear no semantic content in their shape produced by speakers when they want to emphasize relevant information or develop an argument with successive related points. They belong to what could be considered as visual prosody, as they are temporally aligned with the prosodic structure of the verbal utterance, just like eyebrow, shoulders and head nods (McNeill, 1992; Krahmer and Swerts, 2007; Leonard and Cummins, 2011). Yasinnik (2004) showed that beats' apexes (i.e., the maximum extension point of the arm before retraction, corresponding to the functional phase of the gesture) align quite precisely with pitch-accented syllables (peaks of the F0 fundamental frequency). In other words, the kinematics of beats match with spectro-temporal modulation of auditory speech envelope and are thought to modulate both the acoustic properties and the perceived saliency of the affiliated utterance (Munhall et al., 2004; Krahmer and Swerts, 2007). Albeit simple, beats have been found to modulate syntactic parsing (Holle et al., 2012; Guellai et al., 2014), semantic processing (Wang and Chu, 2013) and encoding (So et al., 2012) during audiovisual speech perception. In a previous ERP study, we showed that the sight of beats modulate the ERPs produced by the corresponding spoken words at early phonological stages by reducing negativity of the waveform within the 200–300 ms time window (Biau and Soto-Faraco, 2013). Since the onsets of the beats systematically preceded affiliated words onsets by around 200 ms, we concluded that the order of perception and congruence between pitch accents and apexes attracted the focus of local attention on relevant acoustic cues in the signal (i.e., words onsets), possibly modulating speech processing from early stages.

Based on these previous studies and the stable spatio-temporal relationship between beats and auditory prosody, we argued that continuous speech segmentation should not be limited to the auditory modality, but also take into account visual congruent information both from lip movements and the rest of the body. Recently, Skipper (2014) proposed that listeners use the visual context provided by gestures as predictive information because of learned preceding timing with associated auditory information. Gestures may pre-activate words associated with their kinematics, to process inferences that are compared with following auditory information. In the present context, the idea behind was that if gestures provide robust prosodic information that listeners can use to anticipate associated speech segments, then beats may have an impact on the entrainment mechanisms capitalizing on rhythmic aspects of speech, discussed above (Arnal and Giraud, 2012; Giraud and Poeppel, 2012; Peele and

Davis, 2012). More precisely, we expected that if gestures provide a useful anticipatory signal for particular words in the sentence, this might reflect in phase synchronization of low frequency at relevant moments in the signal, coinciding with the acoustic onsets of the associated words (see **Figure 1**). This is exactly what we have tested in a recent EEG study, by presenting a naturally spoken, continuous AV speech in which the speaker spontaneously produced beats while addressing the audience (Biau et al., 2015). We recorded the EEG signal of participants during AV speech perception, and compared the phase-locking value (PLV) of low-frequency activity at the onset of words pronounced with or without a beat gesture (see **Figure 1**). The PLV analysis revealed strong phase synchronization in the theta 5–6 Hz range with a concomitant desynchronization in the alpha 8–10 Hz range, mainly at left fronto-temporal sites (see **Figure 2**). The gesture-induced synchronization in theta started to increase around 100 ms before the onset of the corresponding affiliate word, and was maintained for around 60 ms thereafter. Given that gestures initiated approximately 200 ± 100 ms before word onsets, we thought that this delay was enough for beat to effectively engage the oscillation-based temporal prediction of speech in preparation for the upcoming word onset (Arnal and Giraud, 2012). Crucially, when visual information was removed (that is, speech was presented in audio modality only), our results showed no difference in PLV or amplitude between words that had been pronounced with or without a beat gesture in the original discourse. Such pattern suggested that the effects observed in the AV modality could be attributed to the sight of gestures, and not just acoustic differences between gesture and no gesture words in the continuous speech. We interpreted these results within the following framework: beats are probably perceived as communicative rather than simple body movements disconnected from the message (McNeill, 1992; Hubbard et al., 2009). Through daily social experience, listeners learn to attribute

linguistic relevance to beats because they gesture when they speak (McNeill, 1992; So et al., 2012), and seem to have an understanding of the sense of a beat at a precise moment. Consequently, listeners may rely on beats to anticipate associated speech segmentation that is reflected through an increase of low-frequency phase resetting at relevant onsets of accompanied words. In addition, it is possible that this prediction engages local attentional mechanisms, reflected by early ERP effects and the alpha activity reduction seen around word onsets with gesture. As far as we know, Biau et al. (2015) was the first study investigating the impact of spontaneous hand gestures on speech processing through low-frequency oscillatory activities in a close-to-natural approach. Further investigations are definitely needed to increase data and set new experimental procedures combining behavioral measures with EEG analyses.

A recent study by He et al. (2015) has investigated AV speech processing through low-frequency activity, albeit with a very different category of speech gestures. He et al. (2015) used intrinsically-meaningful gestures (IMG) conveying semantic content, such as when the speaker makes a “thumbs-up” gesture while uttering “the actor did a good job”. The authors investigated the oscillatory signature of gesture-speech integration by manipulating the relationship between gesture and auditory speech modalities: AV integration (IMG produced in the context of an understandable sentence in the listener’s native language), V (IMG produced in the context of a sentence in a foreign language incomprehensible for the listener) and A (an understandable sentence in the listener’s native language without gestures). The results of a conjunction analysis showed that the AV condition induced a significant centrally-distributed power decrease in the alpha band (7–13 Hz; from 700–1400 ms after the onset of the critical word associated with the gesture in the sentence), as compared to the V and A conditions that contained only semantic inputs from one modality (respectively: in the V condition only the gesture was understandable and in the A condition only the utterance was understandable). The authors concluded that the alpha power decrease reflected an oscillatory correlate of the meaningful gesture–speech integration process.

Investigations on the neural dynamics of hand gesture-speech integration during continuous AV speech perception have just begun but the results reported in both studies Biau et al. (2015) and He et al. (2015) already suggest two important conclusions for the present perspective. First, whereas auditory speech seems at first glance to attract all the listeners’ attention, hand gestures count as well, and may definitely be considered as visual linguistic information for online AV speech segmentation. If the delta-theta rhythmic aspects in the auditory signal can play the role of anchors for predictive coding during speech segmentation (Arnal and Giraud, 2012; Peelle and Davis, 2012; Park et al., 2015), then preceding visual gestural information, naturally present in face to face conversations, may convey very useful information for decoding the signal and thus, be taken into account. For instance, beats are not only exquisitely tuned to the prosodic aspects of the auditory spectro-temporal structure, but also engage language-related brain areas during continuous AV speech perception (Hubbard et al., 2009). This idea is in

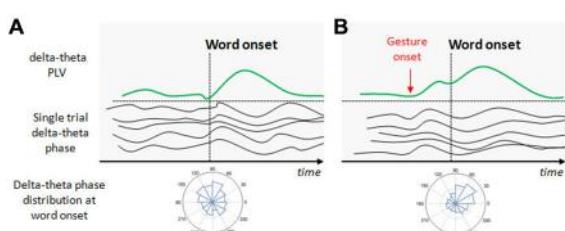


FIGURE 1 | Illustration of the potential effect of beat gestures on the delta-theta phase resetting. **(A)** At the beginning of speech, neural populations in the auditory cortex spontaneously discharge at delta-theta rates but not at the same phase for a given time point (this is illustrated by the single trial delta-theta band phase before the word onset). At the first word onset, a phase distribution in the auditory sensors shows no preferred angle in the delta-theta band. In consequence, the delta-theta phase locking value (PLV) at the first word onset is weak. With progressive entrainment, delta-theta phase synchronizes, increasing PLV with a preferred angle at relevant syllable/word onsets. **(B)** Beat onsets systematically precede word onsets and potentially increase the delta-theta entrainment before the arriving word onset. When the relevant gesture onset occurs, delta-theta activity synchronizes with a preferred angle in the phase, increasing PLV before the associated word onset arrives to anticipate its processing.

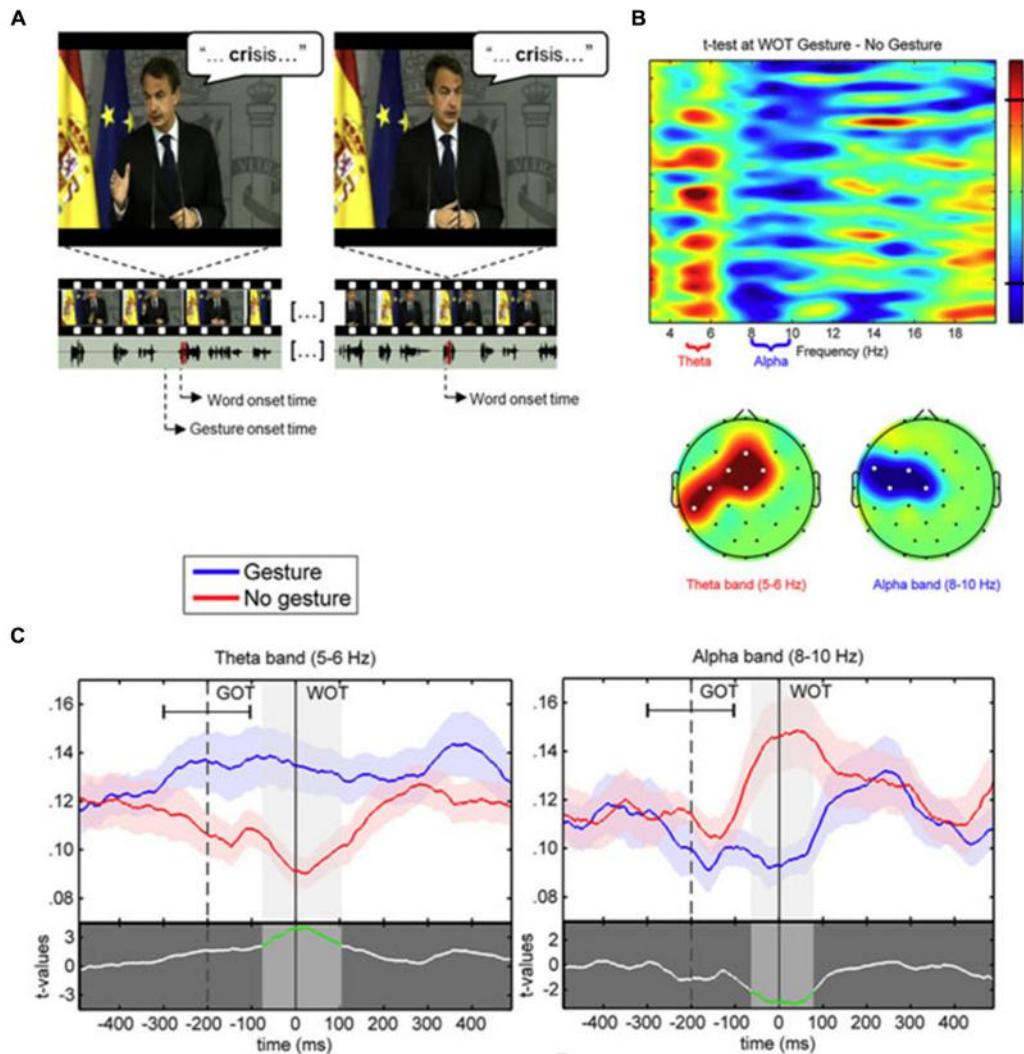


FIGURE 2 | (A) Example of video-frames for the gesture (left) and no gesture (right) conditions associated to the same stimulus word “crisis”. The speaker is the former Spanish President Luis Rodríguez Zapatero, recorded at the palace of La Moncloa, and the video is freely available on the official website (Balance de la acción de Gobierno en 2010, 12-30-2010; <http://www.lamoncloa.gob.es>). Below, the oscillogram of corresponding audio track fragments (section corresponding to the target word shaded in red). The onsets of both the gesture and corresponding word (gesture condition) are marked. **(B)** (Top) Representation of paired *t*-test values for the comparison between PLV at word onset in the gesture and no gesture conditions with frequency bands of interest labeled in the x axis. **(B)** (Bottom) Topographic representation of the significant clusters (significant electrodes marked with white dots) for the *t*-tests within the theta and alpha bands. **(C)** PLV time course in 5–6 Hz theta (left) and 8–10 Hz alpha (right) frequency bands at Cz electrode for the gesture (blue line) and no gesture (red line) conditions. The mean average \pm standard deviation of gesture onset time (GOT) is represented respect to word onset time (WOT). The lower part of each plot displays the paired *t*-test values between gesture and no gesture conditions. The shaded bands indicate significant time intervals (highlighted in green in the *t*-test line).

line with earlier arguments considering auditory speech and gestures as two sides of the same common language system (McNeill, 1992; Kelly et al., 2010; for some examples). Gestures may constitute a good candidate to investigate the multisensory integration between natural auditory speech and social postures. For example, Mitchel and Weiss (2014) showed that the simple temporal alignment between V and A information did not fully explain the AV benefit (i.e., multisensory integration) in a segmentation task of artificial speech. Indeed, segmentation

was significantly better when visual information came from a speaker that was previously exposed to the words he had to pronounce during the stimuli recording (then, knowing the prosodic contours of words, i.e., boundaries), compared from a speaker that was unaware of word boundaries when recording. These results suggested that facial movements conveyed helpful visual prosodic contours if the speakers were aware of them. The same conclusion may apply to beat gestures as they synchronize with auditory prosody in communicative intent (and the speaker

knows the prosodic contours of her/his own discourse). For example, it may be interesting to compare delta-theta activity patterns between gestures conveying the proper communicative prosody and simple synchronized hand movements without the adequate prosodic kinematics.

A second interim conclusion from the few current studies addressing the oscillatory correlates of gestures is that low-frequency brain activity appears to be a successful neural marker to investigate gesture-speech integration and continuous AV speech processing in general. Based on the results reported in these two pioneer studies, low frequency activity seemed sensitive to the type of gesture (intrinsically meaningful gestures in He et al., 2015; and beats in Biau et al., 2015). Both studies analyzed a contrast, comparing the low-frequency activity modulations between an AV gesture condition (i.e., words were accompanied with a gesture) and an AV no gesture condition (i.e., words were pronounced without gesture, but the speaker was visible). He et al. (2015) reported a decrease of alpha power (from 400–1400 ms) and a beta power decrease (from 200–1200ms) after the critical word onset, whilst Biau et al. (2015) reported a theta synchronization with a concomitant alpha desynchronization temporally centred on the affiliate word onset (note that the alpha activity modulation was found in both studies). Even if the experimental procedures and stimuli were not the same [in He et al. (2015) the speaker was still in the no gesture condition, whereas moving in Biau and Soto-Faraco (2013)], the distinct patterns of low-frequency modulations in the gesture-no gesture contrasts suggested that different kind of gestures may be associated to different aspects of the verbalization, modulating speech processing diversely. Indeed, IMGs describe a conventionally established meaning and can be understood silently whereas beats do not and need to be

contextualized by speech to become functional. This might explain why the timing of modulations in He et al. (2015) was quite different respect to Biau et al. (2015). Then, oscillations may constitute an excellent tool for further investigations on neural correlate of AV speech perception and associated social cues with different communicative purposes (IMG vs. beats).

Speech is an intrinsically multisensory object of perception, as the act of speaking produces correlates to the ear and to the eye of the listener. The aim of the present short perspective was to bring attention to the fact that conversations engage a whole set of coordinated body movements. Furthermore, we argue that considering the oscillatory brain responses to natural speech may capture an important aspect of how the listeners' perceptual system integrates back all the different aspects of the communicative production from the talker. Future studies may investigate more precisely how this integration occurs, and what is the role of synchronization and desynchronization patterns that we have tentatively interpreted here.

Funding

This research was supported by the Spanish Ministry of Science and Innovation (PSI2013-42626-P), AGAUR Generalitat de Catalunya (2014SGR856) and the European Research Council (StG-2010 263145).

Acknowledgments

We would like to thank Mireia Torralba, Ruth de Diego Balaguer and Lluís Fuentemilla who took part in the project reported in the present perspective (Biau et al., 2015).

References

- Abrams, D. A., Nicol, T., Zecker, S., and Kraus, N. (2008). Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. *J. Neurosci.* 28, 3958–3965. doi: 10.1523/JNEUROSCI.0187-08.2008
- Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., Merzenich, M. M., et al. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc. Natl. Acad. Sci. U S A* 98, 13367–13372. doi: 10.1073/pnas.20140098
- Arnal, L. H., and Giraud, A. L. (2012). Cortical oscillations and sensory predictions. *Trends Cogn. Sci.* 16, 390–398. doi: 10.1016/j.tics.2012.05.003
- Biau, E., and Soto-Faraco, S. (2013). Beat gestures modulate auditory integration in speech perception. *Brain Lang.* 124, 143–152. doi: 10.1016/j.bandl.2012.10.008
- Biau, E., Torralba, M., Fuentemilla, L., de Diego Balaguer, R., and Soto-Faraco, S. (2015). Speaker's hand gestures modulate speech perception through phase resetting of ongoing neural oscillations. *Cortex* 68, 76–85. doi: 10.1016/j.cortex.2014.11.018
- Ghitza, O., and Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica* 66, 113–126. doi: 10.1159/000208934
- Giraud, A. L., and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* 15, 511–517. doi: 10.1038/nn.3063
- Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends Cogn. Sci.* 3, 419–429. doi: 10.1016/s1364-6613(99)01397-2
- Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, S., et al. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biol.* 11:e1001752. doi: 10.1371/journal.pbio.1001752
- Guellai, B., Langus, A., and Nespor, M. (2014). Prosody in the hands of the speaker. *Front. Psychol.* 5:700. doi: 10.3389/fpsyg.2014.00700
- He, Y., Gebhardt, H., Steines, M., Sammer, G., Kircher, T., Nagels, A., et al. (2015). The EEG and fMRI signatures of neural integration: an investigation of meaningful gestures and corresponding speech. *Neuropsychologia* 72, 27–42. doi: 10.1016/j.neuropsychologia.2015.04.018
- Holle, H., Obermeier, C., Schmidt-Kassow, M., Friederici, A. D., Ward, J., and Gunter, T. C. (2012). Gesture facilitates the syntactic analysis of speech. *Front. Psychol.* 3:74. doi: 10.3389/fpsyg.2012.00074
- Hubbard, A. L., Wilson, S. M., Callan, D. E., and Dapretto, M. (2009). Giving speech a hand: gesture modulates activity in auditory cortex during speech perception. *Hum. Brain Mapp.* 30, 1028–1037. doi: 10.1002/hbm.20565
- Kelly, S. D., Creigh, P., and Bartolotti, J. (2010). Integrating speech and iconic gestures in a stroop-like task: evidence for automatic processing. *J. Cogn. Neurosci.* 22, 683–694. doi: 10.1162/jocn.2009.21254
- Krahmer, E., and Swerts, M. (2007). The effects of visual beats on prosodic prominence: acoustic analyses, auditory perception and visual perception. *J. Mem. Lang.* 57, 396–414. doi: 10.1016/j.jml.2007.06.005
- Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., and Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* 320, 110–113. doi: 10.1126/science.1154735

- Leonard, T., and Cummins, F. (2011). The temporal relation between beat gestures and speech. *Lang. Cogn. Process.* 26, 1457–1471. doi: 10.1080/01690965.2010.500218.
- Luo, H., and Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54, 1001–1010. doi: 10.1016/j.neuron.2007.06.004
- McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748. doi: 10.1038/264746a0
- McNeill, D. (1992). *Hand Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.
- Mitchel, A. D., and Weiss, D. J. (2014). Visual speech segmentation: using facial cues to locate word boundaries in continuous speech. *Lang. Cogn. Process.* 29, 771–780. doi: 10.1080/01690965.2013.791703
- Munhall, K. G., Jones, J. A., Callan, D. E., Kuratake, T., and Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility: head movement improves auditory speech perception. *Psychol. Sci.* 15, 133–137. doi: 10.1111/j.0963-7214.2004.01502010.x
- Nourski, K. V., Reale, R. A., Oya, H., Kawasaki, H., Kovach, C. K., Chen, H., et al. (2009). Temporal envelope of time-compressed speech represented in the human auditory cortex. *J. Neurosci.* 29, 15564–15574. doi: 10.1523/JNEUROSCI.3065-09.2009
- Park, H., Ince, R. A. A., Schyns, P. G., Thut, G., and Gross, J. (2015). Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Curr. Biol.* 25, 1649–1653. doi: 10.1016/j.cub.2015.04.049
- Peelle, J. E., and Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Front. Psychol.* 3:320. doi: 10.3389/fpsyg.2012.00320
- Sánchez-García, C., Alsius, A., Enns, J. T., and Soto-Faraco, S. (2011). Cross-modal prediction in speech perception. *PLoS One* 6:e25198. doi: 10.1371/journal.pone.0025198
- Schroeder, C. E., and Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci.* 32, 9–18. doi: 10.1016/j.tins.2008.09.012
- Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., and Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends Cogn. Sci.* 12, 106–113. doi: 10.1016/j.tics.2008.01.002
- Skipper, J. I. (2014). Echoes of the spoken past: how auditory cortex hears context during speech perception. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130297. doi: 10.1098/rstb.2013.0297
- So, W. C., Chen-Hui, C. S., and Wei-Shan, J. L. (2012). Mnemonic effect of iconic gesture and beatgesture in adults and children: is meaning in gesture important for memory recall? *Lang. Cogn. Process.* 27, 665–681. doi: 10.1080/01690965.2011.573220
- Stekelenburg, J. J., and Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *J. Cogn. Neurosci.* 19, 1964–1973. doi: 10.1037/e527342012-225
- Sumby, W., and Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *J. Acoust. Soc. Am.* 26, 212–215. doi: 10.1121/1.1907309
- van Wassenhove, V., Grant, K. W., and Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proc. Natl. Acad. Sci. U S A* 102, 1181–1186. doi: 10.1073/pnas.0408949102
- Vroomen, J., and Stekelenburg, J. J. (2010). Visual anticipatory information modulates multisensory interactions of artificial audiovisual stimuli. *J. Cogn. Neurosci.* 22, 1583–1596. doi: 10.1162/jocn.2009.21308
- Wang, L., and Chu, M. (2013). The role of beat gesture and pitch accent in semantic processing: an ERP study. *Neuropsychologia* 51, 2847–2855. doi: 10.1016/j.neuropsychologia.2013.09.027
- Yasinnik, Y. (2004). *The Timing of Speech-Accompanying Gestures with Respect to Prosody*. MIT: Proceedings of From Sound to Sense, MIT.
- Zoefel, B., and VanRullen, R. (2015). Selective perceptual phase entrainment to speech rhythm in the absence of spectral energy fluctuations. *J. Neurosci.* 35, 1954–1964. doi: 10.1523/JNEUROSCI.3484-14.2015

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Biau and Soto-Faraco. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Atypical coordination of cortical oscillations in response to speech in autism

Delphine Jochaut^{1*†}, Katia Lehongre^{2†}, Ana Saitovitch³, Anne-Dominique Devauchelle⁴, Itsaso Olasagasti¹, Nadia Chabane⁵, Monica Zilbovicius³ and Anne-Lise Giraud¹

¹ Department of Neurosciences, University of Geneva, Geneva, Switzerland, ² Centre de Recherche de l'Institut du Cerveau et de la Moelle Epinière, INSERM UMR 975 – CNRS UMR 7225, Hôpital de la Pitié-Salpêtrière, Paris, France, ³ Unité Inserm 1000, Service de Radiologie Pédiatrique, Hôpital Necker – Enfants-Malades, AP-HP, Université Paris V René-Descartes, Paris, France, ⁴ Inserm U960, Département des Etudes Cognitives, Ecole Normale Supérieure, Paris, France, ⁵ Unité Multidisciplinaire pour la Santé des Adolescents, Centre Cantonal de l'Autisme, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland

OPEN ACCESS

Edited by:

Anne Keitel,
University of Glasgow and Max Planck Institute for Human Cognitive and Brain Sciences, UK

Reviewed by:

Mark T. Wallace,
Vanderbilt University, USA
Joachim Gross,
University of Glasgow, UK

*Correspondence:

Delphine Jochaut,
Department of Neurosciences,
University of Geneva, Campus Biotech, 9 Chemin des Mines,
1211 Geneva, Switzerland
delphine.jochaut@gmail.com

[†]These authors have contributed equally to this work.

Received: 16 December 2014

Accepted: 11 March 2015

Published: 27 March 2015

Citation:

Jochaut D, Lehongre K, Saitovitch A, Devauchelle A-D, Olasagasti I, Chabane N, Zilbovicius M and Giraud A-L (2015) Atypical coordination of cortical oscillations in response to speech in autism. *Front. Hum. Neurosci.* 9:171. doi: 10.3389/fnhum.2015.00171

Subjects with autism often show language difficulties, but it is unclear how they relate to neurophysiological anomalies of cortical speech processing. We used combined EEG and fMRI in 13 subjects with autism and 13 control participants and show that in autism, gamma and theta cortical activity do not engage synergistically in response to speech. Theta activity in left auditory cortex fails to track speech modulations, and to down-regulate gamma oscillations in the group with autism. This deficit predicts the severity of both verbal impairment and autism symptoms in the affected sample. Finally, we found that oscillation-based connectivity between auditory and other language cortices is altered in autism. These results suggest that the verbal disorder in autism could be associated with an altered balance of slow and fast auditory oscillations, and that this anomaly could compromise the mapping between sensory input and higher-level cognitive representations.

Keywords: speech processing, auditory cortex, cortical oscillations, oscillation coupling, autism

Introduction

Expressive and receptive language difficulties are frequently observed in autism and have long been a diagnostic symptom. The reason why children with autism inadequately respond to the speech of their closest relatives remains unexplained, presumably because (i) the genetic pattern in autism involves a complex combination of genetic and epigenetic factors (Peñagarikano et al., 2011; Peñagarikano and Geschwind, 2012; Murdoch and State, 2013; Pu et al., 2013); (ii) there is no consensus about how genetically altered corticogenesis could impact collective neuronal functioning and cognitive operations; and (iii) the neural mechanisms of speech processing are only partially understood. According to the DSM5 (American Psychiatric Association [APA], 2013), subjects with ASD exhibit “hyper- or hypo-reactivity to sensory input,” which could mean that speech and language deficits in autism reflect auditory (Edgar et al., 2013; Kujala et al., 2013) rather than (or in addition to) higher-level linguistic dysfunctions (Stevenson et al., 2014a).

We explored whether subjects with autism exhibit a neurophysiological deficit in speech processing (Eyler et al., 2012), basing some of our hypotheses on recent advances on the role of cortical oscillations in speech segmentation and decoding (Ghitza and Greenberg, 2009; Ghitza, 2011; Giraud and Poeppel, 2012; Gross et al., 2013; Doelling et al., 2014).

In autism, accelerated neocortical maturation (Courchesne et al., 2003) co-occurs with laminar disorganization in the temporal cortex (Jacot-Descombes et al., 2012) where it compromises the development of auditory and language micro- and macro-circuits (Eyler et al., 2012; Williams et al., 2013). Because cortical oscillations arise from laminar-specific interactions between excitatory and inhibitory neurons (Rotstein et al., 2005; Ainsworth et al., 2011; Whittington et al., 2011), the migration anomalies and local alterations of GABA inhibition observed in autism (Bartos et al., 2007; Tyzio et al., 2014) could directly interfere with the generation of important neurophysiological response patterns such as theta and gamma oscillations, preventing them from playing their expected parsing and decoding roles in speech processing. During development, such anomalies could delay language acquisition, because speech would evoke less reliable neural temporal patterns (Dinstein et al., 2012), compromising the interfacing between auditory cortex and the rest of the language network/other cognitive systems (Uhlhaas and Singer, 2007).

The syllabic structure of speech engages auditory cortical responses in the theta (4–7 Hz) frequency (Luo and Poeppel, 2007; Ghizta and Greenberg, 2009; Ding and Simon, 2013), and theta modulations typically influence gamma signals through nesting, a mechanism whereby the energy in gamma activity is controlled by the phase of theta activity (Schroeder et al., 2008). It is assumed that theta/gamma nesting enables speech decoding by orchestrating neural activity into a syllable-based code aligned on key phonemic events (Ghizta, 2011; Giraud and Poeppel, 2012). Critically, this model predicts that speech decoding is compromised if theta activity fails to track speech modulations (Ahissar et al., 2001; Luo and Poeppel, 2007; Ghizta, 2012) and to shape gamma activity (Giraud and Poeppel, 2012). Accordingly, reduced reactivity to voice modulations in autism (Gervais et al., 2004; Abrams et al., 2013) suggests a speech tracking dysfunction reflected in auditory theta activity. A consequence of this anomaly would be reduced down-regulation of gamma by theta activity, and less accurate speech parsing and encoding. Here we test the processing of speech by participants with autism and controls by recording concurrent EEG and fMRI data while they viewed an engaging documentary film.

Materials and Methods

Participants

Thirty-one subjects (adults and adolescents) participated in a combined EEG/fMRI study. Fifteen of these were identified as presenting with primary autism disorder with language impairments, diagnosed according to DSM-IV criteria, and further confirmed with the Autism Diagnostic Interview-Revised (Lord et al., 1994). We excluded subjects with infectious, metabolic, neurological, or genetic diseases, abnormal hearing levels, and those who were unable to stay confined and still in the MRI scanner. All subjects and their legal representative(s) provided written informed consent for participation in the study, which was approved by the local ethics committee (Biomedical Inserm protocol C08-39). We collected IQ measures (short form of the WAIS-III scale, Weschler, 2000) in all subjects, and autism-spectrum quotients

(AQs, Baron-Cohen et al., 2001) and the verbal component of the Autism Diagnostic Interview-Revised in all but three (one deeply dysphasic, one moderately dysphasic, and one for whom the parents did not give consent to the tests). The two ASD subjects with expressive difficulties were not taken into account in the statistics involving clinical data. These subjects are shown on the related figures, in order for readers to assess their relation to the group. Psychometric data are summarized in **Table 1**. Because we focused on low-level properties of auditory cortex as a possible basic sensory dysfunction in autism, we did not restrict our observations to the high IQ subpopulation with autism (Asperger) or to any specific autism subprofile.

Experimental Procedure

We explored auditory cortical processing during a passive, naturalistic task with a relatively low cognitive demand while both EEG and fMRI were being concurrently recorded. Subjects viewed a TV program for youth, selected to engage the interest of participants with ASD. The program was an audio-visual scientific documentary about the dangers of the sun on seashores (see

TABLE 1 | Psychometric data.

Group	Age	IQ	AQ	ADl _b	ADl _{nv}	ADl _v	ADl _d	Total ADI
ASD	20	50	Profoundly dysphasic	16	12	–	7	35
ASD	27	90	–	24	11	17	8	43
ASD	20	82	Moderately dysphasic	13	7	–	6	26
ASD	22	79	26	20	11	15	2	33
ASD	19	110	28	23	11	17	4	38
ASD	17	66	28	28	11	17	5	44
ASD	15	80	21	17	5	12	8	30
ASD	17	124	33	22	14	21	11	47
ASD	15	35	26	23	5	16	10	38
ASD	16	85	43	20	14	23	3	37
ASD	17	120	21	25	17	22	2	44
ASD	40	75	32	36	14	21	7	57
ASD	17	91	27	25	14	18	5	44
Control	20	106	19					
Control	27	97	20					
Control	19	97	13					
Control	23	102	14					
Control	16	114	12					
Control	20	105	6					
Control	17	112	12					
Control	18	123	9					
Control	17	92	11					
Control	38	94	7					
Control	20	97	14					
Control	40	95	10					
Control	16	127	9					

ADI, Autism Diagnostic Interview; ADl_b, stereotyped behavioral component of the Autism Diagnostic Interview; ADl_{nv}, non verbal component of the Autism Diagnostic Interview; ADl_v, verbal component of the Autism Diagnostic Interview; ADl_d, social component of the Autism Diagnostic Interview; AQ, Autism-Spectrum Quotient; IQ, intellectual quotient.

Movie S1), featuring three different speakers (two males) who made scientific demonstrations while talking to the audience, and occasionally to each other. Participants were asked to rest with eyes closed (movie off), or to watch the movie, in alternation, for short periods of 5 min, over three sessions (sessions one and two: 5 min of movie followed by 5 min of rest, and session three: 5 min of rest). To minimize the influence of the movie on the following resting state period we only analyzed the last 4.5 min of rest. The subjects were instructed to attentively watch the program and were informed that they would have to give a brief report about its content after the MRI sessions. They were also instructed to refrain from moving or falling asleep during the resting periods. Attention was controlled using EEG monitoring of the alpha rhythms and in some subjects by eye tracking. We also used EEG to track movement artifacts, and excluded three of the 31 subjects who exhibited more than one movement artifact per minute. We had to exclude two other subjects due to technical problems during the recordings (malfunction of the sound system and of the amplifier). The remaining 26 participants were comprised of 13 subjects with autism (mean age = 20.67 ± 6.77 years, mean IQ = 83.61 ± 25.27) and 13 control participants (mean age = 22.92 ± 8.14 , mean IQ = 104 ± 11.28) matched for age and not for IQ (**Table 1**). This sample size remains theoretically sufficient to detect medium to large effect sizes (Friston, 2012).

At the end of the scanning sessions, subjects were asked to report what the TV program was about and what the speakers' names were. All participants except for two subjects with autism (the two subjects with dysphasia) correctly reported that the movie was about the dangers of sun exposure, and correctly provided the names of the main speakers. The two subjects who did not provide satisfactory answers were excluded from analyses involving clinical variables, and were included in the neurophysiological analyses only after verification that they were not outliers [Grubbs' test for the theta and gamma parameter estimate variables, theta: mean = 0.017, SD = 0.035, $G(0.05) < 2.84$; gamma: mean = -0.0017 , SD = 0.109, $G(0.05) < 2.84$]. The neurophysiological effects were then related to clinical variables (AQ, ADI verbal communication component), while the non-verbal communication component served as a control variable.

MRI and EEG Acquisition and Preprocessing

Six hundred eighty echoplanar fMRI image volumes (Tim-Trio; Siemens, 40 transverse slices, voxel size = $3 \text{ mm} \times 3 \text{ mm} \times 3 \text{ mm}$; repetition time = 2,000 ms; echo time = 50 ms; field of view = 192) were acquired during the first two sessions, and 310 image volumes during the third one. Continuous EEG was simultaneously recorded with a 5 kHz sampling rate from 12 scalp sites (Easycap electrode cap, International 10–20 system: F3, F4, C3, C4, T7, T8, P3, P4, O1, O2, reference in Cz, Ground in AFz) using MR compatible amplifiers (BrainAmp MR and Brain Vision Recorder software; Brainproducts). One additional electrode for the electrocardiogram was placed under the left shoulder blade. Impedances were kept under $10 \text{ k}\Omega$, and EEG was time-locked with the scanner clock, which further reduced artifacts and resulted in higher EEG quality in the gamma band. A 7-min

anatomical T1-weighted magnetization-prepared rapid acquisition gradient echo sequence (176 slices, field of view = 256, voxel size = $1 \text{ mm} \times 1 \text{ mm} \times 1 \text{ mm}$) was acquired at the end of scanning.

We used statistical parametric mapping (SPM8; Wellcome Department of Imaging Neuroscience, UK¹) for fMRI standard preprocessing (realignment, coregistration with structural images, segmentation, and normalization in the Montreal Neurological Institute stereotactic space). The images were spatially smoothed using a 10-mm full-width half-maximum isotropic Gaussian kernel. Gradient and pulse artifacts were first detected and then marked using in-house software² that correlated the data with automatically (for gradient) or manually (for pulse) defined templates. Artifacts were corrected using PCA, using FASST v111017³ for gradient artifacts, and EEGLab v0.9⁴ for pulse artifacts. We excluded F3 and F4 from the analyses, as this pair of electrodes mostly captures the frontal eye field (Amiez and Petrides, 2009). Data were subsequently down-sampled to 250 Hz and re-referenced to a common average reference. The original reference electrode was recalculated as FCz, resulting in a total of 13 cortical electrodes. For each subject, periods with head movement artifacts were detected by visual inspection, and excluded as described in the EEG informed-fMRI section.

Analyses of the fMRI Dataset

We first analyzed the fMRI data set alone, using a general linear model (GLM) implemented in SPM8. We initially assessed whole brain activity at the single-subject level. The Gaussian distribution of the data allowed us to perform parametric tests. We included motion parameters and their first and second derivatives, the averaged signal of three separate brain compartments (white-matter, gray-matter, and CSF), and all out-of-brain voxels as nuisance covariates. In a second step, we selectively explored speech-related cortical responses by modeling the acoustic envelope of the speech part of the audiovisual sequence in the statistical analysis. The speech envelope was obtained by calculating the Hilbert transform of the stimuli and then filtering the magnitude of the result with a passband of 2–30 Hz. We verified for outliers showing task-related motion artifacts⁵, and further minimized spurious effects of head motion (Chase, 2014) by modeling head motion parameters and their first and second derivatives as covariates of no interest.

Contrast images (movie/rest) were created for each subject and entered into a second level analysis in which IQ was used as a nuisance variable (covariate). As the variance between the two groups was unequal, group differences between subjects with and without autism were assessed using 2-tailed two-sample *t*-tests for each condition. Each group comparison was masked by the relevant main effect of group. Due to a priori predictions of findings within Heschl's gyrus, we performed small volume corrections (SVCs) on the results within this region. The SVC was

¹www.fil.ion.ucl.ac.uk

²wiki.cenir.org/doku.php/datahandler

³www.montefiore.ulg.ac.be/~phillips/FASST.html

⁴[scrn.ucsd.edu/eeqlab](http://sccn.ucsd.edu/eeqlab)

⁵<http://cibsr.stanford.edu/tools/human-brain-project/artrepair-software.html>

done using an independently defined region of interest, anatomically defined with the aal atlas (implemented in xjview⁶). False positives in auditory cortex were further eliminated using an extend threshold >30 voxels for all analyses. For display purposes, we show whole-brain uncorrected statistics. All brain maps are displayed using MRIcron software⁷.

EEG-Informed fMRI

In a second step, we used combined fMRI and EEG to measure power fluctuations of rhythmic cortical activity and its topography in subjects with and without autism spectrum disorder. We used this approach to localize regions where blood oxygen-level dependent (BOLD) fluctuations systematically covary with EEG power fluctuations (Laufs et al., 2006; Giraud et al., 2007; Morillon et al., 2010). While the BOLD effect reflects overall synaptic activity (Logothetis, 2010), cortical oscillations – and in particular theta and gamma oscillations, as recorded with EEG, primarily denote activity involving pyramidal cells (Buzsáki et al., 2012). By combining the two recording techniques we determine the fraction of the BOLD effect that is linked to pyramidal cell activity, at theta and gamma rhythms, which are hypothesized to underpin speech parsing and syllable encoding (Giraud and Poeppel, 2012).

We used EEG power fluctuations in specific frequency bands of interest (averaged for theta over 4–7 Hz, and for low gamma over 30–40 Hz) to inform the fMRI analysis using a GLM. We performed time-frequency (TF) analyses on the EEG signal using a Morlet wavelets approach (Fieldtrip⁸). The TF structure of signals was computed at each channel for frequencies from 1 to 70 Hz, with a frequency step of 1 Hz and a time step of 0.1 s. The power time course of each channel and each frequency was converted to Z-scores after replacing values of previously detected periods of movement by NaNs (Not a Number). We removed further residual artifacts by also rejecting Z-values above 4. The transformed signal was then averaged over channels, Z-transformed a second time and NaNs were replaced by zeros. Finally, we averaged the transformed signal across frequencies and channels (but F3 and F4), and we used this signal in the subsequent EEG/fMRI analyses. This procedure is state-of-the-art and prevents the issue of having to make source inferences prior to the correlation with fMRI (Laufs et al., 2006). The log-transformed data were normally distributed, which allowed us to use standard parametric statistical tests (for example, paired *t*-tests and Pearson's correlations).

As both the theta- and gamma-informed MRI models showed significant effects in left auditory cortex during movie viewing, we assessed gamma and theta oscillations engagement during movie viewing (rest vs. movie) in each group in this region. We extracted the parameter estimates for each subject and each condition from the two regions where there were significant group effects at rest, and ran a two-way ANOVA (group × condition) for each model (theta and gamma).

⁶<http://www.alivelearn.net/xjview8/>

⁷www.sph.sc.edu/comd/rorden/mricron

⁸<http://fieldtrip.fcdonders.nl/>

fMRI-Informed EEG (Partial Correlations)

The previous analysis required that we specify frequency bands of interests. To establish the frequency specificity of the effects found with EEG-informed fMRI for the gamma and theta bands, we explored EEG-BOLD coupling across the whole EEG spectrum in the left auditory region that was more activated in control than ASD subjects during the movie in the fMRI-only analysis. We also explored this coupling in the left visual region that was over-activated during the movie (fusiform gyrus) as a control for the specificity of auditory effects.

For both these regions, we correlated the BOLD time course with EEG power fluctuations across the 1–70 Hz spectrum [resulting from the TF analyses and convolved with the hemodynamic response function (HRF) after concatenation of the three-rest or two-movie sessions]. We modeled head-motion parameters, their derivatives, the averaged signal of white-matter, gray-matter and CSF and out-of-brain voxels as covariates of no interest. Resulting correlation values were Fisher Z-transformed, and standard statistics were performed on a near Gaussian population.

Correlation of Neurophysiological and Clinical Variables

We assessed the covariation of theta and gamma informed-BOLD responses in the left auditory cortex (and in the right auditory cortex as a control), where we detected a group difference in both theta and gamma models. We tested for a dependence of gamma and theta activity in each group using the Pearson's correlation test. For each hemisphere, we then performed a univariate analysis of covariance (ANCOVA) with gamma-BOLD parameter estimates as the dependent factor and theta-BOLD parameter estimates as covariates (as we assume gamma activity to be controlled by theta activity). We used the theta × gamma interaction term to test for correlations (Pearson's correlation test) with clinical variables (AQ, Baron-Cohen et al., 2001, the verbal component of the ADI-R and the non-verbal communication component of the ADI-R). Finally we addressed whether the relation between the theta-gamma interaction variable and the AQ was different between groups, using an ANCOVA with AQ as the dependent factor and theta-gamma variable as a covariate. All analyses were carried out with SPSS (IBM Corp. Released 2011. IBM SPSS Statistics for Windows, Version 18.0, Armonk, NY, USA).

Oscillation-Based Connectivity Analyses

Finally, we explored oscillation-based connectivity (Morillon et al., 2010) within the language network in each hemisphere. The underlying assumption is that the broad-spectrum oscillatory pattern at rest in one region determines the oscillatory pattern during movie viewing in another region only if the two regions interact functionally by exchanging information in specific frequency bands (Fries, 2009; Morillon et al., 2010). We assessed the degree of similarity of EEG power-BOLD broad-spectrum between rest and movie across nine cortical language regions. The primary motor regions (BA4a and BA4p), the planum temporale (Wernicke's region: Te3), the ventral prefrontal cortex (Broca's

region: BA44 and BA45), and the rostral inferior parietal cortex BA40 (merged PFop, PFt, PF, PFm, and PFcm) were spatially defined using probabilistic cytoarchitectonic maps using the SPM anatomy toolbox v.1.6. To delineate auditory regions, including Heschl's gyrus (BA41/BA42), the middle temporal gyrus (BA21) and the caudal inferior parietal cortex (BA39), we used the aal atlas implemented in xjview based on a macroscopic anatomical parcellation of the MNI MRI Single-Subject Brain⁶.

Pearson's correlations across the nine regions were computed between rest and movie conditions, from the EEG-BOLD partial correlation values (1–70 Hz) obtained for each region and subject [see fMRI-Informed EEG (Partial Correlations)]. We obtained two matrices (one per group) consisting of one correlation value per region and subject, reflecting the spectrum similarity between conditions. Statistical significance of the correlation values of each matrix was tested using one-sample *t*-tests. The resulting two matrices of significant (positive and negative) correlations were then compared between groups using two-tailed two-sample *t*-tests (**Figure 4C**). We previously argued (Morillon et al., 2010) that such a matrix may be interpreted in a directional way, under the double assumption that (i) the oscillatory profile observed in a given region at rest determines the oscillatory profile observed in regions that receive its input during the movie, and (ii) the resting profile in one region cannot be explained out by the movie profile in another region of the same functional network. Significant differences between groups are represented in **Figure 4D**. This matrix can be interpreted in

a directional way, as we hypothesize that the resting state profile determines lateralization of the language network during the movie. Arrows pointing from one brain region A to another brain region B indicate significant differences between the EEG-BOLD spectrum at rest in area A and the pattern in area B during movie viewing between groups. Note that in **Figure 4** the different territories corresponding to one functional area were pooled together to facilitate visualization (i.e., BA4, Broca). All statistical analyses were performed using Matlab v11b (The MathWorks Inc., Natick, MA, USA).

Results

We first analyzed the fMRI data using a simple contrast of movie vs. rest in each group. BOLD responses to the movie occurred in visual and auditory brain areas in both groups, yet were less pronounced in the ASD group in left superior parietal and superior temporal cortices (auditory cortex, **Figure 1A**). Conversely, movie-related BOLD activity was enhanced in autism relative to controls in bilateral non-primary visual cortex and the right posterior superior temporal sulcus (**Figure 1B**).

To more precisely characterize the reduced auditory cortical response in ASD, we computed a regressor from the temporal envelope of the movie soundtrack. This regressor primarily indexes syllable boundaries in the speakers' discourse (Ghitza,

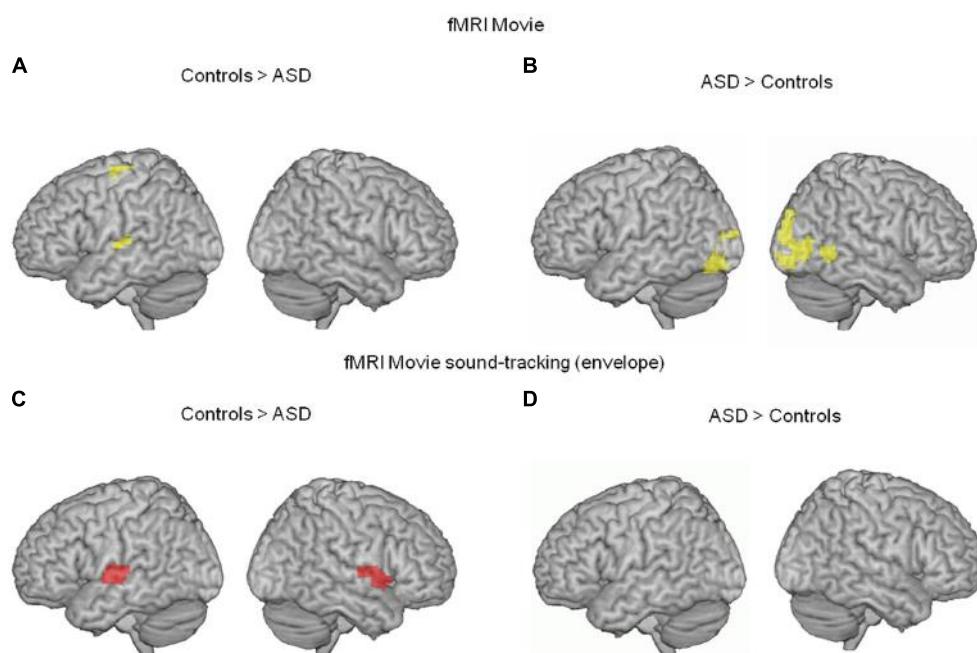


FIGURE 1 | (A,B) Comparison of neural activity blood oxygen-level dependent (BOLD) in 13 subjects with autism and 13 unaffected controls, when they were watching a scientific TV program (vs. rest). In autism, neural activity was reduced in the left supplementary motor area and left auditory cortex (whole brain interaction $p < 0.02$, uncorrected; Heschl's gyrus, $p = 0.02$ FWE corrected; 54, -19, 7 MNI coordinates), and enhanced in bilateral visual cortex (whole brain

$p < 0.01$ uncorrected, 3, -91, -8 MNI coordinates). **(C,D)** To further specify the effect, we compared the BOLD response to the movie soundtrack envelope, i.e., the syllabic envelope of speakers' speech, between the two groups. In ASD, the left auditory cortex showed reduced syllable tracking (-63, -13, 7 MNI coordinates, left Heschl's gyrus $p = 0.016$ FWE corrected; $p = 0.156$ uncorrected in right Heschl's gyrus, 62, -22, 7 MNI coordinates).

2012). Critically, because there was continuous speech throughout the movie with an alternation between off-voices and speakers facing the audience, the regressor was specific to speech and controlled for concurrent visual processing of faces. ASD participants showed a deficit in speech envelope tracking, as assessed by the BOLD signal, in a region of auditory cortex that overlapped with the region showing a global response deficit to the movie (**Figures 1C,D**). These initial two analyses of the fMRI data alone indicate deficient auditory processing in ASD, and show that this deficit is related to atypical speech tracking at the syllabic timescale.

A quantitative reduction in speech tracking as observed in the fMRI data could be a consequence of the failure of slow speech modulations to engage theta-range activity in auditory cortex during speech stimulation (Ghitza, 2012; Peelle et al., 2013). We therefore next addressed whether in ASD EEG anomalies in the theta range were associated with the inability of auditory cortex to optimally represent the soundtrack envelope. The simultaneous EEG and fMRI recordings allowed us to explore how theta power fluctuations driven by the movie correlate with local synaptic activity in auditory cortex, as indexed by the BOLD signal (Magri et al., 2012; see Materials and Methods). In both groups,

theta-BOLD coupling localized to bilateral superior temporal gyri (**Figure S1**).

Stronger theta-BOLD coupling in young adults with autism relative to controls was detected during the movie in left Heschl's gyrus [$p = 0.03$, familywise error (FWE) corrected in Heschl's gyrus] at the anterior border of the auditory cortex (**Figure 2A**, blue). This effect spatially overlapped with the envelope-tracking deficit as defined using fMRI responses to the movie (**Figures 1C** and **2A**). We then went on to compare theta EEG-BOLD coupling at rest and during the movie, in the auditory cortex region where there was a significant theta EEG-BOLD effect in controls during the movie (anterior to auditory cortex). In this region subjects with autism had enhanced resting theta-BOLD coupling relative to controls, and theta-BOLD coupling did not increase when they were exposed to speech (**Figure 2B**, top panel). In sum, unlike in controls, theta activity was already present in auditory cortex at rest and did not increase with speech stimulation. Note, however, that we observed a non-significant theta-BOLD coupling increase at 8 Hz in subjects with ASD during the movie. This small effect was hence outside the typical 4–7 Hz theta range. Taken together, our data indicate that subjects with autism have abnormal theta responses to speech. As it has been established

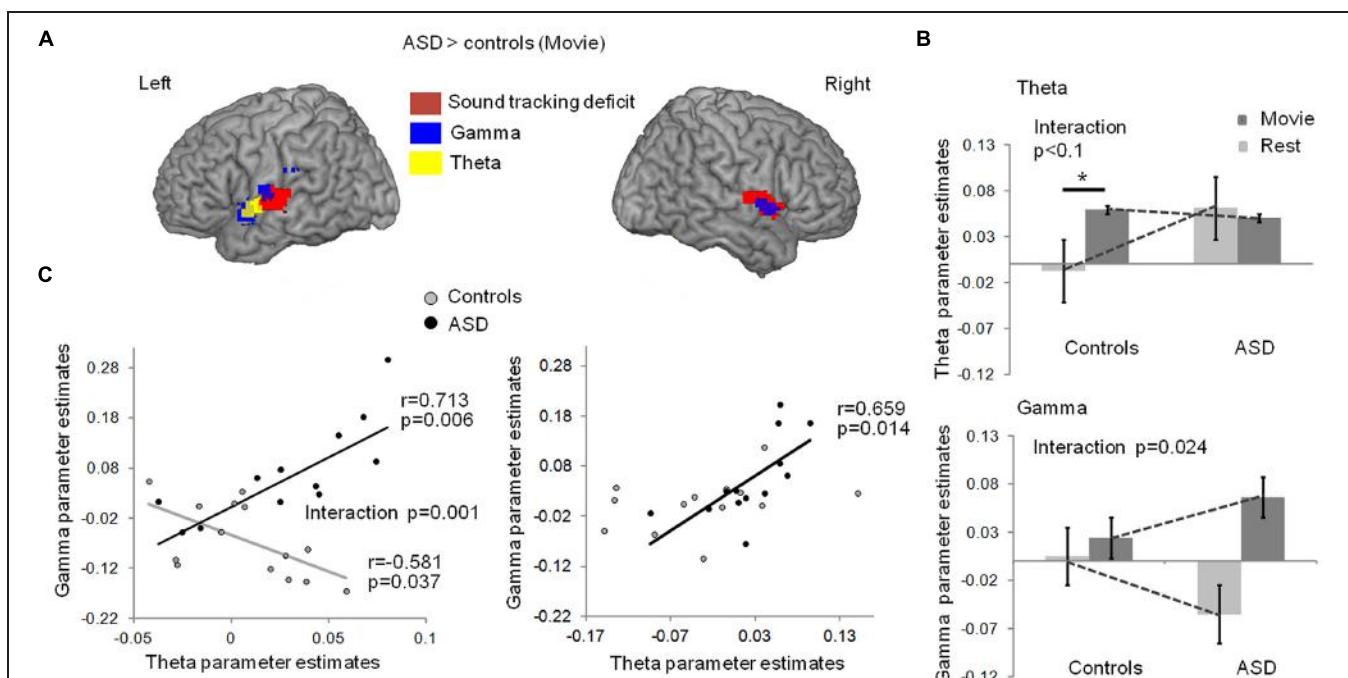


FIGURE 2 | (A) Comparison of EEG-BOLD coupling between groups with and without autism, in theta and low-gamma bands, during movie viewing. Subjects with autism had enhanced theta-BOLD (blue, whole brain $p < 0.01$ uncorrected, $-48, -1, -5$ MNI coordinates; left Heschl's gyrus, $p = 0.034$ FWE) and gamma-BOLD (green, left panel, $p < 0.01, -54, -7, 10$ MNI coordinates; left Heschl's gyrus $p = 0.007$ FWE) coupling in the left superior temporal lobe relative to controls; subjects with ASD had enhanced gamma-BOLD coupling (green, right panel, $p < 0.05, 51, -1, 1$ MNI coordinates) in the right temporal lobe relative to controls. **(B)** EEG-BOLD coupling at rest and during movie viewing in each group, within the theta (up panel) and gamma (bottom panel) frequency bands. The regions were

sampled from the left auditory cortex, at the location where there was a significant theta EEG-BOLD effect during the movie in controls (up panel), and a significant decrease in gamma-BOLD coupling at rest in the ASD group (bottom panel). **(C)** Left panel: in controls, gamma- and theta-BOLD coupling in left auditory cortex were negatively related, in line with a control of gamma by theta activity. In autism, an inverted relation suggests atypical theta/gamma interaction. The group interaction was significant at $p = 0.001$; Right panel: in the right temporal lobe, the anomaly in autism was less pronounced and the negative correlation between theta and gamma was not present in controls ($p = 0.243$). * indicates a significant difference with $p < 0.05$.

that speech intelligibility depends on the strength of theta phase-locking to the most prominent modulations in speech (Peelle et al., 2013) that typically occurs at 4 Hz, atypical theta engagement in response to speech could be one key contributing factor to explain anomalies of language processing in autism (Eyler et al., 2012).

