NeuroImage xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Delta, theta, beta, and gamma brain oscillations index levels of auditory sentence processing

Guangting Mai [a,b], James W. Minett [a,c,*], William S.-Y. Wang [a,c,d]

[a] Language Engineering Laboratory, Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong Special Administrative Region
[b] Speech Research Lab, Experimental Psychology, Faculty of Brain Sciences, University College London, London, United Kingdom
[c] Joint Research Centre for Language and Human Complexity, Faculty of Arts, The Chinese University of Hong Kong, Hong Kong Special Administrative Region
[d] Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region

ARTICLE INFO

Article history:
Received 23 July 2015
Accepted 21 February 2016
Available online xxxx

Keywords:
Brain oscillations
EEG
Auditory sentence
Phonological processing
Semantic processing
Syntactic processing

ABSTRACT

A growing number of studies indicate that multiple ranges of brain oscillations, especially the delta (δ, <4 Hz), theta (θ, 4–8 Hz), beta (β, 13–30 Hz), and gamma (γ, 30–50 Hz) bands, are engaged in speech and language processing. It is not clear, however, how these oscillations relate to functional processing at different linguistic hierarchical levels. Using scalp electroencephalography (EEG), the current study tested the hypothesis that phonological and the higher-level linguistic (semantic/syntactic) organizations during auditory sentence processing are indexed by distinct EEG signatures derived from the δ, θ, β, and γ oscillations. We analyzed specific EEG signatures while subjects listened to Mandarin speech stimuli in three different conditions in order to dissociate phonological and semantic/syntactic processing: (1) sentences comprising valid disyllabic words assembled in a valid syntactic structure (real-word condition); (2) utterances with morphologically valid syllables, but not constituting valid disyllabic words (pseudo-word condition); and (3) backward versions of the real-word and pseudo-word conditions. We tested four signatures: band power, EEG–acoustic entrainment (EAE), cross-frequency coupling (CFC), and inter-electrode renormalized partial directed coherence (rPDC). The results show significant effects of band power and EAE of δ and θ oscillations for phonological, rather than semantic/syntactic processing, indicating the importance of tracking δ- and θ-rate phonetic patterns during phonological analysis. We also found significant β-related effects, suggesting tracking of EEG to the acoustic stimulus (high-β EAE), memory processing (θ–low-β CFC), and auditory-motor interactions (20-Hz rPDC) during phonological analysis. For semantic/syntactic processing, we obtained a significant effect of γ power, suggesting lexical memory retrieval or processing grammatical word categories. Based on these findings, we confirm that scalp EEG signatures relevant to δ, θ, β, and γ oscillations can index phonological and semantic/syntactic organizations separately in auditory sentence processing, compatible with the view that phonological and higher-level linguistic processing engage distinct neural networks.

Introduction

Cortical oscillatory activity plays a key role in conveying and controlling neural information across the brain, whereby various fundamental cognitive functions, such as attention, learning, memory, and decision-making, are realized (Ward, 2003; Siegel et al., 2012). Brain oscillations are conventionally divided into several frequency ranges: delta (δ, <4 Hz), theta (θ, 4–8 Hz), alpha (α, 8–13 Hz), beta (β, 13–30 Hz), and gamma (γ, >30 Hz) (Ward, 2003). Numerous studies have shown that certain cognitive functions are related to oscillations in multiple frequency ranges. For example, attention is related to changes in α and γ activities (Klimesch, 2012; Jensen et al., 2007), whereas working memory and long-term memory processes involve θ, β, and γ activities (Ward, 2003; Jensen et al., 2007; Fell and Axmacher, 2011). An important topic of human cognitive neuroscience in recent years considers how language is processed via coordination of brain oscillations. The current paper focuses on the auditory modality, and deals with how brain oscillations underpin auditory sentence processing. Previous studies have accumulated evidence that speech and auditory sentence processing are associated with multiple ranges of brain oscillations, including both low-frequency components, such as δ and θ oscillations, and high-frequency components, such as β and γ oscillations (see reviews: Giraud and Poeppel, 2012; Lewis et al., 2015).

Abbreviations: EEG, Electroencephalography; MEG, Magnetoencephalography; EAE, EEG–acoustic entrainment; CFC, Cross-frequency coupling; rPDC, Renormalized partial directed coherence; MI, Modulation index; SUS, Semantically unpredictable sentence; WM, Working memory.