Theta activity has been argued to be important in speech decoding (Luo and Poeppel, 2007; Henry and Obleser, 2012) because, among other reasons, it orchestrates gamma activity and the timing of cortical population spiking (Kayser et al., 2012). Mechanistically, this orchestration might serve to package information in time frames that can be read out and decoded at the next hierarchical stage (Shamir et al., 2009; Ghitza, 2011). We therefore, in a next step, addressed the distribution of gamma power/BOLD correlations throughout the brain during the movie (**Figure S1**). We found that gamma power/BOLD correlations were enhanced in subjects with autism relative to controls in bilateral auditory cortices, in particular in the left auditory cortex, at its junction with the supramarginal region in the upper bank of the Sylvian fissure and the insula (**Figure 2A**, green). In these regions, the group difference was significant at $p = 0.007$ (FWE corrected). This effect overlapped with the region where (i) BOLD activity was reduced in ASD during the movie (**Figure 1A**), (ii) speech envelope tracking by fMRI responses was deficient (**Figures 1B** and **2A**), and (iii) theta-correlated BOLD signal was atypical (**Figure 2A**). Controls displayed a weak gamma-BOLD coupling at rest that only moderately increased during the movie. This suggests that the movie induced a temporal reorganization of gamma activity, presumably via theta activity, rather than strong power variations (Benchenane et al., 2010; Kayser et al., 2012). By contrast, subjects with autism showed a marked negative gamma-BOLD coupling at rest and a stronger than normal positive gamma-BOLD coupling during the movie (**Figure 2B**, bottom panel, group \times condition interaction $p = 0.024$), confirming abnormal gamma generation (Edgar et al., 2013) and reactivity to sound modulations.

To ascertain the specificity of these effects for the theta and gamma bands, we explored EEG-BOLD coupling across the full recorded EEG spectrum, focusing principally on the left auditory region that was more activated in control than ASD subjects during movie viewing. We observed significantly enhanced EEG-BOLD coupling in autism during movie viewing between 25 and 35 Hz (**Figure S2**, left panel), i.e., in a range previously related to phonemic processing by auditory cortex (Lehongre et al., 2011). As an additional control for the auditory specificity of theta and gamma effects, we computed correlations between whole spectrum EEG and BOLD signal in the left visual region that was over-activated during the movie (fusiform gyrus, **Figure 1**). In this occipital region, we observed a non-significant reduction in gamma-BOLD correlations in ASD relative to controls (**Figure S2**, right panel). This control offers qualitative support to recent studies showing that gamma activity is reduced in ASD relative to controls in response to faces (Khan et al., 2013). Importantly, such data show that synaptic activity as indexed by the BOLD signal does not systematically translate into strong oscillatory effects (Logothetis, 2010), such as those we observe in the left auditory cortex.

Critically, as our speech processing model assumes that the modulation of gamma activity by theta activity is essential to speech comprehension (Giraud and Poeppel, 2012), we explored how theta and gamma power fluctuations covaried during movie viewing, in Heschl's gyrus. Because scanner noise and motion artifacts more strongly affected phase than power signals, we could not directly assess theta/gamma phase-amplitude coupling. Instead we approximated theta-gamma power relationship by regressing the gamma-BOLD parameter estimates onto the theta-BOLD ones. We observed a negative relationship in controls in left auditory cortex [$r(13) = -0.58$, $p = 0.037$, **Figure 2C**, left], confirming a functional dependency between theta and gamma under physiological conditions, compatible with gamma activity being down-regulated by theta activity. In autism this dependency was reversed [$r(13) = 0.7$; $p = 0.006$, group \times frequency-range interaction significant at $F(1,22) = 15.767$; $p = 0.001$], suggesting atypical coordination between gamma and theta activity, presumably in relation to an absence of down-regulation. The group interaction was not significant in the right temporal cortex [$F(1,22) = 0.872$; $p = 0.361$] where controls showed no theta-gamma dependency (**Figure 2C**, right), in line with the specificity of left auditory cortex for speech processing (Gross et al., 2013).

We next investigated the relation between the severity of autism clinical symptoms and the observed anomalies of oscillatory responses to speech in auditory cortex. We constructed a neurophysiological variable that combined theta and gamma activity. Because theta and gamma variables were not independent, we excluded a linear combination of gamma and theta parameter estimates (theta + gamma + theta \times gamma), but correlated the behavioral data with the interaction term (theta \times gamma), which is sensitive to the sign of the correlation. This composite variable predicted subjects' verbal scores on the ADI test [$r(11) = 0.746$; $p = 0.008$], but only weakly correlated with non-verbal scores in the ASD group (**Figures 3A,B**). This observation is consistent with the view that the absence of a canonical theta-gamma dependency is specifically related to language difficulties. Note that no such effects were present in right auditory cortex (group \times hemisphere interaction $p < 0.001$). Interestingly, the theta \times gamma variable also predicted the AQ across groups [$r(23) = 0.68$, $p = 0.000$, **Figure 3C**], reflecting the large and significant group difference in theta/gamma coupling (**Figure 3D**). Most importantly, the neurophysiological index of theta-gamma dependency was strongly tied to the autism symptoms; within the ASD group the correlation attained $r(10) = 0.924$; $p = 0.000$, with a group interaction of $F(1,19) = 10.135$; $p = 0.005$.

Finally, we assessed how the oscillatory spectral profiles and theta-gamma relationship observed at rest in left auditory cortex related to effects in the distributed language network during the movie. We computed EEG (1–70 Hz)-BOLD coupling from nine language regions of the left hemisphere during rest and movie viewing and correlated it across the two conditions (**Figure 4**). We interpret these findings as directional oscillation-based connectivity, under the hypothesis that EEG-BOLD coupling at rest predicts EEG-BOLD coupling during movie viewing (see Materials and Methods and Morillon et al., 2010). The notion of connectivity is based on the capacity of one region to inherit,

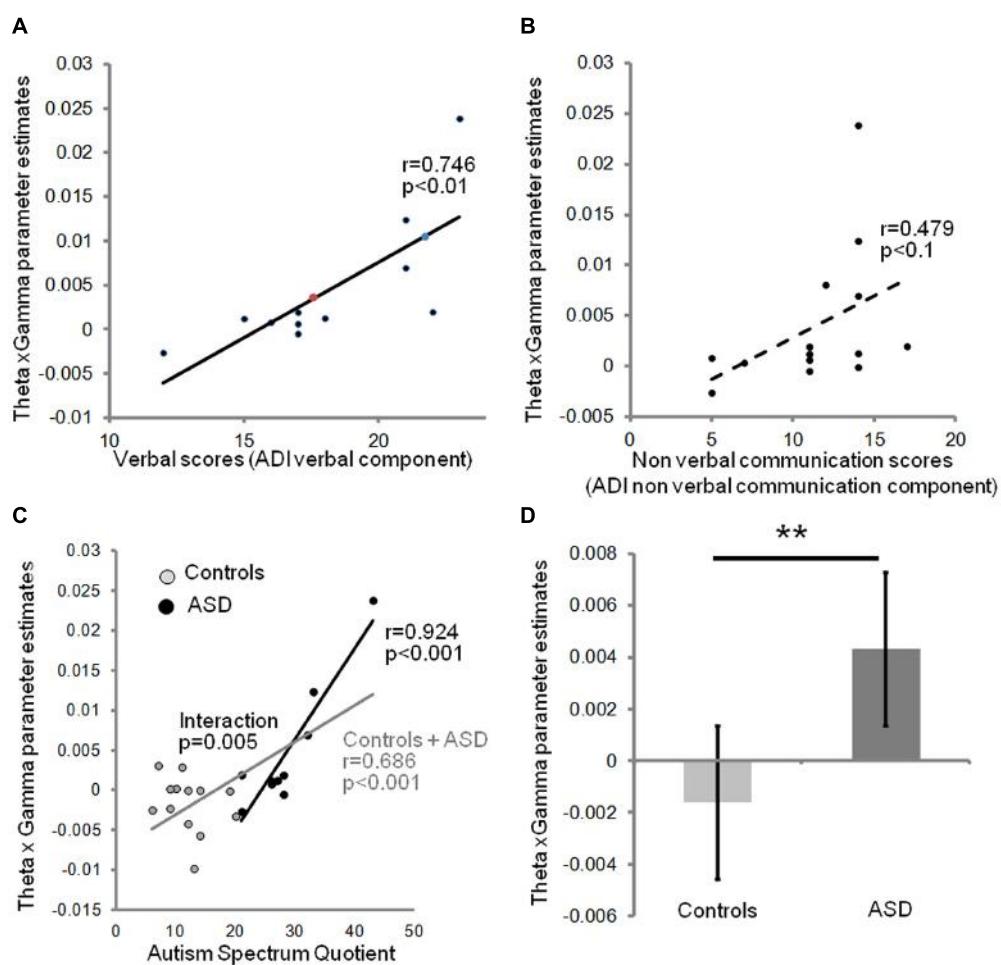


FIGURE 3 | Relationship between theta/gamma-BOLD interaction term, and clinical data: the neurophysiological theta \times gamma variable predicts the verbal component of ADI (A), but not the non-verbal communication component (B). Colored dots indicate the predicted ADI score of the dysphasic subjects: the blue one was deeply dysphasic and the red one was moderately dysphasic; note the high ADI predicted score of the non-verbal one. The theta \times gamma variable closely predicts autism severity

assessed by autism spectrum quotients (AQs) in the affected group, and autism traits across groups, but not within the control group (C). There is a significant group interaction ($p = 0.005$) confirming a non-linear shift in the dependency between autism traits and auditory theta/gamma coupling when moving to AQ scores associated with ASD. (D) Theta \times gamma parameter significantly distinguishes the two groups (Spearman t -test, $p = 0.01$). Error bars represent SEM. ** indicates a significant difference with $p < 0.01$.

during the movie part of the experiment, the oscillatory profile observed at rest in another region. Using this approach, we observed that left auditory cortex was more weakly coupled to Broca's area (BA 44/45), BA39 and 40, and the premotor cortex in ASD than in controls. This pattern suggests that the propagation of the broad-spectrum oscillatory profile in auditory cortex to key regions of the language network was reduced in subjects with autism relative to controls (Figure 4). Importantly, there was reduced connectivity from A1 (BA41/BA42) to Broca's area and motor cortex, but not from Broca's area and motor cortex to A1, indicating that the anomaly is likely primarily auditory. Because the oscillatory profile determines the time constants with which speech is segmented, and the neural code presented to higher order language brain regions due to temporally spike reorganization, functional isolation of auditory cortex should strongly impair on-line speech decoding.

Discussion

The current findings show severe anomalies of auditory cortical activity at rest and in response to speech in subjects with ASD, affecting conjointly the theta and the low-gamma frequency bands. Cortical oscillations arise from excitatory-inhibitory interactions within and across specific cortical laminae (Cannon et al., 2014), and auditory oscillation anomalies represent a plausible functional counterpart to the structural disorganization of language cortices and the disruption of cortical inhibition previously shown in ASD (Rojas et al., 2013). When subjects with ASD engaged in natural activities that do not place specific emphasis on social functions, their left speech processing regions manifested a primary deficit. In ASD, the auditory cortex reacted less to speech syllabic modulations, which were also weakly tracked by theta oscillations. Such an anomaly could

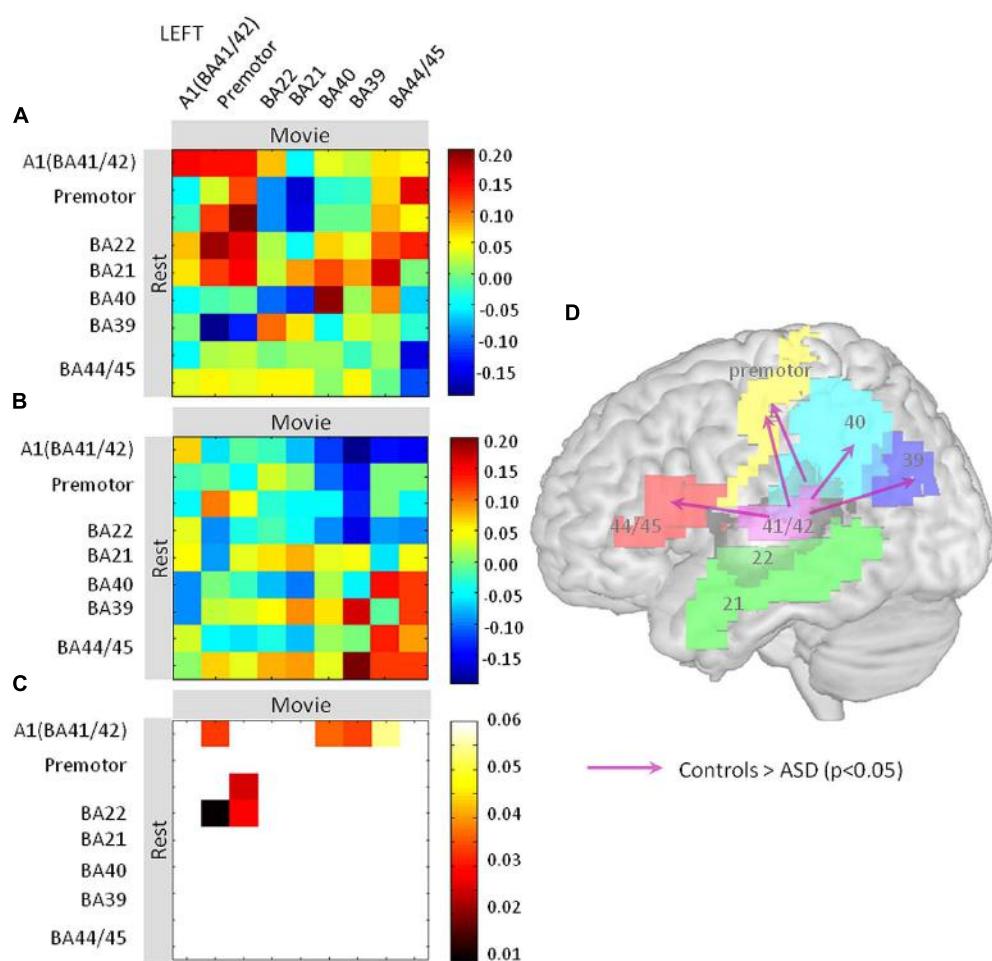


FIGURE 4 | Oscillation-based functional disconnection of auditory cortex in autism. **(A,B)** We quantified how the oscillatory spectrum in auditory cortex at rest propagated to the other regions of the left lateralized language network during the movie, in controls **(A)** and in ASD subjects **(B)**. We used Pearson's correlation matrices of the EEG-BOLD correlations over the whole spectrum (1–70 Hz) across nine regions of the language system. The results are

directional under the non-reciprocal assumption that the resting oscillatory profile is a determinant of the oscillatory profile during the movie. **(C)** Similarity matrices were compared between groups and statistics are reported at $p < 0.05$. **(D)** Statistics are reported on a brain template; they show that the primary auditory cortex (BA41/42) has reduced connectivity with Brodmann Areas 44, 45, 40, 39, and the premotor cortex relative to controls.

have severe functional consequences on speech perception, since disrupted theta tracking of speech modulations results in less efficient syllable encoding and reduced intelligibility (Ahissar et al., 2001; Henry and Obleser, 2012; Peelle et al., 2013; Doelling et al., 2014). From a theoretical perspective, intelligibility difficulties could occur because atypical theta tracking compromises syllable parsing, a process by which theta oscillations locked on syllable onsets determine syllable-based windows of integration, and temporally organize the neural activity that is passed to higher hierarchical processing levels (Ghitza, 2011). The current results are consistent with recent findings showing enlarged temporal windows of integration in audio-visual speech in autism (Stevenson et al., 2014b). In control subjects, there is a 250–300 ms tolerance to audio and visual asynchrony in speech, suggesting that visual and sound tracks could be integrated via a theta-based mechanism (Luo et al., 2010). In subjects with autism, the sensitivity to audio-visual speech asynchrony is dramatically blurred, with

temporal windows of integration reaching up to 1 s (Stevenson et al., 2014b). These observations converge with ours to suggest a severe disruption of theta-based speech integration mechanisms in autism.

Our data further show that speech-driven theta and gamma neural oscillations lack the typical physiological coordination. Unlike in controls, there was no sign of down-regulation of gamma by theta activity during speech processing in ASD, but rather an opposite dependency such that gamma and theta-BOLD coupling jointly increased out of physiological ranges. According to recent oscillation-based models of speech processing (Giraud and Poeppel, 2012), dysfunctional theta/gamma coordination should disrupt the alignment of neuronal excitability with syllabic onset, and compromise speech decoding. The observation that theta/gamma balance was not merely disrupted in ASD but reversed relative to controls could signal pathological wiring patterns within/across cortical microcircuits in autism.

We also found that the atypical interaction between theta and gamma responses to speech strongly correlated with clinical variables. The oscillatory anomalies matched not only the verbal impairment, but more broadly the severity of autism symptoms. These findings underscore the central place of sensory anomalies in ASD (Marco et al., 2011) and open up a possibility to consider sensory disturbances in relation to the complex spectrum of cognitive symptoms. By reducing the ability to temporally organize speech information, altered coordinated neural activity in auditory cortex and disrupted oscillation-based connectivity with Broca's area and motor cortex likely compromise the ability of ASD subjects to respond appropriately to speech signals and to interact with their peers. As illustrated by the increased prevalence of the autism phenotype in children with profound hearing loss (Snowling et al., 2003), auditory-based communication appears of crucial importance for normal cognitive development, and dysfunctional auditory processing could contribute to the social isolation of subjects with autism.

On the other hand, dysfunctional speech-related neural processing in the autistic brain might also denote a deficiency of oscillation coordination, based on temporal integration deficits, that reaches beyond the auditory modality. Given the broad spectrum of sensory and cognitive symptoms in autism, anomalies of oscillatory entrainment and coupling may be more pervasive than currently appreciated. Adjudicating between a primary auditory deficit and a generic deficit of oscillatory function in autism (Dinstein et al., 2011) would require in-depth investigations of oscillatory brain responses in other functional domains besides speech processing. In our present study, some preliminary evidence in favor of a primary impairment of auditory integration may come from the observation that abnormal synaptic activity levels (fMRI) and oscillation anomalies co-occurred in auditory cortex, but were dissociated in visual regions. This observation could indicate a compensatory role of visual processing in autism during speech perception, as supported by the observation that subjects with ASD extensively explore the mouth region in face-to-face situations (Klin et al., 2002), and use specific attention modes to enhanced local visual processing (Schwarzkopf et al., 2014).

Although the current findings point to a primary dysfunction of oscillatory activity resulting in a speech-tracking deficit, this study, unsurprisingly, has some important limitations. First, given that we had to select subjects who could stay confined in an MRI scanner with the EEG equipment, our ASD sample is relatively small, which necessarily limits the generalizability of the findings. A second limitation might be seen in the fact that the two groups of subjects were not matched for IQ as is usually the case in cognitive studies of autism. However, because our hypotheses focused on low-level automatic auditory tracking by auditory cortex, we chose to include all levels of autism, and IQs, only excluding subjects with Asperger syndrome. Our aim here was to work with a sample representative of the diversity of the autism population, which implies IQ and speech proficiency differences with the control group. However, by correcting the statistics for IQ we finally report group effects that are not primarily explained by this factor.

That the results were confined to auditory cortices indicate that this strategy was useful. The broad spectrum of autism severity used here also provided a good sensitivity in correlation analyses, and even the dysphasic ASD subjects were not detected as outliers (**Figures 2 and 3**). A third potential confound when comparing groups lies in head motion, which is invoked as a strong bias in neuroimaging findings (Pelphrey and Deen, 2012). Thanks to EEG recordings, we most likely circumvented this potentially serious issue by excluding subjects and recording periods showing motion artifacts, and we further corrected for head motion parameters in the statistics after verifying that there was no residual outlier for motion. Combining EEG with fMRI makes it unlikely that artifacts of dual origin are reflected in a differential activity precisely in auditory cortices. Finally, although we used EEG to check that subjects were not asleep during the experiment, we cannot ensure that subjects with and without autism maintained comparable levels of auditory attention. There is even reason to believe that – if our hypothesis that subjects with autism have reduced ability to follow speech signals due to oscillatory dysfunction is correct – there should be detrimental consequences for implementing auditory attentional control. It has been shown that auditory attention acts by phase resetting slow oscillations (Kayser, 2009; Zion Golumbic et al., 2013), which in turn enhances the control of gamma by theta oscillations (Sauseng et al., 2008). From a neurophysiological perspective, one can therefore expect speech tracking and attentional mechanisms to be inherently intertwined.

The current findings support a model that relates the coordination of cortical oscillations to temporal integration of the sensory input. The data could be useful for understanding the exact pathogenetic mechanisms of abnormal sensory reactivity in autism. Whether restricted to the auditory modality or more widespread, the lack of coordination across slow (theta) and fast (gamma) oscillations suggests a deficit in information integration at two timescales that could also have important consequences on the ability to manipulate mental representations of different orders, here phonemes and syllables. The present study should be considered as a first attempt at understanding whether speech related oscillatory activity was impaired in autism, and is expected to be followed by others that should clarify the relationship between phenotype and neurophysiology, using detailed evaluation of linguistic skills. Finally, speech reception disturbances in ASD could constitute an interesting possible entry point to clinical handling, as oscillatory activity can be focally modulated, e.g., by neuro-feedback or non-invasive transcranial stimulation (Engelhard et al., 2013).

Acknowledgments

We thank David Poeppel and Narly Golestani for their comments on the manuscript. This work was funded by the European Research Council (Grant 260347 to A-LG) and the Orange Foundation (PhD grant to DJ), and the ANR-10-LABX-0087 IEC; ANR-10-IDEX-0001-02 PSL.

Supplementary Material

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fnhum.2015.00171/abstract>

Figure S1 | Whole brain topography of theta-BOLD (top) and low-gamma-BOLD (bottom) effects in controls (left) and subjects with autism during the movie (whole brain, $p < 0.01$ uncorrected; whole brain,

References

- Abrams, D. A., Lynch, C. J., Cheng, K. M., Phillips, J., Supekar, K., Ryali, S., et al. (2013). Underconnectivity between voice-selective cortex and reward circuitry in children with autism. *Proc. Natl. Acad. Sci. U.S.A.* 110, 12060–12065. doi: 10.1073/pnas.1302982110
- Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., and Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc. Natl. Acad. Sci. U.S.A.* 98, 13367–13372. doi: 10.1073/pnas.201400998
- Ainsworth, M., Lee, S., Cunningham, M. O., Roopun, A. K., Traub, R. D., Kopell, N. J., et al. (2011). Dual γ rhythm generators control interlaminar synchrony in auditory cortex. *J. Neurosci.* 31, 17040–17051. doi: 10.1523/JNEUROSCI.2209-11.2011
- American Psychiatric Association [APA]. (2013). *Diagnostic and Statistical Manual of Mental Disorders*, 5th Edn. (Arlington: American Psychiatric Association).
- Amiez, C., and Petrides, M. (2009). Anatomical organization of the eye fields in the human and non-human primate frontal cortex. *Prog. Neurobiol.* 89, 220–230. doi: 10.1016/j.pneurobio.2009.07.010
- Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., and Clubley, E. (2001). The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J. Autism Dev. Disord.* 31, 5–17. doi: 10.1023/A:1005653411471
- Bartos, M., Vida, I., and Jonas, P. (2007). Synaptic mechanisms of synchronized gamma oscillations in inhibitory interneuron networks. *Nat. Rev. Neurosci.* 8, 45–56. doi: 10.1038/nrn2044
- Benchenane, K., Peyrache, A., Khamassi, M., Tierney, P. L., Gioanni, Y., Battaglia, F. P., et al. (2010). Coherent theta oscillations and reorganization of spike timing in the hippocampal-prefrontal network upon learning. *Neuron* 66, 921–936. doi: 10.1016/j.neuron.2010.05.013
- Buzsáki, G., Anastassiou, C. A., and Koch, C. (2012). The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. *Nat. Rev. Neurosci.* 13, 407–420. doi: 10.1038/nrn3241
- Cannon, J., McCarthy, M. M., Lee, S., Lee, J., Börgers, C., Whittington, M. A., et al. (2014). Neurosystems: brain rhythms and cognitive processing. *Eur. J. Neurosci.* 39, 705–719. doi: 10.1111/ejn.12453
- Chase, A. (2014). Brain imaging: white matter disruption in autism spectrum disorder is exaggerated by head movements during neuroimaging. *Nat. Rev. Neurosci.* 10, 122. doi: 10.1038/nrn.2014.20
- Courchesne, E., Carper, R., and Akshoomoff, N. (2003). Evidence of brain overgrowth in the first year of life in autism. *JAMA* 290, 337–344. doi: 10.1001/jama.290.3.337
- Ding, N., and Simon, J. Z. (2013). Robust cortical encoding of slow temporal modulations of speech. *Adv. Exp. Med. Biol.* 787, 373–381. doi: 10.1007/978-1-4614-1590-9
- Dinstein, I., Heeger, D. J., Lorenzi, L., Minshew, N. J., Malach, R., and Behrmann, M. (2012). Unreliable evoked responses in autism. *Neuron* 75, 981–991. doi: 10.1016/j.neuron.2012.07.026
- Dinstein, I., Pierce, K., Eyler, L., Solso, S., Malach, R., Behrmann, M., et al. (2011). Disrupted neural synchronization in toddlers with autism. *Neuron* 70, 1218–1225. doi: 10.1016/j.neuron.2011.04.018
- Doelling, K. B., Arnal, L. H., Ghitza, O., and Poeppel, D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage* 85(Pt 2), 761–768. doi: 10.1016/j.neuroimage.2013.06.035
- $p < 0.005$ uncorrected; left Heschl's gyrus $p < 0.05$ FWE corrected.**
- Figure S2 | (A)** Neural activity (fMRI only, top panels) in subjects with autism (right) and controls (left) during audio-visual presentation of the documentary ($p < 0.05$, corrected). **(B)** Partial correlations between the EEG power spectrum (1–70 Hz) and fMRI data at the left hemispheric locations where we found reduced (auditory cortex) and enhanced (visual cortex) neural activity. Note that, unlike the effect in left auditory cortex, the effect in the right posterior superior temporal sulcus is not explained by differences in EEG-BOLD correlations.
- Movie S1: Video 1.**

- Klin, A., Jones, W., Schultz, R., Volkmar, F., and Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Arch. Gen. Psychiatry* 59, 809–816. doi: 10.1001/archpsyc.59.9.809
- Kujala, T., Lepistö, T., and Näätänen, R. (2013). The neural basis of aberrant speech and audition in autism spectrum disorders. *Neurosci. Biobehav. Rev.* 37, 697–704. doi: 10.1016/j.neubiorev.2013.01.006
- Laufs, H., Holt, J. L., Elfont, R., Kramm, M., Paul, J. S., Krakow, K., et al. (2006). Where the BOLD signal goes when alpha EEG leaves. *Neuroimage* 31, 1408–1418. doi: 10.1016/j.neuroimage.2006.02.002
- Lehongre, K., Ramus, F., Villiermet, N., Schwartz, D., and Giraud, A.-L. (2011). Altered low- γ sampling in auditory cortex accounts for the three main facets of dyslexia. *Neuron* 72, 1080–1090. doi: 10.1016/j.neuron.2011.11.002
- Logothetis, N. K. (2010). Bold claims for optogenetics. *Nature* 468, E3–E4. doi: 10.1038/nature09532
- Lord, C., Rutter, M., and Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders 1. 24. *J. Autism Dev. Disord.* 24, 659–685. doi: 10.1007/BF02172145
- Luo, H., Liu, Z., and Poeppel, D. (2010). Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biol.* 8:e1000445. doi: 10.1371/journal.pbio.1000445
- Luo, H., and Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54, 1001–1010. doi: 10.1016/j.neuron.2007.06.004
- Magri, C., Schridde, U., Murayama, Y., Panzeri, S., and Logothetis, N. K. (2012). The amplitude and timing of the BOLD signal reflects the relationship between local field potential power at different frequencies. *J. Neurosci.* 32, 1395–1407. doi: 10.1523/JNEUROSCI.3985-11.2012
- Marco, E. J., Hinkley, L. B. N., Hill, S. S., and Nagarajan, S. S. (2011). Sensory processing in autism: a review of neurophysiologic findings. *Pediatr. Res.* 69, 48R–54R. doi: 10.1203/PDR.0b013e3182130c54
- Morillon, B., Lehongre, K., Frackowiak, R. S. J., Ducorps, A., and Kleinschmidt, A. (2010). Neurophysiological origin of human brain asymmetry for speech and language. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18688–18693. doi: 10.1073/pnas.1007189107
- Murdoch, J. D., and State, M. W. (2013). Recent developments in the genetics of autism spectrum disorders. *Curr. Opin. Genet. Dev.* 23, 310–315. doi: 10.1016/j.gde.2013.02.003
- Peelle, J. E., Gross, J., and Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cereb. Cortex* 23, 1378–1387. doi: 10.1093/cercor/bhs118
- Pelphrey, K., and Deen, B. (2012). Brain scans need a rethink. *Nature* 491, S20. doi: 10.1038/491S20a
- Peñagarikano, O., Abrahams, B. S., Herman, E. I., Winden, K. C., Gdalyahu, A., Dong, H., et al. (2011). Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities and core autism-related deficits. *Cell* 147, 235–246. doi: 10.1016/j.cell.2011.08.040. Absence
- Peñagarikano, O., and Geschwind, D. H. (2012). What does CNTNAP2 reveal about autism spectrum disorder? *Trends Mol. Med.* 18, 156–163. doi: 10.1016/j.molmed.2012.01.003
- Pu, D., Shen, Y., and Wu, J. (2013). Association between MTHFR gene polymorphisms and the risk of Autism spectrum disorders: a meta-analysis. *Autism Res.* 6, 384–392. doi: 10.1002/aur.1300
- Rojas, D. C., Singel, D., Steinmetz, S., Hepburn, S., and Brown, M. S. (2013). Decreased left perisylvian GABA concentration in children with autism and unaffected siblings. *Neuroimage* 86, 28–34. doi: 10.1016/j.neuroimage.2013.01.045
- Rotstein, H. G., Pervouchine, D. D., Acker, C. D., Gillies, M. J., White, J. A., Buhl, E. H., et al. (2005). Slow and fast inhibition and an H-current interact to create a theta rhythm in a model of CA1 interneuron network. *J. Neurophysiol.* 94, 1509–1518. doi: 10.1152/jn.00957.2004
- Sauseng, P., Klimesch, W., Gruber, W. R., and Birbaumer, N. (2008). Cross-frequency phase synchronization: a brain mechanism of memory matching and attention. *Neuroimage* 40, 308–317. doi: 10.1016/j.neuroimage.2007.11.032
- Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., and Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends Cogn. Sci.* 12, 106–113. doi: 10.1016/j.tics.2008.01.002
- Schwarzkopf, D. S., Anderson, E. J., de Haas, B., White, S. J., and Rees, G. (2014). Larger extrastriate population receptive fields in autism spectrum disorders. *J. Neurosci.* 34, 2713–2724. doi: 10.1523/JNEUROSCI.4416-13.2014
- Shamir, M., Ghitza, O., Epstein, S., and Kopell, N. (2009). Representation of time-varying stimuli by a network exhibiting oscillations on a faster time scale. *PLoS Comput. Biol.* 5:e1000370. doi: 10.1371/journal.pcbi.1000370
- Snowling, M. J., Gallagher, A., and Frith, U. (2003). Family risk of dyslexia is continuous: individual differences in the precursors of reading skill. *Child Dev.* 74, 358–373. doi: 10.1111/1467-8624.7402003
- Stevenson, R. A., Segers, M., Ferber, S., Barense, M. D., and Wallace, M. T. (2014a). The impact of multisensory integration deficits on speech perception in children with autism spectrum disorders. *Front. Psychol.* 5:379. doi: 10.3389/fpsyg.2014.00379
- Stevenson, R. A., Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M., et al. (2014b). Multisensory temporal integration in autism spectrum disorders. *J. Neurosci.* 34, 691–697. doi: 10.1523/JNEUROSCI.3615-13.2014
- Tyizio, R., Nardou, R., Ferrari, D. C., Tsintsadze, T., Shahrokhi, A., Eftekhari, S., et al. (2014). Oxytocin-mediated GABA inhibition during delivery attenuates autism pathogenesis in rodent offspring. *Science* 343, 675–679. doi: 10.1126/science.1247190
- Uhlhaas, P. J., and Singer, W. (2007). What do disturbances in neural synchrony tell us about autism? *Biol. Psychiatry* 62, 190–191. doi: 10.1016/j.biopsych.2007.05.023
- Weschler, D. (2000). *WAIS-III: Echelle de L'intelligence de Wechsler Pour Adultes*, ed. Troisième. (Paris: Centre de Psychologie Appliquée).
- Whittington, M. A., Cunningham, M. O., LeBeau, F. E. N., Racca, C., and Traub, R. D. (2011). Multiple origins of the cortical γ rhythm. *Dev. Neurobiol.* 71, 92–106. doi: 10.1002/dneu.20814
- Williams, D. L., Cherkassky, V. L., Mason, R. A., Keller, T. A., Minshew, N. J., and Just, M. A. (2013). Brain function differences in language processing in children and adults with autism. *Autism Res.* 6, 288–302. doi: 10.1002/aur.1291
- Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., et al. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party.” *Neuron* 77, 980–991. doi: 10.1016/j.neuron.2012.12.037

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Jochaut, Lehongre, Saitovitch, Devauchelle, Olasagasti, Chabane, Zilbovicius and Giraud. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Oscillopathic Nature of Language Deficits in Autism: From Genes to Language Evolution

Antonio Benítez-Burraco¹ and Elliot Murphy^{2*}

¹ Department of Spanish Philology and its Didactics, University of Huelva, Huelva, Spain, ² Division of Psychology and Language Sciences, University College London, London, UK

Autism spectrum disorders (ASD) are pervasive neurodevelopmental disorders involving a number of deficits to linguistic cognition. The gap between genetics and the pathophysiology of ASD remains open, in particular regarding its distinctive linguistic profile. The goal of this article is to attempt to bridge this gap, focusing on how the autistic brain processes language, particularly through the perspective of brain rhythms. Due to the phenomenon of pleiotropy, which may take some decades to overcome, we believe that studies of brain rhythms, which are not faced with problems of this scale, may constitute a more tractable route to interpreting language deficits in ASD and eventually other neurocognitive disorders. Building on recent attempts to link neural oscillations to certain computational primitives of language, we show that interpreting language deficits in ASD as oscillopathic traits is a potentially fruitful way to construct successful endophenotypes of this condition. Additionally, we will show that candidate genes for ASD are overrepresented among the genes that played a role in the evolution of language. These genes include (and are related to) genes involved in brain rhythmicity. We hope that the type of steps taken here will additionally lead to a better understanding of the comorbidity, heterogeneity, and variability of ASD, and may help achieve a better treatment of the affected populations.

OPEN ACCESS

Edited by:

Anne Keitel,
University of Glasgow, UK

Reviewed by:

Sam McLeod Doesburg,
Hospital for Sick Children, Canada
Delphine Jochaut,
University of Geneva, Switzerland

***Correspondence:**

Elliot Murphy
elliottmurphy91@gmail.com

Received: 23 December 2015

Accepted: 07 March 2016

Published: 18 March 2016

Citation:

Benítez-Burraco A and Murphy E (2016) The Oscillopathic Nature of Language Deficits in Autism: From Genes to Language Evolution. *Front. Hum. Neurosci.* 10:120. doi: 10.3389/fnhum.2016.00120

Keywords: autism, neural oscillations, RUNX2, evo-devo, biolinguistics, language evolution

INTRODUCTION

Autism spectrum disorders (ASD) are pervasive neurodevelopmental disorders involving several social and cognitive deficits. Usually, people with ASD exhibit stereotypical and repetitive behavior, an inability for social interaction, and communicative problems (Bailey et al., 1996). Interestingly, close connections have been made between ASD and specific language impairment (SLI; see Crespi and Badcock, 2008 for discussion and “From Language Deficits to the Brain in ASD” Section below). ASD involves atypical brain wiring during growth, which results in its distinctive cognitive profile, and which has been linked or associated to mutations in an extensive number of genes (Veenstra-VanderWeele and Cook, 2004). Recent advances in genome-wide technology have resulted in a long list of candidate genes for this condition (Geschwind and State, 2015). Although they point to specific pathways and neural mechanisms underlying its associated deficits (Willsey and State, 2015), the gap between the pathophysiology of ASD and genes still remains open, in particular regarding its distinctive linguistic profile (see Jeste and Geschwind, 2014 for discussion).

In truth, the polygenism seen in ASD is somewhat commensurable with the polygenism displayed in language more generally. The goal of this article is to contribute to bridging this gap between genes and ASD, focusing on how the autistic brain processes language, particularly through the perspective of brain rhythms and how language evolved in the species.

Brain rhythms are primitive components of brain function. Because the hierarchy of brain oscillations has remained remarkably preserved during the course of mammalian evolution, it has been hypothesized that cognitive disorders can be conceived as variations (i.e., dysrhythmias or oscillopathies) within the network constellation that constitutes a universal brain syntax (see Cobb and Davies, 2005; Buzsáki et al., 2013 for discussion). And because brain rhythms are connected to some computational primitives of language (see Murphy, 2015a for discussion), we believe that interpreting language deficits in ASD as oscillopathic features is a potentially fruitful way to construct successful endophenotypes of this condition. Current understanding suggests that ASD is characterized, amongst other things, by asynchronous neural oscillations (Tierney et al., 2012), although as discussed below a number of studies have reported both underconnectivity and overconnectivity in the autistic brain. We hope that the type of steps taken here will additionally lead to better understanding of the comorbidity, heterogeneity, and variability of ASD, and may help achieve a better treatment of the affected populations. Since children can progress from one sub-type of disorder to another during development, long-term studies of children with ASD involving perennial oscillopathic monitoring could ideally provide a more accurate diagnosis (refining the disorder boundaries) and prognosis. Additionally, a comprehensive picture of the rhythms and networks implicated in linguistic deficits in ASD should also contribute to the growing understanding of the human cognitive phenotype, set against an evolving dynamic model of mental computation. This is why we also expect our approach to cast light on the neurobiological basis of language and the way in which language evolved in our species. In doing so, we will heavily rely on our attempts of translating language into a grammar of brain rhythms, as developed in Murphy (2015a), which represents a new approach to the problem that the categories of linguistic theory do not easily map onto clinical typologies. This also builds on ongoing developments in neurolinguistics which move beyond the outdated classical production-comprehension division, and which bring to the forefront of discussion the dynamics of brain networks; that is, the collections of brain areas jointly engaged by some cognitive operation (Fedorenko and Thompson-Schill, 2014). But we will also rely on an evolutionary-developmental (evo-devo) perspective in our ongoing research into the origin and development of language, which aims to find a robust link between the changes that occurred in the human brain during our recent speciation and the changes occurring in the child brain during language growth. According to our view, our language-readiness (i.e., our species-specific ability to acquire and use languages) depends greatly on a proper pattern of cortical inhibition and a specific pattern of long-distance

connections across the brain (which was brought about by mutations in genes controlling the development of the skull and the brain), that enable us to form and exploit cross-modular concepts (see Boeckx and Benítez-Burraco, 2014a,b; Benítez-Burraco and Boeckx, 2015a for details), both of which are aspects that are targeted in ASD (Khan et al., 2013; Zikopoulos and Barbas, 2013).

The article is structured as follows. First, we provide a general account of language deficits in ASD. Afterwards we describe the anomalies in brain activity observed in ASD and advance a tentative oscillopathic model of language deficits in ASD. Then we move to the genes, focusing on candidates for ASD that may help to explain this abnormal profile of brain rhythmicity. The last section of the article examines the whole case for the oscillopathic nature of language deficits in ASD from an evolutionary perspective, arguing that the overrepresentation of candidates for ASD among the genes that changed after our split from extinct hominins may help understand how our language-readiness evolved, but also the nature and the prevalence of ASD among modern populations.

FROM LANGUAGE DEFICITS TO THE BRAIN IN ASD

The Language Phenotype in ASD

Language deficits in ASD are not always confined to the pragmatic component of language, which is also compromised in other intellectually impaired children (Abbeduto and Hesketh, 1997). On the contrary, some children with ASD undergo a real linguistic regression from age 12–24 months (Lainhart et al., 2002; Lord et al., 2004), or never acquire functional speech (Tager-Flusberg et al., 2005). Moreover, phonological and morphosyntactic problems are observed in nearly one third of children with ASD (Tager-Flusberg and Cooper, 1999; Rapin and Dunn, 2003; Tager-Flusberg and Joseph, 2003). Interestingly, as noted, ASD is sometimes comorbid with language disorders like SLI (Norbury, 2005; Tager-Flusberg, 2006). It has been hypothesized that a core language deficit exists in ASD, although this is difficult to prove because of the noteworthy variability in the linguistic and communicative abilities of people with ASD, and of the masking effect of their variable IQ and the degree of functionality achieved in the domain of language (see Volden and Lord, 1991, on MLU, and Roberts et al., 2004 on person and tense marking). The impairment of the oromotor function has been claimed to account for expressive language deficits in a subgroup of people with ASD (Belmonte et al., 2013). Nonetheless ASD also entails problems in language comprehension. Moreover, Theory of Mind and related pragmatic faculties do not amount to an explanatory account of language deficits in ASD, and even documented cases of ASD which do not result in formal language impairments still display patterns of linguistic processing and acquisition deviating from typically developing (TD) controls (Howlin, 2003; Bourguignon et al., 2012).

Among the distinctive profile of atypical language development in children with ASD, one observes that syntactic

complexity grows at a similar rate than in children with Down syndrome (although the former employ fewer functional words; Tager-Flusberg et al., 1990). Additionally, children with ASD rely less than their unaffected peers on prosodic cues to disambiguate sentences (Diehl et al., 2015), and less on semantic plausibility to understand passives (Paul et al., 1988). It has been further claimed that they integrate semantic information differently when interpreting syntactic constructions (Eigsti et al., 2007) and that semantic knowledge seems to consolidate in a dissimilar manner. Accordingly, they choose less prototypical words and are less primed by semantically-related words; moreover, even if they seem to be constrained by the “Principle of Mutual exclusivity” (i.e., “expect that each thing in the real world is referred to with one and the same word only”), they are less shape-biased than controls when categorizing items (Dunn et al., 1996; Kamio et al., 2007; Tek et al., 2008). Ultimately, children with ASD exhibit linguistic features that are not observed (or that are much less frequent) in children without ASD, such as echolalia (Tager-Flusberg and Calkins, 1990) or neologism formation (Volden and Lord, 1991; Eigsti et al., 2007). Overall, language acquisition in ASD proceeds in a more scattered way (e.g., the structure of new sentences is less predictable based on previous constructions; Eigsti et al., 2007), whereas the degree of variability and heterogeneity of language growth is greater than in the TD population (see Kjelgaard and Tager-Flusberg, 2001; Luyster et al., 2007; Norbury et al., 2010 on vocabulary development). As pointed out by Eigsti et al. (2007: p. 688), in ASD we often observe a “dramatic variability across and between social, cognitive, and language domains, [while] in studies of typical development [...] there is generally a much smaller range of individual differences”.

Importantly, ASD also entails differences regarding the way in which other cognitive abilities are put into use during language acquisition and usage. For instance, because of the reduced role played by semantic priming during word learning, some children with high-functioning ASD rely more than controls on their (enhanced) capacities for auditory/phonological processing and for statistical/associative learning (Tager-Flusberg, 2006; Kuhl, 2007; Preissler, 2008). At the same time, children with ASD showing deeper language deficits are thought to suffer from some deficit in phonological processing (Lindgren et al., 2009; see Norbury et al., 2010 for a discussion). Delays, asynchronies, and/or deviances are also observed at this cognitive level throughout development. For instance, because word learning depends more on an (enhanced) capacity for associative learning (at least in some children with ASD), it has been claimed that in ASD the declarative memory may play a more central role in language acquisition (Walenski et al., 2006). Similarly, problems with syntax are suggestive of an impairment of the procedural memory. Language impairment in ASD also involves problems with binding, relative clauses, *wh*-questions, raising and passives (Perovic and Janke, 2013).

In DSM-V, unlike DSM-IV, the number of ASD symptom domains has been reduced to two: “Social communication domain” (created by the merger of key symptoms from the DSM-IV Social and Communication domains) and “Fixated

interests and repetitive behavior or activity”. Although language deficits are no longer explicitly defined as a central feature of ASD (because deficits in communication are intimately related to social deficits), an examination of the fixated interests and repetitive behavior described in DSM-V reveals that “stereotyped or repetitive speech” is nevertheless attributed a major role in ASD criteria; pronoun reversal, abnormal self-reference, repetitive vocalizations, unusually formal language, echolalia and neologisms are considered criteria exemplars. Moreover, we believe that this does not preclude the centrality of language deficits within the ASD phenotype. In truth, in the TD population communication is always intimately related to the social domain. But at the same time, it makes sense to claim that human beings are endowed with a special faculty for acquiring and using languages even if it relies on cognitive devices that are not specific to language (Hauser et al., 2002; Boeckx, 2011). Likewise, we believe that language deficits in pathological conditions can be described and interpreted on their own, in spite of the fact that they result from the impairment of biological components that are not specific to language (see Benítez-Burraco, 2016 for discussion).

The Linguistic Brain in ASD

Recent neurobiological findings give support to these perspectives. Accordingly, many intriguing structural differences are observed in the brains of people with ASD when they are compared to non-affected subjects. These include regional differences in brain volumes, variations in gray matter and white matter volumes and thickness across many brain regions, and differences in inter- and intra-hemispheric connection patterns (reviewed in Stefanatos and Baron, 2011; Bourguignon et al., 2012). As expected, most of these regions and nerve tracks are relevant for language processing. For example, children with ASD show an increment of gray matter density in the primary and associative auditory and visual cortices (Hyde et al., 2010). Likewise, high-functioning children with ASD have smaller gray matter volumes in the frontostriatal regions, whereas children with Asperger syndrome (which exhibit milder language deficits) still show reduced volumes of the caudate and the thalamus (McAlonan et al., 2008). As Radalešcu et al. (2013) discuss, frontostriatal connectivity is a core component of healthy linguistic cognition, and, as explored below, the thalamus may be also. Finally, microstructural anomalies and reduced lateralization patterns are characteristically observed in the arcuate fasciculus of people with ASD (Fletcher et al., 2010), suggesting that language impairment in these individuals may result in part from a constraint of the integrative processes during development (Schipul et al., 2011). Not surprisingly then, functional differences important for language processing have been attested as well in the brain of people with ASD. Hence, in children with ASD networks involved in the statistical analysis of speech respond abnormally to artificial languages (Scott-Van Zeeland et al., 2010b); this finding can be related to their attested deficits for implicit learning (Mostofsky et al., 2000). In such cases, the basal ganglia and the left temporo-parietal

cortex are impaired (Scott-Van Zeeland et al., 2010a), which reinforces the view that a cognitive hallmark (or endophenotype) of ASD is a dysfunction of procedural memory. Likewise, Courchesne and Pierce (2005) describe ASD neurocognition as the frontal cortex unconsciously “talking to itself” due to the impairment of normal language functions. In ASD, the frontal cortex networks are usually underactive during language comprehension tasks whenever sentences are not congruent with reality, suggesting that the integration of linguistic and encyclopedic knowledge is also impaired in this condition (Tesink et al., 2011).

Finally, regarding the atypical course of language development in ASD, Stefanatos and Baron (2011: pp. 262–263) point out that “functional anomalies early in development can have crucial implications for neural networking and environmental transactions that, in turn, prompt other potentially more widespread perturbations in cognitive structure or neural architecture”. As noted by Crespi and Badcock (2008: p. 244), deficits in the so-called maternal brain, largely the neocortex, alongside normal functioning in the paternal brain, largely the limbic system, can “lead to the loss of language, mental retardation, and repetitive behavior typical of infantile (Kanner) autism, whereas increased paternal-brain effects, but relatively spared maternal-brain function, may lead to high-functioning autism or Asperger syndrome”. Not surprisingly, some candidate genes for ASD are subject to imprinting (see for example Bonora et al., 2002).

That said, ASD cannot simply be reduced to deficient neural development and connectivity; Hahamy et al.’s (2015: p. 302) study of resting-state activity in adult ASD subjects revealed both increased and decreased intra- and inter-hemispheric connectivity. To account for this, they suggest that ASD can be characterized by idiosyncratic distortions of functional connectivity patterns, since the “magnitude of an individual’s pattern distortion in homotopic interhemispheric connectivity correlated significantly with behavioral symptoms of ASD”. In the next section, we will focus on brain rhythmicity and will try to (re)interpret the linguistic deficits observed in ASD in terms of abnormal patterns in the integration of brain oscillations. As expected, ASD also involves differences at the functional level during language processing by the brain, for example, anomalies in activation patterns or changes in mismatch negativity responses to linguistic elements (reviewed in Stefanatos and Baron, 2011: pp. 259–262), and an abnormal pattern of brain rhythmicity.

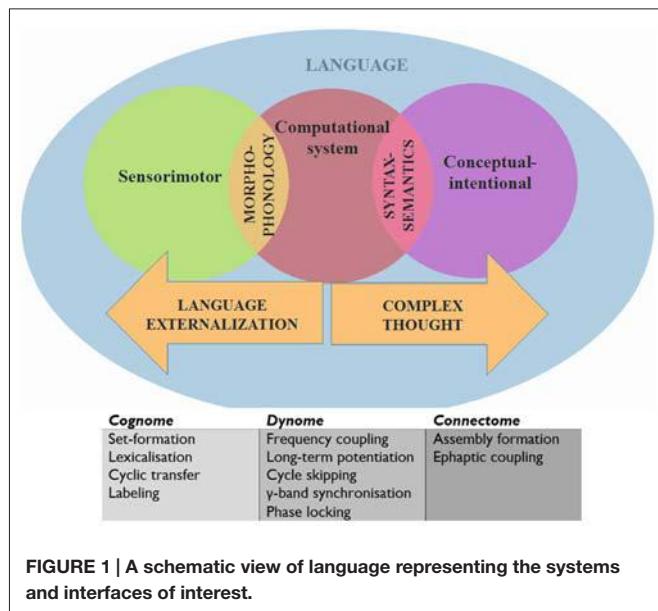
FROM BRAIN RHYTHMICITY TO LANGUAGE DEFICITS IN ASD

As noted above, ASD has been associated with a number of abnormal structural and functional patterns, which likely contribute to the emergence of the disorder (see also Welsh et al., 2005; Pineda et al., 2012). But if Hahamy et al. (2015) are correct in pointing to the idiosyncratic nature of connectivity patterns in ASD, functional localization studies may not provide the right basis from which to construct

linking hypotheses between neural and linguistic operations. According to Uhlhaas et al. (2010), we should expect a strong relationship between the emergence of altered oscillatory patterns during childhood and the appearance of a number of neurocognitive disorders at distinct developmental stages, to the extent that brain oscillation self-regulation has been proposed as a potential treatment for ASD (see Pineda et al., 2012 for details). A brain that grows differently and assumes an unusual size is also differently wired and, ultimately, exhibits altered oscillatory behavior, which we will argue alters their cognitive phenotype (see Buzsáki et al., 2013 for how rhythms can be both preserved and altered with changing brain size). Accordingly, we believe that ASD should not only be seen as a cognitive, rather than purely social disorder (as argued by Bourguignon et al., 2012), but it should also be seen more specifically as an oscillopathic one. Under our view, subjects with ASD might therefore “*construe* language differently, reflecting a linguistic style different from that inherent in neurotypical cognition” (Hinzen et al., 2015). Once ASD is seen as an oscillopathy, and once brain dynamics can be shown to be a plausible candidate for linguistic computation (Murphy, 2015b), various predictions can be generated about the etiology of language-related disorders, and specifically, our understanding of language deficits in ASD (as reviewed above) and why language in ASD subjects is indeed construed differently (as suggested by the impairments in communication and pragmatic cognition documented in Mandy and Skuse, 2008) will surely develop.

Language Processing under an Oscillatory Lens

Before generating such predictions we will provide with a brief outline of our translational proposal of language into a syntax of brain oscillations. In doing so we will rely on the model of the human cognome-dynome outlined in Murphy (2015a), where “cognome” refers to the operations available to the human nervous system (Poeppel, 2012) and “dynome” refers to brain dynamics (Kopell et al., 2014). In this model various cross-frequency couplings and regions were attributed distinct computational roles under a research program termed “Dynamic Cognomics” (**Figure 1**). As Petersen and Sporns (2015: p. 207) emphasize, most accounts of cognition have “focused on computational accounts of cognition while making little contact with the study of anatomical structures and physiological processes”. Since “[t]he fact that a theory is computationally explicit does not automatically render it biologically plausible” (Bornkessel-Schlesewsky et al., 2014: p. 365), Murphy (2015a) attempted to correct for this imbalance by decomposing the computational operations of language down to a small set of (potentially generic) sub-operations, attempting to achieve an appropriate level of granularity from which computational-implementational connections could be constructed. Considerations of systems level computation were addressed, departing from the standard focus on single-neuron computation (Fitch, 2014). By connecting these dynamic insights to the “connectome”



level, a more powerful computational model of dynamic brain activity was proposed. This model embraced the minimalist conception of language (Chomsky, 1995; Narita, 2014) as a *computational system* linking generated, hierarchically structured expressions to two interfaces; the *conceptual-intentional system* (responsible for “thought” and interpretation) and the *sensorimotor system* (responsible for externalization; **Figure 1**).

The computational system requires some elaboration. This is composed of the operations Merge (which generates sets and moves objects to different positions), Label (which assigns a built set an independent categorial identity, such as a Noun Phrase or Tense Phrase), Agree/Search (which establishes relations between objects, such as in Person or Gender agreement; in the sentence “There *seems* to be likely to be *a man* in the garden”, the italicized constituents exhibit syntactic covariance, with their number features agreeing) and Spell-Out/Transfer (the part of the memory buffer which sends structures to be externalized and interpreted in “chunks”, or “phases”; see Narita, 2014).

We will assume, following Murphy (2015a) and Theofanopoulou and Boeckx (forthcoming), that the α band embeds γ rhythms generated cross-cortically, the dynamic realization of inter-modular conceptual combinations, or set-formation (“Simplest Merge”, for Epstein et al., 2015). This form of variable binding may also arise from precisely controlled recurrent interactions between the basal ganglia and prefrontal cortex (Kriete et al., 2013). This perspective is supported by recent findings that α is responsible for visuo-spatial feature-binding, a form of representation “merging” (Roux and Uhlhaas, 2014). Since syntactic theory typically assumes that set-formation purely involves the combination of two representations without modification of either (Narita, 2014), these oscillatory mechanisms are perhaps the best implementational candidates. We will assume also that Spell-Out/Transfer is realized through embedding such γ rhythms inside the θ band, the source

of which is found in the hippocampus. This perspective is supported by the recent finding that γ bursts “reflect the binding of temporal variables to the values allowed by constraints introduced by temporal expressions in discourse” (Brederoo et al., 2015) and by Meyer et al.’s (2015) EEG study suggesting that frontal-posterior θ oscillations reflect memory retrieval during sentence comprehension. Murphy (2015a) speculated that Transfer is also likely supported by the corpus callosum, following insights in Theofanopoulou (2015). More broadly, γ has been associated with lexical processing (Hannemann et al., 2007). Finally, we will assume that labeling (holding in memory one of the items before coupling it with another to generate an independent syntactic identity) involves the slowing down of γ to β before β - α coupling, implicating a basal ganglia-thalamic-cortical loop (see Hyafil, 2015 for discussion of the different types of cross-frequency coupling, such as phase-amplitude, phase-phase, and phase-frequency coupling). Assume also that Agree/Search is implemented via cross-cortical evoked γ due to the role of this band in attention and perceptual “feature binding” (Bartos et al., 2007; Sohal et al., 2009) and the distributed nature of the inter-modular representations Agree/Search operates over (see **Figure 2** for predictions about rhythmic disruptions for agreement relations). Overall, we find this model compatible with recent findings that damage to the basal ganglia and thalamus can lead to various forms of aphasia and other linguistic deficits (Alamri et al., 2015); in turn contributing to an emerging rejection of the classical Wernicke-Lichtheim-Geschwind approach to neurolinguistics (Hagoort, 2014). This is why we expect it to contribute as well to a better understanding of language deficits in ASD.

As we have shown, decomposing linguistic processes into generic sub-operations permits a certain degree of alignment between the basic computational properties of the human nervous system and the neural oscillations which are increasingly being theorized as having major functional roles in memory, attention and perception. The next task is to use these cognome-dynome linking hypotheses to interpret the range of complex data accumulated from electrophysiological and magnetoencephalographic studies of ASD.

Brain Rhythmicity and Language Processing in ASD

Dynamic investigations of ASD are still relatively young, but we believe enough has been learned to begin the construction of an oscillopathic model of this condition. The first physiological study of brain connectivity in ASD children under conscious conditions was conducted by Kikuchi et al. (2013), revealing aberrant brain activity and a rightward-lateralized neurophysiological network in ASD not present in TD children. They detected increased γ power—claimed to be related to the degree of developmental delay (Orehkova et al., 2007)—but reduced α and β (see also Cornew et al., 2012 for increased δ , α and θ in resting state exams). Similar γ -related findings were reported by Rojas et al. (2008), who argued that “gamma-band phase consistency ... may be [a]

Frequency band	Role in the present model of language computation	Observed and predicted differences in Autism Spectrum Disorder
Delta (~0.5-4Hz)	Involved in phrasal processing and possibly labeling.	Increased in eyes-closed resting state exam; predicted to be disrupted in processing phrases involving raising and passives.
Theta (~4-10Hz)	Hippocampal source; embeds γ to generate cyclic transfer of syntactic objects; involved more generally in memory retrieval.	Reduced coherence in children; does not synergistically engage with γ during speech; predicted to be disrupted in certain memory retrieval processes.
Alpha (~8-12Hz)	Synchronizes distant cortical regions; embeds γ generated cross-cortically to yield inter-modular set-formation; involved in lexical decision making.	Reduced cross-cortically; reduced resting-state α - γ phase amplitude coupling; increased in resting state; predicted to be disrupted during certain lexicalisations.
Beta (~10-30Hz)	When γ is slowed to β and coupled with α via a basal ganglia-thalamic-cortical loop, syntactic objects are labeled; holds objects in memory.	Reduced cross-frequency coupling with γ ; predicted to be disrupted in the maintenance of syntactic objects in raising, passives and <i>wh</i> -questions.
Gamma (~30-100Hz)	Generates syntactic objects before β holds them in memory; central role in a number of linguistic operations; involved in lexical processing.	Over-connectivity gives rise to increased γ ; reduced in rSTG and lIFG during picture naming; predicted to be disrupted quite generally in linguistic cognition.

FIGURE 2 | Summary of the present cognome-dynome model of linguistic computation and the observed differences in autism spectrum disorders (ASD); rSTG denotes right superior temporal gyrus, lIFG denotes left inferior frontal gyrus (see Cornew et al., 2012; Buard et al., 2013; Berman et al., 2015; Jochaut et al., 2015; Kikuchi et al., 2015; Murphy, 2015a).

potentially useful [endophenotype] for autism". The central role attributed here to β in storing and sustaining multiple items in memory (part of labeling) generated by fast γ supports Bangel et al.'s (2014: p. 202) interpretation of their MEG study of number estimation in ASD. They found reduced long-range β phase synchronization in ASD subjects at 70–145 ms during the presentation of globally coherent dot patterns, providing "the first evidence for inter-regional phase synchronization during numerosity estimation, as well as its alteration in ASD". If these patterns reflect problems with inter-regional communication, this would lend weight to the present view that β is deployed in the extraction and labeling of meaningful units (in the case of Bangel et al., 2014 the units included animal pictures), centered on basal ganglia activity in particular, which is strongly implicated in β activity by Khanna and Carmena (2015). Since children with ASD often struggle with communicative intent, the finding that imperative pointing implicates greater cross-cortical β activity than declarative pointing (Brunetti et al., 2014) would seem to support the current dynamic model, with Hinzen and Sheehan (2013) proposing a strong connection between linguistic cognition and imperative gestures. More generally, the disruption of rhythmic coordination (the basis of neural network communication, mediated through the synchronization of presynaptic potentials in a given neuronal population, enhancing postsynaptic impact on the target region; Donner and Siegel, 2011) seen in ASD supports the image of this condition as an oscillopathic disorder. Relatedly, as noted above, ASD has "naturally suppressed private speech", which Frawley (2008: p. 269) claims are "best understood in the context of the control processes of cognitive-computational architectures". We believe such control processes should in turn be best understood in terms of cross-frequency couplings between rhythms of distinct cortical and subcortical regions; a form of dataflow management (Pylyshyn, 1985); see Figure 2 for a summary.

Chattopadhyaya and Cristo's (2012) exploration of GABAergic circuit dysfunctions in ASD additionally reinforces the central connectomic role attributed to these interneurons in Murphy (2015a). Numerous studies have discovered GABA_A and GABA_B receptor alterations in the brains of people with ASD (Fatemi et al., 2010; Oblak et al., 2010), with emerging developments in electron microscopy, permitting the 3D

modeling of dendritic networks, likely being able to contribute to a development of these connectomic topics (Fua and Knott, 2015). Additionally, the noted altered γ activity detected in ASD children by Kikuchi et al. (2013) may have arisen from GABAergic or glutamatergic mediator system disturbances, claimed to be involved in producing this rhythm (Sohal et al., 2009). With Blatt et al. (2001) reporting significantly reduced GABA_A-receptor binding in high hippocampal binding areas in ASD subjects, this may suggest deviant applications of syntactic Transfer operations in this condition. This perspective is supported by the following data: the smaller corpus callosum typically found in autistic brains (Waiter et al., 2005; Alexander et al., 2007); the reduced connectivity between left-anterior and right-posterior areas found in Kikuchi et al.'s (2015) MEG study of children with ASD, who exhibited a decrease in θ coherence; and the reduced θ inter-regional synchronization during a set-shifting task in children with ASD (Doesburg et al., 2013; although much more needs to be learnt about the rhythmogenesis of Transfer operations). Further, because the narrow syntactic operations of set-formation, Label and Transfer appear to be preserved in language disorders, we would predict that any dynamic variations detected in ASD would most likely reflect representational/conceptual and interface disruptions. For instance, although the dynamic basis has not been investigated, it is well documented that individuals with ASD have difficulty understanding abstract concepts, metaphors, and often cannot plan ahead and consider multiple options before selecting an appropriate response (Dodd, 2005: p. 47; Jordan, 2010).

Murphy (2015a) also discussed the over-arching role of the "Communication through Coherence" (CTC) hypothesis in rhythm synchronization (Fries, 2005). Coherent brain oscillations are a central mechanism for carving the temporal coordination of neural activity in a global network. More recent work by Fries (2015) expands on the general claim that synchronization affects communication between neuronal groups. Communication, or "the transfer of one representation in a presynaptic, or sending, group to a new representation in a postsynaptic, or receiving, group" (Fries, 2015: p. 220), is the process that implements neural computations and thus creates novel representations. If communication is disrupted, representation construction is consequently disrupted too, as

in cases of rhythmic disruption in, for instance, monkey visual cortex (Tan et al., 2014). The emerging literature reviewed here suggests that this is precisely what happens in autistic brains. Neuronal communication, contrary to classical conceptions, is not limited to structural anatomical determination, but can also be achieved through emergent dynamic activity of neuronal groups (that is, at the level of the *dynome*). Rhythmic synchronization is widespread across the nervous system, and “[p]atterns of synchronization change dynamically with stimulation and behavioral context in a way that strongly suggests that selective coherence implements selective communication” (Fries, 2015: p. 222). As Uhlhaas et al. (2010) also argue, CTC suggests that neural synchrony is not an epiphenomenon but rather an integral part of cortical network functioning.

To take a case in which this rhythmic coherence is disrupted, we can turn to the abnormal forms of semantic organization documented in ASD (Harris et al., 2006). An MEG study by Braeutigam et al. (2008) recorded responses in adults with ASD to sentences ending with a semantically incongruous word. N400 responses following the incongruous word were weaker over left temporal cortices, and late positivity component responses to incongruous words and long-latency γ oscillations following congruous words were stronger in subjects with ASD relative to TD controls over prefrontal and central regions. The long-lasting γ possibly indicates “unusual strategies for resolving semantic ambiguity in autism” (Braeutigam et al., 2008: p. 1026), and this finding may relate to the claim that an inability to delimit activation within an abnormally wired network is a core neural marker of ASD (Polleux and Lauder, 2004). There may consequently be a lack of spatio-temporal and rhythmic activation constraints in ASD (possibly caused by an enlarged braincase, discussed below); a suggestion which speaks to particular communication deficits noted in ASD, such as general difficulties with spoken language and gestures, problems initiating and sustaining suitable conversation, and the use of inappropriate, repetitive speech (Lord et al., 2000). Gandal et al. (2010) discovered reduced γ phase-locking across hemispheres in ASD subjects during auditory pure-tone presentation, while induced and evoked γ were not significantly different from TD adults, again suggesting that problems with frequency synchronization is a central neural characteristic of ASD (see also Wilson et al., 2007 and their MEG study reporting reduced left-hemispheric steady state γ in adolescents with ASD in response to non-speech sounds). A more recent pure-tone study by Edgar et al. (2015) found pre-stimulus abnormalities across multiple frequencies for auditory superior temporal gyrus processes in ASD, followed by early high-frequency, and then low-frequency, abnormalities. For our purposes, the conclusion of Edgar et al. (2015: p. 395), that “elevations in oscillatory activity [suggests] an inability to maintain an appropriate “neural tone” and an inability to rapidly return to a resting state prior to the next stimulus”, speaks to the present hypothesis that ASD may be characterized by abnormal applications of language-related processes of representation maintenance (including labeling).

Deficits in the rhythmic profile of the core syntax-semantics regions were also found in a picture-naming task conducted

by Board et al. (2013), which used MEG and observed reduced evoked high γ in the right superior temporal gyrus and reduced evoked high β and low γ in the left inferior frontal gyrus in subjects with ASD compared to TD controls. Resting-state MEG data collected from adolescents with ASD also demonstrated that alterations of functional connectivity are dependent on region and frequency, and that frontal over-connectivity is “expressed in the gamma band, whereas posterior brain regions exhibit a disconnection to widespread brain areas in slower delta, theta and alpha bands” (Ye et al., 2014: p. 6062). In addition, resting-state $\alpha\text{-}\gamma$ phase amplitude coupling, a basic process in TD children, was recently found to be abnormal in children with ASD (Berman et al., 2015). Jochaut et al. (2015) also discovered atypical coordination of cortical oscillations in autistic linguistic cognition, combining EEG and fMRI to show that γ and θ cortical activity do not synergistically engage in response to speech. Oscillation-based connectivity between auditory and other language cortices was also found to be altered in ASD, compromising the mapping between sensation and higher-level cognitive representations.

We believe that these speech deficits in ASD can be investigated in oscillopathic terms. It has recently been shown, for instance, that there exists a strong correspondence between the average length of speech units and the hierarchy of cortical oscillation frequencies. Syllable sequences and phrases correspond with δ , syllables correspond with θ , and phonetic features correspond with γ and β (Schroeder et al., 2008; Giraud and Poeppel, 2012). These rhythms reflect a computational mechanism such that the brain “sets time intervals for analysis of individual speech components by intrinsic oscillations pretuned to an expected speech rate and retuned during continuous speech processing by locking to the temporal envelope” (Poeppel and Hickok, 2015: p. 252). In general, low frequency oscillations such as δ and θ parse speech streams into temporal units of granularities determined by the particular rhythm, while high frequency oscillations appear to decode these streams and access stored templates from memory. As a result, and considering the above findings of degraded θ and γ synergy in ASD, it may be the case that the lack of coherence between low and high rhythms leads to the documented problems with speech perception, tone recognition, and parsing phonemic representations at the rate seen in TD controls.

Towards an Oscillopathic Characterization of Language Deficits in ASD

It seems, then, that a combination of slow, “global” oscillations and faster, “regional” oscillations can be impaired in ASD (Kikuchi et al., 2015), and that the coupling of such rhythms can also be impaired as a result. Even when hyperconnectivity is detected, as in Ye et al. (2014), dramatic losses of synchrony in the face of such increased oscillatory activity (documented, for instance, by Castelhano et al. (2015) in their EEG study of disrupted perceptual coherence) presents strong evidence that synchrony underlies central coherence. Since these cases of brain rhythm abnormalities likely result from the neural mechanisms underlying ASD, the spectral and spatio-temporal

content distinguishing such rhythms can provide pertinent data for understanding its neurophysiological basis and oscillopathic profile. The idiosyncratic nature of connectivity patterns documented in ASD leads us to suggest that localization studies need to be supplemented by a (computationally explicit and informed) oscillopathic perspective.

The present perspective is therefore something of a refinement and expansion of Brock et al.'s (2002) seminal temporal binding hypothesis. In response to the wealth of data suggesting that ASD subjects showed "weak central coherence" (a bias toward piecemeal, and not configurational, processing), Brock et al. (2002) suggested that this feature arises from temporal binding deficits and a reduction in γ synchronization between local networks processing discrete visuoperceptual features. The picture, as we hope to have shown, has become more complicated in the intervening years, but our oscillopathic model should nevertheless be placed within the same tradition as Brock et al. (2002).

Finally, given that all of the oscillatory mechanisms we have invoked to construct our initial model of linguistic computation in Murphy (2015a) are general to a number of distinct cognitive systems, the oscillopathic profile we have constructed here is likely not unique to the language deficits in ASD. We would further predict that slight modifications to the rhythmic and connectivity patterns discussed above would result in different symptoms; for instance, when set at a particular level of disruption the oscillations responsible for various attentional mechanisms might lead to language-related attentional problems, but when disrupted in a different manner difficulties attending to socially relevant information might arise instead. Although we have restricted our attention to linguistic deficits, we see no reason to assume that our oscillopathic approach cannot also be fruitfully applied to an understanding of pragmatic deficits in ASD.

Overall, it seems that language deficits in ASD (as ASD itself) can be properly characterized as an oscillopathic condition. We believe that this translational effort, which aims to link the ASD dyname and cognome following the lines of Murphy (2015a), may result in a better understanding of language processing by people with ASD. Now we turn to genetics, which provides additional support for this view.

ASD-RELATED GENES AND SOME EVOLUTIONARY CONCERN

As noted in the introduction, the number of candidate genes for ASD has been growing over the time (Geschwind and State, 2015); at the same time, genetic studies also suggest that many of these candidates are related to specific pathways and aspects of brain function associated with susceptibility to ASD. Also several candidates for language impairment in ASD have been identified to date. Among the most promising genes one finds *MET* (Campbell et al., 2006), *CTTNBP2* (Cheung et al., 2001), *EN2* (Benayed et al., 2005), *NBEA* (Castermans et al., 2003), *HRAS* (Comings et al., 1996) and *PTEN* (Naqvi et al., 2000). It is possible too that somatic mutations affecting a subset of neurons cause ASD and language deficits in ASD (Poduri et al.,

2013; Sahin and Sur, 2015). In the first part of this section we will focus on genes related to networks and pathways that seem to be crucial for the maintenance of the adequate balance between neuronal excitation and inhibition and more generally, of brain rhythmicity. In the second part we will re-examine candidate genes for ASD from an evolutionary perspective in the context of language evolution studies with a special focus on brain connectivity and function.

ASD-Candidates and Brain Rhythmicity

As noted above, brain oscillation components and patterns are highly heritable traits, to the extent that we should expect that differences in cognition and behavior result in part from genetic variation affecting oscillatory activity. However, we don't know how pathogenic genetic diversity results in altered pathological patterns of brain oscillations and in desynchronization of neuronal activity. Here we will provide with some recent insights regarding ASD (and language). Because of the concerns discussed in the previous section, candidate genes for ASD that are related to GABAergic activity are of great interest for us. For instance, the loss of *Mecp2* from GABAergic interneurons results in ASD-like repetitive movements and auditory event-related potential deficits in mice (Goffin et al., 2014). Importantly, in response to auditory stimulation *Mecp2^{+/−}* mice recapitulate specific latency differences as well as select γ and β band abnormalities associated with ASD that may help to explain high-order deficits in this condition (Liao et al., 2012). ASD has been associated as well to mutations in *GABRB3* (this gene encodes the β -3 subunit of the GABA receptor A; Cook et al., 1998; Shao et al., 2002, 2003). Interestingly, differences in the expression levels of genes encoding some of the GABA_A-receptor subunits (particularly of $\beta 2$ and $\beta 3$) has been related to differences in the rhythm of hippocampal pyramidal neuron firing and the activity of fast networks (Heistik et al., 2010). More generally, genetic variation in GABA_A receptor properties have been linked to differences in β and γ oscillations, plausibly impacting on network dynamics and cognition (Porjesz et al., 2002). Documented alteration of the GABA catabolism also results in brain and behavioral anomalies that mimic the problems observed in people with ASD, including language deficits (Gibson et al., 1997; Pearl et al., 2003). Moreover, through exploring the functional relationships between ASD candidate genes by using the *BrainSpan* human transcriptome database, Mahfouz et al. (2015) discovered modules of such genes with neurobiologically pertinent co-expression dynamics, enriched for functional ontologies related to synaptogenesis and GABAergic neurons. Within the interaction networks identified by Mahfouz et al., a number of hub genes were detected, including *PROCA1*, *TBC1D22B*, *PPP2R2D* and *HACE1*. Other potential gene of interest is *PDGFRB*, which encodes the subunit β of the receptor of the platelet-derived growth factor (PDGF), a potent mitogen involved in the development of the central nervous system. Both PDGF and PDGFRB have been associated with ASD (Kajizuka et al., 2010). PDGFR- β KO mice show reduced auditory-evoked γ oscillation plausibly resulting from reduced number of GABAergic neurons, as observed in the

amygdala, the hippocampus, and the medial prefrontal cortex, which in turn give rise to problems with social interaction and spatial memory (Nguyen et al., 2011; Nakamura et al., 2015). Thus, phase-locked γ oscillations could be a useful physiological biomarker for ASD (see Nakamura et al., 2015 for discussion).

On a related note, if our model of Dynamic Cognomics is accurate, mutations in genes involved in establishing connections between the cortex and the basal ganglia may also lead to particular aspects of ASD. Among them, we wish to highlight *NLGN1* and *SHANK3*. *Nlgn1* knockout mice exhibit repetitive behaviors and abnormal corticostriatal synapses (Blundell et al., 2010). In turn, *SHANK3* is expressed in the basal ganglia, and when knocked out in mice it leads to abnormal social interactions and also to repetitive grooming behavior (Peça et al., 2011). *SHANK3* is a postsynaptic scaffolding protein which appears to be crucial for the maintenance of functional synapses as well as the adequate balance between neuronal excitation and inhibition. Interestingly, among the factors that seemingly modulate this aspect of brain function are the genes controlling circadian rhythms (Bourgeron, 2007). Another gene of interest is *NAV1.1* which encodes a sodium channel. In mice *Nav1.1* down-regulation in the medial septum and diagonal band of Broca dysregulates hippocampal oscillations and results in a spatial memory deficit (Bender et al., 2013).

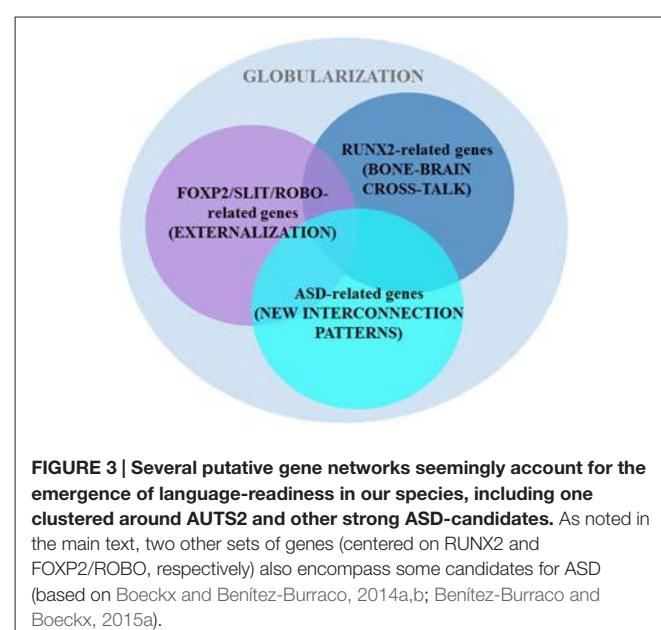
ASD-Candidates and Language Evolution

A more systematic account of genes related to language deficits in ASD may emerge from the new directions for exploring the genetic basis of language-readiness in humans proposed in Boeckx and Benítez-Burraco (2014a,b) and Benítez-Burraco and Boeckx (2015a). In these articles, it is noted that the unusually globular braincase of anatomically modern humans (AMH) may have led to changes in the wiring patterns and oscillatory behavior of the hominin brain. Specifically, an increase of cortical matter across anterior and posterior sites is expected to have followed the observed changes in the skull. In turn these changes may have provided the hominin brain with greater working memory resources and ultimately, with enhanced cross-modular connections (being mediated by the more central role played by the thalamus as a strong modulator of fronto-parietal activity and a connector of distant areas). This new neuronal workspace resulting from globularization also involved the cerebellum and particularly, the corpus callosum (Theofanopoulou, 2015), since interhemispheric integration is crucial for language (Poeppel, 2012). The AMH-specific rewiring of the brain had allowed us to transcend (better than other species) the signature limits of core knowledge systems and thus go beyond modular boundaries (Mithen, 1996; Boeckx, 2011). As discussed in Boeckx and Benítez-Burraco (2014a), our language-readiness (that is, our species-specific ability to learn and use languages) boils down to this enhanced cognitive ability. However, for language to exist this ability has to be further embedded inside the cognitive systems responsible for interpretation and externalization (**Figure 1**). Importantly, as noted in “From Brain Rhythmicity to Language Deficits

in ASD” Section, this embedding involves the embedding of high frequency oscillations inside oscillations operating at a lower frequency (see Boeckx and Benítez-Burraco, 2014a for details). This is why we expect that the emergence of our language-also readiness involved new patterns of long-distance connections among distributed neurons and thus new patterns of brain rhythmicity, aspects that are disturbed in ASD.

Boeckx and Benítez-Burraco put forward a putative gene network (involved in skull morphogenesis and thalamic development, but also in the regulation of GABAergic neurons within the forebrain) that was hypothesized to have been modified after our split from Neanderthals and Denisovans, providing the scaffolding for our species-specific mode of cognition. The network is centered around *RUNX2*, a gene showing strong signals of a selective sweep after our split from Neanderthals (Green et al., 2010; Perdomo-Sabogal et al., 2014). Nonetheless, it also encompasses some *DLX* genes (*DLX1*, *DLX2*, *DLX5*, and *DLX6*) and some *BMP* genes (*BMP2* and *BMP7*; see Boeckx and Benítez-Burraco, 2014a for details). This network is assigned the role of determining parts of language’s syntax-semantics interface. However, it is functionally related to the *FOXP2* and *ROBO/SLIT* interactomes that are claimed to be involved in the externalization of language (see Boeckx and Benítez-Burraco, 2014b for details; **Figure 3**). Overall, the *FOXP2*-*ROBO*/*SLIT*-*RUNX2* connections are interpreted as the result of an evolutionary convergence between the ancient externalization component and the emerging conceptual-intentional component (although see Balari and Lorenzo, 2015 for an argument that these latter networks are more heavily involved in the computational system).

Interestingly, changes in some ASD-candidate genes have been selected in AMHs after our split from extinct hominins, paradigmatically, in *AUTS2*, a gene that according to Green et al. (2010) displays the strongest signal of positive selection



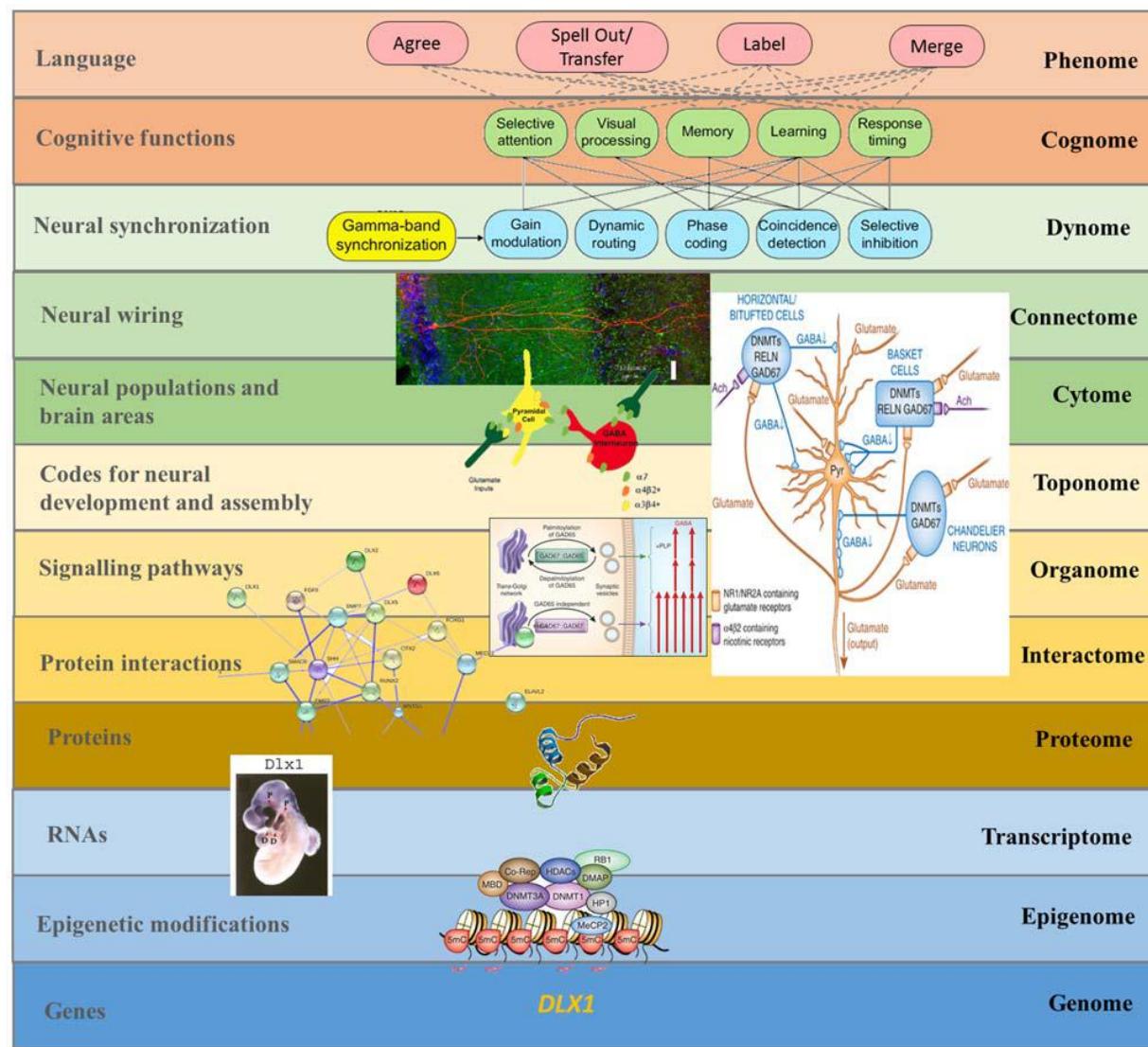


FIGURE 4 | A multilevel approach to language deficits in ASD from an oscillatory perspective. Understanding language problems in ASD demands a systems biology approach that seeks to unravel the nature and links between all biological factors involved. The figure shows one possible line of research focused on brain oscillations, although many others will need to be explored in the future to gain a comprehensive view of this complex issue. As noted in the main text, among the candidates for ASD we find several genes that have changed during recent human evolution and that are believed to be important for the emergence of language. One of them is DLX1, known to control aspects of skull and brain development. As discussed in Boeckx and Benítez-Burraco (2014a), DLX1 is expressed in neocortical GABAergic neurons and it regulates thalamic differentiation and interconnection with the cortex. As the String 9.1 network shows, DLX1 interacts with other core candidates for globularization, like RUNX2 and DLX2. RUNX2, DLX1, and DLX2 are key components of the GAD67 regulatory network, which is important for the normal development of GABAergic neurons within the hippocampus. As noted in the text, disturbances in GABAergic mediator system may contribute to the altered γ activity detected in the hippocampus of ASD children, which may impact on syntactic operations in ASD language. The expression pattern in the forebrain of transcription factors like DLX1 (here exemplified with an *in situ* hybridization of the Dlx1 in E10.5 mouse embryo) has changed over the course of our history. This may explain some of the changes that reshaped our species-specific program for the generation of neocortical local circuit neurons and, ultimately, the changes in GABAergic input to several brain areas (including the hippocampus). In turn, this may have contributed to the retuning of brain oscillations that brought about modern cognitive functions like language (at the top of the figure), although the exact role played by these basic cognitive operations in language processing is still unknown. This, then, is a composite figure elaborated by the authors. The dynamic-cognomic aspects of γ -oscillations (on the top of the figure) are from Bosman et al. (2014). The micrography of the single hippocampal CA1 pyramidal neuron is from <http://basulab.us/research/goals>. The schematic view of the hippocampal GABAergic neurons (below) is from Feduccia et al. (2012). The scheme of the two distinct mechanisms targeting GAD67 to vesicular pathways and presynaptic clusters is from Kanaani et al. (2010). The String 9.1 network is from Boeckx and Benítez-Burraco (2014a). The schematic view of the structure of the DLX1 protein has been taken from www.uscnk.com. The *in situ* hybridization of Dlx1 in E10.5 mouse embryo is from Panganiban and Rubenstein (2002). The schematic representation of a transcriptionally inactive promoter is from Grayson and Guidotti (2013). Finally, the large scheme on the right of the picture shows the effect of DNA methyltransferase overexpression on GABAergic neurons and is also from Grayson and Guidotti (2013) (of interest is that GABAergic promoter downregulation is observed in some cognitive disorders like schizophrenia, resulting in increased levels of some DNA methyltransferases like DNMT1 and 3A, and reduced GAD67, RELN and a variety of interneuron markers).

in AMHs compared to Neanderthals. Importantly, some of the candidates for ASD belong to the two set of genes highlighted above, including *CTNNB1*, *HRAS*, *DLX1*, *DLX5*, *PTEN*, and *SMURF1* (which are part of the set centered on *RUNX2*), and *ROBO2*, *FOXPI*, *POU3F2*, and *CNTNAP2* (which belong to the second set of genes; see Benítez-Burraco and Boeckx, 2015b for details), whereas some other candidates (including *AUTS2*, *PAX6* and some of its partners, like *TBR1* and *FEZF2*) provide additional links between the *RUNX2* and the ROBO-FOXP2 interactomes (**Figure 3**; see Benítez-Burraco and Boeckx, 2015a for details). Because the involvement of these three sets of genes in both interfaces of language, this overlapping may account for the observed deficits in ASD regarding language abilities (see Benítez-Burraco and Boeckx, 2015b for discussion). Furthermore, some of the candidates for the ASD dyname interact with some of the genes encompassing these interactomes. For example, according to Kuhlwilm et al. (2013) the expression of both *PDGFRB* and *NLGN1* changes after *RUNX2* transfection.

We wish to end by highlighting several similarities between the presumed Neanderthal head/brain/mind and the observed ASD phenotype that may be explained by the evidence presented above (and overall, by all these changes and new connections that contributed to the emergence of language-readiness in our species). As noted, the Neanderthal brain(case) was more elongated; the temporal pole, the orbitofrontal cortex, and the olfactory bulbs were smaller (Bastir et al., 2011); and they didn't show a uniform parietal surface enlargement (Bruner, 2010). The fact that ASD often appears alongside an abnormal head shape and higher rates of macrocephaly (Lainhart et al., 2006; Cheung et al., 2011), resulting from enhancements in frontal white matter and minicolumn pathology (Casanova et al., 2002; Vargas et al., 2005) speaks to the present globular perspective on linguistic computation, with our species-specific skull morphology influencing our distinctive brain wiring and cognitive abilities, including language. This perspective goes back to Kanner's (1943) seminal insights concerning individuals with ASD often having "intelligent" and "pensive" physiognomies. This overgrowth results in the forms of frontal over-connectivity discussed above, and stymies the enhancement of anterior to posterior brain region synchronization (Supekar et al., 2009; Coben et al., 2013, 2014). Moreover, as also noted above, both interhemispheric connectivity and language abilities are altered in ASD (Verly et al., 2014). ASD has been characterized in terms of a hyper-modular mind (Kenett et al., 2015) that lacks the cognitive flexibility observed in non-affected people. Interestingly, the Neanderthals mind has been described as a conglomerate of specialized intelligences lacking the cognitive flexibility of AMHs (Mithen, 1996, 2005). All of this does not suggest that ASD is an atavistic trait. However, the study of the etiopathogenesis of ASD (including candidate genes and abnormal brain rhythms) may benefit from ongoing studies of language evolution in the species (and vice versa). Recent research has linked the emergence of complex, highly prevalent conditions like ASD to the uncovering of cryptic genetic variation resulting from our evolutionary history (Gibson, 2009).

Finally, we wish to note that under the globularity hypothesis, visual abilities are claimed to be different in humans and Neanderthals, plausibly because of the selected differences in *PAX6* and related genes (see Benítez-Burraco and Boeckx, 2015a for discussion). It is of interest, then, that atypical visual processing in higher cognition has been extensively documented in ASD, with a general increased reliance on visual imagery perhaps being a compensatory mechanism in lexical processing. For instance, greater occipital activation is seen in ASD compared to TD controls during embedded figure tasks, where subjects must detect geometric figures within a larger visual pattern (Ring et al., 1999). Baron-Cohen et al. (2005) also document enhanced visual-spatial skills relative to verbal skills. However, Bertone et al. (2003) presented evidence that the otherwise high visual integration abilities of subjects with ASD degenerate when second-order visual information is introduced, suggesting that any high performance visual skills they have are limited to basic, non-hierarchical stimuli. Problems with hierarchies also appear to arise in the performance of people with ASD during the perception of biological motion and facial masks (Blake et al., 2003; Deruelle et al., 2004). Relatedly, the resistance of people with ASD to McGurk effects (McGurk and MacDonald, 1976), implicating speech and facial articulatory integration, may possibly be explained not by invoking social comprehension skills, but rather by pointing to their impaired facial-speech hierarchical predicting abilities. Broadly speaking, "autistic perceptual processes are primarily not hierarchical, favoring fragmentary over holistic processing" (Bourguignon et al., 2012: p. 139). While word-level comprehension is either intact or enhanced in ASD (likely a consequence of the increased posterior temporal and occipital activation noted above), sentence-level hierarchical processing is typically impaired, with superior visual processing being insufficient for the processing of hierarchical syntactic principles like c-command (involved in binding relations between pronouns and their antecedents) and A-movement (involved in the formation of passives; Perovic et al., 2007). This suggests a more general deficit in hierarchical processing in ASD, which, under the present cognome-dynome model, is a consequence of oscillatory impairments and the resultant coupling restrictions. ASD also often leads to reduced verbal imagination and inner speech (Whitehouse et al., 2006), along with reduced symbolic play (Honey et al., 2006). Fries (2015: p. 232) documents that visual scenes induce multiple γ rhythms with varying frequencies, yielding a wide "gamma landscape" which, we believe, the enhanced visual cognition of autistic individuals can easily find its place. In contrast, the landscapes of linguistic computation rely on the coupling of a range of frequency fields, leaving certain properties of linguistic cognition susceptible to disruption.

CONCLUSION

Overall, these considerations may provide a suitable response to Dehaene et al. (2015: p. 2) observation that linguistic computation requires "a specific recursive neural code, as yet unidentified by electrophysiology, possibly unique to humans, and which may explain the singularity of human language

and cognition". Our "rhythmic" project will be long-lasting, encompassing a top-down approach to language processing in the brain, from linguistic features to brain rhythms to genetics (**Figure 4**), but we wish to highlight that it encompasses a number of neurodegenerative and neurodevelopmental disorders that should help yield insights into the structure and function of language and mind. In fact, we expect disorders to be particular areas within the whole morpho-space or adaptive landscape of language development in the species, defined by the site of basic brain rhythms (see Benítez-Burraco, 2016 for discussion). As we have argued, ASD is of particular interest in virtue of it representing a mode of cognition and perception distinct from, but plainly related to, normally functioning linguistic cognition.

We further expect that the present perspective of Dynamic Cognomics has the potential to provide robust endophenotypes of ASD. Importantly for ongoing research into the biological underpinnings of ASD, the process of generating and testing dynamic predictions will permit the falsification of a large number of possible dyname-cognome linking hypotheses, allowing the refinement of the present model of ASD, while simultaneously granting more comprehensive and earlier diagnoses of language deficits in this condition. We expect to extend this perspective to other conditions such as schizophrenia, which has been claimed to be at the opposite pole to ASD within a continuum of modes of cognition also encompassing TD cognition (see Crespi and Badcock, 2008 for discussion). Interestingly, the abnormal γ documented in schizophrenics by Xu et al. (2013) is likely a cause of the problems with producing pronounceable nonwords and confusion of antonyms seen in this disorder (Stephane et al., 2007), in contrast to this rhythm's implication in over-connectivity in ASD. Inner speech is also potentiated in schizophrenia through auditory verbal hallucinations, unlike in ASD (Moseley et al., 2013). Moreover, a thalamocortical "dysrhythmia" has been documented in subjects with schizophrenia, obsessive-compulsive disorder and depression by Schulman et al. (2011). Evidence from EEG, MEG and anatomical studies suggests

that oscillatory synchronization abnormalities may play a core role in the pathophysiology of schizophrenia (Uhlhaas and Singer, 2010). In particular, synchronization between γ and β appears to be abnormal in several studies examining visuo-perceptual organization and auditory processing (Spencer et al., 2003; Symond et al., 2005; Uhlhaas et al., 2006), while γ activity displays severe widespread deficiency during perceptual organization tasks (Tillmann et al., 2008). Overall, these findings suggest a distinct oscillopathic profile from that of ASD.

To conclude, we wish to note that these considerations also speak against the newly emerging view in the literature (Berwick and Chomsky, 2016) that language evolution will either remain a mystery or should be explored at the neurological level purely through functional localization studies. As discussed, abnormal cognitive/linguistic development in our species should help unravel the evolutionary path followed by our faculty of language, as the high number of candidates for ASD selected in AMH nicely illustrates. In this respect, schizophrenia is again a natural target (see Berlim et al., 2003 or Crow, 2008 for discussion). It remains to be seen how far the present dynamic model of linguistic computation can be used to enhance understanding of other language-related oscillopathies.

AUTHOR CONTRIBUTIONS

Both authors contributed to all sections of the article.

ACKNOWLEDGMENTS

Preparation of this work was supported in part by funds from the Spanish Ministry of Economy and Competitiveness (grant numbers FFI-2013-43823-P and FFI2014-61888-EXP to AB-B). This work was also supported by an Economic and Social Research Council scholarship (1474910). The authors would also like to thank the two reviewers for their valuable comments.

REFERENCES

- Abbeduto, L., and Hesketh, L. J. (1997). Pragmatic development in individuals with mental retardation: leaning to use language in social interactions. *Ment. Retard. Dev. Disabil. Res. Rev.* 3, 323–333. doi: 10.1002/(sici)1098-2779(1997)3:4<323::aid-mrdd7>3.0.co;2-0
- Alamri, S., Zhang, E., Theofanopoulou, C., Castillo, G., Shi, E., Martins, P. T., et al. (2015). Implications of subcortical structures in aphasia. Frontiers in Psychology Conference Abstract: Academy of Aphasia 53rd Annual Meeting. Tucson, USA. doi: 10.3389/conf.fpsyg.2015.65.00081
- Alexander, A. L., Lee, J. E., Lazar, M., Boudos, R., DuBray, M. B., Oakes, T. R., et al. (2007). Diffusion tensor imaging of the corpus callosum in autism. *Neuroimage* 34, 61–73. doi: 10.1016/j.neuroimage.2006.08.032
- Bailey, A., Phillips, W., and Rutter, M. (1996). Autism: towards an integration of clinical, genetic, neuropsychological and neurobiological perspectives. *J. Child Psychol. Psychiatry* 37, 89–126. doi: 10.1111/j.1469-7610.1996.tb01381.x
- Balari, S., and Lorenzo, G. (2015). It is an organ, it is new, it is not a new organ. Conceptualizing language from a homological perspective. *Front. Ecol. Evol.* 3:58. doi: 10.3389/fevo.2015.00058
- Bangel, K. A., Batty, M., Ye, A. X., Meaux, E., Taylor, M. J., and Doesburg, S. M. (2014). Reduced bet band connectivity during number estimation in autism. *Neuroimage Clin.* 6, 202–213. doi: 10.1016/j.nicl.2014.08.020
- Baron-Cohen, S., Knickmeyer, R. C., and Belmonte, M. K. (2005). Sex differences in the brain: implications for explaining autism. *Science* 310, 819–823. doi: 10.1126/science.1115455
- Bartos, M., Vida, I., and Jonas, P. (2007). Synaptic mechanisms of synchronized gamma oscillations in inhibitory interneuron networks. *Nat. Rev. Neurosci.* 8, 45–56. doi: 10.1038/nrn2044
- Bastir, M., Rosas, A., Gunz, P., Peña-Melian, A., Manzi, G., Harvati, K., et al. (2011). Evolution of the base of the brain in highly encephalized human species. *Nat. Commun.* 2:588. doi: 10.1038/ncomms1593
- Belmonte, M. K., Saxena-Chandhok, T., Cherian, R., Munneer, R., George, L., and Karanth, P. (2013). Oral motor deficits in speech-impaired children with autism. *Front. Integr. Neurosci.* 7:47. doi: 10.3389/fnint.2013.00047
- Benayed, R., Gharani, N., Rossman, I., Mancuso, V., Lazar, G., Kamdar, S., et al. (2005). Support for the homeobox transcription factor gene ENGRAILED 2 as an autism spectrum disorder susceptibility locus. *Am. J. Hum. Genet.* 77, 851–868. doi: 10.1086/497705

- Bender, A. C., Natola, H., Ndong, C., Holmes, G. L., Scott, R. C., and Lenck-Santini, P. P. (2013). Focal *Scn1a* knockdown induces cognitive impairment without seizures. *Neurobiol. Dis.* 54, 297–307. doi: 10.1016/j.nbd.2012.12.021
- Benítez-Burraco, A. (2016). “A biolinguistic approach to language disorders: towards a paradigm shift in clinical linguistics,” in *Advances in Biolinguistics: The Human Language Faculty and its Biological Basis*, eds K. Fujita and C. Boeckx (London: Routledge), 396–420.
- Benítez-Burraco, A., and Boeckx, C. (2015a). Possible functional links among brain- and skull-related genes selected in modern humans. *Front. Psychol.* 6:794. doi: 10.3389/fpsyg.2015.00794
- Benítez-Burraco, A., and Boeckx, C. (2015b). Approaching motor and language deficits in autism from below: a biolinguistic perspective. *Front. Integr. Neurosci.* 9:25. doi: 10.3389/fnint.2015.00025
- Berlim, M. T., Mattevi, B. S., Belmonte-de-Abreu, P., and Crow, T. J. (2003). The etiology of schizophrenia and the origin of language: overview of a theory. *Compr. Psychiatry* 44, 7–14. doi: 10.1053/comp.2003.50003
- Berman, J. I., Liu, S., Bloy, L., Blaskey, L., Roberts, T. P. L., and Edgar, J. C. (2015). Alpha-to-gamma phase-amplitude coupling methods and application to autism spectrum disorder. *Brain Connect.* 5, 80–90. doi: 10.1089/brain.2014.0242
- Bertone, A., Mottron, L., Jelenic, P., and Faubert, J. (2003). Motion perception in autism: a ‘complex’ issue. *J. Cogn. Neurosci.* 15, 218–225. doi: 10.1162/089892903321208150
- Berwick, R., and Chomsky, N. (2016). *Why Only Us: Language and Evolution*. Cambridge, MA: MIT Press.
- Blake, R., Turner, L. M., Smoski, M. J., Pozdol, S. L., and Stone, W. L. (2003). Visual recognition of biological motion is impaired in children with autism. *Psychol. Sci.* 14, 151–157. doi: 10.1111/1467-9280.01434
- Blatt, G. J., Fitzgerald, C. M., Guptill, J. T., Booker, A. B., Kemper, T. L., and Bauman, M. L. (2001). Density and distribution of hippocampal neurotransmitter receptors in autism: an autoradiographic study. *J. Autism Dev. Disord.* 31, 537–543. doi: 10.1023/A:1013238809666
- Blundell, J., Blaiss, C. A., Etherton, M. R., Espinosa, F., Tabuchi, K., Walz, C., et al. (2010). Neuroligin-1 deletion results in impaired spatial memory and increased repetitive behavior. *J. Neurosci.* 30, 2115–2129. doi: 10.1523/JNEUROSCI.4517-09.2010
- Boeckx, C. (2011). *Language in Cognition: Uncovering Mental Structures and the Rules Behind Them*. Malden, MA: Wiley-Blackwell.
- Boeckx, C., and Benítez-Burraco, A. (2014a). The shape of the human language-ready brain. *Front. Psychol.* 5:282. doi: 10.3389/fpsyg.2014.00282
- Boeckx, C., and Benítez-Burraco, A. (2014b). Globularity and language-readiness: generating new predictions by expanding the set of genes of interest. *Front. Psychol.* 5:1324. doi: 10.3389/fpsyg.2014.01324
- Bonora, E., Bacchelli, E., Levy, E. R., Blasi, F., Marlow, A., Monaco, A. P., et al. (2002). Mutation screening and imprinting analysis of four candidate genes for autism in the 7q32 region. *Mol. Psychiatry* 7, 289–301. doi: 10.1038/sj.mp.4001004
- Bornkessel-Schlesewsky, I., Schlesewsky, M., and Small, S. L. (2014). Implementation is crucial but must be neurobiologically grounded. Comment on “Toward a computational framework for cognitive biology: unifying approaches from cognitive neuroscience and comparative cognition” by W. Tecumseh Fitch. *Phys. Life Rev.* 11, 365–366. doi: 10.1016/j.plrev.2014.06.007
- Bosman, C. A., Lansink, C. S., and Pennartz, C. M. A. (2014). Functions of gamma-band synchronization in cognition: from single circuits to functional diversity across cortical and subcortical systems. *Eur. J. Neurosci.* 39, 1982–1999. doi: 10.1111/ejn.12606
- Bourgeron, T. (2007). The possible interplay of synaptic and clock genes in autism spectrum disorders. *Cold Spring Harb. Symp. Quant. Biol.* 72, 645–654. doi: 10.1101/sqb.2007.72.020
- Bourguignon, N., Nadig, A., and Valois, D. (2012). The biolinguistics of autism: emergent perspectives. *Biolinguistics* 6, 124–165.
- Braeutigam, S., Switthenby, S. J., and Bailey, A. J. (2008). Contextual integration the unusual way: magnetoencephalographic study of responses to semantic violation in individuals with autism spectrum disorder. *Eur. J. Neurosci.* 27, 1026–1036. doi: 10.1111/j.1460-9568.2008.06064.x
- Brederoo, S. G., Bos, L. S., Dragoy, O., Bastiaanse, R., and Baggio, G. (2015). Gamma oscillations as a neural signature of shifting times in narrative language. *PLoS One* 10:e0121146. doi: 10.1371/journal.pone.0121146
- Brock, J., Brown, C. C., Boucher, J., and Rippon, G. (2002). The temporal binding deficit hypothesis of autism. *Dev. Psychopathol.* 14, 209–224. doi: 10.1017/s0954579402002018
- Bruner, E. (2010). Morphological differences in the parietal lobes within the human genus: a neurofunctional perspective. *Curr. Anthropol.* 51, S77–S88. doi: 10.1086/650729
- Brunetti, M., Zappasodi, F., Marzetti, L., Perrucci, M. G., Cirillo, S., Romani, G. L., et al. (2014). Do you know what I mean? Brain oscillations and the understanding of communicative intentions. *Front. Hum. Neurosci.* 8:36. doi: 10.3389/fnhum.2014.00036
- Buard, I., Rogers, S. J., Hepburn, S., Kronberg, E., and Rojas, D. C. (2013). Altered oscillation patterns and connectivity during picture naming in autism. *Front. Hum. Neurosci.* 7:742. doi: 10.3389/fnhum.2013.00742
- Buzsáki, G., Logothetis, N., and Singer, W. (2013). Scaling brain size, keeping timing: evolutionary preservation of brain rhythms. *Neuron* 80, 751–764. doi: 10.1016/j.neuron.2013.10.002
- Campbell, D. B., Sutcliffe, J. S., Ebert, P. J., Militerni, R., Bravaccio, C., Trillo, S., et al. (2006). A genetic variant that disrupts MET transcription is associated with autism. *Proc. Natl. Acad. Sci. U.S.A.* 103, 16834–16839. doi: 10.1073/pnas.0605296103
- Casanova, M. F., Buxhoeveden, D. P., Switala, A. E., and Roy, E. (2002). Minicolumnarpathology in autism. *Neurology* 58, 428–432. doi: 10.1212/WNL.58.3.428
- Castelhano, J., Bernardino, I., Rebola, J., Rodriguez, E., and Catelo-Branco, M. (2015). Oscillations or synchrony? Disruption of neural synchrony despite enhanced gamma oscillations in a model of disrupted perceptual coherence. *J. Cogn. Neurosci.* 27, 2416–2426. doi: 10.1162/jocn_a_00863
- Castermans, D., Wilquet, V., Parthoens, E., Huysmans, C., Steyaert, J., Swinnen, L., et al. (2003). The neurobeachin gene is disrupted by a translocation in a patient with idiopathic autism. *J. Med. Genet.* 40, 352–356. doi: 10.1136/jmg.40.5.352
- Chatteropadhyaya, B., and Cristo, G. D. (2012). GABAergic circuit dysfunctions in neurodevelopmental disorders. *Front. Psychiatry* 3:51. doi: 10.3389/fpsyg.2012.00051
- Cheung, C., McAlonan, G. M., Fung, Y. Y., Fung, G., Yu, K. K., Tai, K. S., et al. (2011). MRI study of minor physical anomaly in childhood autism implicates aberrant neurodevelopment in infancy. *PLoS One* 6:e20246. doi: 10.1371/journal.pone.0020246
- Cheung, J., Petek, E., Nakabayashi, K., Tsui, L. C., Vincent, J. B. Y., and Scherer, S. W. (2001). Identification of the human cortactin-binding protein-2 gene from the autism candidate region at 7q31. *Genomics* 78, 7–11. doi: 10.1006/geno.2001.6651
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.
- Cohen, R., Chabot, R. J., and Hirshberg, L. (2013). “EEG analyses in the assessment of autistic disorders,” in *Imaging the Brain in Autism*, eds M. F. Casanova, A. S. El-Baz, and J. S. Suri (NewYork: Springer), 349–370.
- Cobb, S. R., and Davies, C. H. (2005). Cholinergic modulation of hippocampal cells and circuits. *J. Physiol.* 562, 81–88. doi: 10.1113/jphysiol.2004.076539
- Cohen, R., Mohammad-Rezazadeh, I., and Cannon, R. I. (2014). Using quantitative and analytic EEG methods in the understanding of connectivity in autism spectrum disorders: a theory of mixed over- and under-connectivity. *Front. Hum. Neurosci.* 8:45. doi: 10.3389/fnhum.2014.00045
- Comings, D. E., Wu, S., Chiu, C., Muhleman, D. Y., and Sverd, J. (1996). Studies of the c-Harvey-Ras gene in psychiatric disorders. *Psychiatry Res.* 63, 25–32. doi: 10.1016/0165-1781(96)02829-6
- Cook, E. H., Courchesne, R. Y., Cox, N. J., Lord, C., Gonen, D., Guter, S. J., et al. (1998). Linkage-disequilibrium mapping of autistic disorder, with 15q11–13 markers. *Am. J. Hum. Genet.* 62, 1077–1083. doi: 10.1086/301832
- Cornew, L., Roberts, T. P. L., Blaskey, L., and Edgar, J. C. (2012). Resting-state oscillatory activity in autism spectrum disorders. *J. Autism Dev. Disord.* 42, 1884–1894. doi: 10.1007/s10803-011-1431-6
- Courchesne, E., and Pierce, K. (2005). Why the frontal cortex in autism might be talking only to itself: local over-connectivity but long-distance disconnection. *Curr. Opin. Neurobiol.* 15, 225–230. doi: 10.1016/j.conb.2005.03.001
- Crespi, B., and Badcock, C. (2008). Psychosis and autism as diametrical disorders of the social brain. *Behav. Brain Sci.* 31, 241–320; discussion 261–320. doi: 10.1017/S0140525X08004214
- Crow, T. J. (2008). The ‘big bang’ theory of the origin of psychosis and the faculty of language. *Schizophr. Res.* 102, 31–52. doi: 10.1016/j.schres.2008.03.010

- Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., and Pallier, C. (2015). The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. *Neuron* 88, 2–19. doi: 10.1016/j.neuron.2015.09.019
- Deruelle, C., Rondan, C., Gepner, B., and Tardif, C. (2004). Spatial frequency and face processing in children with autism and Asperger syndrome. *J. Autism Dev. Disord.* 34, 199–210. doi: 10.1023/b:ajad.0000022610.09668.4c
- Diehl, J. J., Friedberg, C., Paul, R., and Snedeker, J. (2015). The use of prosody during syntactic processing in children and adolescents with autism spectrum disorders. *Dev. Psychopathol.* 27, 867–884. doi: 10.1017/S0954579414000741
- Dodd, E. (2005). *Understanding Autism*. Sydney: Elsevier Australia.
- Doesburg, S. M., Vidal, J., and Taylor, M. J. (2013). Reduced theta connectivity during set-shifting in children with autism. *Front. Hum. Neurosci.* 7:785. doi: 10.3389/fnhum.2013.00785
- Donner, T. H., and Siegel, M. (2011). A framework for local cortical oscillation patterns. *Trends Cogn. Sci.* 15, 191–199. doi: 10.1016/j.tics.2011.03.007
- Dunn, M., Gomes, H., and Sebastian, M. (1996). Prototypicality of responses of autistic, language disordered and normal children in a word fluency task. *Child Neuropsychol.* 2, 99–108. doi: 10.1080/09297049608401355
- Edgar, J. C., Khan, S. Y., Blaskey, L., Chow, V. Y., Rey, M., Gaetz, W., et al. (2015). Neuromagnetic oscillations predict evoked-response latency delays and core language deficits in autism spectrum disorders. *J. Autism Dev. Disord.* 45, 395–405. doi: 10.1007/s10803-013-1904-x
- Eigsti, I. M., Bennetto, L., and Dadlani, M. (2007). Beyond pragmatics: morphosyntactic development in autism. *J. Autism Dev. Disord.* 37, 1007–1023. doi: 10.1007/s10803-006-0239-2
- Epstein, S. D., Kitahara, H., and Seely, T. D. (2015). “From *Aspects*’ ‘daughterless mothers’ (aka delta nodes) to *POP*’s ‘motherless sets’ (aka non-projection): a selective history of the evolution of simplest merge,” in *50 Years Later: Reflections on Chomsky’s Aspects.*, eds Á. Gallego and D. Ott (Cambridge, MA: MITWPL), 99–112.
- Fatemi, S. H., Reutiman, T. J., Folsom, T. D., Rooney, R. J., Patel, D. H., and Thuras, P. D. (2010). mRNA and protein levels for GABA $\alpha 4$, $\alpha 5$, $\beta 1$ and GABABR1 receptors are altered in brains from subjects with autism. *J. Autism Dev. Disord.* 40, 743–750. doi: 10.1007/s10803-009-0924-z
- Fedorenko, E., and Thompson-Schill, S. L. (2014). Reworking the language network. *Trends Cogn. Sci.* 18, 120–126. doi: 10.1016/j.tics.2013.12.006
- Feduccia, A. A., Chatterjee, S., and Bartlett, S. E. (2012). Neuronal nicotinic acetylcholine receptors: neuroplastic changes underlying alcohol and nicotine addictions. *Front. Mol. Neurosci.* 5:83. doi: 10.3389/fnmol.2012.00083
- Fitch, W. T. (2014). Toward a computational framework for cognitive biology: unifying approaches from cognitive neuroscience and comparative cognition. *Phys. Life Rev.* 11, 329–364. doi: 10.1016/j.plrev.2014.04.005
- Fletcher, P. T., Whitaker, R. T., Tao, R., DuBray, M. B., Froehlich, A., Ravichandran, C., et al. (2010). Microstructural connectivity of the arcuate fasciculus in adolescents with high-functioning autism. *Neuroimage* 51, 1117–1125. doi: 10.1016/j.neuroimage.2010.01.083
- Frawley, W. (2008). Private speech, cognitive-computational control and the autism-psychosis continuum. *Behav. Brain Sci.* 31, 269–270. doi: 10.1017/s0140525x08004305
- Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn. Sci.* 9, 474–480. doi: 10.1016/j.tics.2005.08.011
- Fries, P. (2015). Rhythms for cognition: communication through coherence. *Neuron* 88, 220–235. doi: 10.1016/j.neuron.2015.09.034
- Fua, P., and Knott, G. W. (2015). Modeling brain circuitry over a wide range of scales. *Front. Neuroanat.* 9:42. doi: 10.3389/fnana.2015.00042
- Gandal, M. J., Edgar, J. C., Ehrlichman, R. S., Mehta, M., Roberts, T. P. L., and Siegel, S. J. (2010). Validating γ oscillations and delayed auditory responses as translational biomarkers of autism. *Biol. Psychiatry* 68, 1100–1106. doi: 10.1016/j.biopsych.2010.09.031
- Geschwind, G. H., and State, M. W. (2015). Gene hunting in autism spectrum disorder: on the path to precision medicine. *Lancet Neurol.* 14, 1109–1120. doi: 10.1016/s1474-4422(15)00044-7
- Gibson, G. (2009). Decanalization and the origin of complex disease. *Nat. Rev. Genet.* 10, 134–140. doi: 10.1038/nrg2502
- Gibson, K. M., Christensen, E., Jakobs, C., Fowler, B., Clarke, M. A., Hammersen, G., et al. (1997). The clinical phenotype of succinic semialdehyde dehydrogenase deficiency (4-hydroxybutyric aciduria): case reports of 23 new patients. *Pediatrics* 99, 567–574. doi: 10.1542/peds.99.4.567
- Giraud, A.-L., and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* 15, 511–517. doi: 10.1038/nn.3063
- Goffin, D., Brodkin, E. S., Blendy, J. A., Siegel, S. J., and Zhou, Z. (2014). Cellular origins of auditory event-related potential deficits in Rett syndrome. *Nat. Neurosci.* 17, 804–806. doi: 10.1038/nn.3710
- Grayson, D. R., and Guidotti, A. (2013). The dynamics of DNA methylation in schizophrenia and related psychiatric disorders. *Neuropsychopharmacology* 38, 138–166. doi: 10.1038/npp.2012.125
- Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., et al. (2010). A draft sequence of the Neanderthal genome. *Science* 328, 710–722. doi: 10.1126/science.1188021
- Hagoort, P. (2014). Nodes and networks in the neural architecture for language: broca’s region and beyond. *Curr. Opin. Neurobiol.* 28, 136–141. doi: 10.1016/j.conb.2014.07.013
- Hahamy, A., Behrmann, M., and Malach, R. (2015). The idiosyncratic brain: distortion of spontaneous connectivity patterns in autism spectrum disorder. *Nat. Neurosci.* 18, 302–309. doi: 10.1038/nn.3919
- Hannemann, R., Obleser, J., and Eulitz, C. (2007). Top-down knowledge supports the retrieval of lexical information from degraded speech. *Brain Res.* 1153, 134–143. doi: 10.1016/j.brainres.2007.03.069
- Harris, G. J., Chabris, C. F., Clark, J., Urban, T., Aharon, I., Steele, S., et al. (2006). Brain activation during semantic processing in autism spectrum disorders via functional magnetic resonance imaging. *Brain Cogn.* 61, 54–68. doi: 10.1016/j.bandc.2005.12.015
- Hauser, M., Chomsky, N., and Fitch, W. T. (2002). The faculty of language: what is it, who has it and how did it evolve. *Science* 298, 1569–1579. doi: 10.1126/science.298.5598.1569
- Heistek, T. S., Jaap Timmerman, A., Spijker, S., Brussaard, A. B., and Mansveld, H. D. (2010). GABAergic synapse properties may explain genetic variation in hippocampal network oscillations in mice. *Front. Cell Neurosci.* 4:18. doi: 10.3389/fncel.2010.00018
- Hinzen, W., Rosselló, J., Mattos, O., Schroeder, K., and Vila, E. (2015). The image of mind in the language of children with autism. *Front. Psychol.* 6:841. doi: 10.3389/fpsyg.2015.00841
- Hinzen, W., and Sheehan, M. (2013). *The Philosophy of Universal Grammar*. Oxford: Oxford University Press.
- Honey, E., Leekam, S. R., Turner, M., and McConachie, H. (2006). Repetitive behaviour and play in typically developing children and children with autism. *J. Autism Dev. Disord.* 37, 1107–1115. doi: 10.1007/s10803-006-0253-4
- Howlin, P. (2003). Outcome in high-functioning adults with autism with and without early language delays: implications for the differentiation between autism and Asperger syndrome. *J. Autism Dev. Disord.* 33, 3–13. doi: 10.1023/A:1022270118899
- Hyafil, A. (2015). Misidentifications of specific forms of cross-frequency coupling: three warnings. *Front. Neurosci.* 9:370. doi: 10.3389/fnins.2015.00370
- Hyde, K. L., Samson, F., Evans, A. C., and Mottron, L. (2010). Neuroanatomical differences in brain areas implicated in perceptual and other core features of autism revealed by cortical thickness analysis and voxel-based morphometry. *Hum. Brain Mapp.* 31, 556–566. doi: 10.1002/hbm.20887
- Jeste, S. S., and Geschwind, D. H. (2014). Disentangling the heterogeneity of autism spectrum disorder through genetic findings. *Nat. Rev. Neurol.* 10, 74–81. doi: 10.1038/nrneurol.2013.278
- Jochaut, D., Lehongre, K., Saitovitch, A., Devauchelle, A.-D., Olasagasti, I., Chabane, N., et al. (2015). Atypical coordination of cortical oscillations in response to speech in autism. *Front. Hum. Neurosci.* 9:171. doi: 10.3389/fnhum.2015.00171
- Jordan, R. (2010). *Autism Spectrum Disorders: An Introductory Handbook for Practitioners*. Oxfordshire: David Fulton Publishers.
- Kajizuka, M., Miyachi, T., Matsuzaki, H., Iwata, K., Shinmura, C., Suzuki, K., et al. (2010). Serum levels of platelet-derived growth factor BB homodimers are increased in male children with autism. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 34, 154–158. doi: 10.1016/j.pnpbp.2009.10.017
- Kamio, Y., Robins, D., Kelley, E., Swainson, B., and Fein, D. (2007). Atypical lexical/semantic processing in high-functioning autism spectrum disorders without early language delay. *J. Autism Dev. Disord.* 37, 1116–1122. doi: 10.1007/s10803-006-0254-3

- Kanaani, J., Kolibachuk, J., Martinez, H., and Baekkeskov, S. (2010). Two distinct mechanisms target GAD67 to vesicular pathways and presynaptic clusters. *J. Cell. Biol.* 190, 911–925. doi: 10.1083/jcb.200912101
- Kanner, L. (1943). Autistic disturbances of affective contact. *Nerv. Child* 2, 217–250.
- Kenett, Y. N., Gold, R., and Faust, M. (2015). The hyper-modular associative mind: a computational analysis of associative responses of persons with Asperger syndrome. *Lang. Speech* doi: 10.1177/0023830915589397 [Epub ahead of print].
- Khan, S., Gramfort, A., Shetty, N. R., Kitzbichler, M. G., Ganeshan, S., Moran, J. M., et al. (2013). Local and long-range functional connectivity is reduced in concert in autism spectrum disorders. *Proc. Natl. Acad. Sci. U S A* 110, 3107–3112. doi: 10.1073/pnas.1214533110
- Khanna, P., and Carmena, J. M. (2015). Neural oscillations: beta band activity across motor networks. *Curr. Opin. Neurobiol.* 32, 60–67. doi: 10.1016/j.conb.2014.11.010
- Kikuchi, M., Shitamichi, K., Yoshimura, Y., Ueno, S., Hiraishi, H., Hirosawa, T., et al. (2013). Altered brain connectivity in 3-to 7-year-old children with autism spectrum disorder. *Neuroimage Clin.* 2, 394–401. doi: 10.1016/j.nicl.2013.03.003
- Kikuchi, M., Yoshimura, Y., Hiraishi, H., Munehisa, T., Hashimoto, T., Tsubokawa, T., et al. (2015). Reduced long-range functional connectivity in young children with autism spectrum disorder. *Soc. Cogn. Affect. Neurosci.* 10, 248–254. doi: 10.1093/scan/nsu049
- Kjelgaard, M. M., and Tager-Flusberg, H. (2001). An investigation of language impairment in autism: implications for genetic subgroups. *Lang. Cogn. Process.* 16, 287–308. doi: 10.1080/01690960042000058
- Kopell, N., J. Gritton, H. J., Whittington, M. A., and Kramer, M. A. (2014). Beyond the connectome: the dynome. *Neuron* 83, 1319–1328. doi: 10.1016/j.neuron.2014.08.016
- Kriete, T., Noelle, D. C., Cohen, J. D., and O'Reilly, R. C. (2013). Indirection and symbol-like processing in the prefrontal cortex and basal ganglia. *Proc. Natl. Acad. Sci. U S A* 110, 16390–16395. doi: 10.1073/pnas.1303547110
- Kuhl, P. (2007). Is speech learning 'gated' by the social brain? *Dev. Sci.* 10, 110–120. doi: 10.1111/j.1467-7687.2007.00572.x
- Kuhlwilm, M., Davierwala, A., and Pääbo, S. (2013). Identification of putative target genes of the transcription factor RUNX2. *PLoS One.* 8:e83218 doi: 10.1371/journal.pone.0083218
- Lainhart, J. E., Bigler, E. D., Bocijan, M., Coon, H., Dinh, E., Dawson, G., et al. (2006). Head circumference and height in autism: a study by the collaborative program of excellence in autism. *Am. J. Med. Genet. A* 140, 2257–2274. doi: 10.1002/ajmg.a.31465
- Lainhart, J. E., Ozonoff, S., Coon, H., Krasny, L., Dinh, E., Nice, J., et al. (2002). Autism, regression and the broader autism phenotype. *Am. J. Med. Genet.* 113, 231–237. doi: 10.1002/ajmg.10615
- Liao, W., Gandal, M. J., Ehrlichman, R. S., Siegel, S. J., and Carlson, G. C. (2012). MeCP2+/- mouse model of RTT reproduces auditory phenotypes associated with Rett syndrome and replicate select EEG endophenotypes of autism spectrum disorder. *Neurobiol. Dis.* 46, 88–92. doi: 10.1016/j.nbd.2011.12.048
- Lindgren, K., Folstein, S., Tomblin, J. B., and Tager-Flusberg, H. (2009). Language and reading abilities of children with autism spectrum disorders and specific language impairment and their first-degree relatives. *Autism Res.* 2, 22–38. doi: 10.1002/aur.63
- Lord, C., Risi, S., Lambrecht, L., Cook, E. H. Jr., Leventhal, B. L., DiLavore, P. C., et al. (2000). The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J. Autism Dev. Disord.* 30, 205–223. doi: 10.1023/A:1005592401947
- Lord, C., Shulman, C., and DiLavore, P. (2004). Regression and word loss in autistic spectrum disorders. *J. Child Psychol. Psychiatry* 45, 936–955. doi: 10.1111/j.1469-7610.2004.t01-1-00287.x
- Luyster, R., Lopez, K., and Lord, C. (2007). Characterizing communicative development in children referred for autism spectrum disorders using the MacArthur-Bates Communicative Development Inventory (CDI). *J. Child Lang.* 34, 623–654. doi: 10.1017/s0305000907008094
- Mahfouz, A., Ziats, M. N., Rennert, O. M., Lelieveldt, B. P. F., and Reinders, M. J. T. (2015). Shared pathways among autism candidate genes determined by co-expression network analysis of the developing human brain transcriptome. *J. Mol. Neurosci.* 57, 580–594. doi: 10.1007/s12031-015-0641-3
- Mandy, W., and Skuse, D. (2008). What is the association between the social-communication element of autism and repetitive interests, behaviours and activities? *J. Child Psychol. Psychiatry* 49, 795–808. doi: 10.1111/j.1469-7610.2008.01911.x
- McAlonan, G. M., Suckling, J., Wong, N., Cheung, V., Lienkenkaemper, N., Cheung, C., et al. (2008). Distinct patterns of grey matter abnormality in high-functioning autism and Asperger's syndrome. *J. Child Psychol. Psychiatry* 49, 1287–1295. doi: 10.1111/j.1469-7610.2008.01933.x
- McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748. doi: 10.1038/264746a0
- Meyer, L., Grigutsch, M., Schmuck, N., Gaston, P., and Friederici, A. D. (2015). Frontal-posterior theta oscillations reflect memory retrieval during sentence comprehension. *Cortex* 71, 205–218. doi: 10.1016/j.cortex.2015.06.027
- Mithen, S. (2005). *The Singing Neanderthals: The Origins of Music, Language, Mind and Body*. London: Weidenfeld and Nicholson.
- Mithen, S. J. (1996). *The Prehistory of the Mind: A Search for the Origins of Art, Science and Religion*. London: Thames and Hudson.
- Moseley, P., Fernyhough, C., and Ellison, A. (2013). Auditory verbal hallucinations as atypical inner speech monitoring and the potential of neurostimulation as a treatment option. *Neurosci. Biobehav. Rev.* 37, 2794–2805. doi: 10.1016/j.neubiorev.2013.10.001
- Mostofsky, S. H., Goldberg, M. C., Landa, R. J., and Denckla, M. B. (2000). Evidence for a deficit in procedural learning in children and adolescents with autism: implications for cerebellar contribution. *J. Int. Neuropsychol. Soc.* 6, 752–759. doi: 10.1017/s1355617700677020
- Murphy, E. (2015a). The brain dynamics of linguistic computation. *Front. Psychol.* 6:1515. doi: 10.3389/fpsyg.2015.01515
- Murphy, E. (2015b). Labels, cognomes and cyclic computation: an ethological perspective. *Front. Psychol.* 6:715. doi: 10.3389/fpsyg.2015.00715
- Nakamura, T., Matsumoto, J., Takamura, Y., Ishii, Y., Sasahara, M., Ono, T., et al. (2015). Relationships among parvalbumin-immunoreactive neuron density, phase-locked gamma oscillations and autistic/schizophrenic symptoms in PDGFR- β knock-out and control mice. *PLoS One* 10:e0119258. doi: 10.1371/journal.pone.0119258
- Naqvi, S., Cole, T. Y., Graham, J. M. Jr. (2000). Cole-Hughes macrocephaly syndrome and associated autistic manifestations. *Am. J. Med. Genet.* 94, 149–152. doi: 10.1002/1096-8628(20000911)94:2<149::aid-ajmg7>3.0.co;2-#
- Narita, H. (2014). *Endocentric Structuring of Projection-Free Syntax*. Amsterdam: John Benjamins.
- Nguyen, P. T., Nakamura, T., Hori, E., Urakawa, S., Uwano, T., Zhao, J., et al. (2011). Cognitive and socio-emotional deficits in platelet-derived growth factor receptor- β gene knockout mice. *PLoS One* 6:e18004. doi: 10.1371/journal.pone.0018004
- Norbury, C. F., Griffiths, H., and Nation, K. (2010). Sound before meaning: word learning in autistic disorders. *Neuropsychologia* 48, 4012–4019. doi: 10.1016/j.neuropsychologia.2010.10.015
- Norbury, C. F. (2005). The relationship between theory of mind and metaphor: evidence from children with language impairment and autistic spectrum disorder. *Br. J. Dev. Psychol.* 23, 383–399. doi: 10.1348/026151005x26732
- Oblak, L., Gibbs, T. T., and Blatt, G. J. (2010). Decreased GABA-B receptors in the cingulated cortex and fusiform gyrus in autism. *J. Neurochem.* 114, 1414–1423. doi: 10.1111/j.1471-4159.2010.06858.x
- Orekhova, E. V., Stroganova, T. A., Nygren, G., Tsetlin, M. M., Posikera, I. N., Gillberg, C., et al. (2007). Excess of high frequency electroencephalogram oscillations in boys with autism. *Biol. Psychiatry* 62, 1022–1029. doi: 10.1016/j.biopsych.2006.12.029
- Panganiban, G., and Rubenstein, J. L. (2002). Developmental functions of the *Distal-less/Dlx* homeobox genes. *Development* 129, 4371–4386.
- Paul, R., Fischer, M., and Cohen, D. (1988). Brief report: sentence comprehension strategies in children with autism and specific language disorders. *J. Autism Dev. Disord.* 18, 669–679. doi: 10.1007/bf02211884
- Pearl, P. L., Gibson, K. M., Acosta, M. T., Vezina, L. G., Theodore, W. H., Rogawski, M. A., et al. (2003). Clinical spectrum of succinic semialdehyde dehydrogenase deficiency. *Neurology* 60, 1413–1417. doi: 10.1212/01.wnl.0000059549.70717.80
- Peça, J., Feliciano, C., Ting, J. T., Wang, W., Wells, M. F., Venkatraman, T. N., et al. (2011). Shank3 mutant mice display autistic-like behaviours and striatal dysfunction. *Nature* 472, 437–442. doi: 10.1038/nature09965

- Perdomo-Sabogal, A., Kanton, S., Walter, M. B., and Nowick, K. (2014). The role of gene regulatory factors in the evolutionary history of humans. *Curr. Opin. Genet. Dev.* 29, 60–67. doi: 10.1016/j.gde.2014.08.007
- Perovic, A., and Janke, V. (2013). Issues in the acquisition of binding, control and raising in high-functioning children with autism. *UCL Work. Pap. Linguist.* 25, 131–143.
- Perovic, A., Modyanova, N., and Wexler, K. (2007). Knowledge of c-command A-movement in children and adolescents with autism and with Asperger syndrome. Paper presented at 2007 Generative Approaches to Language Acquisition (GALA), Barcelona, Spain.
- Petersen, S. E., and Sporns, O. (2015). Brain networks and cognitive architectures. *Neuron* 88, 207–219. doi: 10.1016/j.neuron.2015.09.027
- Pineda, J. A., Juavinett, A., and Datko, M. (2012). Self-regulation of brain oscillations as a treatment for aberrant brain connections in children with autism. *Med. Hypotheses* 79, 790–798. doi: 10.1016/j.mehy.2012.08.031
- Poduri, A., Evrony, G. D., Cai, X., and Walsh, C. A. (2013). Somatic mutation, genomic variation and neurological disease. *Science* 341:1237758. doi: 10.1126/science.1237758
- Poeppel, D. (2012). The maps problem and the mapping problem: two challenges for a cognitive neuroscience of speech and language. *Cogn. Neuropsychol.* 29, 34–55. doi: 10.1080/02643294.2012.710600
- Poeppel, D., and Hickok, G. (2015). Electromagnetic recording of the auditory system. *Handb. Clin. Neurol.* 129, 245–255. doi: 10.1016/B978-0-444-62630-1.00014-7
- Polleux, F., and Lauder, J. M. (2004). Toward a developmental neurobiology of autism. *Ment. Retard. Dev. Disabil. Res. Rev.* 10, 303–317. doi: 10.1002/mrdd.20044
- Porjesz, B., Almasy, L., Edenberg, H. J., Wang, K., Chorlian, D. B., Foroud, T., et al. (2002). Linkage disequilibrium between the beta frequency of the human EEG and a GABAA receptor gene locus. *Proc. Natl. Acad. Sci. U S A* 99, 3729–3733. doi: 10.1073/pnas.052716399
- Preissler, M. A. (2008). Associative learning of pictures and words by low-functioning children with autism. *Autism* 12, 231–248. doi: 10.1177/1362361307088753
- Pylyshyn, Z. (1985). *Computation and Cognition*. Cambridge, MA: MIT Press.
- Radalescu, E., Minati, L., Ganeshan, B., Harrison, N. A., Gray, M. A., Beacher, F. D., et al. (2013). Abnormalities in fronto-striatal connectivity within language networks relate to differences in grey-matter heterogeneity in Asperger syndrome. *Neuroimage Clin.* 27, 716–726. doi: 10.1016/j.nicl.2013.05.010
- Rapin, I., and Dunn, M. (2003). Update on the language disorders of individuals on the autistic spectrum. *Brain Dev.* 25, 166–172. doi: 10.1016/s0387-7604(02)00191-2
- Ring, H. A., Baron-Cohen, S., Wheelwright, S., Williams, S. C. R., Brammer, M., Andrew, C., et al. (1999). Cerebral correlates of preserved cognitive skills in autism. *Brain* 122, 1305–1315. doi: 10.1093/brain/122.7.1305
- Roberts, J. A., Rice, M. L., and Tager-Flusberg, H. (2004). Tense marking in children with autism. *Appl. Psycholinguist.* 25, 429–448. doi: 10.1017/s0142716404001201
- Rojas, D. C., Maharaj, K., Teale, P., and Rogers, S. J. (2008). Reduced neural synchronization of gamma-band MEG oscillations in first-degree relatives of children with autism. *BMC Psychiatry* 8:66. doi: 10.1186/1471-244X-8-66
- Roux, F., and Uhlhaas, P. J. (2014). Working memory and neural oscillations: alpha-gamma versus theta-gamma codes for distinct WM information? *Trends Cogn. Sci.* 18, 16–25. doi: 10.1016/j.tics.2013.10.010
- Sahin, M., and Sur, M. (2015). Genes, circuits and precision therapies for autism and related neurodevelopmental disorders. *Science* 350:aab3897. doi: 10.1126/science.aab3897
- Schipul, S. E., Keller, T. A., and Just, M. A. (2011). Inter-regional brain communication and its disturbance in autism. *Front. Syst. Neurosci.* 5:10. doi: 10.3389/fnsys.2011.00010
- Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., and Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends Cogn. Sci.* 12, 106–113. doi: 10.1016/j.tics.2008.01.002
- Schulman, J. J., Cancro, R., Lowe, S., Lu, F., Walton, K. D., and Llinás, R. R. (2011). Imaging of thalamocortical dysrhythmia in neuropsychiatry. *Front. Hum. Neurosci.* 5:69. doi: 10.3389/fnhum.2011.00069
- Scott-Van Zeeland, A. A., Abrahams, B. S., Alvarez-Retuerto, A. I., Sonnenblick, L. I., Rudie, J. D., Ghahremani, D., et al. (2010a). Altered functional connectivity in frontal lobe circuits is associated with variation in the autism risk gene. *Sci. Transl. Med.* 2:56ra80. doi: 10.1126/scitranslmed.3001344
- Scott-Van Zeeland, A. A., McNealy, K., Wang, A. T., Sigman, M., Bookheimer, S. Y., and Dapretto, M. (2010b). No neural evidence of statistical learning during exposure to artificial languages in children with autism spectrum disorders. *Biol. Psychiatry* 68, 345–351. doi: 10.1016/j.biopsych.2010.01.011
- Shao, Y., Cuccaro, M. L., Hauser, E. R., Raiford, K. L., Menold, M. M., Wolpert, C. M., et al. (2003). Fine mapping of autistic disorder to chromosome 15q11–q13 by use of phenotypic subtypes. *Am. J. Hum. Genet.* 72, 539–548. doi: 10.1086/367846
- Shao, Y., Wolpert, C. M., Raiford, K. L., Menold, M. M., Donnelly, S. L., Ravan, S. A., et al. (2002). Genomic screen and follow-up analysis for autistic disorder. *Am. J. Med. Genet.* 114, 99–105. doi: 10.1002/ajmg.10153
- Sohal, V. S., Zhang, F., Yizhar, O., and Deisseroth, K. (2009). Parvalbumin neurons and gamma rhythms enhance cortical circuit performance. *Nature* 459, 698–702. doi: 10.1038/nature07991
- Spencer, K. M., Nestor, P. G., Niznikiewicz, M. A., Salisbury, D. F., Shenton, M. E., and McCarley, R. W. (2003). Abnormal neural synchrony in schizophrenia. *J. Neurosci.* 23, 7407–7411.
- Stefanatos, G. A., and Baron, I. S. (2011). The ontogenesis of language impairment in autism: a neuropsychological perspective. *Neuropsychol. Rev.* 21, 252–270. doi: 10.1007/s11065-011-9178-6
- Stephanie, M., Pellizzari, G., Fletcher, C. R., and McClannahan, K. (2007). Empirical evaluation of language disorder in schizophrenia. *J. Psychiatry Neurosci.* 32, 250–258.
- Supekar, K., Mesin, M., and Menon, V. (2009). Development of large-scale functional brain networks in children. *PLoS Biol.* 7:e1000157. doi: 10.1371/journal.pbio.1000157
- Symond, M. P., Harris, A. W., Gordon, E., and Williams, L. M. (2005). “Gamma synchrony” in first-episode schizophrenia: a disorder of temporal connectivity? *Am. J. Psychiatry* 162, 459–465. doi: 10.1176/appi.ajp.162.3.459
- Tager-Flusberg, H. (2006). Defining language phenotypes in autism. *Clin. Neurosci. Res.* 6, 219–224. doi: 10.1016/j.cnr.2006.06.007
- Tager-Flusberg, H., and Calkins, S. (1990). Does imitation facilitate acquisition of grammar? Evidence from the study of autistic, Down's syndrome and normal children. *J. Child Lang.* 17, 591–606. doi: 10.1017/s0305000900009898
- Tager-Flusberg, H., Calkins, S., Nolin, T., Baumberger, T., Anderson, M., and Chadwick-Dias, A. (1990). A longitudinal study of language acquisition in autistic and Downs syndrome children. *J. Autism Dev. Disord.* 20, 1–21. doi: 10.1007/bf02206853
- Tager-Flusberg, H., and Cooper, J. (1999). Present and future possibilities for defining a phenotype for specific language impairment. *J. Speech Lang. Hear. Res.* 42, 1275–1278. doi: 10.1044/jslhr.4205.1275
- Tager-Flusberg, H., and Joseph, R. M. (2003). Identifying neurocognitive phenotypes in autism. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 303–314. doi: 10.1098/rstb.2002.1198
- Tager-Flusberg, H., Paul, R., and Lord, C. E. (2005). “Language and communication in autism,” in *Handbook of Autism and Pervasive Developmental Disorders: Vol. 1. Diagnosis, Development, Neurobiology and Behavior*, eds F. Volkmar, R. Paul, A. Klin and D. J. Cohen 3rd Edn. (New York: Wiley), 335–364.
- Tan, A. Y., Chen, Y., Scholl, B., Seidemann, E., and Priebe, N. J. (2014). Sensory stimulation shifts visual cortex from synchronous to asynchronous states. *Nature* 509, 226–229. doi: 10.1038/nature13159
- Tek, S., Jaffrey, G., Fein, D., and Naigles, L. R. (2008). Do children with autism spectrum disorders show a shape bias in word learning? *Autism Res.* 1, 208–222. doi: 10.1002/aur.38
- Tesink, C. M., Buitelaar, J. K., Petersson, K. M., van der Gaag, R. J., Teunisse, J. P., and Hagoort, P. (2011). Neural correlates of language comprehension in autism spectrum disorders: when language conflicts with world knowledge. *Neuropsychologia* 49, 1095–1104. doi: 10.1016/j.neuropsychologia.2011.01.018
- Theofanopoulou, C. (2015). Brain asymmetry in the white matter making and globularity. *Front. Psychol.* 6:1355. doi: 10.3389/fpsyg.2015.01355
- Theofanopoulou, C., and Boeckx, C. (forthcoming). “(Neural) syntax,” in *Language, Syntax and the Natural Sciences*, eds R. Martin and Á. Gallego (Cambridge: Cambridge University Press).

- Tierney, A. L., Gabard-Durnam, L., Vogel-Farley, V., Tager-Flusberg, H., and Nelson, C. A. (2012). Developmental trajectories of resting EEG power: an endophenotype of autism spectrum disorder. *PLoS One* 7:e39127. doi: 10.1371/journal.pone.0039127
- Tillmann, C., Wibral, M., Leweke, M., Kohler, A., Singer, W., Koethe, D., et al. (2008). Source localization of high-frequency oscillations reveals widespread reductions in gamma-band activity during perceptual organisation in chronic and first-episode schizophrenia. *Soc. Neurosci. Abstr.* 34:54.
- Uhlhaas, P. J., Linden, D. E. J., Singer, W., Haenschel, C., Lindner, M., Maurer, K., et al. (2006). Dysfunctional long-range coordination of neural activity during gestalt perception in schizophrenia. *J. Neurosci.* 26, 8168–8175. doi: 10.1523/JNEUROSCI.2002-06.2006
- Uhlhaas, P. J., Rouz, F., Rodriguez, E., Rotarska-Jagiela, A., and Singer, W. (2010). Neural synchrony and the development of cortical networks. *Trends Cogn. Sci.* 14, 72–80. doi: 10.1016/j.tics.2009.12.002
- Uhlhaas, P. J., and Singer, W. (2010). Abnormal neural oscillations and synchrony in schizophrenia. *Nat. Rev. Neurosci.* 11, 100–113. doi: 10.1038/nrn2774
- Vargas, D. L., Nascimbene, C., Krishnan, C., Zimmerman, A. W., and Pardo, C. A. (2005). Neuroglial activation and neuroinflammation in the brain of patients with autism. *Ann. Neurol.* 57, 67–81. doi: 10.1002/ana.20315
- Veenstra-VanderWeele, J., and Cook, E. H. (2004). Molecular genetics of autism spectrum disorder. *Mol. Psychiatry* 9, 819–832. doi: 10.1038/sj.mp.4001505
- Verly, M., Verhoeven, J., Zink, I., Mantini, D., Peeters, R., Deprez, S., et al. (2014). Altered functional connectivity of the language network in ASD: role of classical language areas and cerebellum. *Neuroimage Clin.* 4, 374–382. doi: 10.1016/j.nicl.2014.01.008
- Volden, J., and Lord, C. (1991). Neologisms and idiosyncratic language in autistic speakers. *J. Autism Dev. Disord.* 21, 109–130. doi: 10.1007/bf02284755
- Waiter, G. D., Williams, J. H. G., Murray, A. D., Gilchrist, A., Perrett, D. I., and Whiten, A. (2005). Structural white matter deficits in high-functioning individuals with autistic spectrum disorder: a voxel-based investigation. *Neuroimage* 24, 455–461. doi: 10.1016/j.neuroimage.2004.08.049
- Walenski, M., Tager-Flusberg, H., and Ullman, M. (2006). “Language in autism,” in *Understanding Autism: From Basic Neuroscience to Treatment*, eds S. O. Moldin and J. L. R. Rubenstein (Boca Raton: Taylor and Francis Books), 175–203.
- Welsh, J. P., Ahn, E. S., and Placantonakis, D. G. (2005). Is autism due to brain desynchronization? *Int. J. Dev. Neurosci.* 23, 253–263. doi: 10.1016/j.ijdevneu.2004.09.002
- Whitehouse, A., Maybery, M. T., and Durkin, K. (2006). Inner speech impairments in autism. *J. Child Psychol. Psychiatry* 47, 857–865. doi: 10.1111/j.1469-7610.2006.01624.x
- Willsey, A. J., and State, M. W. (2015). Autism spectrum disorders: from genes to neurobiology. *Curr. Opin. Neurobiol.* 30C, 92–99. doi: 10.1016/j.conb.2014.10.015
- Wilson, T. W., Rojas, D. C., Reite, M. L., Teale, P. D., and Rogers, S. J. (2007). Children and adolescents with autism exhibit reduced MEG steady-state gamma responses. *Biol. Psychiatry* 62, 192–197. doi: 10.1016/j.biopsych.2006.07.002
- Xu, T., Stephane, M., and Parhi, K. K. (2013). Multidimensional analysis of the abnormal neural oscillations associated with lexical processing in schizophrenia. *Clin. EEG Neurosci.* 44, 135–143. doi: 10.1177/1550059412465078
- Ye, A. X., Leung, R. C., Schäfer, C. B., Taylor, M. J., and Doesburg, S. M. (2014). Atypical resting synchrony in autism spectrum disorder. *Hum. Brain Mapp.* 35, 6049–6066. doi: 10.1002/hbm.22604
- Zikopoulos, B., and Barbas, H. (2013). Altered neural connectivity in excitatory and inhibitory cortical circuits in autism. *Front. Hum. Neurosci.* 7:609. doi: 10.3389/fnhum.2013.00609

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Benítez-Burraco and Murphy. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Bridging the Gap between Genes and Language Deficits in Schizophrenia: An Oscillopathic Approach

Elliot Murphy^{1*} and Antonio Benítez-Burraco²

¹ Division of Psychology and Language Sciences, University College London, London, UK, ² Department of Philology, University of Huelva, Huelva, Spain

Schizophrenia is characterized by marked language deficits, but it is not clear how these deficits arise from the alteration of genes related to the disease. The goal of this paper is to aid the bridging of the gap between genes and schizophrenia and, ultimately, give support to the view that the abnormal presentation of language in this condition is heavily rooted in the evolutionary processes that brought about modern language. To that end we will focus on how the schizophrenic brain processes language and, particularly, on its distinctive oscillatory profile during language processing. Additionally, we will show that candidate genes for schizophrenia are overrepresented among the set of genes that are believed to be important for the evolution of the human faculty of language. These genes crucially include (and are related to) genes involved in brain rhythmicity. We will claim that this translational effort and the links we uncover may help develop an understanding of language evolution, along with the etiology of schizophrenia, its clinical/linguistic profile, and its high prevalence among modern populations.

OPEN ACCESS

Edited by:

Anne Keitel,
University of Glasgow, UK

Reviewed by:

Carrie E. Bearden,
University of California, USA
Peter Uhlhaas,
University of Glasgow, UK

*Correspondence:

Elliot Murphy
elliottmurphy91@gmail.com

Received: 16 January 2016

Accepted: 08 August 2016

Published: 23 August 2016

Citation:

Murphy E and Benítez-Burraco A (2016) Bridging the Gap between Genes and Language Deficits in Schizophrenia: An Oscillopathic Approach.
Front. Hum. Neurosci. 10:422.
doi: 10.3389/fnhum.2016.00422

INTRODUCTION

Schizophrenia is a pervasive neurodevelopmental disorder entailing several (and severe) social and cognitive deficits (van Os and Kapur, 2009). Usually, people with schizophrenia exhibit language problems at all levels, from phonology to pragmatics, which coalesce into problems for speech perception (auditory verbal hallucinations), abnormal speech production (formal thought disorder), and production of abnormal linguistic content (delusions, commonly understood to be distinct from thought disorders), which are the hallmarks of the disease in the domain of language (Stephane et al., 2007, 2014; Bakhshi and Chance, 2015). Importantly, although schizophrenia is commonly defined as a disturbance of thought or selfhood, some authors claim that most of its distinctive symptoms may arise from language dysfunction; in particular, from failures in language-mediated forms of meaning (Hinzen and Rosselló, 2015).

There is ample evidence that schizophrenia is caused by a complex interaction between genetic, epigenetic, and environmental factors. To date, schizophrenia has been related to mutations, copy number variation, or changes in the expression pattern of an extensive number of genes (see O'Tuathaigh et al., 2012; Flint and Munafò, 2014; McCarthy et al., 2014; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014 for recent reviews). Many of them point to specific regulatory and signaling pathways (like dopaminergic, glutamatergic, GABAergic, and cholinergic pathways, the neuregulin signaling pathway, and the Akt/GSK-3 pathway) and to specific neural mechanisms (like those involving dendritic spines and synaptic terminals,

synapses, gray matter development, and neural plasticity, Buonanno, 2010; Karam et al., 2010; Bennett, 2011; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014; Hall et al., 2015). However, the gap between genes, brain abnormalities, and cognitive dysfunction in schizophrenia still remains open, particularly regarding its distinctive linguistic profile.

The goal of this Perspective article is to suggest new ways of bridging the gap between genes and schizophrenia. Cognitive disorders are increasingly being conceived as oscillopathies, or pathological variations of the normal profile of brain rhythmicity (Buzsáki and Watson, 2012; Buzsáki et al., 2013). Current understanding suggests that schizophrenia is characterized by asynchronous neural oscillations, and particularly, by an inhibitory interneuron dysfunction (Moran and Hong, 2011; Pittman-Polletta et al., 2015). Importantly, brain rhythms are heritable components of brain function (Linkenkaer-Hansen et al., 2007), also in pathological conditions (see Hall et al., 2011 for schizophrenia). At the same time, there seems to be a robust link between language disorders and language evolution: recently evolved cognitive abilities are preferably disturbed in disorders because of the reduced resilience of the neural networks (see Benítez-Burraco, 2016 for discussion). The human pattern of brain activity can be conceived of as a slight variation of the patterns observed in other primates (Buzsáki et al., 2013). Accordingly, our species-specific ability to acquire and use languages (aka *language-readiness*) plausibly resulted from the emergence of a new pattern of cortical inhibition and of long-distance connections across the brain (see Boeckx and Benítez-Burraco, 2014a for details), both of which are aspects that are targeted in schizophrenia (Morice and McNicol, 1985; Horn et al., 2012; Jiang et al., 2015). If we are on the right track, we expect that examining language deficits in schizophrenia from this oscillopathic and evolutionary perspective will help us understand its distinctive neurocognitive profile, but also its origins and its prevalence among modern populations.

FROM LANGUAGE DEFICITS TO THE BRAIN IN SCHIZOPHRENIA

Schizophrenics have been known to have disordered speech (McKenna and Oh, 2005), but the most severe linguistic changes occur at the internal, conceptual level, where studies frequently examine patients who experience thoughts being “inserted” into them from outside sources or “broadcast” out of their minds and into other people’s (Crow, 1980; Frith, 1992). Patients also sometimes hear their thoughts “echoed,” or spoken aloud, and are also known to experience third-person and second-person auditory hallucinations, with an external voice either discussing them or commenting on their actions (Ramsden, 2013, pp. 234–265). Frith and Allen’s (1988) review observed “a failure to structure discourse at higher levels.” Abnormalities can also be detected with syntax, however, and this is where we will focus most of our attention. Schizophrenic patients exhibit fewer relative clauses (as their discourse difficulties would predict), shorter utterances, and less clausal embedding (Fraser et al., 1986;

Thomas et al., 1987). Importantly, this relative lack of clausal embedding implies that patients do not engage in thoughts about mental states or Theory of Mind (Morice and Ingram, 1982; Morice and McNicol, 1986).

In contrast to normal left-lateralization of activity in fronto-temporal regions during language processing, a wide range of schizophrenic patients exhibit bilateral and right-lateralized activity (Weiss et al., 2005; Diederich et al., 2010). Angrilli et al. (2009) have relatedly proposed that, judging by evoked potentials, certain features of schizophrenia appear to be (partly) a failure of phonological left hemispheric dominance, since the above deficit in lateralization is specific to phonological processing, being absent in semantic and word recognition tasks.

FROM BRAIN RHYTHMICITY TO LANGUAGE DEFICITS IN SCHIZOPHRENIA

Although schizophrenia was for a time deemed “the graveyard of neuropathology” (Plum, 1972) due to its unusually subtle neurophysiological markers, we believe that research in neuronal dynamics (particularly over the past half-decade) has the potential to carve a clearer image of the abnormally-developing brain. Oscillations play a central role in selectively enhancing neural assembly interconnectivity and information processing through the provision of spatio-temporal windows of enhanced or reduced patterns of excitability (Jensen et al., 2014; Weisz et al., 2014), and are consequently strong candidates for the origin of certain cognitive faculties.

If the translational approach taken in Murphy (2015a, 2016) toward the brain dynamics of language is accurate, and if Hinzen and Rosselló (2015) are correct in claiming that linguistic disorganization in schizophrenia “plays a more central role in the pathogenesis of this disease than commonly supposed,” then it is appropriate to inform our understanding of schizophrenia by focusing on the central role of brain rhythms in linguistic computation. If schizophrenia represents a breakdown in normal linguistic cognition, then we would expect to see disruptions in the model of brain dynamics of language processing outlined in Murphy (2015a) when examining the recent, burgeoning literature concerning the oscillatory profile of schizophrenics.

To briefly summarize previous work, it was claimed in Murphy (2015a,b) that set-formation amounts to the α rhythm embedding cross-cortical γ rhythms, with α reflecting long-range cortical interactions (Nunez et al., 2001) and thalamocortical loop activity (Nunez and Srinivasan, 2006). The syntactic operation of “Transfer” (which “chunks” constructed objects into short-term memory) was claimed to amount to the embedding of these γ rhythms inside the θ band, generated in the hippocampus. It was also claimed that labeling (maintaining an item in memory before coupling it with another, yielding an independent syntactic identity) amounts to the slowing down of γ to β before β - α coupling, likely involving a basal ganglia-thalamic-cortical loop. These suggestions are in line with Mai et al.’s (2016) finding of γ -related modulations during semantic and syntactic processing (our claims should also not be conflated with the well-known *phonological* oscillatory investigations of Giraud and Poeppel

(2012). We will adopt these assumptions here when interpreting the rhythmic literature on schizophrenia.

Since schizophrenia, like other cognitive impairments, appears not to be the result of a locally delimited neural deficit but rather emerges from distributed impairments, neural oscillations, and their role in flexible brain connectivity have recently become the target of research. Investigating the frequency and brain location of the neural oscillations involved in lexical processing in schizophrenia, Xu et al. (2013) conducted an MEG study in which patients discriminated correct from incorrect visually presented stimuli. This lexical decision task revealed that the patients, relative to healthy controls, showed abnormal oscillatory activity during periods of lexical encoding and post-encoding, particularly in the occipital and left frontal-temporal areas (see also Sun et al., 2014). Since a broad range of rhythms were implicated, we will avoid speculation about the specific operations impaired and instead suggest that the results imply familiar problems with semantic memory. However, the results did reveal reduced temporal lobe α and left frontal lobe β activity during lexical processing, suggesting difficulties in assigning lexical classes (labels) to items and successful categorization (findings corroborating the cartographic profile presented above, which included reduced activation during complex sentence processing at left superior frontal cortex). These results corroborate the more general findings of reduced α and β in schizophrenia by Moran and Hong (2011) and Uhlhaas et al. (2008). A level of thalamocortical dysrhythmia was also detected by Schulman et al.'s (2011). MEG study; a discovery which bears on the claim that thalamocortical axons also likely play a role in language externalization (Boeckx and Benítez-Burraco, 2014b). These suggested problems with the mechanisms responsible for phrase structure building also gain support from Ghorashi and Spencer's (2015) findings that attentional load increases β phase-locking factor at frontal, parietal and occipital sites in healthy controls during a visual oddball task but not in schizophrenic patients (although this varied across individuals of different abilities), with the latter group having difficulty attending to and maintaining relevant objects in memory (perhaps as a result of their semantic memory deficits). β -generating circuits may well be responsible, then, for the types of computations attributed to them in Murphy (2015a).

An earlier MEG sentence presentation task by Xu et al. (2012) also found reduced α and β in left temporal-parietal sites, along with reduced δ at left parietal-occipital and right temporal sites, and reduced θ at occipital and right frontal lobe sites, suggesting problems with phrase structure chunking; that is, problems with word movement and phrasal embedding, as attested above (see Ferrarelli et al., 2012). Schizophrenic patients also displayed reduced δ synchrony at left frontal lobe sites after sentence presentation, suggesting semantic processing dysfunctions. These findings are consistent with Hirayasu et al.'s (1998). MRI study of schizophrenic and bipolar individuals, which reported relatively reduced gray matter volumes in the left superior temporal gyrus for schizophrenics. Their results also give some support to the present hypothesis about chunking difficulties in schizophrenia, since they also reported reduced hippocampal volumes. Altogether, these studies are in agreement

the findings of Hoffman et al. (1999), who suggested that the core schizophrenic deficit is not centered on attentional-perceptual cognitive processes, but rather verbal working memory (and, hence, difficulties with syntactic computation, given the “chunking” nature of linguistic phrase structure building; see Narita, 2014), mediated by oscillations generated in the hippocampus and left temporal regions (Murphy, 2015a). Başar-Eroğlu et al. (2011) also documented reduced anterior α in response to simple auditory input, suggesting less efficient processing power.

Power and synchrony reductions in evoked γ have also been documented in chronic, first-episode and early-onset schizophrenia (Williams and Boksa, 2010). Given the role of this band in feature binding and object representation (Uhlhaas et al., 2008) and its functional significance in the present model (Murphy, 2015a, 2016), this suggests that schizophrenics have difficulties generating the correct category of semantic objects to employ in successful phrase structure building, as the behavioral results of lexical decision and related tasks appear to verify (likely explaining the features of delusions and formal thought disorder reviewed above). More recent studies appear to support this perspective. The amplitude of EEG γ was measured during phonological, semantic, and visuo-perceptual tasks by Spironelli and Angrilli (2015). Schizophrenic patients, relative to normal controls, exhibited a significantly weaker hemispheric asymmetry across all tasks and reduced frontal γ . Ferrarelli et al. (2008) also found a decreased γ response in schizophrenic patients after TMS stimulation to the frontal cortex, suggesting an impaired ability to efficiently generate this rhythm. This is of particular significance given that γ amplitude has been shown to scale with the number of items held in working memory (Roux et al., 2012), and the limited phrase structure building and syntactic embedding capacities of schizophrenic patients would follow naturally from these results.

Recall also that the model of linguistic computation adopted here invokes a number of cross-frequency coupling operations. It is of interest, then, that schizophrenic patients showed higher γ - α cross-frequency coupling in Popov and Popova's (2015) study of general cognitive performance, despite this co-varying with poorer attention and working memory capacities. The reason for this may be that the increased phase-amplitude-locking likely results in smaller “gamma pockets” of working memory items (as Korotkova et al., 2010 argue on independent grounds) and hence low total γ power. In this instance, the size and order of working memory sequences outputted by the conceptual systems is not optimally compatible with the oscillatory profile, leading to greater rhythmic excitability, and yet inhibited linguistic functionality. Global rhythmicity is consequently disrupted due to unusually strong fronto-parietal interconnectivity. We believe that this represents a genuine neural mechanism of an “interface” between syntactically generated conceptual representations and external (memory) systems; a highly significant finding if corroborated by further experimental studies.

Corroborating Angrilli et al.'s (2009) above hypothesis about schizophrenia being a failure of left-hemispheric phonological dominance, an MEG study of the oscillatory differences between bipolar disorder and schizophrenia revealed that schizophrenic

patients showed delayed phase-locking in response to speech sounds in the left hemisphere, relative to bipolar individuals and normal controls (Oribe et al., 2010). This lack of left-hemispheric dominance may trigger confusion about internal and external voices and bring about a number of delusions, with language's normal computational functioning being derailed. The left hypofrontality documented by Spironelli et al. (2011), with schizophrenic patients showing greater δ amplitudes over language-relevant sites (that is, greater functional inhibition), similarly point to a general functional deficit at the core memory sites of linguistic representations. It is also significant that the role attributed to θ in the present dynamic model gains support from the finding that this rhythm has greater amplitude in left superior temporal cortex during auditory hallucinations in schizophrenia (Ishii et al., 2000), as opposed to steady θ during resting state, with patients being seemingly incapable of regulating chunking operations. Given the identification of such dysrhythmias in schizophrenia, repetitive TMS (rTMS) could be used as a therapeutic intervention to modulate the oscillations responsible for the abnormal linguistic profile documented above, as has been done to improve performance on visual tasks (Farzan et al., 2012; Barr et al., 2013). The oscillopathic profile constructed here is presented in **Table 1**.

SCHIZOPHRENIA-RELATED GENES AND SOME EVOLUTIONARY CONCERN

As noted in the introduction, the number of genes related to schizophrenia has been growing over recent years. Interestingly, some of them are involved in the maintenance of the adequate balance between neuronal excitation and inhibition and/or have been related to language dysfunction. Likewise, as we also noted above and will discuss in detail below, a robust link exists between evolution and abnormal development and, in particular, between language evolution and schizophrenia. In this section we focus on candidate genes for schizophrenia that are involved in brain rhythmicity and that have been related to language impairment or to the dysfunction of basic cognitive abilities involved in language processing, but also on genes important for language evolution that play a role in brain rhythmicity and that are candidates for schizophrenia. The genes we highlight seem to us robust candidates for language deficits in this condition.

Schizophrenia-Candidates and Brain Rhythmicity

Among the genes related to schizophrenia that play a role in brain oscillations and that have been associated to language dysfunction one finds *ZNF804A*. This gene encodes a zinc finger binding protein important for cortical functioning and neural connectivity, involved in growth cone function and neurite elongation (Hinna et al., 2015). GWAs analyses have identified a SNP tagging an intronic region of the gene (Gurung and Prata, 2015) which have been found to impact on white matter microstructure (Mallas et al., 2016). Schizophrenia risk polymorphisms of *ZNF804A* have been also related to differences in performance in the domain of phonology, such as in reading

and spelling tasks (Becker et al., 2012), but also in the domain of semantics, specifically in task evaluating category fluency (Nicodemus et al., 2014). *ZNF804A* modulates hippocampal γ oscillations and, ultimately, the co-ordination of distributed networks belonging to the hippocampus and the prefrontal cortex (Cousijn et al., 2015), which are aspects known to be impaired in schizophrenia, as noted above (Uhlhaas et al., 2008; Godsil et al., 2013). Likewise, both *NRG1* and its receptor *ERBB4*, which have been posited as promising candidates for schizophrenia as resulting from next-generation sequencing analyses (Agim et al., 2013; Hatzimanolis et al., 2013), enhance synchronized oscillations of neurons in the prefrontal cortex, known to be reduced in schizophrenia, via inhibitory synapses (Fisahn et al., 2009; Hou et al., 2014). Specifically *NRG1* increases the synchrony of pyramidal neurons via presynaptic interneurons and the synchrony between pairs of interneurons through their mutually-inhibitory synapses (Hou et al., 2014). Risk polymorphisms of *NRG1* are associated with increased IQs as well as memory and learning performance, along with language in subjects with bipolar disorder (Rilstad et al., 2015). Moreover, risk alleles for the gene correlate with reduced left superior temporal gyrus volumes (a robust imaging finding in schizophrenia, Tosato et al., 2012), a region related to language abilities (Aeby et al., 2013). Another gene of interest is *PDGFR*, which encodes the β subunit of the platelet-derived growth factor (PDGF) receptor, known to be involved in the development of the central nervous system. *Pdgfr- β* knocked-out mice show reduced auditory phase-locked γ oscillations, which correlates with anatomical (e.g., reduced density of GABAergic neurons in the amygdala, hippocampus, and medial prefrontal cortex), physiological (alterations of prepulse inhibition) and behavioral (reduced social behavior, impaired spatial memory and problems with conditioning) hallmarks of schizophrenia (Nguyen et al., 2011; Nakamura et al., 2015). Additional evidence of the involvement of this gene in schizophrenia comes from risk polymorphisms analyses (Kim et al., 2008). Interestingly, *PDGFRA* has been found to act downstream of *FOXP2*, the renowned "language gene," to promote neuronal differentiation (Chiu et al., 2014, more on *FOXP2* below).

Other genes of interest encode ion channels. Genome-wide analyses (GWAs) have identified the schizophrenia risk gene *CACNA1I* as one of the genes that may contribute to sleep spindle deficits (Manoach et al., 2015). Sleep spindles are a type of brain rhythm that recurs during non-rapid eye movement sleep and that constrains aspects of the thalamocortical crosstalk, impacting on sensory transmission, cortical plasticity, memory consolidation, and learning (Manoach et al., 2015). *CACNA1I* encodes a calcium channel and is abundantly expressed in the spindle generator of the thalamus. Likewise *CACNA1C* encodes the alpha 1C ($\alpha 1C$) subunit of the Cav1.2 voltage-dependent L-type calcium channel, a calcium channel involved in the generation of β to γ waves during wakefulness and rapid eye movement (REM) sleep, and ultimately in sleep modulation; all of which are aspects known to be altered in schizophrenics (Kumar et al., 2015). Intriguingly, *CACNA1C* is related to semantic (but not lexical) verbal fluency in healthy individuals; conversely, risk alleles of this gene correlate with

TABLE 1 | Summary of the present cognome-dynome model of linguistic computation and the observed differences in schizophrenia, where “cognomen” refers to the operations available to the human nervous system (Poeppel, 2012) and “dynome” refers to brain dynamics (Kopell et al., 2014); ISTG denotes left superior temporal gyrus, AVH denotes auditory verbal hallucination.

Frequency band	Role in the present model of language computation	Observed and predicted differences in schizophrenia
Delta (~0.5–4 Hz)	Involved in phrasal processing and possibly labeling.	Reduced at left parietal-occipital sites during sentence processing; predicted to be disrupted in processing phrasal embedding and relative clauses.
Theta (~4–10 Hz)	Hippocampal source; embeds γ to generate cyclic transfer of syntactic objects; involved more generally in memory retrieval.	Reduced at occipital and frontal lobe sites during sentence processing; increased at ISTG during AVHs; predicted to be reduced in deictic and definite NPs.
Alpha (~8–12 Hz)	Synchronizes distant cortical regions; embeds γ generated cross-cortically to yield inter-modular set-formation; involved in lexical decision making.	Reduced at left temporal lobe during lexical and sentence processing; predicted to be disrupted during certain lexicalisations.
Beta (~10–30 Hz)	When γ is slowed to β and coupled with α via a basal ganglia-thalamic-cortical loop, syntactic objects are labeled; holds objects in memory.	Reduced at left frontal lobe during lexical processing; predicted to be disrupted in the maintenance of syntactic objects in embedded clauses.
Gamma (~30–100 Hz)	Generates syntactic objects before β holds them in memory; central role in a number of linguistic operations; involved in lexical processing.	Reduced at frontal sites during semantic tasks; higher cross-frequency coupling with occipital α ; predicted to be disrupted in language-related memory tasks.

lower performance scores, and thus with non-fluent verbal performance of schizophrenics (Krug et al., 2010). Two proteins associated with ion channels are also worth considering, namely DPP10 and CNTNAP2. DPP10 is a membrane protein that binds specific K⁺ channels and modifies their expression and biophysical properties (Djurovic et al., 2010). Also CNTNAP2 is associated with K⁺ voltage-gated channels, particularly, in the axon initial segment of pyramidal cells in the temporal cortex, that are mostly innervated by GABAergic interneurons (Inda et al., 2006). Several studies have correlated CNTNAP2 with schizophrenia, including CNV and SNPs studies (Friedman et al., 2008; Ji et al., 2013). The gene is also a candidate for several types of language disorders, including child apraxia of speech (Worthey et al., 2013), dyslexia (Peter et al., 2011), SLI (Newbury et al., 2011), language delay, and language impairment (Petrin et al., 2010; Sehested et al., 2010). CNTNAP2 additionally affects language development in the normal population (Whalley et al., 2011; Whitehouse et al., 2011; Kos et al., 2012), apparently because of its effects on brain connectivity and cerebral morphology (Scott-Van Zeeland et al., 2010; Tan et al., 2010; Dennis et al., 2011) and dendritic arborization and spine development (Anderson et al., 2012). CNTNAP2 is also a target of FOXP2 (Vernes et al., 2008).

Several genes encoding neurotransmitter receptors have been also related to both abnormal brain oscillation patterns and language deficits in schizophrenia. *HTR1A* encodes the receptor 1A of serotonin and modulates hippocampal γ oscillations, seemingly impacting on behavioral and cognitive functions, such as learning and memory linked to serotonin function (Johnston et al., 2014). Several studies involving common polymorphisms of this gene highlight *HTR1A* as a promising candidate for schizophrenia risk, treatment response to the disease, and cognitive dysfunction in this condition (Gu et al., 2013; Lin et al., 2015; Takekita et al., 2015). Similarly, receptors of NMDA, particularly those containing the subunit NR2A, encoded by *GRIN2A*, are known to be reduced in fast-firing

interneurons in schizophrenics, which plays a critical role in γ oscillation formation; a blockade of NR2A-containing receptors gives rise to strong increases in γ power and a reduction in low-frequency γ modulation (Kocsis, 2012). More generally, functional (GT)n polymorphisms in the promoter of the gene have been associated with the disease (Iwayama-Shigeno et al., 2005; Tang et al., 2006; Liu et al., 2015), and genome-wide association analyses has identified *GRIN2A* as a risk factor for schizophrenia (Lencz and Malhotra, 2015), emerging as a promising candidate because of its expression in the adult neocortex (Ohi et al., 2016). Additionally, mutations in *GRIN2A* cause epilepsy-aphasia spectrum disorders, including Landau-Kleffner syndrome and continuous spike and waves during slow-wave sleep syndrome (CSWSS), in which speech impairment and language regression are prominent symptoms (Carvill et al., 2013; Lesca et al., 2013). The gene has been related as well to rolandic epilepsies, the most frequent epilepsies in childhood, in which cognitive, speech, language, and reading problems are commonly observed (Dimassi et al., 2014). Speech problems linked to *GRIN2A* mutations include imprecise articulation, impaired pitch and prosody, and hypernasality, as well as poor performance on maximum vowel duration and repetition of monosyllables and trisyllables, resulting in lifelong dysarthria and dyspraxia (Turner et al., 2015). Finally, cannabinoid-1 receptor, encoded by *CNR1*, modulates θ and γ oscillations in several areas of the brain, including the hippocampus, impacting on sensory gating function in the limbic circuitry (Hajós et al., 2008). *CNR1*-positive GABA-ergic interneurons have been also involved in several aspects of behavior, including response to auditory cues (Brown et al., 2014). Translational convergent functional genomics studies have highlighted *CNR1* as an important gene for schizophrenia onset (Ayalew et al., 2012). Several risk polymorphisms of the gene have been related to the disease, and specifically, to brain changes and metabolic disturbances in schizophrenics (Yu et al., 2013; Suárez-Pinilla et al., 2015). Interestingly, *CNR1* has also been linked to cases

of complete absence of expressive speech (Poot et al., 2009). *CNR1* is functionally linked to the last gene we wish to highlight, namely, *DISC1* (Xie et al., 2015). *DISC1* encodes a protein involved in neurite outgrowth, cortical development and callosal formation (Brandon and Sawa, 2011; Osbun et al., 2011). *DISC1* is a historical candidate for schizophrenia (but also to other cognitive disorders like ASD), although its status as a candidate is controversial, provided that most GWAs and CNV analyses have been unable to independently implicate it in the disease (see Farrell et al., 2015 for discussion; see Ayalew et al., 2012 for a promising result). Nonetheless, in hippocampal area CA1 of a transgenic mouse that expresses a truncated version of *Disc1* mimicking the schizophrenic genotype, θ burst-induced long-term potentiation (and ultimately, long-term synaptic plasticity) has been found altered (Booth et al., 2014). The ability of *DISC1* to regulate excitatory-inhibitory synapse formation by cortical interneurons depends on its inhibitory effect on NRG1-induced ERBB4 activation and signaling, ultimately effecting the spiking interneuron-pyramidal neuron circuit (Seshadri et al., 2015). *DISC1* is also a target of *FOXP2* (Walker et al., 2012).

Schizophrenia-Candidates and Language Evolution

As pointed out above, there exists a robust link between evolution and abnormal development. Because, as noted in the introduction, brain rhythms are heritable components of brain function, and because patterns of brain rhythmicity are species-specific and disorder-specific, we hypothesized that new candidates for language dysfunction in schizophrenia under our oscillopathic view may emerge from the examination of candidate genes for the evolution of language-readiness in our species. As we also pointed out in the introduction, our distinctive ability for acquiring and using language has been hypothesized to have resulted from the emergence of new patterns of cortical rhythmic coupling that habilitated the neuronal workspace needed for transcending the boundaries of core knowledge systems and being able to form cross-modular concepts (known to be affected in schizophrenia); in turn, these changes may have resulted from the brain changes linked to the globularization of the anatomically-modern human (AMH) skull (see Boeckx and Benítez-Burraco, 2014a for details). In a series of related papers, we have put forth a list of tentative candidates for globularization and language-readiness (Boeckx and Benítez-Burraco, 2014a,b; Benítez-Burraco and Boeckx, 2015; see Table 2). As discussed there, core candidates for globularization and language readiness fulfill the following criteria: they show (or are functionally related to genes showing) differences with extinct hominin species, particularly, with Neanderthals/Denisovans, which affect their regulatory regions, their coding regions, and/or their methylation patterns; they play some role in brain growth, regionalization, and/or neural interconnection; they have been associated (or are functionally related to genes associated) to conditions in which language, or cognitive abilities important for language, are impaired; and they are candidates (or are functionally related to candidates) for craniosynostosis or some other conditions affecting skull development. Our list

of candidates encompasses genes involved in bone development, brain development (specifically of GABAergic neurons), and more generally, brain-skull cross-talk, like *RUNX2*, some *DLX* genes (including *DLX1*, *DLX2*, *DLX5*, and *DLX6*), and some *BMP* genes (like *BMP2* and *BMP7*). It also includes genes that regulate subcortical-cortical axon pathfinding and that are involved in the externalization of language (such as *FOXP2*, *ROBO1*, and the genes encoding the SLITs factors). Finally, it also comprises genes connecting the former two interactomes, including *AUTS2* and some of its partners. We have found ample evidence, *in silico* and in the available literature, supporting the biological reliability of these interactomes. Moreover, we have collected some empirical evidence suggesting that many of the genes we regard important for language evolution are dysregulated in clinical conditions involving skull, brain, and cognitive anomalies. Accordingly, differential expression of several of our candidates (*DLX5*, *ROBO1*, *SLIT2*, *NCAM1*, *TGFB2*, *DCN*, *RUNX2*, and *SFRP2*) was found *in vivo* in the sutures of people with non-syndromic craniosynostosis, which are prematurely ossified, and also *in vitro* in cells induced toward osteogenic differentiation (Lattanzi et al., 2016).

Interestingly, we have found that candidates for schizophrenia are overrepresented among the genes highlighted by Benítez-Burraco and Boeckx (Table 2). Accordingly, nearly 5% of the human genes are expected to be related to the disease [assuming that the human genome contains about 20,000 protein-coding genes and that about 1000 of them have been associated to schizophrenia, according to the Schizophrenia Gene repository (<http://www.szgene.org/>)]. In turn, around 30% of candidates for language readiness are also candidates for schizophrenia (42 out of 153 in Table 2). Because the involvement of these genes in language development and evolution, this overlapping may account for the observed deficits in schizophrenia regarding language abilities. These genes are discussed in detail in the Supplementary Materials to this paper. Moreover, several of these common candidates for language-readiness and schizophrenia also play a role in brain rhythmicity, including *AKT1*, *APOE*, *DLX5*, *DLX6*, *EGR1*, *FMR1*, *GAD1*, *MAPK14*, *MECP2*, and *SIRT1* (Table 2). These genes attracted our attention as promising new candidates for the oscillopathic nature of language deficits in schizophrenia. Finally, some of the candidates for the schizophrenia dynome interact with some of the genes encompassing these interactomes important for our language-readiness (Figure S2). In our opinion, all these findings reinforce the view that language impairment in schizophrenia results from (and can be confidently construed in terms of) abnormal patterns of brain connectivity and dynamics.

This overrepresentation of candidates for schizophrenia among the genes involved in language evolution is an intriguing finding. It has been hypothesized that schizophrenia candidate genes were involved in the evolution of the human brain and that the processes they contributed to improving are identical to those impaired in schizophrenics. For example, the human prefrontal cortex, which is responsible for many human-specific cognitive abilities, is differently organized in humans compared to great apes as a result of a recent reorganization of the frontal

TABLE 2 | Genes discussed in Section Schizophrenia-related genes and some evolutionary concerns.

Gene	LR	BR	SZ	Selected references	Gene	LR	BR	SZ	Selected references
<i>ABL1</i>	+				<i>KDM5B</i>	+			
<i>AKT1</i>	+	+	++	Emamian et al., 2004; Wockner et al., 2014	<i>LHX2</i>	+			
<i>ANAPC10</i>	+				<i>MAPK1</i>	+	+		
<i>APOE</i>	+	+	+	Verbrugghe et al., 2012	<i>MAPK14</i>	+	+	++	Onwuameze et al., 2013; Chang et al., 2015
<i>ARHGAP32</i>	+				<i>MECP2</i>	+	+	++	Cohen et al., 2002; McCarthy et al., 2014
<i>ARHGEF6</i>	+				<i>MEF2A</i>	+		++	Thygesen et al., 2015
<i>ARX</i>	+	+			<i>MET</i>	+		++	Burdick et al., 2010
<i>ASCL1</i>	+				<i>NCAM1</i>	+		++	Vawter et al., 2001; Atz et al., 2007; Ayalew et al., 2012
<i>ASPM</i>	+				<i>NCOA6</i>	+			
<i>AUTS2</i>	+		+	Zhang et al., 2014	<i>NFASC</i>	+			
<i>BAZ2A</i>	+				<i>NKX2-1</i>	+			
<i>BGLAP</i>	+				<i>NODAL</i>	+			
<i>BMP2</i>	+				<i>NOTCH1</i>	+			
<i>BMP7</i>	+				<i>NOVA1</i>	+			
<i>CACNA1C</i>		+	++	Krug et al., 2010	<i>NR1H2</i>	+			
<i>CACNA1I</i>		+	++	Manoach et al., 2015	<i>NRG1</i>	+	+	++	Ayalew et al., 2012; Agim et al., 2013; Hatzimanolis et al., 2013
<i>CBL</i>	+	+			<i>NRG3</i>	+		+	Kao et al., 2010; Hatzimanolis et al., 2013; Zeledón et al., 2015
<i>CDC42</i>	+		++	Gilks et al., 2012; Datta et al., 2015	<i>NTN1</i>	+			
<i>CDC42BPB</i>	+		+	Narayan et al., 2008	<i>OTX2</i>	+			
<i>CDC42EP4</i>	+		+	Datta et al., 2015	<i>PAK5</i>	+			
<i>CDH1</i>	+				<i>PAK6</i>	+			
<i>CDKN1A</i>	+				<i>PARP1</i>	+			
<i>CEBPB</i>	+				<i>PAX3</i>	+			
<i>CEP192</i>	+				<i>PAX6</i>	+			
<i>CITED2</i>	+				<i>PCDH11</i>	+			
<i>CKAP5</i>	+				<i>PCM1</i>	+		+	Kamiya et al., 2008
<i>CLOCK</i>	+		+	Zhang et al., 2011; Jung et al., 2014	<i>PCNT</i>	+			
<i>CMIP</i>	+				<i>PDGFR</i>		+	++	Kim et al., 2008
<i>CNR1</i>		+	++	Ayalew et al., 2012; Onwuameze et al., 2013	<i>PDX1</i>	+			
<i>CNTNAP2</i>	+	+	++	Friedman et al., 2008; Ji et al., 2013	<i>PIN1</i>	+			
<i>CREBBP</i>	+				<i>PITPNA</i>	+			
<i>CTIP2</i>	+				<i>PLAUR</i>	+	+		
<i>CTNNB1</i>	+		+	Levchenko et al., 2015	<i>POU3F2</i>	+		++	Huang et al., 2005; Potkin et al., 2009
<i>DCC</i>	+		+	Grant et al., 2007, 2012	<i>PQBP1</i>	+			
<i>DIP2A</i>	+				<i>PTEN</i>	+	+		
<i>DISC1</i>	+	+	++	Johnstone et al., 2011; Ayalew et al., 2012	<i>PTPRB</i>	+			
<i>DISP1</i>	+				<i>PVALB</i>	+			
<i>DLL1</i>	+				<i>RELN</i>	+		++	Shifman et al., 2008; Ayalew et al., 2012; Li et al., 2015
<i>DLX1</i>	+		+	Kromkamp et al., 2003	<i>ROBO1</i>	+		++	Potkin et al., 2009, 2010
<i>DLX2</i>	+				<i>ROBO2</i>	+		++	Potkin et al., 2009, 2010
<i>DLX5</i>	+	+	+	Cho et al., 2015	<i>RUNX1</i>	+			
<i>DLX6</i>	+	+	+	Cho et al., 2015	<i>RUNX2</i>	+		+	Benes et al., 2007

(Continued)

TABLE 2 | Continued

Gene	LR	BR	SZ	Selected references	Gene	LR	BR	SZ	Selected references
<i>DPP10</i>		+			<i>RUNX3</i>	+			
<i>DUSP1</i>	+				<i>SATB2</i>	+			
<i>DYRK1A</i>	+	+			<i>SFRP2</i>	+			
<i>DYX1C1</i>	+				<i>SHH</i>	+		+	Boyd et al., 2015
<i>EGFR</i>	+		+	Benzel et al., 2007	<i>SIRT1</i>	+	+	++	Kishi et al., 2011; Wang et al., 2015
<i>EGR1</i>	+	+	+	Pérez-Santiago et al., 2012	<i>SIX3</i>	+			
<i>ELavl2</i>	+		++	Yamada et al., 2011	<i>SLIT1</i>	+			
<i>ELP4</i>	+	+			<i>SLIT2</i>	+			
<i>EMX2</i>	+				<i>SLTRK3</i>	+	+		
<i>EP300</i>	+				<i>SMAD9</i>	+			
<i>ERBB4</i>	+	+	++	Agim et al., 2013; Hatzimanolis et al., 2013	<i>SMURF1</i>	+			
<i>ETV4</i>	+				<i>SOLH</i>	+			
<i>EXOC6B</i>	+				<i>SOX10</i>	+		++	Iwamoto et al., 2005; Drerup et al., 2009; Yuan et al., 2013; Wockner et al., 2014
<i>FEZF2</i>	+				<i>SOX2</i>	+			
<i>FGF7</i>	+				<i>SOX3</i>	+			
<i>FGF8</i>	+				<i>SOX9B</i>	+			
<i>FGFR1</i>	+		+	Mor et al., 2013	<i>SPAG5</i>	+			
<i>FLNA</i>	+	+			<i>SPC7</i>	+			
<i>FMR1</i>	+	+	+	Kovács et al., 2013; Folsom et al., 2015	<i>SPP1</i>	+			
<i>FOXA1</i>	+				<i>SRGAP2</i>	+	+		
<i>FOXA2</i>	+				<i>SRGAP3</i>	+		+	Wilson et al., 2011; Waltereit et al., 2012
<i>FOXG1</i>	+	+			<i>SRPX2</i>	+	+		
<i>FOXO1</i>	+				<i>TBR1</i>	+			
<i>FOXP1</i>	+		++	Ingason et al., 2015	<i>TGFB</i>	+		+	Frydecka et al., 2015,
<i>FOXP2</i>	+		+	Španiel et al., 2011; Li et al., 2013	<i>TLE2</i>	+			
<i>FRP</i>	+				<i>TLE3</i>	+			
<i>GAD1</i>	+	+	++	Ayalew et al., 2012; Mitchell et al., 2015; Lett et al., 2016	<i>TP53</i>	+		+	Ni et al., 2005; Molina et al., 2011
<i>GADD45G</i>	+				<i>TSC1</i>	+	+		
<i>GBX2</i>	+				<i>USF1</i>	+			
<i>GLI3</i>	+				<i>USH2A</i>	+			
<i>GRIN2A</i>		+	++	Lencz and Malhotra, 2015; Ohi et al., 2016	<i>VCAM1</i>	+			
<i>GTF2I</i>	+				<i>VCAN</i>	+			
<i>GTF3C3</i>	+				<i>VDR</i>	+			
<i>HES1</i>	+				<i>WNT5A</i>	+			
<i>HOXA2</i>	+				<i>YAP1</i>	+			
<i>HRAS</i>	+	+			<i>ZBTB20</i>	+			
<i>HTR1A</i>	+	+	+	Gu et al., 2013; Lin et al., 2015; Takekita et al., 2015	<i>ZFHXB1</i>	+			
<i>ITGB4</i>	+				<i>ZNF804A</i>		+	++	Becker et al., 2012; Nicodemus et al., 2014
<i>KATNA1</i>	+								

The first column contains the official name of the genes according to the Hugo Gene Nomenclature Committee (<http://www.genenames.org/>). The three remaining columns show whether the genes are candidates for language readiness according to Boeckx and Benítez-Burraco (2014a,b) and Benítez-Burraco and Boeckx (2015) (column 2: LR), are involved in brain rhythmicity according to the available literature, consulted via PubMed (<http://www.ncbi.nlm.nih.gov/pubmed>) (column 3: BR), or are candidates for schizophrenia (idem.) (column 4: SZ). The last column contains the most relevant papers that are indicative of an association between the gene and the disease. Candidate genes for schizophrenia resulting from GWA and CNV/exome sequencing studies are marked with ++ and should be regarded as more robust candidates than those resulting from candidate gene studies (marked with +) (for further details, see the Supplementary Files).

cortical circuitry; at the same time, these circuits are impaired in schizophrenia and other psychiatric and neurological conditions (Teffer and Semendeferi, 2012). Nonetheless, when it comes to testing this hypothesis, contradictory results have been obtained. Concerning the protein-coding regions of genes associated to psychiatric disorders Ogawa and Vallender (2014) did not find evidence of differential selection in humans compared to non-human primates, although elevated dN/dS was observed in primates and other large-brained taxa like cetaceans (dN/dS is the average number of nucleotide differences between sequences per non-synonymous site referred to the average number of nucleotide differences between sequences per synonymous site; dN/dS values that are significantly higher than 1 are indicative of positive selection). However, recent analyses based on large GWAs of schizophrenia and data of selective sweeps in the human genome compared to Neanderthals suggest that brain-related genes showing signals of recent positive selection in AMHs are also significantly associated with schizophrenia (Srinivasan et al., 2016), supporting the view that schizophrenia may be a by-product of the changes in the human brain that led to modern cognition. Interestingly, among the loci highlighted by Srinivasan et al. (2016), we have found several genes related to language development, language impairment, and language evolution, which strike us as new promising candidates for language dysfunction in schizophrenia. Among them, we wish highlight: *FOXP1*, *GATAD2B*, *MEF2C*, *NRG3*, *NRXN1*, and *ZNF804A* (see Supplementary Materials for details). We wish also highlight that some of the genes involved in brain rhythmicity (reviewed above) also show differences in the human lineage. *DPP10* shows signals of differential expression in the human brain compared to primates and sequences at *DPP10* show regulatory motifs absent in archaic hominins and signals of strong selection in modern human populations (Shulha et al., 2012). Likewise, *DISC1* interacts with *PCNT*, mentioned by Green et al. (2010) as being amongst the proteins that show non-synonymous and non-fixed changes compared to Neanderthals, and a candidate for dyslexia (Poelmans et al., 2011). Finally, the human *CNTNAP2* protein bears a fixed change (I345V) compared to the Denisovan variant (Meyer et al., 2012) and it is related in addition to *NFASC*, a protein involved in postsynaptic development and neurite outgrowth (Kriebel et al., 2012) which also shows a fixed change (T987A) in AMHs compared to Neanderthals/Denisovans (Pääbo, 2014, Table S1).

Some authors have explicitly linked the aetiopathology of schizophrenia and the evolution of language. According to Arbib and Mundhenk (2005) the primate mirror neurons, which fire both when the animal manipulates an object and when it sees another conspecific manipulating it, provided the scaffolding for imitation abilities involved in language acquisition. At the same time, schizophrenics show a spared ability to generate actions, whether manual or verbal, but they lack the ability to attribute the generation of that action to themselves. More drastically, Crow (1997) suggested that schizophrenia is the “price we paid for language.” According to him, schizophrenia represents an extreme of variation of hemispheric specialization and a single genetic mechanism (involving both the X and Y chromosomes) that was modified during recent human history can account for

this variation, because it generates epigenetic diversity related to both the species capacity for language and the predisposition to psychosis (Crow, 2008).

Our findings provide a different causative explanation to the origins and prevalence of schizophrenia, while still supporting the view that the etiopathology of this condition is heavily rooted in the evolution of human cognition. The genes discussed here map onto specific neuronal types (mostly, GABAergic), particular brain areas (several cortical layers, thalamic nuclei), particular physiological processes (the balance between inhibition and excitation), specific developmental processes (inter and interhemispheric axon pathfinding), and particular cognitive abilities (formal thought), all of which are aspects known to be impaired in schizophrenia. At the same time, all of them are involved in language development and processing and many of them have been modified during our recent evolutionary history. Interestingly, schizophrenia associations have been recently proved to be strongly enriched at enhancers that are active in tissues with important immune functions, giving support to the view that immune dysregulation plays a role in schizophrenia (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). Likewise, changes in the brain/immune system crosstalk have been hypothesized to have contributed to the changes in brain connectivity that prompted the emergence of our language-readiness (Benítez-Burraco and Uriagereka, 2015).

Accordingly, instead of thinking of schizophrenia as the “price we paid for language,” we believe a more accurate claim is that schizophrenia is the price we paid for a globular braincase housing more efficient and widespread recursive oscillatory embeddings. Because the more novel a neural network is in evolutionary terms, the less resilient it is (due to its lack of robust compensatory mechanisms, Toro et al., 2010), schizophrenia is found as a high prevalent condition among modern populations. This view is in line with current approaches to the etiology of complex diseases in humans, according to which high prevalent conditions of a multifactorial nature resulted from the de-canalization of the robust primate condition as a consequence of our evolutionary history (involving demographic bottlenecks, specific mutations, and cultural changes that uncovered cryptic variation, see Gibson, 2009 for details).

CONCLUSIONS

The considerations we have made here may provide a suitable response to Dehaene et al.’s (2015, p. 2) observation that linguistic computation requires “a specific recursive neural code, as yet unidentified by electrophysiology, possibly unique to humans, and which may explain the singularity of human language and cognition.” Hierarchical rhythmic coupling operations of the kind proposed in Murphy (2015a, 2016) and discussed here may also provide ways of integrating different forms of hierarchical representations, such as phonological, semantic and syntactic information (see Ding et al., 2016). Disruptions to the present dynamic model of linguistic computation may represent a comprehensive, unifying account of language-related

neurocognitive disorders. As we have argued, schizophrenia is of particular interest because it represents a mode of cognition and externalization of thought distinct from, but plainly related to, normally functioning linguistic cognition. Importantly, this deviance seems construable in terms of an alteration of the cognome-dynome cross-talk. A dynamic perspective cuts across the traditional positive-negative symptom division, being implicated both in abnormal active processes and in the absence of normal functions. This view is in line with more general, recent moves in neuroscience to view psychiatric illnesses as oscillatory connectopathies (Cao et al., 2016; Vinogradov and Herman, 2016). At the same time, the considerations we have presented also reinforce the view that the survey of the evolutionary itinerary followed by our faculty of language should help unravel abnormal cognitive/linguistic development in our species (and vice versa). The high number of candidates for schizophrenia selected in our species ostensibly proves this. We further expect that the present proposal has the potential to provide robust endophenotypes of schizophrenia (in the form of specific brain oscillation patterns and novel gene candidates) and contribute to an improved diagnosis and treatment of the disorder.

REFERENCES

- Aeby, A., De Tiège, X., Creuzil, M., David, P., Balériaux, D., Van Overmeire, B., et al. (2013). Language development at 2 years is correlated to brain microstructure in the left superior temporal gyrus at term equivalent age: a diffusion tensor imaging study. *Neuroimage* 78, 145–151. doi: 10.1016/j.neuroimage.2013.03.076
- Agim, Z. S., Esendal, M., Briollais, L., Uyan, O., Meschian, M., Martinez, L. A., et al. (2013). Discovery, validation and characterization of Erbb4 and Nrg1 haplotypes using data from three genome-wide association studies of schizophrenia. *PLoS ONE* 8:e53042. doi: 10.1371/journal.pone.0053042
- Anderson, G. R., Galfin, T., Xu, W., Aoto, J., Malenka, R. C., and Südhof, T. C. (2012). Candidate autism gene screen identifies critical role for cell-adhesion molecule CASPR2 in dendritic arborization and spine development. *Proc. Natl. Acad. Sci. U.S.A.* 109, 18120–18125. doi: 10.1073/pnas.1216398109
- Angrilli, A., Spironelli, C., Elbert, T., Crow, T. J., Marano, G., and Stegagnino, L. (2009). Schizophrenia as failure of left hemispheric dominance for the phonological component of language. *PLoS ONE* 4:e4507. doi: 10.1371/journal.pone.0004507
- Arbib, M. A., and Mundhenk, T. N. (2005). Schizophrenia and the mirror system: an essay. *Neuropsychologia* 43, 268–280. doi: 10.1016/j.neuropsychologia.2004.11.013
- Atz, M. E., Rollins, B., and Vawter, M. P. (2007). NCAM1 association study of bipolar disorder and schizophrenia: polymorphisms and alternatively spliced isoforms lead to similarities and differences. *Psychiatr. Genet.* 17, 55–67. doi: 10.1097/YPG.0b013e328012d850
- Ayalew, M., Le-Niculescu, H., Levey, D. F., Jain, N., Changala, B., Patel, S. D., et al. (2012). Convergent functional genomics of schizophrenia: from comprehensive understanding to genetic risk prediction. *Mol. Psychiatry* 17, 887–905. doi: 10.1038/mp.2012.37
- Bakhshi, K., and Chance, S. A. (2015). The neuropathology of schizophrenia: a selective review of past studies and emerging themes in brain structure and cytoarchitecture. *Neuroscience* 303, 82–102. doi: 10.1016/j.neuroscience.2015.06.028
- Barr, M. S., Farzan, F., Rajji, T. K., Voineskos, A. N., Blumberger, D. M., Arenovich, T., et al. (2013). Can repetitive magnetic stimulation improve cognition in schizophrenia? Pilot data from a randomized controlled trial. *Biol. Psychiatry* 73, 510–517. doi: 10.1016/j.biopsych.2012.08.020
- Başar-Eroğlu, C., Mathes, B., Brandt, A., and Schmiedt-Fehr, C. (2011). Occipital gamma response to auditory stimulation in patients with schizophrenia. *Int. J. Psychophysiol.* 79, 3–8. doi: 10.1016/j.ijpsycho.2010.10.011
- Becker, J., Czamara, D., Hoffmann, P., Landerl, K., Blomert, L., Brandeis, D., et al. (2012). Evidence for the involvement of ZNF804A in cognitive processes of relevance to reading and spelling. *Transl. Psychiatry* 2, e136. doi: 10.1038/tp.2012.62
- Benes, F. M., Lim, B., Matzilevich, D., Walsh, J. P., Subburaju, S., and Minns, M. (2007). Regulation of the GABA cell phenotype in hippocampus of schizophrenics and bipolars. *Proc. Natl. Acad. Sci. U.S.A.* 104, 10164–10169. doi: 10.1073/pnas.07073806104
- Benítez-Burraco, A. (2016). “A biolinguistic approach to language disorders: towards a paradigm shift in clinical linguistics,” in *Advances in Biolinguistics: The Human Language Faculty and Its Biological Basis*, eds K. Fujita and C. Boeckx (London: Routledge), 256–272.
- Benítez-Burraco, A., and Boeckx, C. (2015). Possible functional links among brain- and skull-related genes selected in modern humans. *Front. Psychol.* 6:794. doi: 10.3389/fpsyg.2015.00794
- Benítez-Burraco, A., and Uriagereka, J. (2015). The immune syntax revisited: opening new windows on language evolution. *Front. Mol. Neurosci.* 8:84. doi: 10.3389/fnmol.2015.00084
- Bennett, M. R. (2011). Schizophrenia: susceptibility genes, dendritic-spine pathology and gray matter loss. *Prog. Neurobiol.* 95, 275–300. doi: 10.1016/j.pneurobio.2011.08.003
- Benzel, I., Bansal, A., Browning, B. L., Galwey, N. W., Maycox, P. R., McGinnis, R., et al. (2007). Interactions among genes in the ErbB-Neuregulin signalling network are associated with increased susceptibility to schizophrenia. *Behav. Brain Funct.* 3:31. doi: 10.1186/1744-9081-3-31
- Boeckx, C., and Benítez-Burraco, A. (2014a). The shape of the human language-ready brain. *Front. Psychol.* 5:282. doi: 10.3389/fpsyg.2014.00282
- Boeckx, C., and Benítez-Burraco, A. (2014b). Globularity and language-readiness: generating new predictions by expanding the set of genes of interest. *Front. Psychol.* 5:1324. doi: 10.3389/fpsyg.2014.01324
- Booth, C. A., Brown, J. T., and Randall, A. D. (2014). Neurophysiological modification of CA1 pyramidal neurons in a transgenic mouse expressing a truncated form of disrupted-in-schizophrenia 1. *Eur. J. Neurosci.* 39, 1074–1090. doi: 10.1111/ejn.12549

AUTHOR CONTRIBUTIONS

EM contributed primarily to Sections From language deficits to the brain in schizophrenia and From brain rhythmicity to language deficits in schizophrenia, AB contributed primarily to Sections Introduction and Schizophrenia-related genes and some evolutionary concerns. Both authors contributed equally to Section Conclusions.

ACKNOWLEDGMENTS

Preparation of this work was supported in part by funds from the Spanish Ministry of Economy and Competitiveness (grant numbers FFI-2013-43823-P and FFI2014-61888-EXP to AB) and an Economic and Social Research Council scholarship (number 1474910 to EM).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fnhum.2016.00422>

- Boyd, P. J., Cunliffe, V. T., Roy, S., and Wood, J. D. (2015). Sonic hedgehog functions upstream of disrupted-in-schizophrenia 1 (discl): implications for mental illness. *Biol. Open* 4, 1336–1343. doi: 10.1242/bio.012005
- Brandon, N. J., and Sawa, A. (2011). Linking neurodevelopmental and synaptic theories of mental illness through DISC1. *Nat. Rev. Neurosci.* 12, 707–722. doi: 10.1038/nrn3120
- Brown, J. A., Horváth, S., Garbett, K. A., Schmidt, M. J., Everheart, M., Gellér, L., et al. (2014). The role of cannabinoid 1 receptor expressing interneurons in behavior. *Neurobiol. Dis.* 63, 210–221. doi: 10.1016/j.nbd.2013.11.001
- Buonanno, A. (2010). The neuregulin signaling pathway and schizophrenia: from genes to synapses and neural circuits. *Brain Res. Bull.* 83, 122–131. doi: 10.1016/j.brainresbull.2010.07.012
- Burdick, K. E., DeRosse, P., Kane, J. M., Lencz, T., and Malhotra, A. K. (2010). Association of genetic variation in the MET proto-oncogene with schizophrenia and general cognitive ability. *Am. J. Psychiatry* 167, 436–443. doi: 10.1176/appi.ajp.2009.09050615
- Buzsáki, G., Logothetis, N., and Singer, W. (2013). Scaling brain size, keeping timing: evolutionary preservation of brain rhythms. *Neuron* 80, 751–764. doi: 10.1016/j.neuron.2013.10.002
- Buzsáki, G., and Watson, B. O. (2012). Brain rhythms and neural syntax: implications for efficient coding of cognitive content and neuropsychiatric disease. *Dialogues Clin. Neurosci.* 14, 345–367.
- Cao, H., Dixson, L., Meyer-Lindenberg, A., and Tost, H. (2016). Functional connectivity measures as schizophrenia intermediate phenotypes: advances, limitations, and future directions. *Curr. Opin. Neurobiol.* 36, 7–14. doi: 10.1016/j.conb.2015.07.008
- Carvill, G. L., Regan, B. M., Yendle, S. C., O’Roak, B. J., Lozovaya, N., Bruneau, N., et al. (2013). GRIN2A mutations cause epilepsy-aphasia spectrum disorders. *Nat. Genet.* 45, 1073–1076. doi: 10.1038/ng.2727
- Chang, S., Fang, K., Zhang, K., and Wang, J. (2015). Network-based analysis of schizophrenia genome-wide association data to detect the joint functional association signals. *PLoS ONE* 10:e0133404. doi: 10.1371/journal.pone.0133404
- Chiu, Y. C., Li, M. Y., Liu, Y. H., Ding, J. Y., Yu, J. Y., and Wang, T. W. (2014). Foxp2 regulates neuronal differentiation and neuronal subtype specification. *Dev. Neurobiol.* 74, 723–738. doi: 10.1002/dneu.22166
- Cho, K. K., Hoch, R., Lee, A. T., Patel, T., Rubenstein, J. L., and Sohal, V. S. (2015). Gamma rhythms link prefrontal interneuron dysfunction with cognitive inflexibility in Dlx5/6(+/-) mice. *Neuron* 85, 1332–1343. doi: 10.1016/j.neuron.2015.02.019
- Cohen, D., Lazar, G., Couvert, P., Desportes, V., Lippe, D., Mazet, P., et al. (2002). MECP2 mutation in a boy with language disorder and schizophrenia. *Am. J. Psychiatry* 159, 148–149. doi: 10.1176/appi.ajp.159.1.148-a
- Cousijn, H., Tunbridge, E. M., Rolinski, M., Wallis, G., Colclough, G. L., Woolrich, M. W., et al. (2015). Modulation of hippocampal theta and hippocampal-prefrontal cortex function by a schizophrenia risk gene. *Hum. Brain Mapp.* 36, 2387–2395. doi: 10.1002/hbm.22778
- Crow, T. J. (1980). Molecular pathology of schizophrenia: more than one disease process? *Br. Med. J.* 280, 66–68. doi: 10.1136/bmjj.280.6207.66
- Crow, T. J. (1997). Is schizophrenia the price that Homo sapiens pays for language? *Schizophr. Res.* 28, 127–141. doi: 10.1016/S0920-9964(97)00110-2
- Crow, T. J. (2008). The ‘big bang’ theory of the origin of psychosis and the faculty of language. *Schizophr. Res.* 102, 31–52. doi: 10.1016/j.schres.2008.03.010
- Datta, D., Arion, D., Corradi, J. P., and Lewis, D. A. (2015). Altered expression of CDC42 signaling pathway components in cortical layer 3 pyramidal cells in schizophrenia. *Biol. Psychiatry* 78, 775–785. doi: 10.1016/j.biopsych.2015.03.030
- Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., and Pallier, C. (2015). The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. *Neuron* 88, 2–19. doi: 10.1016/j.neuron.2015.09.019
- Dennis, E. L., Jahanshad, N., Rudie, J. D., Brown, J. A., Johnson, K., McMahon, K. L., et al. (2011). Altered structural brain connectivity in healthy carriers of the autism risk gene, CNTNAP2. *Brain Connect.* 1, 447–459. doi: 10.1089/brain.2011.0064
- Diederik, K. M. J., De Weijer, A. D., Daalman, K., Blom, J. D., Neggers, S. F. W., Kahn, R. S., et al. (2010). Decreased language lateralization is characteristic of psychosis, not auditory hallucinations. *Brain* 133, 3734–3744. doi: 10.1093/brain/awq313
- Dimassi, S., Labalme, A., Lesca, G., Rudolf, G., Bruneau, N., Hirsch, E., et al. (2014). A subset of genomic alterations detected in rolandic epilepsies contains candidate or known epilepsy genes including GRIN2A and PRRT2. *Epilepsia* 55, 370–378. doi: 10.1111/epi.12502
- Ding, N., Melloni, L., Zhang, H., Tian, X., and Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nat. Neurosci.* 19, 158–164. doi: 10.1038/nn.4186
- Djurovic, S., Gustafsson, O., Mattingdal, M., Athanasiu, L., Bjella, T., Tesli, M., et al. (2010). A genome-wide association study of bipolar disorder in Norwegian individuals, followed by replication in Icelandic sample. *J. Affect. Disord.* 126, 312–316. doi: 10.1016/j.jad.2010.04.007
- Drerup, C. M., Wiora, H. M., Topczewski, J., and Morris, J. A. (2009). Disc1 regulates foxd3 and sox10 expression, affecting neural crest migration and differentiation. *Development* 136, 2623–2632. doi: 10.1242/dev.030577
- Emamian, E. S., Karayiorgou, M., and Gogos, J. A. (2004). Decreased phosphorylation of NMDA receptor type 1 at serine 897 in brains of patients with schizophrenia. *J. Neurosci.* 24, 1561–1564. doi: 10.1523/JNEUROSCI.4650-03.2004
- Farrell, M. S., Werge, T., Sklar, P., Owen, M. J., Ophoff, R. A., O’Donovan, M. C., et al. (2015). Evaluating historical candidate genes for schizophrenia. *Mol. Psychiatry* 20, 555–562. doi: 10.1038/mp.2015.16
- Farzan, F., Barr, M. S., Sun, Y., Fitzgerald, P. B., and Daskalakis, Z. J. (2012). Transcranial magnetic stimulation on the modulation of gamma oscillations in schizophrenia. *Ann. N.Y. Acad. Sci.* 1265, 25–35. doi: 10.1111/j.1749-6632.2012.06543.x
- Ferrarelli, F., Massimini, M., Peterson, M. J., Riedner, B. A., Lazar, M., Murphy, M. J., et al. (2008). Reduced evoked gamma oscillations in the frontal cortex in schizophrenia patients: a TMS/EEG study. *Am. J. Psychiatry* 165, 996–1005. doi: 10.1176/appi.ajp.2008.07111733
- Ferrarelli, F., Sarasso, S., Guller, Y., Riedner, B. A., Peterson, M. J., Belles, M., et al. (2012). Reduced natural oscillatory frequency of frontal thalamocortical circuits in schizophrenia. *Arch. Gen. Psychiatry* 69, 766–774. doi: 10.1001/archgenpsychiatry.2012.147
- Fisahn, A., Neddens, J., Yan, L., and Buonanno, A. (2009). Neuregulin-1 modulates hippocampal gamma oscillations: implications for schizophrenia. *Cereb. Cortex* 19, 612–618. doi: 10.1093/cercor/bhn107
- Flint, J., and Munafò, M. R. (2014). Genetics: finding genes for schizophrenia. *Curr. Biol.* 24, R755–R757. doi: 10.1016/j.cub.2014.07.018
- Folsom, T. D., Thuras, P. D., and Fatemi, S. H. (2015). Protein expression of targets of the FMRP regulon is altered in brains of subjects with schizophrenia and mood disorders. *Schizophr. Res.* 165, 201–211. doi: 10.1016/j.schres.2015.04.012
- Fraser, W., King, K., Thomas, P., and Kendall, R. E. (1986). The diagnosis of schizophrenia by language analysis. *Br. J. Psychiatry* 148, 275–278. doi: 10.1192/bjps.148.3.275
- Friedman, J. I., Vrijenhoek, T., Markx, S., Janssen, I. M., van der Vliet, W. A., Faas, B. H., et al. (2008). CNTNAP2 gene dosage variation is associated with schizophrenia and epilepsy. *Mol. Psychiatry* 13, 261–266. doi: 10.1038/sj.mp.4002049
- Frith, C. D. (1992). *The Cognitive Neuropsychology of Schizophrenia*. Hove: Lawrence Erlbaum Associates.
- Frith, C. D., and Allen, H. A. (1988). “Language disorders in schizophrenia and their implications for neuropsychology,” in *Schizophrenia: The Major Issues*, eds P. Bebbington and P. McGuffin (Oxford: Heinemann), 172–186.
- Frydecka, D., Misiak, B., Pawlak-Adamska, E., Karabon, L., Tomkiewicz, A., Sedlacek, P., et al. (2015). Sex differences in TGFB-β signaling with respect to age of onset and cognitive functioning in schizophrenia. *Neuropsychiatr. Dis. Treat.* 11, 575–584. doi: 10.2147/NDT.S74672
- Ghorashi, S., and Spencer, K. M. (2015). Attentional load effects on beta oscillations in healthy and schizophrenic individuals. *Front. Psychiatry* 6:149. doi: 10.3389/fpsyg.2015.00149
- Gibson, G. (2009). Decanalization and the origin of complex disease. *Nature Rev. Genet.* 10, 134–140. doi: 10.1038/nrg2502
- Gilks, W. P., Hill, M., Gill, M., Donohoe, G., Corvin, A. P., and Morris, D. W. (2012). Functional investigation of a schizophrenia GWAS signal at the CDC42 gene. *World J. Biol. Psychiatry* 13, 550–554. doi: 10.3109/15622975.2012.666359
- Giraud, A.-L., and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* 15, 11–517. doi: 10.1038/nn.3063

- Godsil, B. P., Kiss, J. P., Spedding, M., and Jay, T. M. (2013). The hippocampal-prefrontal pathway: the weak link in psychiatric disorders? *Eur. Neuropsychopharmacol.* 23, 1165–1181. doi: 10.1016/j.euroneuro.2012.10.018
- Grant, A., Fathallli, F., Rouleau, G., Joober, R., and Flores, C. (2012). Association between schizophrenia and genetic variation in DCC: a case-control study. *Schizophr. Res.* 137, 26–31. doi: 10.1016/j.schres.2012.02.023
- Grant, A., Hoops, D., Labelle-Dumais, C., Prévost, M., Rajabi, H., Kolb, B., et al. (2007). Netrin-1 receptor-deficient mice show enhanced mesocortical dopamine transmission and blunted behavioural responses to amphetamine. *Eur. J. Neurosci.* 26, 3215–3228. doi: 10.1111/j.1460-9568.2007.05888.x
- Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., et al. (2010). A draft sequence of the Neandertal genome. *Science* 328, 710–722. doi: 10.1126/science.1188021
- Gu, H., Liu, C., Liu, C., Chen, M., Zhang, Q., Zhai, J., et al. (2013). The combined effects of the 5'-HTTLPR and HTR1A rs6295 polymorphisms modulate decision making in schizophrenia patients. *Genes Brain Behav.* 12, 133–139. doi: 10.1111/j.1601-183X.2012.00866.x
- Gurung, R., and Prata, D. P. (2015). What is the impact of genome-wide supported risk variants for schizophrenia and bipolar disorder on brain structure and function? A systematic review. *Psychol. Med.* 45, 2461–2480. doi: 10.1017/S0033291715000537
- Hajós, M., Hoffmann, W. E., and Kocsis, B. (2008). Activation of cannabinoid-1 receptors disrupts sensory gating and neuronal oscillation: relevance to schizophrenia. *Biol. Psychiatry* 63, 1075–1083. doi: 10.1016/j.biopsych.2007.12.005
- Hall, J., Trent, S., Thomas, K. L., O'Donovan, M. C., and Owen, M. J. (2015). Genetic risk for schizophrenia: convergence on synaptic pathways involved in plasticity. *Biol. Psychiatry* 77, 52–58. doi: 10.1016/j.biopsych.2014.07.011
- Hall, M. H., Taylor, G., Sham, P., Schulze, K., Rijsdijk, F., Picchioni, M., et al. (2011). The early auditory gamma-band response is heritable and a putative endophenotype of schizophrenia. *Schizophr. Bull.* 37, 778–787. doi: 10.1093/schbul/sbp134
- Hatzimanolis, A., McGrath, J. A., Wang, R., Li, T., Wong, P. C., Nestadt, G., et al. (2013). Multiple variants aggregate in the neuregulin signaling pathway in a subset of schizophrenia patients. *Transl. Psychiatry* 3, e264. doi: 10.1038/tp.2013.33
- Hinna, K. H., Rich, K., Fex-Svenningssen, Å., and Benedikz, E. (2015). The rat homolog of the schizophrenia susceptibility gene ZNF804A is highly expressed during brain development, particularly in growth cones. *PLoS ONE* 10:e0132456. doi: 10.1371/journal.pone.0132456
- Hinzen, W., and Rosselló, J. (2015). The linguistics of schizophrenia: thought disturbance as language pathology across positive symptoms. *Front. Psychol.* 6:971. doi: 10.3389/fpsyg.2015.00971
- Hirayasu, Y., Shenton, M. E., Salisbury, D. F., DICKEY, C. C., Fischer, I. A., Mazzoni, P., et al. (1998). Lower left temporal lobe MRI volumes in patients with first-episode schizophrenia compared with psychotic patients with first episode affective disorder and normal subjects. *Am. J. Psychiatry* 155, 1384–1391. doi: 10.1176/ajp.155.10.1384
- Hoffman, R. E., Rapaport, J., Mazure, C. M., and Quinlan, D. M. (1999). Selective speech perception alterations in schizophrenic patients reporting hallucinated "voices." *Am. J. Psychiatry* 156, 393–399.
- Horn, H., Jann, K., Federspiel, A., Walther, S., Wiest, R., Müller, T., et al. (2012). Semantic network disconnection in formal thought disorder. *Neuropsychobiology* 66, 14–23. doi: 10.1159/000337133
- Hou, X. J., Ni, K. M., Yang, J. M., and Li, X. M. (2014). Neuregulin 1/ErbB4 enhances synchronized oscillations of prefrontal cortex neurons via inhibitory synapses. *Neuroscience* 261, 107–117. doi: 10.1016/j.neuroscience.2013.12.040
- Huang, Y. T., Iwamoto, K., Kurosaki, T., Nasu, M., and Ueda, S. (2005). The neuronal POU transcription factor Brn-2 interacts with Jab1, a gene involved in the onset of neurodegenerative diseases. *Neurosci. Lett.* 382, 175–178. doi: 10.1016/j.neulet.2005.03.008
- Inda, M. C., DeFelipe, J., and Muñoz, A. (2006). Voltage-gated ion channels in the axon initial segment of human cortical pyramidal cells and their relationship with chandelier cells. *Proc. Natl. Acad. Sci. U.S.A.* 103, 2920–2925. doi: 10.1073/pnas.0511197103
- Ingason, A., Giegling, I., Hartmann, A. M., Genius, J., Konte, B., Friedl, M., et al. (2015). Expression analysis in a rat psychosis model identifies novel candidate genes validated in a large case-control sample of schizophrenia. *Transl. Psychiatry* 5, e656. doi: 10.1038/tp.2015.151
- Ishii, R., Shinosaki, K., Ikejiri, I., Ukai, S., Yamashita, K., Iwase, M., et al. (2000). Theta rhythm increases in left superior temporal cortex during auditory hallucinations in schizophrenia: a case report. *Neuroreport* 11, 3283–3287. doi: 10.1097/00001756-200009280-00047
- Iwamoto, K., Bundo, M., Yamada, K., Takao, H., Iwayama-Shigeno, Y., Yoshikawa, T., et al. (2005). DNA methylation status of SOX10 correlates with its downregulation and oligodendrocyte dysfunction in schizophrenia. *J. Neurosci.* 25, 5376–5381. doi: 10.1523/JNEUROSCI.0766-05.2005
- Iwayama-Shigeno, Y., Yamada, K., Itokawa, M., Toyota, T., Meerabux, J. M. A., Minabe, Y., et al. (2005). Extended analyses support the association of a functional (GT)n polymorphism in the GRIN2A promoter with Japanese schizophrenia. *Neurosci. Lett.* 378, 102–105. doi: 10.1016/j.neulet.2004.12.013
- Jensen, O., Gips, B., Bergmann, T. O., and Bonnefond, M. (2014). Temporal coding organized by coupled alpha and gamma oscillations prioritize visual processing. *Trends Neurosci.* 37, 357–369. doi: 10.1016/j.tins.2014.04.001
- Ji, W., Li, T., Pan, Y., Tao, H., Ju, K., Wen, Z., et al. (2013). CNTNAP2 is significantly associated with schizophrenia and major depression in the Han Chinese population. *Psychiatry Res.* 207, 225–228. doi: 10.1016/j.psychres.2012.09.024
- Jiang, L., Xu, Y., Zhu, X. T., Yang, Z., Li, H. J., and Zuo, X. N. (2015). Local-to-remote cortical connectivity in early- and adulthood-onset schizophrenia. *Transl. Psychiatry* 5, e566. doi: 10.1038/tp.2015.59
- Johnston, A., McBain, C. J., and Fisahn, A. (2014). 5-Hydroxytryptamine 1A receptor activation hyperpolarizes pyramidal cells and suppresses hippocampal gamma oscillations via Kir3 channel activation. *J. Physiol.* 592, 4187–4199. doi: 10.1113/jphysiol.2014.279083
- Johnstone, M., Thomson, P. A., Hall, J., McIntosh, A. M., Lawrie, S. M., and Porteous, D. J. (2011). DISC1 in schizophrenia: genetic mouse models and human genomic imaging. *Schizophr. Bull.* 37, 14–20. doi: 10.1093/schbul/sbq135
- Jung, J. S., Lee, H. J., Cho, C. H., Kang, S. G., Yoon, H. K., Park, Y. M., et al. (2014). Association between restless legs syndrome and CLOCK and NPAS2 gene polymorphisms in schizophrenia. *Chronobiol. Int.* 31, 838–844. doi: 10.3109/07420528.2014.914034
- Kamiya, A., Tan, P. L., Kubo, K., Engelhard, C., Ishizuka, K., Kubo, A., et al. (2008). Recruitment of PCM1 to the centrosome by the cooperative action of DISC1 and BBS4: a candidate for psychiatric illnesses. *Arch. Gen. Psychiatry* 65, 996–1006. doi: 10.1001/archpsyc.65.9.996
- Kao, W. T., Wang, Y., Kleinman, J. E., Lipska, B. K., Hyde, T. M., Weinberger, D. R., et al. (2010). Common genetic variation in Neuregulin 3 (NRG3) influences risk for schizophrenia and impacts NRG3 expression in human brain. *Proc. Natl. Acad. Sci. U.S.A.* 107, 15619–15624. doi: 10.1073/pnas.1005410107
- Karam, C. S., Ballon, J. S., Bivens, N. M., Freyberg, Z., Girgis, R. R., Lizardi-Ortiz, J. E., et al. (2010). Signaling pathways in schizophrenia: emerging targets and therapeutic strategies. *Trends Pharmacol. Sci.* 31, 381–390. doi: 10.1016/j.tips.2010.05.004
- Kim, H. J., Kim, M. H., Choe, B. K., Kim, J. W., Park, J. K., Cho, A. R., et al. (2008). Genetic association between 5'-upstream single-nucleotide polymorphisms of PDGFRB and schizophrenia in a Korean population. *Schizophr. Res.* 103, 201–208. doi: 10.1016/j.schres.2008.04.031
- Kishi, T., Fukuo, Y., Kitajima, T., Okochi, T., Yamanouchi, Y., Kinoshita, Y., et al. (2011). SIRT1 gene, schizophrenia and bipolar disorder in the Japanese population: an association study. *Genes Brain Behav.* 10, 257–263. doi: 10.1111/j.1601-183X.2010.00661.x
- Kocsis, B. (2012). Differential role of NR2A and NR2B subunits in N-methyl-D-aspartate receptor antagonist-induced aberrant cortical gamma oscillations. *Biol. Psychiatry* 71, 987–995. doi: 10.1016/j.biopsych.2011.10.002
- Kopell, N. J., Gritton, H. J., Whittington, M. A., and Kramer, M. A. (2014). Beyond the connectome: the dynome. *Neuron* 83, 1319–1328. doi: 10.1016/j.neuron.2014.08.016
- Korotkova, T., Fuchs, E. C., Ponomarenko, A., von Engelhardt, J., and Monyer, H. (2010). NMDA receptor ablation on parvalbumin-positive interneurons impairs hippocampal synchrony, spatial representations, and working memory. *Neuron* 68, 557–569. doi: 10.1016/j.neuron.2010.09.017
- Kos, M., van den Brink, D., Snijders, T. M., Rijpkema, M., Franke, B., Fernandez, G., et al. (2012). CNTNAP2 and language processing in healthy individuals as measured with ERPs. *PLoS ONE* 7:e46995. doi: 10.1371/journal.pone.0046995

- Kovács, T., Kelemen, O., and Kéri, S. (2013). Decreased fragile X mental retardation protein (FMRP) is associated with lower IQ and earlier illness onset in patients with schizophrenia. *Psychiatry Res.* 210, 690–693. doi: 10.1016/j.psychres.2012.12.022
- Kriebel, M., Wuchter, J., Trinks, S., and Volkmer, H. (2012). Neurofascin: a switch between neuronal plasticity and stability. *Int. J. Biochem. Cell. Biol.* 44, 694–697. doi: 10.1016/j.biocel.2012.01.012
- Kromkamp, M., Uylings, H. B., Smidt, M. P., Hellemans, A. J., Burbach, J. P., and Kahn, R. S. (2003). Decreased thalamic expression of the homeobox gene DLX1 in psychosis. *Arch. Gen. Psychiatry* 60, 869–874. doi: 10.1001/archpsyc.60.9.869
- Krug, A., Nieratschker, V., Markov, V., Krach, S., Jansen, A., Zerres, K., et al. (2010). Effect of CACNA1C rs1006737 on neural correlates of verbal fluency in healthy individuals. *Neuroimage* 49, 1831–1836. doi: 10.1016/j.neuroimage.2009.09.028
- Kumar, D., Dedic, N., Flachskamm, C., Voulé, S., Deussing, J. M., and Kimura, M. (2015). Cacna1c (Cav1.2) modulates electroencephalographic rhythm and rapid eye movement sleep recovery. *Sleep* 38, 1371–1380. doi: 10.5665/sleep.4972
- Lattanzi, W., Di Pietro, L., Barba, M., Boeckx, C., Massimi, L., Chieffo, D., et al. (2016). "Differential expression of candidate genes for skull globularization in nonsyndromic craniosynostosis patients: a common molecular network in skull shape evolution, craniofacial dysmorphology and cognitive disabilities?" in *The European Human Genetics Conference 2016 (poster presentation)* (Barcelona).
- Lencz, T., and Malhotra, A. K. (2015). Targeting the schizophrenia genome: a fast track strategy from GWAS to clinic. *Mol. Psychiatry* 20, 820–826. doi: 10.1038/mp.2015.28
- Lesca, G., Rudolf, G., Bruneau, N., Lozovaya, N., Labalme, A., Boutry-Kryza, N., et al. (2013). GRIN2A mutations in acquired epileptic aphasia and related childhood focal epilepsies and encephalopathies with speech and language dysfunction. *Nat. Genet.* 45, 1061–1066. doi: 10.1038/ng.2726
- Lett, T. A., Kennedy, J. L., Radhu, N., Dominguez, L. G., Chakravarty, M. M., Nazeri, A., et al. (2016). Prefrontal white matter structure mediates the influence of GAD1 on working memory. *Neuropsychopharmacology* 41, 2224–2231. doi: 10.1038/npp.2016.14
- Levchenko, A., Davtian, S., Freylichman, O., Zagrinaya, M., Kostareva, A., and Malashichev, Y. (2015). Beta-catenin in schizophrenia: possibly deleterious novel mutation. *Psychiatry Res.* 228, 843–848. doi: 10.1016/j.psychres.2015.05.014
- Li, T., Zeng, Z., Zhao, Q., Wang, T., Huang, K., Li, J., et al. (2013). FoxP2 is significantly associated with schizophrenia and major depression in the Chinese Han population. *World J. Biol. Psychiatry* 14, 146–150. doi: 10.3109/15622975.2011.615860
- Li, W., Guo, X., and Xiao, S. (2015). Evaluating the relationship between reelin gene variants (rs7341475 and rs262355) and schizophrenia: a meta-analysis. *Neurosci. Lett.* 609, 42–47. doi: 10.1016/j.neulet.2015.10.014
- Lin, H., Lei, Y., Zhang, B., Dai, Z., and Lu, X. (2015). Common variants of HTR1A and SLC6A4 confer the increasing risk of Schizophrenia susceptibility: a population-based association and epistasis analysis. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 168, 749–755. doi: 10.1002/ajmg.b.32380
- Linkenkaer-Hansen, K., Smit, D. J., Barkil, A., van Beijsterveldt, T. E., Brussaard, A. B., Boomsma, D. I., et al. (2007). Genetic contributions to long-range temporal correlations in ongoing oscillations. *J. Neurosci.* 27, 13882–13889. doi: 10.1523/JNEUROSCI.3083-07.2007
- Liu, R., Dang, W., Du, Y., Zhou, Q., Liu, Z., and Jiao, K. (2015). Correlation of functional GRIN2A gene promoter polymorphisms with schizophrenia and serum d-serine levels. *Gene* 568, 25–30. doi: 10.1016/j.gene.2015.05.011
- Mai, G., Minett, J. W., and Wang, W. S.-Y. (2016). Delta, theta, beta, and gamma brain oscillations index levels of auditory sentence processing. *Neuroimage* 133, 516–528. doi: 10.1016/j.neuroimage.2016.02.064
- Mallas, E. J., Carletti, F., Chaddock, C. A., Woolley, J., Picchioni, M. M., Shergill, S. S., et al. (2016). Genome-wide discovered psychosis-risk gene ZNF804A impacts on white matter microstructure in health, schizophrenia and bipolar disorder. *PeerJ* 4:e1570. doi: 10.7717/peerj.1570
- Manoach, D. S., Pan, J. Q., Purcell, S. M., and Stickgold, R. (2015). Reduced sleep spindles in schizophrenia: a treatable endophenotype that links risk genes to impaired cognition? *Biol. Psychiatry*. doi: 10.1016/j.biopsych.2015.10.003. [Epub ahead of print].
- McCarthy, S. E., Gillis, J., Kramer, M., Lihm, J., Yoon, S., Berstein, Y., et al. (2014). *De novo* mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. *Mol. Psychiatry* 19, 652–658. doi: 10.1038/mp.2014.29
- McKenna, P. J., and Oh, T. M. (2005). *Schizophrenic Speech: Making Sense of Bathrooms and Ponds that Fall in Doorways*. Cambridge: Cambridge University Press.
- Meyer, M., Kircher, M., Gansauge, M. T., Li, H., Racimo, F., Mallick, S., et al. (2012). A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338, 222–226. doi: 10.1126/science.1224344
- Mitchell, A. C., Jiang, Y., Peter, C., and Akbarian, S. (2015). Transcriptional regulation of GAD1 GABA synthesis gene in the prefrontal cortex of subjects with schizophrenia. *Schizophr. Res.* 167, 28–34. doi: 10.1016/j.schres.2014.10.020
- Molina, V., Papiol, S., Sanz, J., Rosa, A., Arias, B., Fatjó-Vilas, M., et al. (2011). Convergent evidence of the contribution of TP53 genetic variation (Pro72Arg) to metabolic activity and white matter volume in the frontal lobe in schizophrenia patients. *Neuroimage* 56, 45–51. doi: 10.1016/j.neuroimage.2011.01.076
- Mor, E., Kano, S., Colantuoni, C., Sawa, A., Navon, R., and Shomron, N. (2013). MicroRNA-382 expression is elevated in the olfactory neuroepithelium of schizophrenia patients. *Neurobiol. Dis.* 55, 1–10. doi: 10.1016/j.nbd.2013.03.011
- Moran, L. V., and Hong, L. E. (2011). High versus low frequency neural oscillations in schizophrenia. *Schizophr. Bull.* 37, 659–663. doi: 10.1093/schbul/sbr056
- Morice, R., and Ingram, J. C. L. (1982). Language analysis in schizophrenia. *Aust. N.Z. J. Psychiatry* 16, 11–21. doi: 10.3109/00048678209161186
- Morice, R., and McNicol, D. (1985). The comprehension and production of complex syntax in schizophrenia. *Cortex* 21, 567–580. doi: 10.1016/S0010-9452(58)80005-2
- Morice, R., and McNicol, D. (1986). Language changes in schizophrenia. *Schizophr. Bull.* 12, 239–250. doi: 10.1093/schbul/12.2.239
- Murphy, E. (2015a). The brain dynamics of linguistic computation. *Front. Psychol.* 6:1515. doi: 10.3389/fpsyg.2015.01515
- Murphy, E. (2015b). Labels, cognomes and cyclic computation: an ethological perspective. *Front. Psychol.* 6:715. doi: 10.3389/fpsyg.2015.00715
- Murphy, E. (2016). The human oscillome and its explanatory potential. *Biolinguistics* 10, 6–20.
- Nakamura, T., Matsumoto, J., Takamura, Y., Ishii, Y., Sasahara, M., Ono, T., et al. (2015). Relationships among parvalbumin-immunoreactive neuron density, phase-locked gamma oscillations, and autistic/schizophrenic symptoms in PDGFR- β knock-out and control mice. *PLoS ONE* 10:e0119258. doi: 10.1371/journal.pone.0119258
- Narayan, S., Tang, B., Head, S. R., Gilmartin, T. J., Sutcliffe, J. G., Dean, B., et al. (2008). Molecular profiles of schizophrenia in the CNS at different stages of illness. *Brain Res.* 1239, 235–248. doi: 10.1016/j.brainres.2008.08.023
- Narita, H. (2014). *Endocentric Structuring of Projection-free Syntax*. Amsterdam: John Benjamins.
- Newbury, D. F., Paracchini, S., Scerri, T. S., Winchester, L., Addis, L., Richardson, A. J., et al. (2011). Investigation of dyslexia and SLI risk variants in reading- and language-impaired subjects. *Behav. Genet.* 41, 90–104. doi: 10.1007/s10519-010-9424-3
- Nguyen, P. T., Nakamura, T., Hori, E., Urakawa, S., Uwano, T., Zhao, J., et al. (2011). Cognitive and socio-emotional deficits in platelet-derived growth factor receptor- β gene knockout mice. *PLoS ONE* 6:e18004. doi: 10.1371/journal.pone.0018004
- Ni, X., Trakalo, J., Valente, J., Azevedo, M. H., Pato, M. T., Pato, C. N., et al. (2005). Human p53 tumor suppressor gene (TP53) and schizophrenia: case-control and family studies. *Neurosci. Lett.* 388, 173–178. doi: 10.1016/j.neulet.2005.06.050
- Nicodemus, K. K., Elvevåg, B., Foltz, P. W., Rosenstein, M., Diaz-Asper, C., and Weinberger, D. R. (2014). Category fluency, latent semantic analysis and schizophrenia: a candidate gene approach. *Cortex* 55, 182–191. doi: 10.1016/j.cortex.2013.12.004
- Nunez, P. L., and Srinivasan, R. (2006). *Electric Fields of the Brain: The Neurophysics of EEG*. 2nd Edn. Oxford: Oxford University Press.
- Nunez, P. L., Wingeier, B. M., and Silberstein, R. B. (2001). Spatial-temporal structures of human alpha rhythms: theory, microcurrent sources, multiscale

- measurements, and global binding of local networks. *Hum. Brain Mapp.* 13, 125–164. doi: 10.1002/hbm.1030
- O'Tuathaigh, C. M., Desbonnet, L., Moran, P. M., and Waddington, J. L. (2012). Susceptibility genes for schizophrenia: mutant models, endophenotypes and psychobiology. *Curr. Top. Behav. Neurosci.* 12, 209–250. doi: 10.1007/7854_2011_194
- Ogawa, L. M., and Vallender, E. J. (2014). Evolutionary conservation in genes underlying human psychiatric disorders. *Front. Hum. Neurosci.* 8:283. doi: 10.3389/fnhum.2014.00283
- Ohi, K., Shimada, T., Nitta, Y., Kihara, H., Okubo, H., Uehara, T., et al. (2016). Specific gene expression patterns of 108 schizophrenia-associated loci in cortex. *Schizophr. Res.* 174, 35–38. doi: 10.1016/j.schres.2016.03.032
- Onwuameze, O. E., Nam, K. W., Epping, E. A., Wassink, T. H., Ziebell, S., Andreasen, N. C., et al. (2013). MAPK14 and CNR1 gene variant interactions: effects on brain volume deficits in schizophrenia patients with marijuana misuse. *Psychol. Med.* 43, 619–631. doi: 10.1017/S0033291712001559
- Oribe, N., Onitsuka, T., Hirano, S., Hirano, Y., Maekawa, T., Obayashi, C., et al. (2010). Differentiation between bipolar disorder and schizophrenia revealed by neural oscillation to speech sounds: an MEG study. *Bipolar Disord.* 12, 804–812. doi: 10.1111/j.1399-5618.2010.00876.x
- Osburn, N., Li, J., O'Driscoll, M. C., Strominger, Z., Wakahiro, M., Rider, E., et al. (2011). Genetic and functional analyses identify DISC1 as a novel callosal agenesis candidate gene. *Am. J. Med. Genet. A* 155A, 1865–1876. doi: 10.1002/ajmg.a.34081
- Pääbo, S. (2014). The human condition: a molecular approach. *Cell* 157, 216–226. doi: 10.1016/j.cell.2013.12.036
- Pérez-Santiago, J., Diez-Alarcia, R., Callado, L. F., Zhang, J. X., Chana, G., White, C. H., et al. (2012). A combined analysis of microarray gene expression studies of the human prefrontal cortex identifies genes implicated in schizophrenia. *J. Psychiatr. Res.* 46, 1464–1474. doi: 10.1016/j.jpsychires.2012.08.005
- Peter, B., Raskind, W. H., Matsushita, M., Lisowski, M., Vu, T., Berninger, V. W., et al. (2011). Replication of CNTNAP2 association with nonword repetition and support for FOXP2 association with timed reading and motor activities in a dyslexia family sample. *J. Neurodev. Disord.* 3, 39–49. doi: 10.1007/s11689-010-9065-0
- Petrin, A. L., Giachetti, C. M., Maximino, L. P., Abramides, D. V., Zanchetta, S., Rossi, N. F., et al. (2010). Identification of a microdeletion at the 7q33-q35 disrupting the CNTNAP2 gene in a Brazilian stuttering case. *Am. J. Med. Genet. A* 152A, 3164–3172. doi: 10.1002/ajmg.a.33749
- Pittman-Polletta, B. R., Kocsis, B., Vijayan, S., Whittington, M. A., and Kopell, N. J. (2015). Brain rhythms connect impaired inhibition to altered cognition in schizophrenia. *Biol. Psychiatry* 77, 1020–1030. doi: 10.1016/j.biopsych.2015.02.005
- Plum, F. (1972). Prospects for research on schizophrenia. 3. Neurophysiology. Neuropathological findings. *Neurosci. Res. Program Bull.* 10, 384–388.
- Poelmans, G., Buitelaar, J. K., Pauls, D. L., and Franke, B. (2011). A theoretical molecular network for dyslexia: integrating available genetic findings. *Mol. Psychiatry* 16, 365–382. doi: 10.1038/mp.2010.105
- Poeppel, D. (2012). The maps problem and the mapping problem: two challenges for a cognitive neuroscience of speech and language. *Cogn. Neuropsychol.* 29, 34–55. doi: 10.1080/02643294.2012.710600
- Poot, M., van't Slot, R., Leupert, R., Beyer, V., Passarge, E., and Haaf, T. (2009). Three *de novo* losses and one insertion within a pericentric inversion of chromosome 6 in a patient with complete absence of expressive speech and reduced pain perception. *Med. Genet.* 52, 27–30. doi: 10.1016/j.ejmg.2008.11.002
- Popov, T., and Popova, P. (2015). Same clock, different time read-out: spontaneous brain oscillations and their relationship to deficient coding of cognitive content. *NeuroImage* 119, 316–324. doi: 10.1016/j.neuroimage.2015.06.071
- Potkin, S. G., Macciardi, F., Guffanti, G., Fallon, J. H., Wang, Q., Turner, J. A., et al. (2010). Identifying gene regulatory networks in schizophrenia. *NeuroImage* 53, 839–847. doi: 10.1016/j.neuroimage.2010.06.036
- Potkin, S. G., Turner, J. A., Guffanti, G., Lakatos, A., Fallon, J. H., Nguyen, D. D., et al. (2009). A genome-wide association study of schizophrenia using brain activation as a quantitative phenotype. *Schizophr. Bull.* 35, 96–108. doi: 10.1093/schbul/sbn155
- Ramsden, P. (2013). *Understanding Abnormal Psychology: Clinical and Biological Perspectives*. London: SAGE Publications.
- Rolstad, S., Pálsson, E., Ekman, C. J., Eriksson, E., Sellgren, C., and Landén, M. (2015). Polymorphisms of dopamine pathway genes NRG1 and LMX1A are associated with cognitive performance in bipolar disorder. *Bipolar Disord.* 17, 859–868. doi: 10.1111/bdi.12347
- Roux, F., Wibral, M., Mohr, H. M., Singer, W., and Uhlhaas, P. J. (2012). Gamma-band activity in human prefrontal cortex codes for the number of relevant items maintained in working memory. *J. Neurosci.* 32, 12411–12420. doi: 10.1523/JNEUROSCI.0421-12.2012
- Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511, 421–427. doi: 10.1038/nature13595
- Schulman, J. J., Cancro, R., Lowe, S., Lu, F., Walton, K. D., and Llinás, R. R. (2011). Imaging of thalamocortical dysrhythmia in neuropsychiatry. *Front. Hum. Neuroscience* 5:69. doi: 10.3389/fnhum.2011.00069
- Scott-Van Zeeland, A. A., Abrahams, B. S., Alvarez-Retuerto, A. I., Sonnenblick, L. I., Rudie, J. D., Ghahremani, D., et al. (2010). Altered functional connectivity in frontal lobe circuits is associated with variation in the autism risk gene CNTNAP2. *Sci. Transl. Med.* 2, 56–80. doi: 10.1126/scitranslmed.3001344
- Sehested, L. T., Møller, R. S., Bache, I., Andersen, N. B., Ullmann, R., Tommerup, N., et al. (2010). Deletion of 7q34-q36.2 in two siblings with mental retardation, language delay, primary amenorrhea, and dysmorphic features. *Am. J. Med. Genet. A* 152A, 3115–3119. doi: 10.1002/ajmg.a.33476
- Seshadri, S., Faust, T., Ishizuka, K., Delevich, K., Chung, Y., Kim, S. H., et al. (2015). Interneuronal DISC1 regulates NRG1-ErbB4 signalling and excitatory-inhibitory synapse formation in the mature cortex. *Nat. Commun.* 6:10118. doi: 10.1038/ncomms10118
- Shiftman, S., Johannesson, M., Bronstein, M., Chen, S. X., Collier, D. A., Craddock, N. J., et al. (2008). Genome-wide association identifies a common variant in the reelin gene that increases the risk of schizophrenia only in women. *PLoS Genet.* 4:e28. doi: 10.1371/journal.pgen.0040028
- Shulha, H. P., Crisci, J. L., Reshetov, D., Tushir, J. S., Cheung, I., Bharadwaj, R., et al. (2012). Human-specific histone methylation signatures at transcription start sites in prefrontal neurons. *PLoS Biol.* 10:e1001427. doi: 10.1371/journal.pbio.1001427
- Španiel, F., Horáček, J., Tintěra, J., and Ibrahim, I., Novák, T., Čermák, J., et al. (2011). Genetic variation in FOXP2 alters grey matter concentrations in schizophrenia patients. *Neurosci. Lett.* 493, 131–135. doi: 10.1016/j.neulet.2011.02.024
- Spironelli, C., and Angrilli, A. (2015). Language-related gamma EEG frontal reduction is associated with positive symptoms in schizophrenia patients. *Schizophr. Res.* 165, 22–29. doi: 10.1016/j.schres.2015.04.003
- Spironelli, C., Angrilli, A., Calogero, A., and Stegagno, L. (2011). Delta EEG band as a marker of left hypofrontality for language schizophrenia patients. *Schizophr. Bull.* 37, 757–767. doi: 10.1093/schbul/sbp145
- Srinivasan, S., Bettella, F., Mattingdal, M., Wang, Y., Witoelar, A., Schork, A. J., et al. (2016). Genetic markers of human evolution are enriched in schizophrenia. *Biol. Psychiatry* 80, 284–292. doi: 10.1016/j.biopsych.2015.10.009
- Stephanie, M., Kuskowski, M., and Gundel, J. (2014). Abnormal dynamics of language in schizophrenia. *Psychiatry Res.* 216, 320–324. doi: 10.1016/j.psychres.2014.02.027
- Stephanie, M., Pellizzer, G., Fletcher, C. R., and McClannahan, K. (2007). Empirical evaluation of language disorder in schizophrenia. *J. Psychiatry Neurosci.* 32, 250–258.
- Suárez-Pinilla, P., Roiz-Santiañez, R., Ortiz-García de la Foz, V., Guest, P. C., Ayesa-Arriola, R., Córdoba-Palomera, A., et al. (2015). Brain structural and clinical changes after first episode psychosis: focus on cannabinoid receptor 1 polymorphisms. *Psychiatry Res.* 233, 112–119. doi: 10.1016/j.psychres.2015.05.005
- Sun, J., Tang, Y., Lim, K. O., Wang, J., Tong, S., Li, H., et al. (2014). Abnormal dynamics of EEG oscillations in schizophrenia patients on multiple time scales. *IEEE Trans. Biomed. Eng.* 61, 1756–1764. doi: 10.1109/TBME.2014.2306424
- Takekita, Y., Fabbri, C., Kato, M., Nonen, S., Sakai, S., Sunada, N., et al. (2015). HTR1A gene polymorphisms and 5-HT1A receptor partial agonist antipsychotics efficacy in schizophrenia. *J. Clin. Psychopharmacol.* 35, 220–227. doi: 10.1097/JCP.0000000000000304
- Tan, G. C., Doke, T. F., Ashburner, J., Wood, N. W., and Frackowiak, R. S. (2010). Normal variation in fronto-occipital circuitry and cerebellar structure with an

- autism-associated polymorphism of CNTNAP2. *Neuroimage* 53, 1030–1042. doi: 10.1016/j.neuroimage.2010.02.018
- Tang, J., Chen, X., Xu, X., Wu, R., Zhao, J., Hu, Z., et al. (2006). Significant linkage and association between a functional (GT)n polymorphism in promoter of the N-methyl-d-aspartate receptor subunit gene (GRIN2A) and schizophrenia. *Neurosci. Lett.* 409, 80–82. doi: 10.1016/j.neulet.2006.09.022
- Teffer, K., and Semendeferi, K. (2012). Human prefrontal cortex: evolution, development, and pathology. *Prog. Brain Res.* 195, 191–218. doi: 10.1016/B978-0-444-53860-4.00009-X
- Thomas, P., King, K., Fraser, W., and Kendall, R. E. (1987). Linguistic performance in schizophrenia: a comparison of patients with positive and negative symptoms. *Acta Psychiatr. Scand.* 76, 144–151. doi: 10.1111/j.1600-0447.1987.tb02877.x
- Thygesen, J. H., Zambach, S. K., Ingason, A., Lundin, P., Hansen, T., Bertalan, M., et al. (2015). Linkage and whole genome sequencing identify a locus on 6q25–26 for formal thought disorder and implicate MEF2A regulation. *Schizophr. Res.* 169, 441–446. doi: 10.1016/j.schres.2015.08.037
- Toro, R., Konyukh, M., Delorme, R., Leblond, C., Chaste, P., Fauchereau, F., et al. (2010). Key role for gene dosage and synaptic homeostasis in autism spectrum disorders. *Trends Genet.* 26, 363–372. doi: 10.1016/j.tig.2010.05.007
- Tosato, S., Bellani, M., Bonetto, C., Ruggeri, M., Perlini, C., Lasalvia, A., et al. (2012). Is neuregulin 1 involved in determining cerebral volumes in schizophrenia? Preliminary results showing a decrease in superior temporal gyrus volume. *Neuropsychobiology* 65, 119–125. doi: 10.1159/000330584
- Turner, S. J., Mayes, A. K., Verhoeven, A., Mandelstam, S. A., Morgan, A. T., and Scheffer, I. E. (2015). GRIN2A: an aptly named gene for speech dysfunction. *Neurology* 84, 586–593. doi: 10.1212/WNL.0000000000001228
- Uhlhaas, P. J., Haenschel, C., Nikolic, D., and Singer, W. (2008). The role of oscillations and synchrony in cortical networks and their putative relevance for the pathophysiology of schizophrenia. *Schizophr. Bull.* 34, 927–943. doi: 10.1093/schbul/sbn062
- van Os, J., and Kapur, S. (2009). Schizophrenia. *Lancet* 374, 635–645. doi: 10.1016/S0140-6736(09)60995-8
- Vawter, M. P., Usen, N., Thatcher, L., Ladenheim, B., Zhang, P., VanderPutten, D. M., et al. (2001). Characterization of human cleaved N-CAM and association with schizophrenia. *Exp. Neurol.* 172, 29–46. doi: 10.1006/exnr.2001.7790
- Verbrugge, P., Bouwer, S., Wiltshire, S., Carter, K., Chandler, D., Cooper, M., et al. (2012). Impact of the Reelin signaling cascade (ligands-receptors-adaptor complex) on cognition in schizophrenia. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 159B, 392–404. doi: 10.1002/ajmg.b.32042
- Vernes, S. C., Newbury, D. F., Abrahams, B. S., Winchester, L., Nicod, J., Groszer, M., et al. (2008). A functional genetic link between distinct developmental language disorders. *N. Engl. J. Med.* 359, 2337–2345. doi: 10.1056/NEJMoa0802828
- Vinogradov, S., and Herman, A. (2016). Psychiatric illnesses as oscillatory connectomopathies. *Neuropsychopharmacology* 41, 387–388. doi: 10.1038/npp.2015.308
- Walker, R. M., Hill, A. E., Newman, A. C., Hamilton, G., Torrance, H. S., Anderson, S. M., et al. (2012). The DISC1 promoter: characterization and regulation by FOXP2. *Hum. Mol. Genet.* 21, 2862–2872. doi: 10.1093/hmg/ddz111
- Waltereit, R., Leimer, U., von Bohlen und Halbach, O., Panke, J., Höller, S. M., Garrett, L., et al. (2012). Srgap3^{-/-} mice present a neurodevelopmental disorder with schizophrenia-related intermediate phenotypes. *FASEB J.* 26, 4418–4428. doi: 10.1096/fj.11-202317
- Wang, Y., Huang, Y., Peng, M., Cong, Z., Li, X., Lin, A., et al. (2015). Association between Silent Information Regulator 1 (SIRT1) gene polymorphisms and schizophrenia in a Chinese Han population. *Psychiatry Res.* 225, 744–745. doi: 10.1016/j.psychres.2014.11.027
- Weiss, A. P., Dewitt, I., Goff, D., Ditman, T., and Heckers, S. (2005). Anterior and posterior hippocampal volumes in schizophrenia. *Schizophr. Res.* 73, 103–112. doi: 10.1016/j.schres.2004.05.018
- Weisz, N., Wuhle, A., Monittola, G., Demarchi, G., Frey, J., Popov, T., et al. (2014). Prestimulus oscillatory power and connectivity patterns predispose conscious somatosensory perception. *Proc. Natl. Acad. Sci. U.S.A.* 111, E417–E425. doi: 10.1073/pnas.1317267111
- Whalley, H. C., O'Connell, G., Sussmann, J. E., Peel, A., Stanfield, A. C., Hayiou-Thomas, M. E., et al. (2011). Genetic variation in CNTNAP2 alters brain function during linguistic processing in healthy individuals. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 156B, 941–948. doi: 10.1002/ajmg.b.31241
- Whitehouse, A. J., Bishop, D. V., Ang, Q. W., Pennell, C. E., and Fisher, S. E. (2011). CNTNAP2 variants affect early language development in the general population. *Genes Brain Behav.* 10, 451–456. doi: 10.1111/j.1601-183X.2011.00684.x
- Williams, S., and Boksa, P. (2010). Gamma oscillations and schizophrenia. *J. Psychiatry Neurosci.* 35, 75–77. doi: 10.1503/jpn.100021
- Wilson, N. K., Lee, Y., Long, R., Hermetz, K., Rudd, M. K., Miller, R., et al. (2011). A novel microduplication in the neurodevelopmental gene SRGAP3 that segregates with psychotic illness in the family of a COS proband. *Case Rep. Genet.* 2011:585893. doi: 10.1155/2011/585893
- Wockner, L. F., Noble, E. P., Lawford, B. R., Young, R. M., Morris, C. P., Whitehall, V. L., et al. (2014). Genome-wide DNA methylation analysis of human brain tissue from schizophrenia patients. *Transl. Psychiatry* 4, e339. doi: 10.1038/tp.2013.111
- Worthey, E. A., Raca, G., Laffin, J. J., Wilk, B. M., Harris, J. M., Jakielski, K. J., et al. (2013). Whole-exome sequencing supports genetic heterogeneity in childhood apraxia of speech. *J. Neurodev. Disord.* 5:29. doi: 10.1186/1866-1955-5-29
- Xie, J., Gizatullin, R., Vukojevic, V., and Leopardi, R. (2015). The CCDC55 couples cannabinoid receptor CNR1 to a putative DISC1 schizophrenia pathway. *Neuroscience* 310, 723–730. doi: 10.1016/j.neuroscience.2015.10.012
- Xu, T., Stephane, M., and Parhi, K. K. (2012). Selection of abnormal neural oscillation patterns associated with sentence-level language disorder in schizophrenia. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2012, 4923–4926. doi: 10.1109/EMBC.2012.6347098
- Xu, T., Stephane, M., and Parhi, K. K. (2013). Multidimensional analysis of the abnormal neural oscillations associated with lexical processing in schizophrenia. *Clin. EEG Neurosci.* 44, 135–143. doi: 10.1177/1550059412465078
- Yamada, K., Iwayama, Y., Hattori, E., Iwamoto, K., Toyota, T., Ohnishi, T., et al. (2011). Genome-wide association study of schizophrenia in Japanese population. *PLoS ONE* 6:e20468. doi: 10.1371/journal.pone.0020468
- Yu, W., De Hert, M., Moons, T., Claes, S. J., Correll, C. U., and van Winkel, R. (2013). CNR1 gene and risk of the metabolic syndrome in patients with schizophrenia. *J. Clin. Psychopharmacol.* 33, 186–192. doi: 10.1097/JCP.0b013e318283925e
- Yuan, A., Li, W., Yu, T., Zhang, C., Wang, D., Liu, D., et al. (2013). SOX10 rs139883 polymorphism is associated with the age of onset in schizophrenia. *J. Mol. Neurosci.* 50, 333–338. doi: 10.1007/s12031-013-9982-y
- Zeledón, M., Eckart, N., Taub, M., Vernon, H., Szymanksi, M., Wang, R., et al. (2015). Identification and functional studies of regulatory variants responsible for the association of NRG3 with a delusion phenotype in schizophrenia. *Mol. Neuropsychiatry* 1, 36–46. doi: 10.1159/000371518
- Zhang, B., Xu, Y. H., Wei, S. G., Zhang, H. B., Fu, D. K., Feng, Z. F., et al. (2014). Association study identifying a new susceptibility gene (AUTS2) for schizophrenia. *Int. J. Mol. Sci.* 15, 19406–19416. doi: 10.3390/ijms151119406
- Zhang, J., Liao, G., Liu, C., Sun, L., Liu, Y., Wang, Y., et al. (2011). The association of CLOCK gene T3111C polymorphism and hPER3 gene 54-nucleotide repeat polymorphism with Chinese Han people schizophrenics. *Mol. Biol. Rep.* 38, 349–354. doi: 10.1007/s11033-010-0114-2
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- The Review Editor PU and handling Editor AK declared their shared affiliation, and the handling Editor states that the process nevertheless met the standards of a fair and objective review.
- Copyright © 2016 Murphy and Benítez-Burraco. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.**



Flicker-Driven Responses in Visual Cortex Change during Matched-Frequency Transcranial Alternating Current Stimulation

Philipp Ruhnau^{1,2*†}, Christian Keitel^{3†}, Chrysa Lithari¹, Nathan Weisz¹ and Toralf Neuling¹

¹ Centre for Cognitive Neuroscience, University of Salzburg, Salzburg, Austria, ² Center for Mind/Brain Science, University of Trento, Mattarello, Italy, ³ Centre for Cognitive Neuroimaging, Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, UK

We tested a novel combination of two neuro-stimulation techniques, transcranial alternating current stimulation (tACS) and frequency tagging, that promises powerful paradigms to study the causal role of rhythmic brain activity in perception and cognition. Participants viewed a stimulus flickering at 7 or 11 Hz that elicited periodic brain activity, termed steady-state responses (SSRs), at the same temporal frequency and its higher order harmonics. Further, they received simultaneous tACS at 7 or 11 Hz that either matched or differed from the flicker frequency. Sham tACS served as a control condition. Recent advances in reconstructing cortical sources of oscillatory activity allowed us to measure SSRs during concurrent tACS, which is known to impose strong artifacts in magnetoencephalographic (MEG) recordings. For the first time, we were thus able to demonstrate immediate effects of tACS on SSR-indexed early visual processing. Our data suggest that tACS effects are largely frequency-specific and reveal a characteristic pattern of differential influences on the harmonic constituents of SSRs.

Keywords: alpha rhythm, brain oscillation, entrainment, frequency tagging, MEG, NIBS, steady-state response, tACS

OPEN ACCESS

Edited by:

Johanna Maria Rimmele,
Max-Planck-Institute for Empirical
Aesthetics, Germany

Reviewed by:

Till R. Schneider,
University Medical Center
Hamburg-Eppendorf, Germany
Viola Stoermer,
Harvard University, USA

*Correspondence:

Philipp Ruhnau
mail@philipp-ruhnau.de

[†]These authors have contributed
equally to this work.

Received: 22 December 2015
Accepted: 11 April 2016
Published: 26 April 2016

Citation:

Ruhnau P, Keitel C, Lithari C, Weisz N
and Neuling T (2016) Flicker-Driven
Responses in Visual Cortex Change
during Matched-Frequency
Transcranial Alternating
Current Stimulation.
Front. Hum. Neurosci. 10:184.
doi: 10.3389/fnhum.2016.00184

INTRODUCTION

Neural rhythms are prime candidates for a universal means of communication within and across brain regions and may code information from bits up to full objects (Engel et al., 2001; Buzsáki and Draguhn, 2004). A number of recent studies have thus attempted to entrain brain rhythms with external pacemakers by means of non-invasive brain stimulation (NIBS). An NIBS method widely applied in current cognitive neuroscience is transcranial alternating current stimulation (tACS; Thut et al., 2011; Antal and Paulus, 2013; Herrmann et al., 2013). Compared to classic electrophysiological research, tACS is in principle a more direct means to probe the role of brain oscillations in cognition: a strictly periodically alternating current is applied to modify brain rhythms directly that have been previously implicated with cognitive function. This way, different parameters of brain oscillations (e.g., amplitude, phase, frequency) become the independent variable and behavioral measures the dependent variable, which in turn allows for causal interpretations. Oscillations of various frequencies have been found to show tACS after-effects that appear brain state dependent. For instance, alpha band power (~ 10 Hz) was increased after 10 min of individual alpha frequency (IAF) stimulation (Zahle et al., 2010),

an effect lasting up to 30 min after stimulation (Neuling et al., 2013). tACS targeting different frequency bands and brain functions has also been shown to influence behavioral performance. As an example, stimulation within the theta frequency band (3–8 Hz) affects working memory performance (Polanía et al., 2012; Vosskuhl et al., 2015). Alpha tACS phase influences detection of near threshold stimuli in a phasic manner (Neuling et al., 2012a), while the IAF can be modulated by tACS, which in turn affects the multisensory double flash illusion (Cecere et al., 2015).

Although event-related activity and modulations in other frequencies has been successfully demonstrated during tACS using electroencephalography (EEG), attempts at investigating intrinsic brain oscillations at the stimulation frequency have proven to be challenging (Helfrich et al., 2014a,b; Voss et al., 2014). The main reason for this limitation is a heavy electrical artifact introduced by tACS that disallows analyses of spectral components of EEG/magnetoencephalographic (MEG) time series that are close to the stimulation frequency.

Recently, however, Soekadar et al. (2013) demonstrated that artifacts introduced by another NIBS method, transcranial direct current stimulation (tDCS), can be effectively suppressed by means of a beamformer source reconstruction of MEG sensor data. We successfully extended their approach to reconstruct brain activity during alpha-band tACS (Neuling et al., 2015). In that study we were able to demonstrate that two classes of mass neural activity, the parieto-occipital alpha rhythm and event-related responses, can be reconstructed from tACS-contaminated MEG-recorded data. Most importantly, in both cases the reconstructed activity was virtually identical with the same neural signal when no tACS was applied.

These advances generally allow an investigation of any oscillatory brain response during concurrent tACS. Here, we put our approach to a new test by probing online tACS effects on a special type of rhythmic brain activity known as steady-state responses (SSRs) that are driven by, and thus strictly time-locked, to periodic visual flicker stimulation (Regan, 1989; Norcia et al., 2015).

SSRs have been studied since the early days of EEG research (Adrian and Matthews, 1934). To date, their exact neurophysiological origin is still under debate (Keitel et al., 2014). Whereas some researchers treat SSRs as externally entrained intrinsic neural rhythms, such as the alpha rhythm (Mathewson et al., 2012; Spaak et al., 2014), others suggest that they mainly compose of successive transient sensory evoked responses that add to the ongoing electrophysiological signal (Shah et al., 2004; Capilla et al., 2011). For the purpose of the present study we refrain from endorsing either perspective but simply treat SSRs as stimulus-driven brain oscillations with unique properties outlined below that make them an ideal candidate for a combination with tACS research.

In the spectral domain SSRs can be considered narrow-band responses whose bandwidth can be neglected when considering multiple SSR cycles, i.e., longer stimulation periods. SSRs further comprise a number of (equally narrow-band)

higher order harmonics, i.e., spectral components at multiples of the driving frequency that are typically found in frequency-tagging experiments (Appelbaum et al., 2006; Kim et al., 2011; Porcu et al., 2013) and point towards non-linear properties of the visual system (Roberts and Robinson, 2012). A body of research on visual processing has employed SSRs to study, for instance, attentional influences (Müller and Hillyard, 2000; Kim et al., 2007; Störmer and Alvarez, 2014), cognitive load (Jacoby et al., 2012), perceptual segregation (Appelbaum et al., 2006), the aging brain (Quigley and Müller, 2014), inter-stimulus competition (Porcu et al., 2014), as well as object- (Kaspar et al., 2010; Koenig-Robert and VanRullen, 2013), and face processing (Rossion and Boremanse, 2011; Rossion et al., 2012).

In comparison with relatively broadband intrinsic rhythms that are typically targeted in tACS experiments, such as alpha (8–13 Hz), SSRs may be a better fit to the strictly sinusoidally alternating current and the implicit underlying stationarity assumption of cortical oscillations. Online effects of tACS on stimulus-driven oscillatory responses might be more readily observable because the spectral profile of the stimulation, and thus in principle the resulting waveforms, are precisely set by the experimenter.

In the present study we administered tACS while concurrently recording SSRs. To this end, we developed a novel protocol that was optimized to deliver tACS in 2 s intervals concurrently with matched- and different-frequency visual flicker in a classical trial-by-trial experimental paradigm. Based on our previous success in recovering the alpha rhythm during alpha band tACS by means of a beamformer source projection (Neuling et al., 2015), we expected a similar outcome with regard to SSRs during tACS. In line with studies showing alpha power increases after alpha band tACS (Zaehle et al., 2010; Neuling et al., 2013; Vossen et al., 2015), we hypothesized that matching flicker and tACS frequency would lead to pronounced SSR power. The latter hypothesis further entailed the assumption that no effects would be observed when flicker and tACS frequencies did not match.

MATERIALS AND METHODS

Participants

Seventeen healthy participants volunteered for the current study (4 female, mean age 26 years, one left handed). Two had to be excluded due to hardware problems with the stimulation setup, resulting in a final group of 15 analyzed subjects (4 female, 25.5 years, one left handed). The experiment was approved by the local ethics committee of the University of Trento and adhered to the tenets of the Declaration of Helsinki. All participants signed an informed consent prior to the beginning of the experiment.

Visual Stimuli

Participants viewed experimental stimuli back-projected on a translucent screen by a Propixx DLP projector

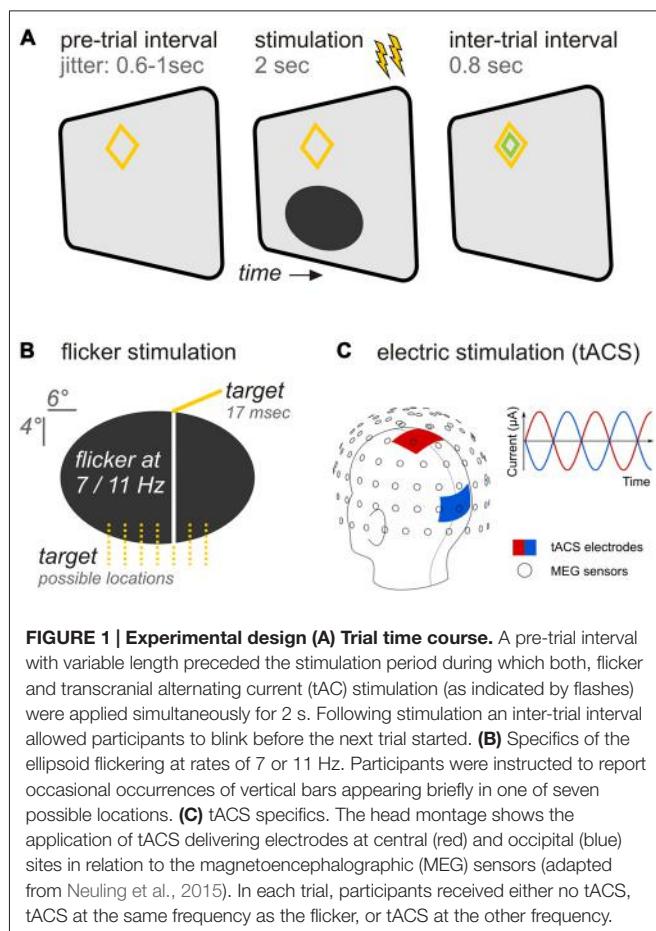


FIGURE 1 | Experimental design (A) Trial time course. A pre-trial interval with variable length preceded the stimulation period during which both, flicker and transcranial alternating current (tAC) stimulation (as indicated by flashes) were applied simultaneously for 2 s. Following stimulation an inter-trial interval allowed participants to blink before the next trial started. **(B)** Specifics of the ellipsoid flickering at rates of 7 or 11 Hz. Participants were instructed to report occasional occurrences of vertical bars appearing briefly in one of seven possible locations. **(C)** tACS specifics. The head montage shows the application of tACS delivering electrodes at central (red) and occipital (blue) sites in relation to the magnetoencephalographic (MEG) sensors (adapted from Neuling et al., 2015). In each trial, participants received either no tACS, tACS at the same frequency as the flicker, or tACS at the other frequency.

(VPixx technologies, Saint-Bruno, Canada), employing a refresh rate of 120 frames per second and a resolution of 1920×1080 pixel (width \times height). The stimulation comprised an ellipse (horizontal/vertical diameter = $6.6^\circ/3.3^\circ$ of visual angle) positioned in the lower visual field at a center-to-center eccentricity of 3° below fixation (Figures 1A,B). A diamond shape (maximum eccentricity = 0.9°) served as a central fixation point. Stimuli were presented against a gray background (RGB: 85, 85, 85).

The ellipse underwent periodic luminance changes (= *flicker*) at rates of either 7 or 11 Hz in the course of each trial: Relative luminance to background oscillated between a minimum of 0% (total black, RGB: 0,0,0) and a maximum of 100% (background gray). Ellipse luminance changed in small increments on each presentation frame to approximate sinusoidal modulations.

We chose our two frequencies within a range that is typically used in frequency-tagging experiments (see Norcia et al., 2015). Both frequencies were hence known to produce SSRs of high signal-to-noise ratios. Further vital to the design of our study was that 7 and 11 Hz SSRs did not produce harmonics that coincided spectrally within the range of frequencies that we analyzed (<50 Hz).

TACS Parameters

A battery-operated stimulator system connected to rubber stimulation electrodes (DC-Stimulator Plus, NeuroConn GmbH, Ilmenau, Germany) controlled by the stimulation computer was placed outside the magnetically shielded room. It was connected to the stimulation electrodes inside the MEG cabin via the magnetic resonance imaging (MRI) module (NeuroConn). Using the remote input of the stimulator to control the stimulation signal on a trial-by-trial basis, an alternating, sinusoidal current at either 7 or 11 Hz was delivered for 2 s. Stimulation electrodes were centered at electrode positions Cz and Oz of the international 10–20 system (Figure 1C). These positions were chosen for maximal stimulation intensity in the parieto-occipital cortex (Neuling et al., 2012b). The electrodes had a size of 7 by 5 cm and were attached to each participant's head with a conductive paste (Ten20, D.O. Weaver, Aurora, CO, USA) resulting in impedance values lower than $10\text{ k}\Omega$. The electrode cables were located on the left side of the participant's head. To keep participants unaware of the electrical stimulation during the experiment, the stimulation intensity was kept below each participant's sensation and phosphene threshold. To obtain the threshold, the participants were first familiarized with the skin sensation. Afterwards, an intensity of $400\text{ }\mu\text{A}$ (peak-to-peak) was applied at 7 Hz for 30 cycles (4.29 s). Intensity was increased in steps of $100\text{ }\mu\text{A}$ until participants indicated skin sensation or phosphene perception or an intensity of $1500\text{ }\mu\text{A}$ was reached. In the five cases in which the participant already reported a skin sensation at $400\text{ }\mu\text{A}$, the intensity was reduced to a start level of $100\text{ }\mu\text{A}$. The staircase procedure resulted in stimulation intensities of $M = 613$ SD $128\text{ }\mu\text{A}$. The net tACS stimulation time during the experiment was 10 min 40 s (5'20" for each stimulation frequency) when summing individual trial stimulation (2 s each).

Note that an inherent difficulty in combining tACS and SSRs lies in the fact that measured effects may depend on the phase relationship of both types of stimulation. Starting electrical and visual stimulation simultaneously and in phase will inadvertently lead to a phase lag in the periodic modulation of neural activity induced by the two methods: whereas electrically induced oscillations will likely have a near-zero phase lag with regard to the driving tACS (Fröhlich and McCormick, 2010; Reato et al., 2010), SSRs will show a substantial phase lag relative to the driving flicker stimulation that depends on the synaptic conduction delays of the visual system from eyes to visual cortex. In the present study, we neglected this tACS-SSR phase lag because: (1) to date, it needs to be shown that an SSR phase lag relative to flicker stimulation can be reliably estimated and remains constant during concurrent tACS; and (2) it went past the scope of our study, namely, demonstrating the feasibility of reconstructing SSRs from MEG recordings contaminated with tACS artifacts at identical frequencies.

Procedure and Task

We manipulated the factors *ellipse flicker frequency* (7 vs. 11 Hz) and *tACS frequency* (7 vs. 11 Hz) in a fully balanced

design. For both flicker frequencies, a sham condition (no-tACS) served as a control condition: while all other parameters remained constant, the stimulator did not receive a signal in the sham tACS trials. Trials of the resulting six conditions were presented in a pseudo-randomized order. In total, each participant ran 480 trials (= 80 trials per condition) divided into eight blocks (~5 min each), separated by self-paced breaks.

During the experiment participants were seated in a comfortable chair and directed gaze towards a screen positioned 1 m in front of them. Experimental trials started with ellipse onset. During the following 2 s the ellipse flickered at a constant rate of either 7 or 11 Hz, dependent on experimental condition. At the end of each trial, a smaller green diamond appeared within the orange fixation diamond for 800 ms indicating participants a favorable time-range to blink before the next trial started (**Figure 1A**).

Participants were instructed to press a button with the right hand after occasional brief occurrences (16.6 ms/2 frames) of a vertical line superimposed on the ellipse at one of seven pseudo-randomly chosen locations (**Figure 1B**). Target events appeared in 40% of all trials and, if so, once per trial at a pseudo-randomly chosen time point within an interval starting 500 ms after ellipse onset and ending two frames before stimulus offset. Responses were recorded with an MRI compatible response collector (RESPONSEPpixx, VPixx technologies, Saint-Bruno, Canada).

Prior to the main experiment, participants performed at least one prolonged training block (~10 min). After each block, participants received feedback regarding average task performance in terms of hit rate and response speed.

MEG Data Recording

Electrophysiological data were recorded using a whole head Elekta Neuromag MEG (ElektaOy, Helsinki, Finland) placed in a magnetically shielded room (Vacuumschmelze, Hanau, Germany). Magnetic brain activity was recorded from 102 positions above the head, each comprising a sensor triplet (one magnetometer, two orthogonal planar gradiometers) and sampled at 1000 Hz with an on-line band-pass filter (0.1–330 Hz) active. Before the experiment individual head shapes were acquired for each participant, including fiducials (nasion, left/right pre-auricular point), and around 200 digitized points on the scalp acquired with a Polhemus Fastrak digitizer (Polhemus, VT, USA). During the recording five head position indicator coils (HPIs) tracked the position of the participants' head.

MEG Data Analysis

Continuous data were high-pass filtered off-line (Finite Impulse Response (FIR), Kaiser window, cut-off 1 Hz, pass-band 2 Hz) and down-sampled to 512 Hz. Then, epochs of 4 s were cut out, starting 1 s before and ending 3 s after flicker onset. Epochs without tACS stimulation were visually inspected to identify flat or noisy channels

as well as epochs containing physiological artifacts (e.g., caused by blinks or muscle activity). Bad channels identified in these trials were excluded from the whole data set.

Because tACS creates a massive electro-magnetic artifact, several orders of magnitude larger than the brain signal (see Neuling et al., 2015), sensor space epochs were projected into source space using linearly constrained minimum variance (LCMV) beamformer filters (Van Veen et al., 1997) before further analyses. To do this, we followed a procedure described here for individual virtual sensors¹ and extended it to an equally spaced 1.5 cm grid covering the whole brain (see also Neuling et al., 2015, for a similar procedure).

In short, epochs were low-pass filtered at 45 Hz and single epoch covariances estimated and averaged. With the help of the acquired head shapes (see above), individual subjects' structural magnetic resonance images were aligned to the MEG space, which was subsequently used to create single-shell head models (Nolte, 2003) and lead field matrices. The average covariance, head model, and lead field matrix were used to obtain beamformer filters. This was done separately for each tACS condition—no tACS, 7 Hz tACS, 11 Hz tACS—to optimize the suppression of the artifact. The filters were subsequently multiplied with the individual epochs resulting in source level epochs. We used a 1.5 cm equally spaced grid (889 grid points covering the brain) in Montreal Neurological Institute (MNI) space and warped these positions into individual headspace, which allowed us to average and compute statistics across participants without further interpolation.

Spectral Analysis

We analyzed SSRs in the frequency domain using two complementary approaches. First, and in accordance with typical SSR analyses (e.g., Appelbaum et al., 2006; Andersen and Müller, 2010), source-level time series were averaged for each participant and condition separately. Fast Fourier Transforms (MATLAB function *fft*) of averaged data within an interval of 0.5–1.5 s relative to SSR onset² yielded complex spectra. Power spectra were obtained by squaring the absolute values of the complex Fourier coefficients. Statistical analyses were performed on SSR amplitudes (square-root of SSR power) divided by the individual mean amplitude across conditions for each frequency. This normalization procedure removed the substantial inter-individual variance in absolute SSR amplitude while retaining the net effects of tACS.

Secondly, we estimated phase locking values (PLV, also referred to as inter-trial phase coherence; Lachaux et al., 1999)

¹http://www.fieldtriptoolbox.org/tutorial/shared/virtual_sensors

²The steady-state response has a certain build-up time and is initially overlapped by the event-related response. To get an estimate of the “true” steady-state the initial part of the epoch is typically ignored (e.g., Regan, 1989; Kim et al., 2011; Keitel et al., 2013). Similar reasoning holds for the late part of the epoch where offset responses can contaminate the frequency estimate.

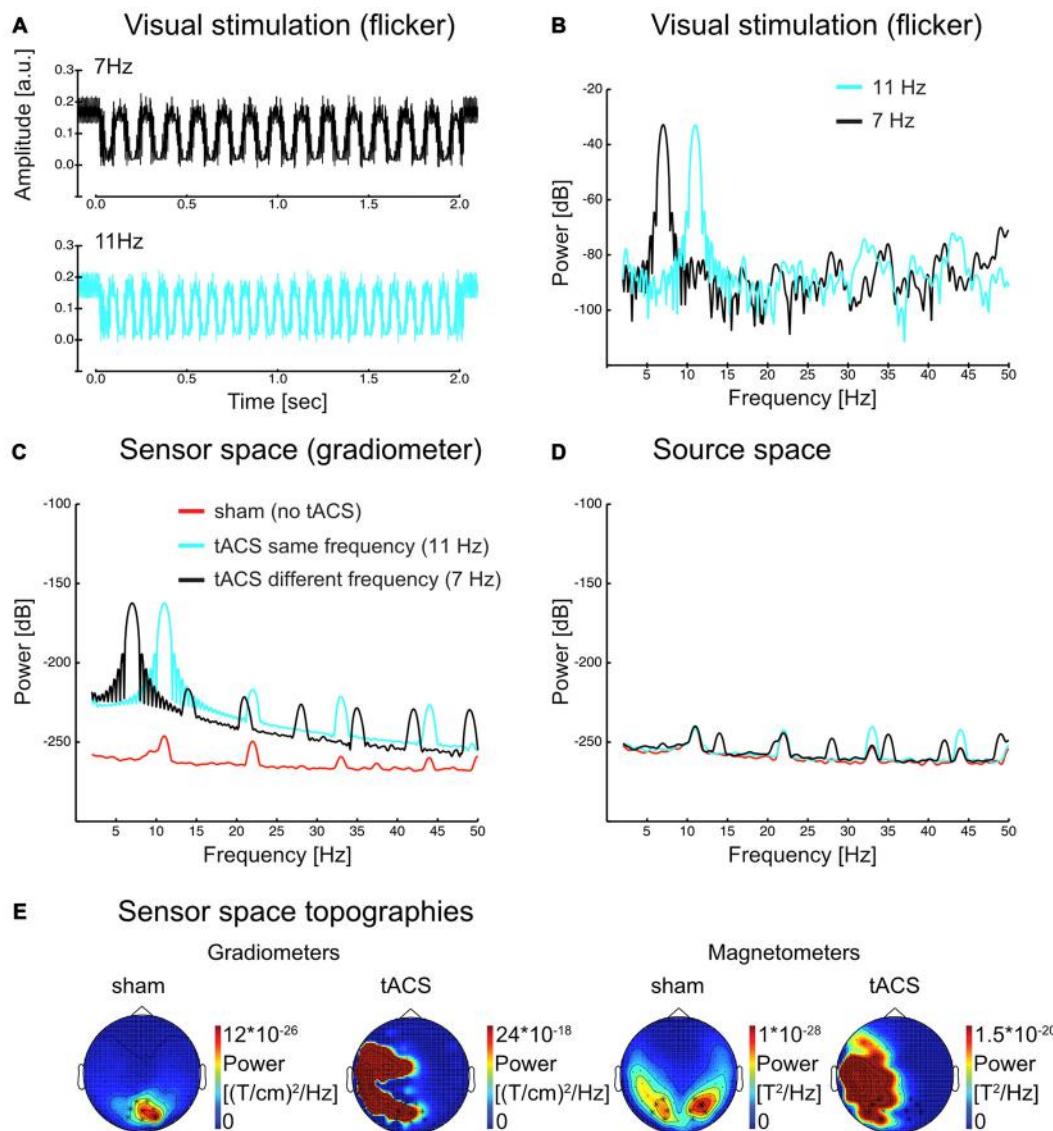


FIGURE 2 | (A) Visual flicker as recorded with a photo diode (average over 50 trials in each flicker frequency) and **(B)** power spectrum of the same signal for each flicker frequency. **(C)** Sensor space (gradiometer) MEG data power spectrum in sham and tACS conditions for 11 Hz steady-state response (SSR). Harmonic activity is visible, yet considerably reduced compared to the fundamental. **(D)** Source space data of the same conditions as in **(C)** in visual cortex regions (see **Figure 4** for the region of interest). The scale in **(B–D)** is dB ($10^{\log 10(\text{power})}$). **(E)** Sensor space topographies. SSR topographies show occipital activity at the flicker frequency (11 Hz) in the sham condition. During tACS, the artifact is dominating the signal. Note [that] the left lateralized activity is a result of the tACS cables, which were placed on the left side of the participant's head. Scales change drastically from sham to tACS (factor of around 10^7 – 10^8).

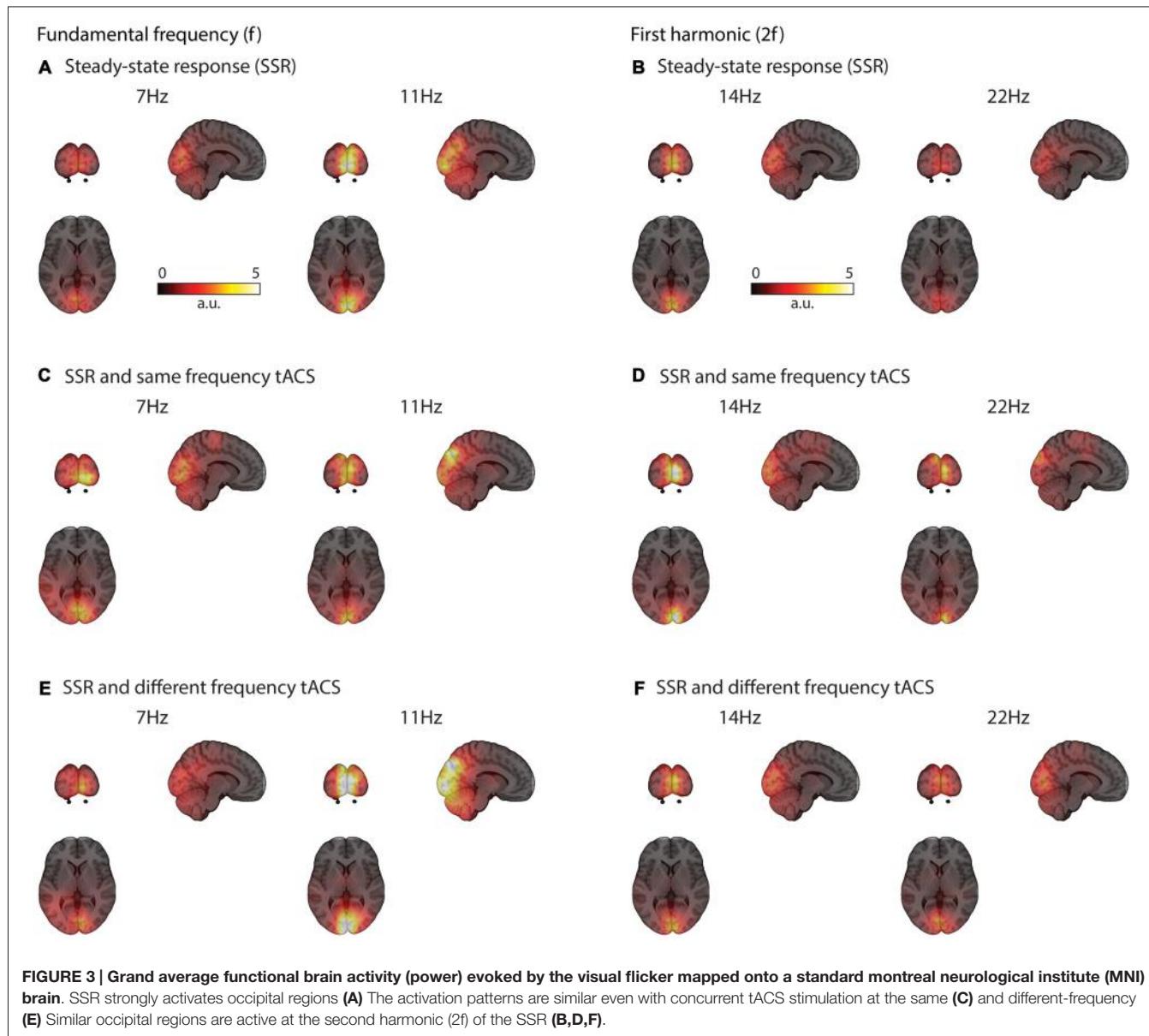
for each condition by Fourier-transforming individual epochs first (again selecting data within an interval of 0.5–1.5 s post SSR onset), and then taking the absolute value of the complex mean of the Fourier coefficients for each condition, normalized to unit length:

$$PLV(f) = \left| \frac{1}{N} \sum_{n=1}^N \frac{c_n(f)}{|c_n(f)|} \right|$$

where $c_n(f)$ is the complex Fourier coefficient of trial n at frequency f . Phase locking (= phase synchrony) as a

measure of SSR modulation has been introduced to SSR analyses more recently (e.g., Kim et al., 2007). Previous findings indicate differential sensitivities of SSR amplitude and phase synchrony to top-down influences on sensory processing (Kashiwase et al., 2012; Porcu et al., 2013). We thus included SSR phase synchrony to provide a comprehensive description of SSR modulation by concurrent tACS.

In both analyses, we investigated frequencies from 2–50 Hz. The data were zero padded to a length of 8 s to achieve a 0.125 Hz frequency resolution.



Statistical Analysis

For the behavioral data, responses were considered a “hit” when a button press occurred between 200–1200 ms after target onset. When participants responded in the absence of target presentations responses were classified as false alarms. Behavioral data analyses revealed that participants produced only few false alarms on average (1.3 ± 0.2 per condition). Thus, we based statistical analyses on hit rates (= number of hits divided by total number of targets per condition). Individual hit rates were subjected to a two-way repeated measures analysis of variances (ANOVA) with factors of *flicker frequency* (7 Hz; 11 Hz) and *tACS frequency* relative to flicker frequency (no tACS; same; different).

Reaction times (RTs) of correct responses were analyzed accordingly. Note that RT analyses were based on median RTs

per participant and condition to account for the typical left skew of RT distributions.

Spectral source space MEG data (power and PLV) were analyzed with a 3-way repeated-measures ANOVA comprised of the factors *flicker frequency* (7 Hz; 11 Hz), *tACS frequency* relative to flicker frequency (no tACS; same; different) and *SSR harmonic* (fundamental [f]; 2f; 3f; 4f). Amplitudes were normalized per participant and frequency by dividing them by the mean of the three tACS conditions to reduce individual SSR amplitude variability. For all significant main effects and interactions, probabilities were corrected to control for sphericity violations by adjusting the degrees of freedom (Greenhouse and Geisser, 1959). We report original degrees of freedom, corrected *p*-values (p_{GG}) and the correction coefficient epsilon (ϵ_{GG}). *Post hoc* tests were conducted where appropriate and

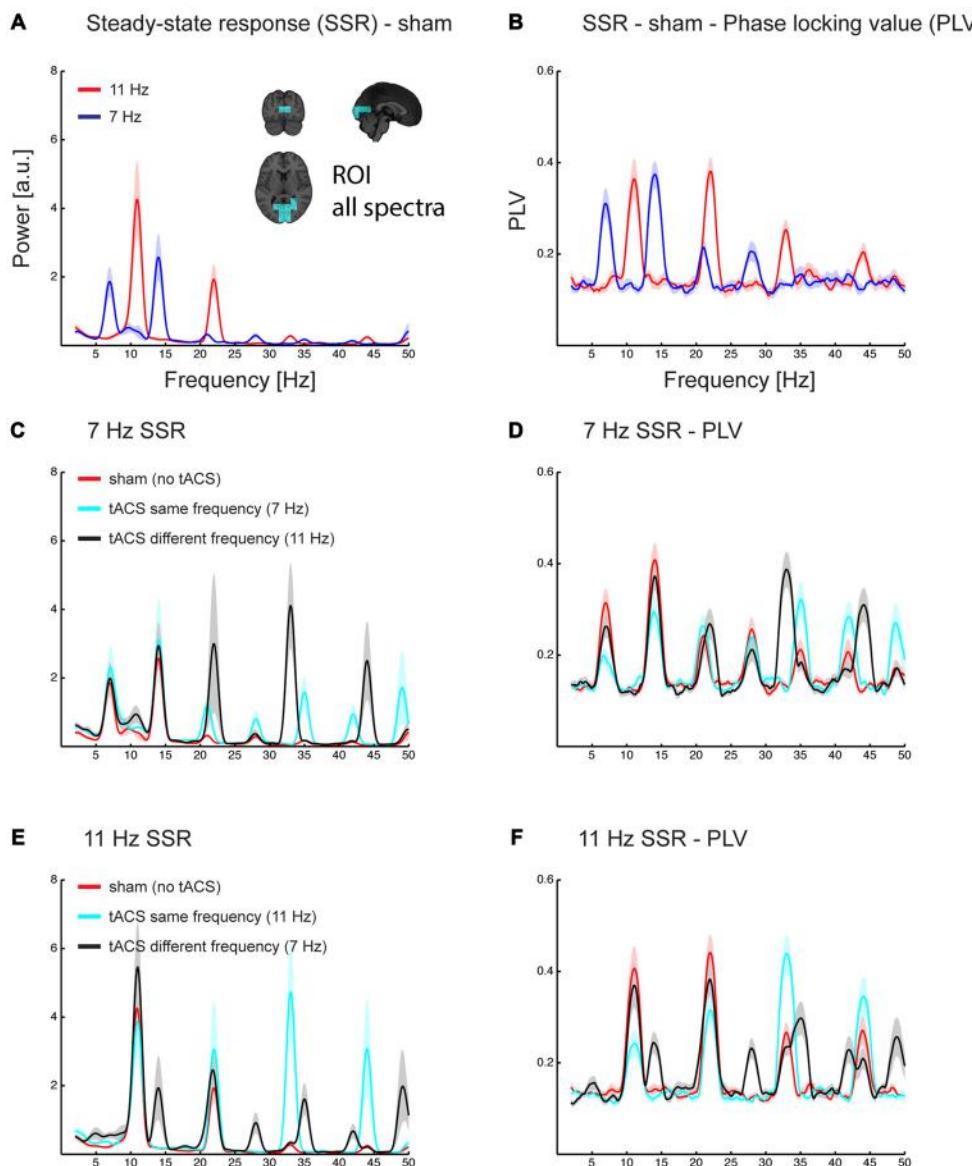


FIGURE 4 | Grand average spectra of evoked power (A,C,E) and phase locking value (PLV) across trials (B,D,F). All spectra are the average of estimates in a region of interest covering early sensory visual cortex areas along the Calcarine fissure (see inlet in **A**). Power and PLV are shown for 7/11 Hz SSR without applying tACS (**A,B**) and for all three tACS conditions (no-tACS, same-frequency tACS, and different-frequency tACS) for 7 Hz (**C,D**) and 11 Hz SSR (**E,F**). Note that tACS seems to differently affect the magnitude for the fundamental and harmonics frequencies. Shaded areas represent standard error of the mean.

controlled for multiple comparisons using the false discovery rate (FDR) procedure across all analyses (Benjamini and Hochberg, 1995).

RESULTS

Behavioral Measures Independent of tACS Manipulation

Participants detected target events (briefly flashed vertical lines) with comparable hit rates (Table 1) on

ellipses flickering at 7 and 11 Hz (main effect *flicker frequency*: $F_{(1,14)} = 2.14$; $p > 0.05$) and different frequencies of simultaneously administered tACS (main effect *tACS frequency*: $F_{(2,28)} = 1.47$, $p_{GG} > 0.05$, $\varepsilon_{GG} = 0.890$). Reaction time analyses revealed a similar pattern: Neither ellipse *flicker frequency* ($F_{(1,14)} = 0.04$, $p > 0.05$) nor *tACS frequency* ($F_{(2,28)} = 0.49$, $p_{GG} > 0.05$, $\varepsilon_{GG} = 0.822$) influenced response speed (Table 1).

In both analyses, interactions of the factors *flicker frequency* and *tACS frequency* were insignificant (F 's < 1).

TABLE 1 | Average behavioral performance in the visual detection task ($N = 15$).

SSR frequency	7 Hz			11 Hz			
	No	Same	Diff	No	Same	Diff	
Hit rate (%)	M	68.1	69.4	71.7	66.0	67.1	69.2
	$\pm SEM$	2.2	3.2	3.2	2.5	3.3	2.4
Reaction time (ms)	M	437	428	426	429	432	427
	$\pm SEM$	14	11	13	10	12	13

M = mean, SEM = standard error of the mean. *Relative to flicker frequency.

Sensor Level Data Cannot be Analyzed Because of the tACS Artifact

Figures 2A–C illustrate that visual and electrical stimulation signals were dominated by strong fundamental frequency components indicating that both signals were principally sinusoidal (Power spectra in Figure 2 were acquired in a similar manner as for the source space time series, see Materials and Methods section “Spectral analysis”). As Figure 2C demonstrates, the tACS artifact dominated the spectrum at the sensor level and made an analysis of the interaction of SSR and tACS impossible. Source reconstruction by means of LCMV beamforming however suppressed the artifact: in the spectrum in Figure 2D peaks corresponding to the stimulation frequencies were of similar magnitude (compare with Figure 2C). Scalp maps in Figure 2E give an impression of the topographical distribution of the tACS artifact at the fundamental frequency (exemplarily shown here for 11 Hz). Note the massive differences in topography and scale between sham (i.e., SSR only) and tACS conditions. Further note the lateralized topographies during tACS that were caused by currents in the electrode cables fastened to the left side of the participants’ head.

Interestingly, especially in case of tACS the electrical artifact picked up at the sensor level (Figure 2C) also contained higher order harmonic components. These harmonics were several orders of magnitude smaller than the driving frequency component ($\sim 60 \text{ dB} = 40:1$). In source-projected data, however, fundamental and harmonic responses were of similar magnitude (Figure 2D). Ultimately, our experiment alone did not allow a further investigation into whether it was the minute stimulation of harmonic components itself or non-linear responses to tACS at the fundamental frequency in the brain that gave rise to neural harmonics (as proposed for SSRs, see Roberts and Robinson, 2012). Considerable tACS harmonics in artifact-removed source reconstructions speak for the latter option, nevertheless. Given the data at hand, in the following, we regard them as genuine brain responses in either case.

Visual SSRs can be Reconstructed Even with tACS at the Same Frequency

Visual flicker drives brain response at the stimulation frequency and also at harmonics mainly in early visual areas (see Figures 3–5). These responses could be clearly reconstructed with concurrent same- and different-frequency tACS. The neural

sources of the SSR were localized to highly comparable regions on the occipital pole (Figure 3).

As mentioned above, spectra of source reconstructed oscillatory activity contained fundamental and harmonic responses elicited by tACS (clearly visible in spectra of conditions in which flicker and tACS frequencies differed; see Figures 4C–F).

SSR Power and Phase Locking—tACS Affects Fundamental and Harmonic Frequencies Differently

SSR Amplitude

An ANOVA, comprised of the factors *flicker frequency* (7 Hz; 11 Hz), *tACS frequency* relative to flicker frequency (no tACS; same; different) and *SSR harmonic* (fundamental [f]; 2f; 3f; 4f) revealed the following effects: a significant main effect of tACS frequency ($F_{(2,28)} = 26.91$, $p_{\text{GG}} < 0.001$, $\epsilon_{\text{GG}} = 0.718$), caused by larger amplitudes in the same-frequency tACS condition compared to no- and different-frequency tACS ($p_{\text{FDR}} < 0.05$), while there were no significant differences between no and different-frequency tACS. Furthermore, a *flicker frequency* \times *tACS frequency* interaction was significant ($F_{(2,28)} = 16.27$, $p_{\text{GG}} < 0.001$, $\epsilon_{\text{GG}} = 0.743$), explained by the fact that at 7 Hz visual flicker (pooled across harmonics) no-tACS showed smaller amplitudes than same-frequency tACS ($p_{\text{FDR}} < 0.05$) but no other significant differences were found, while at 11 Hz both no- and different-frequency tACS were significantly smaller in amplitude than same-frequency tACS (all $p_{\text{FDR}} < 0.05$). Furthermore, an *SSR harmonic* \times *tACS frequency* interaction was significant ($F_{(6,84)} = 36.40$, $p_{\text{GG}} < 0.001$, $\epsilon_{\text{GG}} = 0.564$) caused by tACS frequency effects at 3f and 4f, with larger amplitudes at same-frequency tACS compared to no- and different-frequency tACS (all $p_{\text{FDR}} < 0.05$), while there were no significant tACS frequency effects at the fundamental and 2f (all $p_{\text{FDR}} > 0.05$). This pattern was more pronounced with 11 Hz compared to 7 Hz visual flicker (see Figure 5), which resulted in a significant 3-way interaction ($F_{(6,84)} = 5.35$, $p_{\text{GG}} = 0.002$, $\epsilon_{\text{GG}} = 0.604$). This was evident in larger differences of no- and different-frequency tACS compared to same-frequency tACS in 3f and 4f (7 vs. 11 Hz, all $p_{\text{FDR}} < 0.05$), while contrasts between tACS frequency differences revealed no significant effects at the fundamental and 2f (all $p_{\text{FDR}} > 0.05$).

SSR Phase Locking

A similar three-way ANOVA on SSR phase locking revealed main effects of *flicker frequency* ($F_{(1,14)} = 12.73$, $p = 0.003$), caused by larger PLVs for 11 Hz compared to 7 Hz, and a main effect of *SSR harmonic* ($F_{(3,42)} = 16.31$, $p_{\text{GG}} < 0.001$, $\epsilon_{\text{GG}} = 0.875$), caused by largest PLVs at 2f followed by the fundamental ($p_{\text{FDR}} < 0.05$) and 3f ($p_{\text{FDR}} < 0.001$) and smallest PLVs at 4f (all $p_{\text{FDR}} < 0.05$). Furthermore, the *SSR harmonics* \times *tACS frequency* interaction ($F_{(6,84)} = 35.70$, $p_{\text{GG}} < 0.001$, $\epsilon_{\text{GG}} = 0.562$) and the *flicker frequency* \times *tACS frequency* interaction ($F_{(2,28)} = 7.83$, $p_{\text{GG}} = 0.003$, $\epsilon_{\text{GG}} = 0.954$) was significant, yet they were further explained by the significant three-way interaction ($F_{(6,84)} = 5.06$,

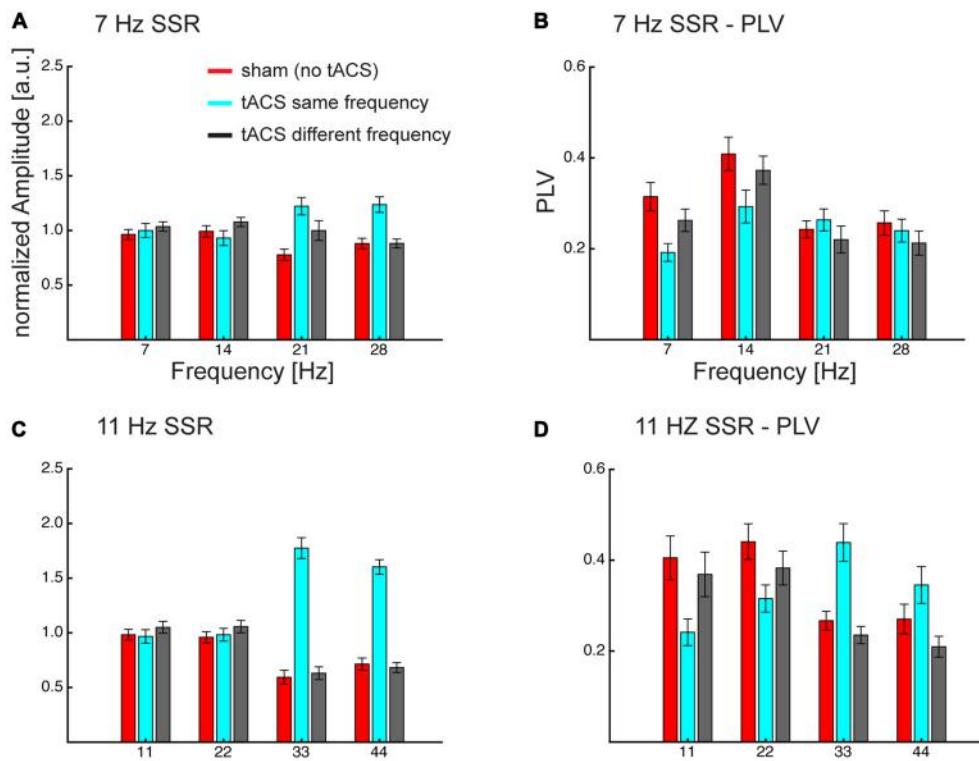


FIGURE 5 | Grand average normalized amplitudes (A,C) and PLV (B,D) at the visual flicker driven fundamental ($f = 7/11$ Hz) and 3 harmonic (2f, 3f, 4f) frequencies. Frequency specificity of tACS can be seen in amplitude and PLV, also tACS seems to affect different parts of the flicker driven brain signal. Both measures show that same-frequency tACS is strongly enhancing the harmonics (3f, 4f) and PLV additionally shows a decrease in the lower harmonics. Overall this pattern seems to be more pronounced for 11 Hz SSRs. Note that neither fundamental nor harmonics seem to be affected when the tACS frequency differs from the flicker/SSR frequency. Error bars represent standard error of the mean.

$p_{GG} = 0.003$, $\epsilon_{GG} = 0.583$). No other effect was significant (all $F < 2.4$, $p > 0.095$).

To resolve the three-way interaction, we conducted two-way ANOVAs on the individual frequencies.

For 7 Hz the ANOVA revealed a significant main effect of tACS frequency ($F_{(2,28)} = 6.89$, $p_{GG} = 0.005$, $\epsilon_{GG} = 0.888$) and of SSR harmonic ($F_{(3,42)} = 14.44$, $p_{GG} < 0.001$, $\epsilon_{GG} = 0.873$). Furthermore, the interaction was significant ($F_{(6,84)} = 9.13$, $p_{GG} < 0.001$, $\epsilon_{GG} = 0.742$). Post hoc tests showed harmonic dependent tACS frequency effects: at the fundamental response the no-tACS condition yielded larger PLVs than same-frequency tACS ($p_{FDR} < 0.01$) and different-frequency tACS ($p_{FDR} < 0.05$) and different-frequency tACS yielded larger PLVs than same-frequency tACS ($p_{FDR} < 0.01$). At 2f no-tACS and different-frequency tACS yielded similar PLVs ($p_{FDR} > 0.05$) but both yielded larger PLVs than same-frequency tACS ($p_{FDR} < 0.01$). At 3f the pattern reversed and showed smaller PLVs for no-tACS compared to same-frequency tACS ($p_{FDR} < 0.05$), but there were no differences between no-tACS and different-frequency tACS and same- and different-frequency tACS (all $p_{FDR} > 0.05$). At 4f there were no differences (all $p_{FDR} > 0.05$).

For 11 Hz the ANOVA revealed a main effect of SSR harmonic ($F_{(3,42)} = 6.39$, $p_{GG} = 0.006$, $\epsilon_{GG} = 0.623$) and a significant tACS frequency \times SSR harmonics interaction

($F_{(6,84)} = 27.55$, $p_{GG} < 0.001$, $\epsilon_{GG} = 0.453$). This interaction was caused by a difference in the overall patterns of tACS effects on SSR harmonics: for the fundamental frequency there were no differences between no- and different-frequency tACS ($p_{FDR} > 0.05$) but both showed larger PLVs than the same-frequency tACS condition (all $p_{FDR} < 0.01$). At 2f no-tACS still showed larger PLVs than same-frequency tACS ($p_{FDR} < 0.05$) but no other comparisons were significant. At 3f the pattern observed at the fundamental inversed; although there were no differences between no- and different-frequency tACS ($p_{FDR} > 0.05$), both showed smaller PLVs than the same-frequency tACS condition (all $p_{FDR} > 0.01$). At 4f, no-tACS showed smaller PLVs than same ($p_{FDR} > 0.01$) and larger PLVs than different-frequency tACS ($p_{FDR} > 0.05$), furthermore, same-frequency tACS showed larger PLVs than different-frequency tACS ($p_{FDR} < 0.01$).

TACS Alters SSR Waveform—An Example

To visualize the specific effects of same-frequency tACS and to illustrate the differential contribution of fundamental and harmonic components to the time domain signal we reconstructed time series waveforms from source-level spectral SSR representations. To this end, we summed the sinusoids described by the (amplitude and phase of) Fourier coefficients

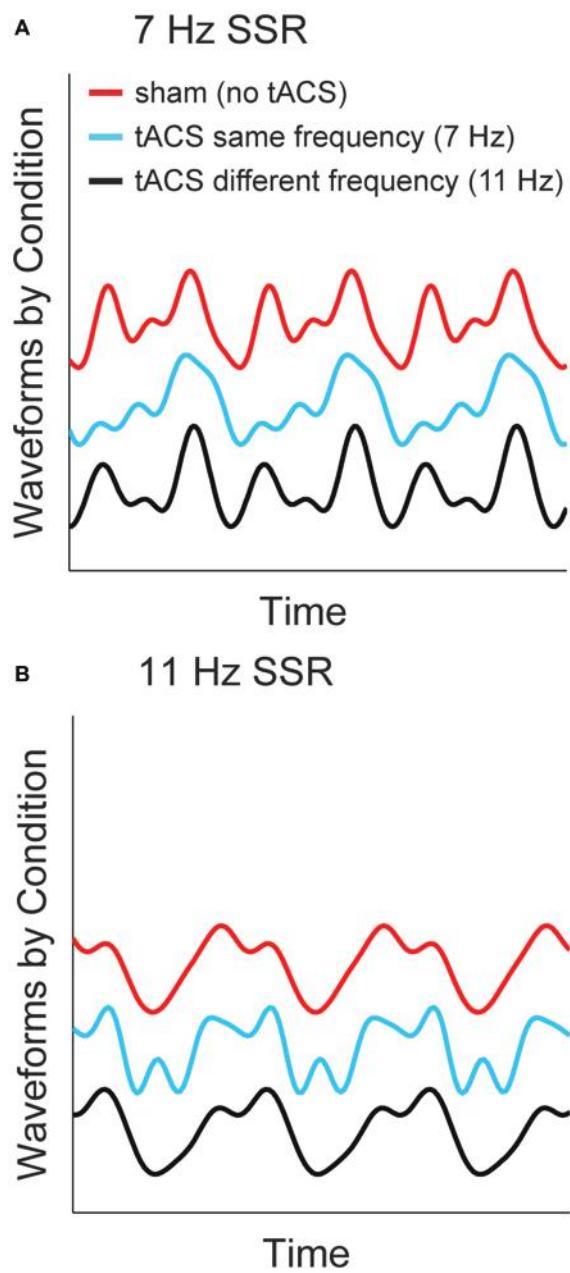


FIGURE 6 | SSR waveforms reconstructed from spectral components at fundamental and harmonic frequencies. Graphs show three cycles of the fundamental response of one representative subject for both stimulation frequencies **(A)** 7 Hz SSR **(B)** 11 Hz SSR. Line colors code the three tACS conditions. Note that the shift along the y-axis is non-informative and only serves illustrative purposes.

at fundamental frequencies and the harmonics up to 4f averaged across voxels in an early visual cortex region of interest (ROI, see **Figure 4A**). Respective complex Fourier coefficients were derived as described above in the analysis of SSR power (see “Materials and Methods” section “Spectral analysis”). **Figure 6** depicts reconstructed waveforms of one representative subject that correspond to three cycles of the respective fundamentals

for the three conditions: no-tACS, same-frequency tACS and different-frequency tACS. The no-tACS condition clearly shows the quasi-sinusoidal morphology that gives rise to strong higher order harmonics in spectral decompositions. Whereas no-tACS and different-frequency tACS waveforms show a strong resemblance, same-frequency tACS has a specific influence on SSR morphology.

DISCUSSION

The present study shows that: (1) recording SSRs in MEG during concurrent tACS, and thus a combination of both methods of brain stimulation, is feasible. To this end, we have implemented a novel tACS protocol that allows intermittent stimulation with frequencies varying in a classical trial-by-trial experimental design; (2) thus recorded SSRs can be reconstructed at the source level by means of LCMV beamforming that effectively removes tACS-introduced artifacts. Importantly, this procedure yields plausible results even when SSR and tACS have identical temporal frequencies; and (3) simultaneous tACS modulates SSRs in a frequency-specific manner: for both stimulation frequencies tested (7 and 11 Hz), same-frequency tACS had the most profound effect on SSRs. The effects of different-frequency tACS instead were largely comparable to a control condition in which no tACS was administered.

In the following we discuss these findings in detail, expose outstanding questions and issues and introduce possible future directions regarding experimental applications.

Combining Two Rhythmic Stimulation Methods

The first aim of the current study was to provide a proof of principle that brain activity evoked by a rhythmic visual stimulation can be reconstructed with concurrent tACS at the same frequency. Using LCMV beamformers on concurrent MEG-tACS data achieved this aim. As pointed out by Van Veen et al. (1997), the beamformer source reconstruction reduced highly correlated noise thus suppressing the massive tACS sensor artifact (Neuling et al., 2015). Crucially, spatially circumscribed generators of SSRs in early visual cortices remained unaffected (see **Figure 3**) independently of the applied tACS frequency. This is particularly remarkable in case of matched-frequency tACS because the artifact removal via beamforming could have resulted in a suppression (if not removal) of SSR power itself.

Typically, in electrical stimulation designs the stimulation is applied for a longer period of time (e.g., Zaehle et al., 2010; Neuling et al., 2013; Helfrich et al., 2014b) to yield stable after-effects (but see Vossen et al., 2015). Here, we investigated direct effects of short 2 s tACS trains on brain activity, showing a modulation of the SSR waveform by electrical stimulation (see **Figures 4–6**). After-effects were out of the scope of the current study. Instead, we aimed to demonstrate immediate and frequency-specific tACS effects following established designs of SSR experiments. Nevertheless, future investigations of on-line effects and additionally registering after-effects and their

relationship might help understanding the mechanism of how tACS is modulating brain oscillations and could clarify whether these are based on entrainment or neural plasticity or a combination of both (Vossen et al., 2015).

tACS Influences a Stimulus Driven Oscillator

Our data suggest a well-circumscribed online modulation of brain oscillations in humans by tACS. Even though in our recent study (Neuling et al., 2015) we showed that endogenous oscillations and their modulations can be recovered during same frequency tACS, modulations of brain activity caused by tACS were not in our focus.

Many studies showed behavioral consequences or electrophysiological changes in other frequencies (Helfrich et al., 2014a; Voss et al., 2014), but a neurophysiological proof for the stimulation frequency is still missing. Note that Helfrich et al. (2014b) did not include a control condition, and thus the 10 Hz increase during stimulation might still be a result of the stimulation artifact itself. This fact is underlined by work from the same group (Helfrich et al., 2014a) in which signals around the stimulation frequency had to be notch-filtered. Here, however, we stimulated at different but spectrally close frequencies (7, 11 Hz) and used a fully balanced design. Thus any artifactual effects caused by the tACS would have been evident in the analysis when SSR and tACS were presented at different frequencies, an artifact which would have additionally spread across the spectrum (see also **Figure 2**). However, we did not observe such an artifact and effects of tACS were limited to matched frequency stimulations.

More specifically, we found that matched frequency stimulation reduced phase synchrony of fundamental (1f) and second harmonic (2f) SSR components while boosting evoked power and phase synchrony of third and fourth harmonic components. This result contrasts with our initial hypotheses of tACS-induced power increases of fundamental SSR components. In the following we suggest that our finding critically depends on the phase relationship between tACS and SSR.

Studies targeting the alpha rhythm with tACS are based on the assumption that the generative neural process underlying alpha will align its phase with the external electrical pacemaker (Fröhlich and McCormick, 2010; Neuling et al., 2012a, 2013). A similar phase alignment for SSRs is unlikely because SSR phase is itself strictly locked to the driving visual stimulation. Therefore, the phase of concurrent matching frequency tACS and SSR phase can differ in principle. For the present study we assumed a fixed phase relationship between tACS and SSRs for both stimulation frequencies and across participants. However, SSRs have been shown to require a number of cycles to fully build up (i.e., reach maximum amplitude e.g., Regan, 1989) whereas the flow of electrical currents introduced by tACS is assumed to have instantaneous effects (Fröhlich and McCormick, 2010; Reato et al., 2010). Due to inter-individual neuro-anatomical differences (e.g., conduction delays in early

visual processing pathways) SSR phase might jitter between participants. Although we have taken into account the build-up time by analyzing data epochs only during which the SSR was fully established we cannot exclude the possibility that SSR and tACS phase differed substantially and with a variable lag between participants.

In fact, a considerable tACS-SSR phase lag is a possible explanation for our finding of reduced phase locking in 1f and 2f SSR components and enhanced contributions of higher order harmonics during matched frequency stimulation. The two stimulation techniques forced entrainment (here phasic alignment) in similar areas but at different times. Put differently, neural activity evoked by a visual stimulus will peak shortly after maximal tACS (i.e., current peak or trough). As slight timing differences will be considerable parts of the SSR cycle this possibly affects the lower harmonics more strongly. The extent to which the SSR alignment is impaired by tACS probably varies from trial to trial, which consequentially leads to lower phase locking. In turn, boosted higher harmonics could be explained by tACS induced distortions of the SSR waveforms towards less sinusoidal morphologies (see **Figure 6**).

Open Questions and Future Directions

Thus far only human experimental study evidence has been provided using the LCMV beamformer approach with tACS (Neuling et al., 2015) and modeling and phantom measurements only exist for synthetic-aperture magnetometry beamformers (Soekadar et al., 2013). To know exactly how well the LCMV beamformer performs (i.e., reducing the tACS artifact and reconstructing the true source) phantom measurements are essential and more methodological studies need to be performed.

Above we laid out consequences of a possible tACS-SSR phase lag in our study. Future studies should thus implement methods to first estimate individual SSR phase and then re-align flicker stimulation with tACS as to minimize and standardize the phase lag. Furthermore, one could systematically vary the phase lag to test, for instance, whether out-of-phase stimulation produces cancellation effects.

Another aspect of the present study was that our tACS mainly targeted SSR components at fundamental frequencies (with only weak tACS at harmonic frequencies, cf. **Figure 2**) although visual stimulation also led to pronounced oscillatory components at harmonic frequencies. Harmonic components are typically found in SSR recordings (Appelbaum et al., 2006; Kim et al., 2011; Porcu et al., 2013) and may have their origin in non-linearities of the visual system (Roberts and Robinson, 2012; Norcia et al., 2015). Considering the fact that our results show a complex relationship of matched-frequency tACS with all corresponding SSR components it might be worthwhile to target specific harmonics driven by the stimulus. Conversely, one could also take into account the harmonic composition of SSRs and use a tACS signal that matches the spectral profile, i.e., consists of a superposition of sines that optimally resembles the SSR waveform.

Here, a detection task was simply employed to keep the subjects' attention on the visual stimuli. The targets were distributed randomly across tACS and SSR phase, thus no

behavioral effects were expected. Yet, many studies showed behavioral consequences of tACS phase on perception (Neuling et al., 2012a; Riecke et al., 2015). Recently, similar effects have been presented using visual (Mathewson et al., 2012; Spaak et al., 2014) and also auditory stimuli (Henry and Obleser, 2012; Henry et al., 2014). Basically, detection performance of a low contrast targets depended on the phase of a rhythmic stimulus in which the targets are embedded. A combination of both lines of research may provide evidence as to whether sensory and tACS entrainment work in a similar manner; whether they can be interactive and even increase the, typically small, behavioral effects. To conduct these studies, however, it will be vital to reliably estimate the stimulus-to-brain phase lag.

CONCLUSION

Our study demonstrated that reconstructing visual SSRs from MEG recordings during simultaneously administered tACS is possible, even when both match in temporal frequency. tACS influenced SSRs mainly by reducing phase synchrony for the fundamental and second harmonic. At the same time higher order harmonic responses were increased in power and phase

synchrony. Importantly, the present results provide further evidence for online effects of tACS on human mass-neuronal rhythmic activity. They open new avenues in studying perception and cognitive influences thereof through causal interference with stimulus-trained brain rhythms.

AUTHOR CONTRIBUTIONS

PR, CK, CL, NW, and TN conceived the experiment. PR, CL, and TN performed the research. PR and CK analyzed the data. All authors interpreted the data and wrote the article. All authors approved the final version of the manuscript.

ACKNOWLEDGMENTS

PR, CL, TN, and NW were supported by the European Research Council (ERC StG 283404, WIN2CON). We thank Dr. Gianpaolo Demarchi and Dr. Gianpiero Monittola for technical support and Dr. Julia N. Frey for help with data acquisition. The data for this study were acquired at the Center for Mind/Brain Science, University of Trento, Italy. We also thank two reviewers for helpful comments on an earlier version of this manuscript.

REFERENCES

- Adrian, E. D., and Matthews, B. H. (1934). The interpretation of potential waves in the cortex. *J. Physiol.* 81, 440–471. doi: 10.1113/jphysiol.1934.sp003147
- Andersen, S. K., and Müller, M. M. (2010). Behavioral performance follows the time course of neural facilitation and suppression during cued shifts of feature-selective attention. *Proc. Natl. Acad. Sci. U S A* 107, 13878–13882. doi: 10.1073/pnas.1002436107
- Antal, A., and Paulus, W. (2013). Transcranial alternating current stimulation (tACS). *Front. Hum. Neurosci.* 7:317. doi: 10.3389/fnhum.2013.00317
- Appelbaum, L. G., Wade, A. R., Vildavski, V. Y., Pettet, M. W., and Norcia, A. M. (2006). Cue-invariant networks for figure and background processing in human visual cortex. *J. Neurosci.* 26, 11695–11708. doi: 10.1523/JNEUROSCI.2741-06.2006
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B* 57, 289–300.
- Buzsáki, G., and Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science* 304, 1926–1929. doi: 10.1126/science.1099745
- Capilla, A., Pazo-Alvarez, P., Darriba, A., Campo, P., and Gross, J. (2011). Steady-state visual evoked potentials can be explained by temporal superposition of transient event-related responses. *PLoS One* 6:e14543. doi: 10.1371/journal.pone.0014543
- Cecere, R., Rees, G., and Romei, V. (2015). Individual differences in alpha frequency drive crossmodal illusory perception. *Curr. Biol.* 25, 231–235. doi: 10.1016/j.cub.2014.11.034
- Engel, A. K., Fries, P., and Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top-down processing. *Nat. Rev. Neurosci.* 2, 704–716. doi: 10.1038/35094565
- Frohlich, F., and McCormick, D. A. (2010). Endogenous electric fields may guide neocortical network activity. *Neuron* 67, 129–143. doi: 10.1016/j.neuron.2010.06.005
- Greenhouse, S. W., and Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika* 24, 95–112. doi: 10.1007/bf02289823
- Helfrich, R. F., Knepper, H., Nolte, G., Strüber, D., Rach, S., Herrmann, C. S., et al. (2014a). Selective modulation of interhemispheric functional connectivity by HD-tACS shapes perception. *PLoS Biol.* 12:e1002031. doi: 10.1371/journal.pbio.1002031
- Helfrich, R. F., Schneider, T. R., Rach, S., Trautmann-Lengsfeld, S. A., Engel, A. K., and Herrmann, C. S. (2014b). Entrainment of brain oscillations by transcranial alternating current stimulation. *Curr. Biol.* 24, 333–339. doi: 10.1016/j.cub.2013.12.041
- Henry, M. J., Herrmann, B., and Obleser, J. (2014). Entrained neural oscillations in multiple frequency bands comodulate behavior. *Proc. Natl. Acad. Sci. U S A* 111, 14935–14940. doi: 10.1073/pnas.1408741111
- Henry, M. J., and Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proc. Natl. Acad. Sci. U S A* 109, 20095–20100. doi: 10.1073/pnas.1213390109
- Herrmann, C. S., Rach, S., Neuling, T., and Strüber, D. (2013). Transcranial alternating current stimulation: a review of the underlying mechanisms and modulation of cognitive processes. *Front. Hum. Neurosci.* 7:279. doi: 10.3389/fnhum.2013.00279
- Jacoby, O., Hall, S. E., and Mattingley, J. B. (2012). A crossmodal crossover: opposite effects of visual and auditory perceptual load on steady-state evoked potentials to irrelevant visual stimuli. *Neuroimage* 61, 1050–1058. doi: 10.1016/j.neuroimage.2012.03.040
- Kashiwase, Y., Matsumiya, K., Kuriki, I., and Shioiri, S. (2012). Time courses of attentional modulation in neural amplification and synchronization measured with steady-state visual-evoked potentials. *J. Cogn. Neurosci.* 24, 1779–1793. doi: 10.1162/jocn_a_00212
- Kaspar, K., Hassler, U., Martens, U., Trujillo-Barreto, N., and Gruber, T. (2010). Steady-state visually evoked potential correlates of object recognition. *Brain Res.* 1343, 112–121. doi: 10.1016/j.brainres.2010.04.072
- Keitel, C., Andersen, S. K., Quigley, C., and Müller, M. M. (2013). Independent effects of attentional gain control and competitive interactions on visual stimulus processing. *Cereb. Cortex* 23, 940–946. doi: 10.1093/cercor/bhs084
- Keitel, C., Quigley, C., and Ruhnau, P. (2014). Stimulus-driven brain oscillations in the alpha range: entrainment of intrinsic rhythms or frequency-following response? *J. Neurosci.* 34, 10137–10140. doi: 10.1523/JNEUROSCI.1904-14.2014
- Kim, Y.-J., Graboweky, M., Paller, K. A., Muthu, K., and Suzuki, S. (2007). Attention induces synchronization-based response gain in steady-state visual evoked potentials. *Nat. Neurosci.* 10, 117–125. doi: 10.1038/nn1821
- Kim, Y.-J., Graboweky, M., Paller, K. A., and Suzuki, S. (2011). Differential roles of frequency-following and frequency-doubling visual responses revealed by evoked neural harmonics. *J. Cogn. Neurosci.* 23, 1875–1886. doi: 10.1162/jocn.2010.21536

- Koenig-Robert, R., and VanRullen, R. (2013). SWIFT: a novel method to track the neural correlates of recognition. *Neuroimage* 81, 273–282. doi: 10.1016/j.neuroimage.2013.04.116
- Lachaux, J. P., Rodriguez, E., Martinerie, J., and Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Hum. Brain Mapp.* 8, 194–208. doi: 10.1002/(sici)1097-0193(1999)8:4<194::aid-hbm4>3.0.co;2-c
- Mathewson, K. E., Prudhomme, C., Fabiani, M., Beck, D. M., Lleras, A., and Gratton, G. (2012). Making waves in the stream of consciousness: entraining oscillations in EEG alpha and fluctuations in visual awareness with rhythmic visual stimulation. *J. Cogn. Neurosci.* 24, 2321–2333. doi: 10.1162/jocn_a_00288
- Müller, M. M., and Hillyard, S. (2000). Concurrent recording of steady-state and transient event-related potentials as indices of visual-spatial selective attention. *Clin. Neurophysiol.* 111, 1544–1552. doi: 10.1016/s1388-2457(00)00371-0
- Neuling, T., Ruhnau, P., Fuscà, M., Demarchi, G., Herrmann, C. S., and Weisz, N. (2015). Friends, not foes: magnetoencephalography as a tool to uncover brain dynamics during transcranial alternating current stimulation. *Neuroimage* 118, 406–413. doi: 10.1016/j.neuroimage.2015.06.026
- Neuling, T., Rach, S., and Herrmann, C. S. (2013). Orchestrating neuronal networks: sustained after-effects of transcranial alternating current stimulation depend upon brain states. *Front. Hum. Neurosci.* 7:161. doi: 10.3389/fnhum.2013.00161
- Neuling, T., Rach, S., Wagner, S., Wolters, C. H., and Herrmann, C. S. (2012a). Good vibrations: oscillatory phase shapes perception. *Neuroimage* 63, 771–778. doi: 10.1016/j.neuroimage.2012.07.024
- Neuling, T., Wagner, S., Wolters, C. H., Zaehle, T., and Herrmann, C. S. (2012b). Finite-element model predicts current density distribution for clinical applications of tDCS and tACS. *Front. Psychiatry* 3:83. doi: 10.3389/fpsyg.2012.00083
- Nolte, G. (2003). The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors. *Phys. Med. Biol.* 48, 3637–3652. doi: 10.1088/0031-9155/48/22/002
- Norgia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B. R., and Rossion, B. (2015). The steady-state visual evoked potential in vision research: a review. *J. Vis.* 15:4. doi: 10.1167/15.4.6
- Polanía, R., Nitsche, M. A., Korman, C., Batsikadze, G., and Paulus, W. (2012). The importance of timing in segregated theta phase-coupling for cognitive performance. *Curr. Biol.* 22, 1314–1318. doi: 10.1016/j.cub.2012.05.021
- Porcu, E., Keitel, C., and Müller, M. M. (2013). Concurrent visual and tactile steady-state evoked potentials index allocation of inter-modal attention: a frequency-tagging study. *Neurosci. Lett.* 556, 113–117. doi: 10.1016/j.neulet.2013.09.068
- Porcu, E., Keitel, C., and Müller, M. M. (2014). Visual, auditory and tactile stimuli compete for early sensory processing capacities within but not between senses. *Neuroimage* 97, 224–235. doi: 10.1016/j.neuroimage.2014.04.024
- Quigley, C., and Müller, M. M. (2014). Feature-selective attention in healthy old age: a selective decline in selective attention? *J. Neurosci.* 34, 2471–2476. doi: 10.1523/JNEUROSCI.2718-13.2014
- Reato, D., Rahman, A., Bikson, M., and Parra, L. C. (2010). Low-intensity electrical stimulation affects network dynamics by modulating population rate and spike timing. *J. Neurosci.* 30, 15067–15079. doi: 10.1523/JNEUROSCI.2059-10.2010
- Regan, D. (1989). *Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine*. New York: Elsevier.
- Riecke, L., Formisano, E., Herrmann, C. S., and Sack, A. T. (2015). 4-Hz transcranial alternating current stimulation phase modulates hearing. *Brain Stimul.* 8, 777–783. doi: 10.1016/j.brs.2015.04.004
- Roberts, J. A., and Robinson, P. A. (2012). Quantitative theory of driven nonlinear brain dynamics. *Neuroimage* 62, 1947–1955. doi: 10.1016/j.neuroimage.2012.05.054
- Rossoni, B., and Boremanse, A. (2011). Robust sensitivity to facial identity in the right human occipito-temporal cortex as revealed by steady-state visual-evoked potentials. *J. Vis.* 11:16. doi: 10.1167/11.2.16
- Rossoni, B., Prieto, E. A., Boremanse, A., Kuefner, D., and Van Belle, G. (2012). A steady-state visual evoked potential approach to individual face perception: effect of inversion, contrast-reversal and temporal dynamics. *Neuroimage* 63, 1585–1600. doi: 10.1016/j.neuroimage.2012.08.033
- Shah, A. S., Bressler, S. L., Knuth, K. H., Ding, M., Mehta, A. D., Ulbert, I., et al. (2004). Neural dynamics and the fundamental mechanisms of event-related brain potentials. *Cereb. Cortex* 14, 476–483. doi: 10.1093/cercor/bhh009
- Soekadar, S. R., Witkowski, M., Cossio, E. G., Birbaumer, N., Robinson, S. E., and Cohen, L. G. (2013). *In vivo* assessment of human brain oscillations during application of transcranial electric currents. *Nat. Commun.* 4:2032. doi: 10.1038/ncomms3032
- Spaak, E., de Lange, F. P., and Jensen, O. (2014). Local entrainment of alpha oscillations by visual stimuli causes cyclic modulation of perception. *J. Neurosci.* 34, 3536–3544. doi: 10.1523/JNEUROSCI.4385-13.2014
- Störmer, V. S., and Alvarez, G. A. (2014). Feature-based attention elicits surround suppression in feature space. *Curr. Biol.* 24, 1985–1988. doi: 10.1016/j.cub.2014.07.030
- Thut, G., Schyns, P. G., and Gross, J. (2011). Entrainment of perceptually relevant brain oscillations by non-invasive rhythmic stimulation of the human brain. *Front. Psychol.* 2:170. doi: 10.3389/fpsyg.2011.00170
- Van Veen, B. D., van Drongelen, W., Yuchtman, M., and Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Trans. Biomed. Eng.* 44, 867–880. doi: 10.1109/10.623056
- Voss, U., Holzmann, R., Hobson, A., Paulus, W., Koppehele-Gossel, J., Klimke, A., et al. (2014). Induction of self-awareness in dreams through frontal low current stimulation of gamma activity. *Nat. Neurosci.* 17, 810–812. doi: 10.1038/nrn.3719
- Vossen, A., Gross, J., and Thut, G. (2015). Alpha power increase after transcranial alternating current stimulation at alpha frequency (α -tACS) reflects plastic changes rather than entrainment. *Brain Stimul.* 8, 499–508. doi: 10.1016/j.brs.2014.12.004
- Vosskuhl, J., Huster, R. J., and Herrmann, C. S. (2015). Increase in short-term memory capacity induced by down-regulating individual theta frequency via transcranial alternating current stimulation. *Front. Hum. Neurosci.* 9:257. doi: 10.3389/fnhum.2015.00257
- Zaehle, T., Rach, S., and Herrmann, C. S. (2010). Transcranial alternating current stimulation enhances individual alpha activity in human EEG. *PLoS One* 5:e13766. doi: 10.1371/journal.pone.0013766

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Ruhnau, Keitel, Lithari, Weisz and Neuling. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Interpretations of Frequency Domain Analyses of Neural Entrainment: Periodicity, Fundamental Frequency, and Harmonics

Hong Zhou¹, Lucia Melloni^{2,3}, David Poeppel^{4,5} and Nai Ding^{1,6,7*}

¹ College of Biomedical Engineering and Instrument Sciences, Zhejiang University, Hangzhou, China, ² Department of Neurology, New York University Langone Medical Center, New York, NY, USA, ³ Department of Neurophysiology, Max-Planck Institute for Brain Research, Frankfurt, Germany, ⁴ Neuroscience Department, Max-Planck Institute for Empirical Aesthetics, Frankfurt, Germany, ⁵ Department of Psychology, New York University, New York, NY, USA, ⁶ Interdisciplinary Center for Social Sciences, Zhejiang University, Hangzhou, China, ⁷ Neuro and Behavior EconLab, Zhejiang University of Finance and Economics, Hangzhou, China

Brain activity can follow the rhythms of dynamic sensory stimuli, such as speech and music, a phenomenon called neural entrainment. It has been hypothesized that low-frequency neural entrainment in the neural delta and theta bands provides a potential mechanism to represent and integrate temporal information. Low-frequency neural entrainment is often studied using periodically changing stimuli and is analyzed in the frequency domain using the Fourier analysis. The Fourier analysis decomposes a periodic signal into harmonically related sinusoids. However, it is not intuitive how these harmonically related components are related to the response waveform. Here, we explain the interpretation of response harmonics, with a special focus on very low-frequency neural entrainment near 1 Hz. It is illustrated why neural responses repeating at f Hz do not necessarily generate any neural response at f Hz in the Fourier spectrum. A strong neural response at f Hz indicates that the time scales of the neural response waveform within each cycle match the time scales of the stimulus rhythm. Therefore, neural entrainment at very low frequency implies not only that the neural response repeats at f Hz but also that each period of the neural response is a slow wave matching the time scale of a f Hz sinusoid.

OPEN ACCESS

Edited by:

Anne Keitel,
University of Glasgow, UK

Reviewed by:

Philipp Ruhnau,
Paris-Lodron-Universität Salzburg,
Austria

Malte Wöstmann,
University of Lübeck, Germany

***Correspondence:**

Nai Ding
ding_nai@zju.edu.cn

Received: 16 March 2016

Accepted: 23 May 2016

Published: 06 June 2016

Citation:

Zhou H, Melloni L, Poeppel D and Ding N (2016) Interpretations of Frequency Domain Analyses of Neural Entrainment: Periodicity, Fundamental Frequency, and Harmonics. *Front. Hum. Neurosci.* 10:274. doi: 10.3389/fnhum.2016.00274

Keywords: neural entrainment, rhythm, harmonics, oscillations, periodicity

INTRODUCTION

Cortical activity, measured by electroencephalography (EEG), magnetoencephalography (MEG), or local field potential recordings (LFP), can synchronize to the rhythm of a sensory stimulus. For example, when the intensity of a sound, e.g., a pure tone, fluctuates at a given frequency (f Hz), a neural response at that frequency (f Hz) is often observed and is referred to as the auditory steady state response (aSSR; Galambos et al., 1981; Ross et al., 2000; Wang et al., 2012). Similarly, when the luminance of a visual stimulus, e.g., a Gabor patch, fluctuates at f Hz, a neural response at f Hz can also be observed and is referred to as the steady state visual evoked response (SSVEP; Norcia et al., 2015). Recently, low-frequency (<3 Hz) neural entrainment has also been observed for abstract stimulus properties such as the rhythms of musical beats and linguistic

constituents (Buiatti et al., 2009; Nozaradan et al., 2011; Ding et al., 2016) and during the processing of natural speech or movies (Ding and Simon, 2012; Zion Golumbic et al., 2013; Koskinen and Seppä, 2014; Lankinen et al., 2014). It has been hypothesized that low-frequency neural synchronization to a stimulus provides a mechanism for selective attention and temporal integration of information (Schroeder et al., 2008; Schroeder and Lakatos, 2009; Giraud and Poeppel, 2012), and is important for parsing the temporal structure of speech and music (Nozaradan et al., 2011; Ding et al., 2016).

Neural entrainment to a stimulus rhythms is often analyzed in the frequency domain, while traditional neurophysiological responses, e.g., the event-related responses, are usually analyzed in the time domain. Therefore, some frequency-domain measures may appear unintuitive for researchers mainly using time domain analysis methods. For example, when the stimulus rhythm is at f Hz, neural responses can often be observed not just at f but also at its harmonics, i.e., at $2f, 3f, 4f$ etc. The harmonics can provide additional insight into the underlying neural encoding mechanisms (O'connell et al., 2015) but their interpretations are not straightforward. In this article, we explain how these harmonics are related to time-domain waveforms. We restrict the frequency-domain analysis method to the Discrete Fourier Transform (DFT), the most classic frequency domain analysis method. We will elaborate how the properties of a signal are represented in the lens of the DFT, and will not discuss whether the DFT is the best method to represent a particular signal.

The article is organized as follows: we first present examples that describe the relationship between time-domain signal periodicity and signal spectrum. Since it remains controversial whether the experimentally observed neural tracking of such low-frequency stimulus rhythms reflects a succession of event-related responses or a proper entrainment of neural oscillators (Ding and Simon, 2014; Keitel et al., 2014). We also describe how to interpret the power spectrum of a series of event-related responses. These discussions are purely based on intuitive examples, avoiding a more formal, mathematical treatment (for formal treatment, see, e.g., Oppenheim et al., 1989). A glossary is provided in **Table 1**.

SIGNAL PERIODICITY AND THE FOURIER SPECTRUM

Here we will consider a signal that has a period of T . In other words, the signal repeats every T seconds. The Fourier transform analyses the frequency content of a signal by decomposing it into sinusoids at different frequencies. A signal with a period T repeats at a rate of $f_0 = 1/T$, which is referred to as the fundamental frequency of the signal. Usually and intuitively, the Fourier spectrum of such a signal shows strong power at f_0 . In other words, the signal can be well explained by a sinusoid at frequency f_0 . Sometimes, the response at f_0 may be the only component in the Fourier spectrum, indicating that the signal is a sinusoid. **Figure 1** illustrates this condition.

TABLE 1 | Glossary table.

Basic definitions

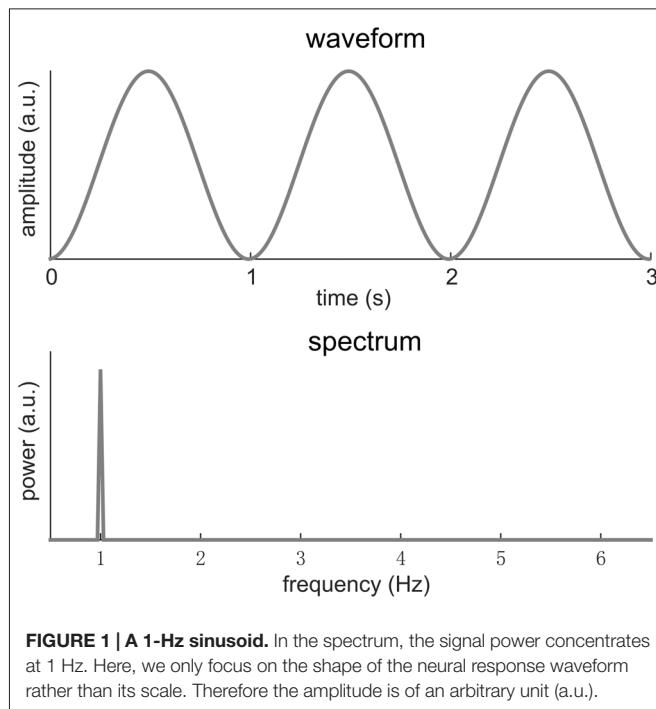
Signal measurement	In principle, a signal could be infinite in duration. In practice, we can only measure a signal in a finite interval, e.g., from T_0 to T_1 . The signal measured in this finite interval is called a measurement.
Periodic signal	A signal $x(t)$ is periodic if $x(t) = x(t + T)$, for any t . If $x(t) = x(t + T)$, then it follows that $x(t + kT) = x(t)$ for any integer k . A periodic signal is infinite in duration and in this article, we only consider a measurement that has a duration of kT .
Period	The period of a signal is T if and only if T the smallest positive number such that $x(t) = x(t + T)$.
Fundamental frequency	The fundamental frequency of a signal is $f_0 = 1/T$, where T is the period of the signal.
Harmonics	The frequency of the k th harmonic is k times the fundamental frequency, i.e., kf_0 .
Discrete-time signals	A continuous-time signal $x(t)$ can be converted into a discrete-time signal by a sampling process. If the sampling period is t_s , the discrete signal measurement will be $x(kt_s)$, where k is an integer and $T_0 < kt_s < T_1$.
Sampling frequency	If the sampling period is t_s , the sampling frequency is $f_s = 1/t_s$.

Discrete Fourier Transform (DFT)

DFT	The DFT represents a discrete-time signal as the sum of harmonically related sinusoids. The DFT coefficients determine the amplitude and phase of these sinusoids.
Frequency extracted by the DFT	If a discrete-time signal measurement contains D samples, i.e., time points, the Fourier transform represents it as the sum of sinusoids at harmonically related frequencies, i.e., k/D , where k is an integer and k/D falls between 0 and half of the sampling rate.
Signal extrapolation	The DFT assumes that a signal measurement constitutes a single period of a periodic signal. In other words, the DFT extrapolates the signal measurement by repeating the same waveform in time. For a periodic signal, such extrapolation is correct as long as an integer number of periods of the signal are measured.

Neuroscience terms

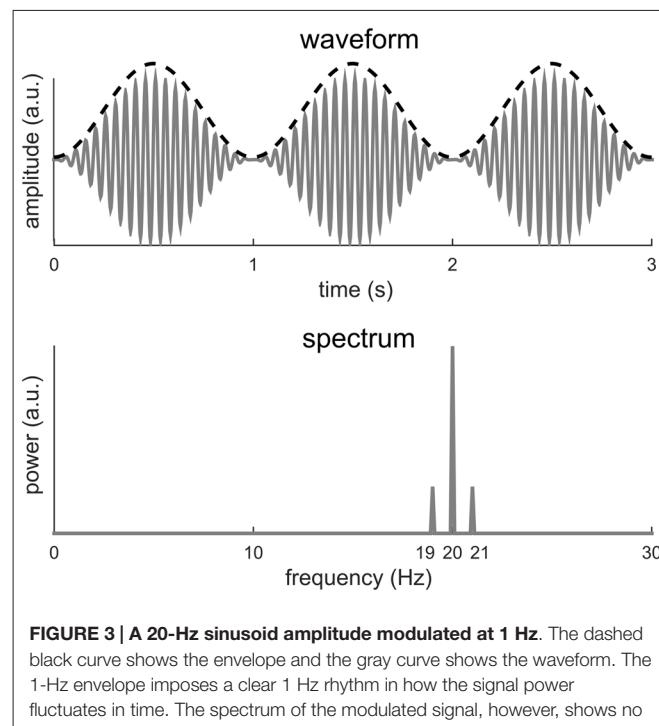
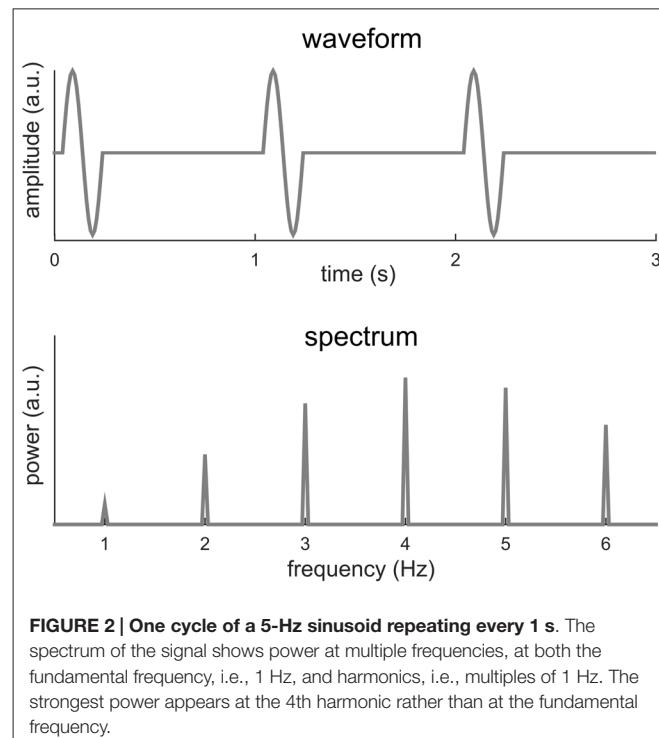
Event-related responses	The event-related response refers to the response waveform precisely following, i.e., time-locked (often called phase-locked) to, a brief sensory stimulus. To measure the event-related response, a brief stimulus is presented repetitively for tens or hundreds of times and the interval between two presentations is usually randomized. The event-related response is obtained by averaging the response waveform over repetitions of the stimulus.
High-gamma activity	Neural activity in the high-gamma band, which is usually defined to be between 70 and 200 Hz. The waveform of high-gamma activity is usually not synchronized to the sensory stimuli while the power envelope of high-gamma activity is sometimes synchronized to sensory stimuli.



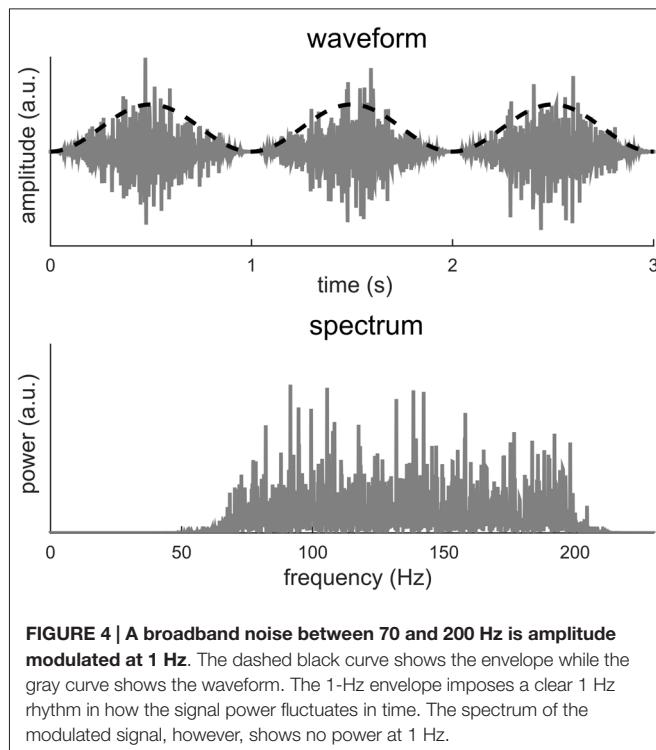
Sinusoids, of course, are mathematical abstractions and in the real world few signals are precisely sinusoidal. When a signal deviates from a sinusoid, its Fourier spectrum will have power not just at f_0 , but also at multiples of f_0 , e.g., $2f_0$, $3f_0$, $4f_0$ etc. **Figure 2** illustrates one such condition, in which a short biphasic signal lasting for 200 ms repeats every 1 s. In this illustration, the biphasic signal is one cycle of a 5 Hz sinusoid while the signal repeats at 1 Hz (i.e., $f_0 = 1$ Hz). The power spectrum of this signal spreads over a number of frequencies, e.g., f_0 , $2f_0$, $3f_0$, $4f_0$... The strongest power in the Fourier spectrum appears at $4f_0$ instead of f_0 . This signal can be viewed as a rough simulation of a sequence of transient event-related responses.

Figure 2 illustrates that a signal with a period T may have its power spread over f_0 and its harmonics. In the following, we show an additional example in which no power even exists at f_0 . In this example (**Figure 3**), a 10-Hz sinusoid is amplitude modulated at 1 Hz. Amplitude modulation involves the product of two signals. The fast signal is called the carrier and the slow signal is called the envelope. In general, the envelope captures how the signal power fluctuates over time. In **Figure 3**, the modulated signal is the product of a 20-Hz sinusoid and a 1-Hz sinusoid. The amplitude modulated signal has a period of 1 s, as is evident from its waveform. Nonetheless, the Fourier spectrum shows no power at 1 Hz. In this example, the envelope signal is a sinusoid, if the signal is not a sinusoid, additional responses will be seen at 20 ± 2 Hz, 20 ± 3 Hz..., on top of the responses at 20 Hz and 20 ± 1 Hz.

The example in **Figure 3** also introduces more general cases in which signal periodicity occurs in the modulation domain, i.e., in the signal envelope. In **Figure 4**, the carrier



is a band-limited white noise between 70 and 200 Hz while the envelope signal remains a 1 Hz sinusoid. A visual inspection of the signal waveform suggests a strong rhythm at 1 Hz while no 1 Hz information can be found in the spectrum. In this case, the apparent 1 Hz periodicity only exists in the signal envelope and can only be revealed



after the envelope signal by itself is extracted. A Fourier analysis of the envelope reduces to the condition illustrated in **Figure 1**. This example can be viewed as a simulation of high-gamma neural activity tracking a 1-Hz rhythm. The signal envelope can be extracted either explicitly using, e.g., the Hilbert transform, or implicitly through a time-frequency analysis, such as the short-term Fourier transform (STFT) or the wavelet transform. One interpretation of the spectrogram obtained by the STFT or wavelet analysis is that the input signal is filtered into narrow frequency bands and in each band the power envelope of the signal is extracted (Vaidyanathan, 1990). Therefore, periodicity in the modulation domain can be revealed by analyzing the time course of the STFT or wavelet spectrogram.

In sum, we show here that if the period of a signal is T , the DFT spectrum of the signal may show power at f_0 and its harmonically related frequencies. Importantly, the power at f_0 may not be the strongest (**Figure 2**) and may not even exist (**Figure 3**). Furthermore, even when the signal is a periodic, it may show higher order regularity including the periodicity in its envelope (**Figure 4**).

FACTORS AFFECTING THE POWER AT HARMONIC FREQUENCY

The previous section shows that a neural signal repeating every T seconds is represented in the frequency domain by harmonically related frequencies at $f_0, 2f_0, 3f_0, 4f_0\dots$. In this section, we discuss in more details about what factors decide the power at each frequency. A periodic signal is fully characterized by a single cycle. **Figure 5** illustrates the

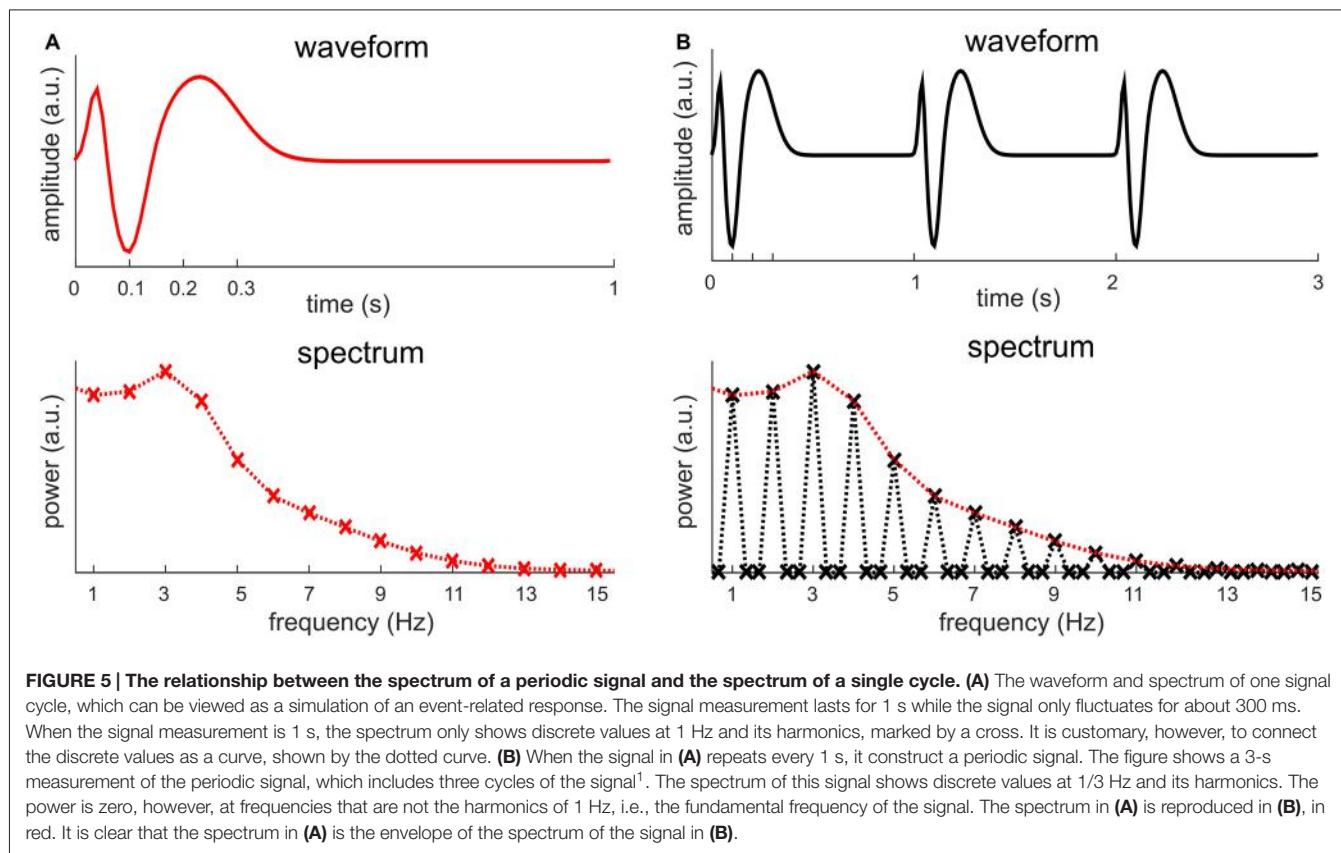
Fourier analysis of a periodic signal (including multiple cycles) and the Fourier analysis of a single cycle. All non-zero values in the spectrum of a periodic signal are captured by the spectrum of a single cycle. The spectrum of a periodic signal can be obtained by inserting zeros into the spectrum of a single cycle. The spectrum of a periodic signal can only have nonzero values at the fundamental frequency and its harmonics, and the spectrum of a single cycle only takes values at these frequencies. In general, the non-zero values in the spectrum of a periodic signal are decided by the waveform of a single cycle while the frequencies at which the spectrum shows nonzero values are decided by its period. In other words, the spectrum of a single cycle provides the spectral envelope of the spectrum of a periodic signal.

Based on the analysis above, when the waveform of a single cycle contains “fast” oscillations or “sharp” edges, the signal will have high power at high-frequency harmonics. Here, “fast” oscillation means oscillations at frequencies much higher than the fundamental frequency of the periodic signal (**Figure 3**). “Sharp” edges mean edges rising/decaying faster relative to how fast a sinusoid at f_0 rises or decays. The power of a periodic signal will concentrate at f_0 only if the stimulus rate matches the spectral resonance, i.e., time scales, of the response in a single stimulus cycle (for details see, the next section and **Figure 6**). Therefore, a neural peak at f_0 Hz in the Fourier spectrum does not only indicate the repetition of a neural waveform at f_0 Hz but also indicates that the waveform being repeated is roughly a cycle of a f_0 Hz sinusoid.

FOURIER ANALYSIS OF A SERIES OF EVENT-RELATED RESPONSES

In this section, we describe the frequency domain representation of a periodic neural response composed of a series of discrete, steady-state event-related responses. Specifically, the sensory stimulus is a sequence of periodically occurring events and each event is modeled as an impulse, which can be viewed as an approximation of a short tone pip or a short flash of light. We assume that, when the neural response reaches steady state, the event-related response to each sensory event is identical (time-invariance), and the measured neural response is a linear superposition of different event-related responses (linearity). Under these assumptions, the neural network generating the measured neural response can be modeled by a linear time-invariant system (Oppenheim et al., 1989). It then follows that intrinsic properties of the neural network are fully characterized by the impulse response, i.e., the event-related response.

Under the linear time-invariant theory, each cycle of the neural response, i.e., the event-related response, is a property of the neural system while periodicity is a property of the stimulus. If the event-related response decays to baseline within each stimulus cycle, the spectral envelope of the periodic neural response is determined by the spectrum of the event-related response (**Figure 5**). This



conclusion, however, also holds if the event-related response does not return to baseline when the next stimulus comes for reasons that will not be elaborated². Furthermore, the linear time-invariant system model also applies to aperiodic stimuli and to continuously changing stimuli. This article, however, only focuses on the response to a periodic sensory input.

In Figure 6, we illustrate how the spectrum of a single event-related response and the stimulus repetition rate jointly determine the spectrum of the neural responses under the linear time-invariant systems theory. In this example, we assume that the event-related response has the strongest power in the theta

band (4–8 Hz) and the stimulus is a sequence of pulses, i.e., very brief sensory events. The resonance frequency range of the neural system, i.e., the theta band, is arbitrarily chosen. When the stimulus rate is below the resonance frequency range, the response is weak and shows power at high-frequency harmonics (e.g., the 1 Hz condition in Figure 6B). When the stimulus rate is within the resonance frequency range, a strong response is seen at f_0 and also harmonic frequencies falling into the resonance frequency range (e.g., the 4 and 6 Hz conditions in Figure 6B). Finally, if the stimulus rate is above the resonance frequency range, the steady state response is very weak (e.g., the 12 Hz condition in Figure 6B). Therefore, if strong neural entrainment is seen at f_0 and not at any harmonic frequency, it indicates that the f_0 is within the range of resonance frequency while $2f_0$ is outside this range.

What needs additional explanation is why the response is so weak at any frequency when the stimulus rate is low (Figure 6A). When the stimulus rate is below the resonance frequency range, the neural system has no difficulty producing a response for every stimulus event. The reason why the response is weak is two-fold. First, the response only deviates from the baseline in a short period after each stimulus event, making the total power of the response, i.e., power summed over time, very low. Second, the power in the frequency domain is distributed over several harmonic frequencies, making the response at each single frequency even weaker.

¹In Figure 5B, we only consider three cycles of the periodic signal, and therefore two zeros are inserted between any two frequencies resolved in Figure 5A (i.e., the frequencies shown by a x in Figure 5A). In general, if the signal contains N cycles, $N-1$ points will be inserted between any two frequencies resolved in Figure 5A. If the signal is truly periodic and is infinite in duration, the spectrum is continuous and is zero at any frequency except for the frequencies resolved in Figure 5A.

²For linear time-invariant systems, the neural response is the stimulus convolving the event-related response. The Fourier transform of the convolution of two signals is the product of the Fourier transform of each signal. The stimulus considered here is a periodic pulse train. The Fourier transform of a pulse train is also a pulse train. The spectrum of a pulse train is only nonzero at $0, f_0, 2f_0, 3f_0 \dots$ and take the same nonzero value at $0, f_0, 2f_0, 3f_0 \dots$. The Fourier transform of the periodic neural response is the product of the spectrum of the event-related response and the spectral domain pulse train. Therefore, the spectrum of the event-related response is the envelope of the spectrum of the periodic response.

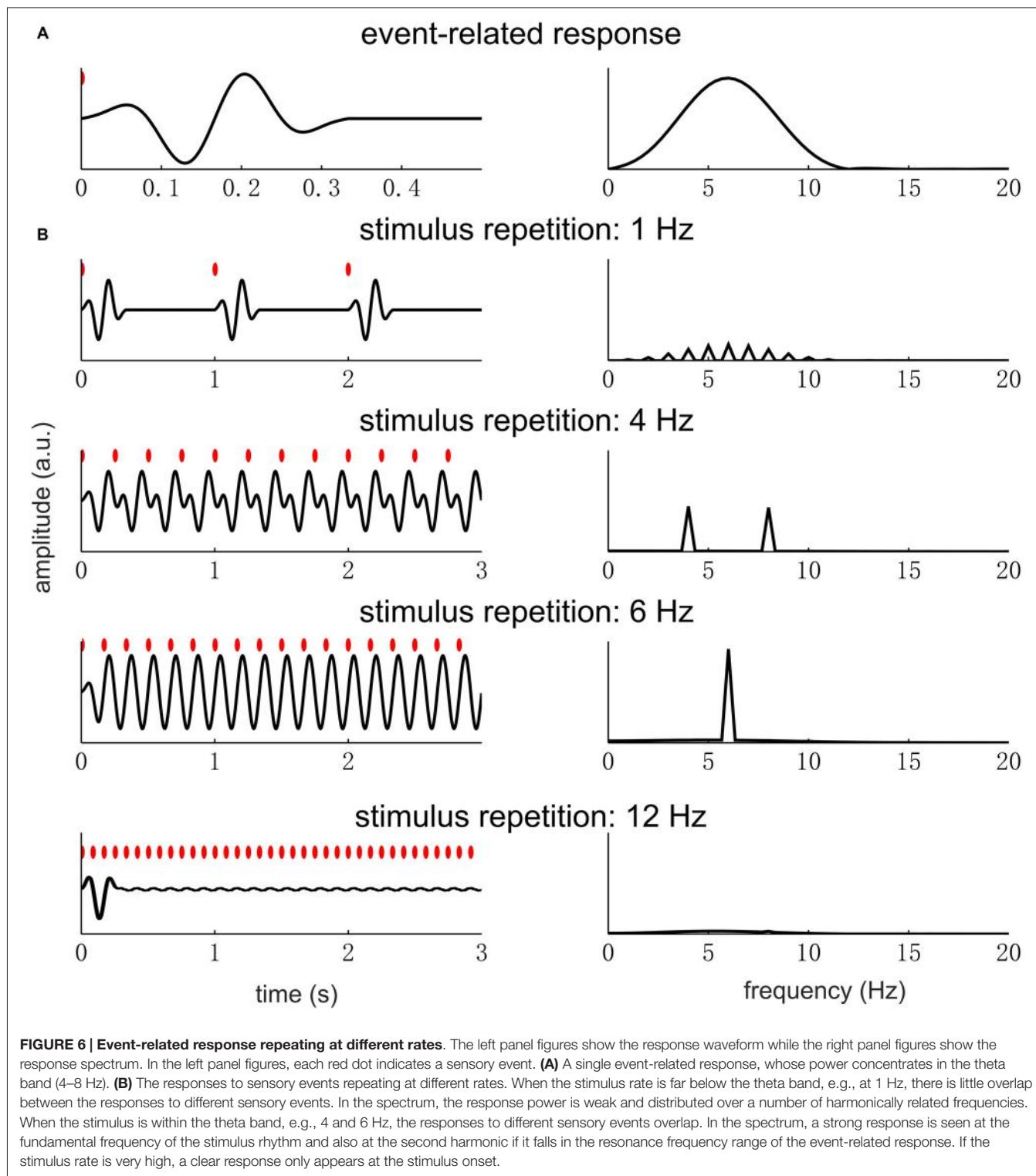


FIGURE 6 | Event-related response repeating at different rates. The left panel figures show the response waveform while the right panel figures show the response spectrum. In the left panel figures, each red dot indicates a sensory event. **(A)** A single event-related response, whose power concentrates in the theta band (4–8 Hz). **(B)** The responses to sensory events repeating at different rates. When the stimulus rate is far below the theta band, e.g., at 1 Hz, there is little overlap between the responses to different sensory events. In the spectrum, the response power is weak and distributed over a number of harmonically related frequencies. When the stimulus is within the theta band, e.g., 4 and 6 Hz, the responses to different sensory events overlap. In the spectrum, a strong response is seen at the fundamental frequency of the stimulus rhythm and also at the second harmonic if it falls in the resonance frequency range of the event-related response. If the stimulus rate is very high, a clear response only appears at the stimulus onset.

In the above discussion, we only consider the neural responses to a sequence of briefly sensory events. Nonetheless, many sensory stimuli, such as speech and music, change continuously. How neural activity follows a continuously changing stimulus is still an unresolved research question (for a review see,

Ding and Simon, 2014). One hypothesis is that the response is still triggered by discrete sensory or perceptual events, e.g., acoustic edges or syllable/sentence onsets, in which case the above discussion still holds. The other hypothesis, however, is that the response follows stimulus continuously. Under this

hypothesis, for a linear time-invariant system, the response is the continuously changing stimulus feature convolved by the event-related response. If the stimulus feature changes smoothly, e.g., sinusoidally, its power will concentrate at the fundamental frequency. In this case, the response power will also concentrate at the fundamental frequency and any response at harmonic frequencies will reflect nonlinear neural processing.

INTERPRETING LOW-FREQUENCY NEURAL ENTRAINMENT

When a response shows strong power at f_0 , it indicates “baseline” fluctuations within each stimulus period. For example, when f_0 is below 1 Hz and a strong neural response is seen at f_0 , it indicates a slow drift in the response “baseline” over the time interval comparable to the duration of a stimulus cycle. If a response is “local”, i.e., lasting for a duration shorter than the stimulus cycle, it can hardly contribute to the neural tracking at the fundamental frequency (**Figure 6**, 1-Hz condition). For example, the mid-latency auditory evoked response lasts for less than 100 ms. If this response repeats every 1 s, it can hardly contribute to 1-Hz neural entrainment. Very low-frequency (e.g., <1 Hz) neural tracking indicates long lasting and slowly changing responses. The slowness of low-frequency entrainment is in fact its core feature. A key hypothesis about low-frequency neural entrainment is that the neural response does not return to baseline when the next stimulus comes, and this “baseline drift” provides a context for the processing of the next stimulus (Schroeder et al., 2008; Schroeder and Lakatos, 2009).

LOW-FREQUENCY NEURAL ENTRAINMENT DURING SPEECH, MUSIC, AND AUDITORY PROCESSING

Low-frequency neural entrainment is often observed during speech and music processing. When listening to discourse-level speech materials, neural entrainment is reliably observed in the delta band (>4 Hz), including the frequency range near or below 1 Hz (Ding and Simon, 2012; Zion Golumbic et al., 2013; Koskinen and Seppä, 2014; Lankinen et al., 2014). It is further demonstrated that neural entrainment in the delta band reflects not only neural encoding of acoustic features but also neural encoding of syntactic structures (Ding et al., 2016). Similar delta-band neural entrainment is observed during music processing. In particular, it has been shown that neural activity can follow the rhythm of musical beat and meter (Nozaradan et al., 2011, 2012; Sturm et al., 2014; Tierney and Kraus, 2014). These results suggest that low-frequency neural entrainment could play a role in parsing the temporal structure of speech and music, forming phrasal-level chunks.

On the other hand, although very low-frequency neural entrainment is reliably observed during speech and music processing, it is not at all a universal phenomenon during the processing of arbitrary auditory stimuli, even ones with a very

low-frequency acoustic rhythm (Lakatos et al., 2013; Doelling and Poeppel, 2015). For example, in a study conducted by Lakatos et al. (2013), tone pips are presented every 1.5 s. Only transient auditory evoked responses, i.e., the P1-N1-P2 complex, are seen after each tone pip during passive listening. When the subjects are performing an outlier detection task, however, a slow drift in baseline throughout the 1.5 s stimulus period emerges (in healthy subjects but not in schizophrenia patients). For another example, Doelling and Poeppel (2015) studied the neural responses to music at a very slow tempo, in some cases below 1 Hz. Neural responses of musicians are entrained at the tempo of a given piece while neural responses of non-musicians only appear at harmonics of the tempo rate. Both examples show that very slow neural entrainment (>1 Hz) is not a natural consequence of sensory evoked responses and only emerges as a consequence of task engagement or expertise.

Slow event-related responses such as the N400, P3, and P600 components can indeed contribute to very low frequency neural entrainment near 1 Hz or below. It is usually difficult to dissociate these slow event-related responses from low-frequency neural entrainment simply based on the response spectrum/waveform (O’Connell et al., 2012). Nevertheless, the neural entrainment paradigm does not need to isolate the response to a single event and therefore provides a more flexible research paradigm. Furthermore, very low-frequency neural entrainment has been observed in primary sensory areas which are not viewed as the generators of long-latency event-related responses (Lakatos et al., 2008, 2009). To show that entrained activity is distinct from a classic long-latency event-related response, another approach is to show that they have distinct functional properties (Ding et al., 2016).

Therefore, low-frequency neural entrainment at the fundamental frequency of a stimulus rhythm generally implies slow fluctuations in the neural response waveform, which can be viewed as a slow drift in the response “baseline” within each stimulus cycle. Neural activity that oscillates at a rate much faster than the rhythm of the stimulus and transient neural responses that dies off within a small portion of a stimulus period are more strongly reflected by harmonic frequencies in the spectrum. When low-frequency neural entrainment emerges at the fundamental frequency of the stimulus rhythm, it indicates that the influences of previous stimuli do not die out when the next stimulus comes. In other words, previous stimuli set the (neural) context in which a new stimulus will be processed.

In sum, low-frequency neural entrainment implies that either the neural generators have slow intrinsic dynamics or that the neural system continuously tracks certain smoothly changing stimulus properties. It is unlikely that very low-frequency neural entrainment (e.g., <1 Hz) is composed of a series of transient responses evoked by discrete sensory/perceptual events.

AUTHOR CONTRIBUTIONS

ND conceived the study. HZ and ND did the simulations. HZ, LM, DP, and ND wrote the article.

FUNDING

Work supported by National Natural Science Foundation of China 31500873 (ND), Fundamental Research Funds

REFERENCES

- Buiatti, M., Peña, M., and Dehaene-Lambertz, G. (2009). Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. *Neuroimage* 44, 509–551. doi: 10.1016/j.neuroimage.2008.09.015
- Ding, N., Melloni, L., Zhang, H., Tian, X., and Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nat. Neurosci.* 19, 158–164. doi: 10.1038/nn.4186
- Ding, N., and Simon, J. Z. (2012). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *J. Neurophysiol.* 107, 78–89. doi: 10.1152/jn.00297.2011
- Ding, N., and Simon, J. Z. (2014). Cortical entrainment to continuous speech: functional roles and interpretations. *Front. Hum. Neurosci.* 8: 311. doi: 10.3389/fnhum.2014.00311
- Doelling, K. B., and Poeppel, D. (2015). Cortical entrainment to music and its modulation by expertise. *Proc. Natl. Acad. Sci. U S A* 112, E6233–E6242. doi: 10.1073/pnas.1508431112
- Galambos, R., Makeig, S., and Talmachoff, P. J. (1981). A 40-Hz auditory potential recorded from the human scalp. *Proc. Natl. Acad. Sci. U S A* 78, 2643–2647. doi: 10.1073/pnas.78.4.2643
- Giraud, A.-L., and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* 15, 511–517. doi: 10.1038/nn.3063
- Keitel, C., Quigley, C., and Ruhnau, P. (2014). Stimulus-driven brain oscillations in the alpha range: entrainment of intrinsic rhythms or frequency-following response? *J. Neurosci.* 34, 10137–10140. doi: 10.1523/JNEUROSCI.1904-14.2014
- Koskinen, M., and Seppä, M. (2014). Uncovering cortical MEG responses to listened audiobook stories. *Neuroimage* 100, 263–270. doi: 10.1016/j.neuroimage.2014.06.018
- Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., and Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* 320, 110–113. doi: 10.1126/science.1154735
- Lakatos, P., Musacchia, G., O'connel, M. N., Falchier, A. Y., Javitt, D. C., and Schroeder, C. E. (2013). The spectrotemporal filter mechanism of auditory selective attention. *Neuron* 77, 750–761. doi: 10.1016/j.neuron.2012.11.034
- Lakatos, P., O'connell, M. N., Barczak, A., Mills, A., Javitt, D. C., and Schroeder, C. E. (2009). The leading sense: supramodal control of neurophysiological context by attention. *Neuron* 64, 419–430. doi: 10.1016/j.neuron.2009.10.014
- Lankinen, K., Saari, J., Hari, R., and Koskinen, M. (2014). Intersubject consistency of cortical MEG signals during movie viewing. *Neuroimage* 92, 217–224. doi: 10.1016/j.neuroimage.2014.02.004
- Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B. R., and Rossion, B. (2015). The steady-state visual evoked potential in vision research: a review. *J. Vis.* 15:4. doi: 10.1167/15.6.4
- Nozaradan, S., Peretz, I., Missal, M., and Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *J. Neurosci.* 31, 10234–10240. doi: 10.1523/jneurosci.0411-11.2011
- for the Central Universities (ND), Zhejiang Provincial Natural Science Foundation of China LR16C090002 (ND), and USA National Institutes of Health grant 2R01DC 05660 (DP).
- Nozaradan, S., Peretz, I., and Mouraux, A. (2012). Selective neuronal entrainment to the beat and meter embedded in a musical rhythm. *J. Neurosci.* 32, 17572–17581. doi: 10.1523/jneurosci.3203-12.2012
- O'connell, M. N., Barczak, A., Ross, D., McGinnis, T., Schroeder, C. E., and Lakatos, P. (2015). Multi-scale entrainment of coupled neuronal oscillations in primary auditory cortex. *Front. Hum. Neurosci.* 9:655. doi: 10.3389/fnhum.2015.00655
- O'connell, R. G., Dockree, P. M., and Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nat. Neurosci.* 15, 1729–1735. doi: 10.1038/nn.3248
- Oppenheim, A. V., Schafer, R. W., and Buck, J. R. (1989). *Discrete-Time Signal Processing*. Upper Saddle River, NJ: Prentice-Hall, Inc.
- Ross, B., Borgmann, C., Draganova, R., Roberts, L. E., and Pantev, C. (2000). A high-precision magnetoencephalographic study of human auditory steady-state responses to amplitude-modulated tones. *J. Acoust. Soc. Am.* 108, 679–691. doi: 10.1121/1.429600
- Schroeder, C. E., and Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci.* 32, 9–18. doi: 10.1016/j.tins.2008.09.012
- Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., and Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends Cogn. Sci.* 12, 106–113. doi: 10.1016/j.tics.2008.01.002
- Sturm, I., Blankertz, B., Potes, C., Schalk, G., and Curio, G. (2014). ECoG high gamma activity reveals distinct cortical representations of lyrics passages, harmonic and timbre-related changes in a rock song. *Front. Hum. Neurosci.* 8: 798. doi: 10.3389/fnhum.2014.00798
- Tierney, A., and Kraus, N. (2014). Neural entrainment to the rhythmic structure of music. *J. Cogn. Neurosci.* 27, 400–408. doi: 10.1162/jocn_a_00704
- Vaidyanathan, P. P. (1990). Multirate digital filters, filter banks, polyphase networks and applications: a tutorial. *Proc. IEEE* 78, 56–93. doi: 10.1109/5.52200
- Wang, Y., Ding, N., Ahmar, N., Xiang, J., Poeppel, D., and Simon, J. Z. (2012). Sensitivity to temporal modulation rate and spectral bandwidth in the human auditory system: MEG evidence. *J. Neurophysiol.* 107, 2033–2041. doi: 10.1152/jn.00310.2011
- Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., Mckhann, G. M., et al. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party”. *Neuron* 77, 980–991. doi: 10.1016/j.neuron.2012.12.037

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Zhou, Melloni, Poeppel and Ding. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read,
for greatest visibility



COLLABORATIVE PEER-REVIEW

Designed to be rigorous
– yet also collaborative,
fair and constructive



FAST PUBLICATION

Average 85 days from
submission to publication
(across all journals)



COPYRIGHT TO AUTHORS

No limit to article
distribution and re-use



TRANSPARENT

Editors and reviewers
acknowledged by name
on published articles



SUPPORT

By our Swiss-based
editorial team



IMPACT METRICS

Advanced metrics
track your article's impact



GLOBAL SPREAD

5'100'000+ monthly
article views
and downloads



LOOP RESEARCH NETWORK

Our network
increases readership
for your article

Frontiers

EPFL Innovation Park, Building I • 1015 Lausanne • Switzerland
Tel +41 21 510 17 00 • Fax +41 21 510 17 01 • info@frontiersin.org
www.frontiersin.org

Find us on