* Corresponding author at: Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong Special Administrative Region.
E-mail address: jminett@ee.cuhk.edu.hk (J.W. Minett).

http://dx.doi.org/10.1016/j.neuroimage.2016.02.064
1053-8119/

For low-frequency components (i.e., δ and θ), recent findings showed that the phase information of the δ and θ oscillations are involved in speech perception. The δ and θ (i.e., 1–8 Hz) phase measured by magnetoencephalography (MEG) can be used to successfully classify different auditory sentences attended to by subjects (θ phase in Luo and Poeppel, 2007; δ and θ phase in Cogan and Poeppel, 2011). In an electroencephalography (EEG) study, the phase restricted to 2–9 Hz (which overlaps the δ and θ bands) can successfully classify different American English consonants (Wang et al., 2012). In connection with such findings on the importance of δ/θ phase, two other recent neurophysiological studies have found that entrainment (i.e., phase-locking) of δ and θ brain oscillations to the speech envelope at the corresponding δ and θ amplitude-modulation rates may underpin speech intelligibility and serve as one of the neural mechanisms of speech processing (Peelle et al., 2013; Doelling et al., 2014). Peelle et al. (2013) found that the degree of θ (4–7 Hz) MEG-envelope entrainment was related to sentence intelligibility observed in the left auditory cortex and middle temporal gyrus. Doelling et al. (2014) artificially removed the δ- and θ-rate (2–9 Hz) envelopes of sentences in various acoustic spectral bands and consequently found that the δ–θ MEG-envelope entrainment was suppressed, accompanied by a reduction in sentence intelligibility. The correlation between brain–acoustic entrainment in the δ–θ range and speech intelligibility thus emphasizes the importance of δ and θ brain oscillations in auditory sentence processing (see review by Ding and Simon, 2014).

Besides involvement in brain–acoustic entrainment, the power of low-frequency components was also found to be important for speech processing. For instance, Peña and Melloni (2012) used a cross-linguistic design to compare the EEG oscillations elicited from Italian and Spanish speakers while listening attentively to Italian, Spanish, and Japanese utterances played both forward and backward. This study found that, in both Italian and Spanish subjects, θ power was significantly higher when listening to forward than to backward utterances, regardless whether or not the language was native. The finding that forward utterances elicit higher θ power than backward utterances, even for a non-native language, thus indicates that θ power may be involved in tracking syllable patterns (Peña and Melloni, 2012). In a more recent MEG study (Ding et al., 2015), similar results were found which showed that, when listening to Chinese sentences with syllable rate of around 4 Hz, both native Chinese or English listeners showed significantly higher 4-Hz MEG power for forward sentences than for the backward versions. Considering that backward utterances preserve properties that are closely matched to the acoustic complexity of speech utterances but cause serious phonological distortions (Binder et al., 2000; Saur et al., 2010; Gross et al., 2013), syllabic tracking in speech utterances may involve a higher degree of phonological analysis compared to backward utterances, even in a non-native language. Studies have found that θ oscillations are also involved in lexical–semantic retrieval (Bastiaansen et al., 2008) and in syntactic processing during sentence perception (Bastiaansen et al., 2002), the former involving retrieval of long-term semantic knowledge and the latter involving working memory processing.

For high-frequency components, such as β and γ oscillations, there is evidence that brain oscillations in this range are involved in different linguistic processes. A recent MEG study (Alho et al., 2014) investigated the inter-areal phase synchronies of high-β (β2, 20–30 Hz) and γ oscillations between the auditory and motor cortices during active and passive listening to phonologically valid but meaningless mono-syllables in both clean and noisy environments. It showed that the left-hemispheric inter-areal β2 synchronies were significantly greater during syllable listening in noisy than in clean environments and that such synchronies were positively correlated with syllable identification accuracy. Furthermore, inter-areal γ synchronies were found to be greater during active than passive listening. This indicates the mediation of phonological categories in speech by inter-areal connectivity between auditory–sensory and motor regions via β2 and γ oscillations. For higher linguistic-level processing, β oscillations were reported to be involved in syntactic processing, showing higher EEG β power for syntactically correct than syntactically unstructured and word category violated sentences (Bastiaansen et al., 2010; also reviews by Lewis and Bastiaansen, 2015; Lewis et al., 2015). In addition, γ oscillations were reported to be involved in lexico-semantic retrieval (Lutzenberger et al., 1994; Pulvermüller et al., 1996). These studies found significant increases in γ oscillations when subjects actively perceived real-word compared to pseudo-word stimuli in both visual (Lutzenberger et al., 1994) and auditory (Pulvermüller et al., 1996) modalities, which is consistent with the critical role of γ activity in long-term memory processing (Ward, 2003).

In addition to the respective roles of δ, θ, β, and γ oscillations, the hierarchical organization between the low-frequency and high-frequency oscillations, termed cross-frequency coupling (CFC), serves as another important parameter for speech processing (Fell and Axmacher, 2011; Lisman and Jensen, 2013). Here, we focus on phase-power CFC, in which the power of high-frequency oscillations is controlled by the phase patterns of low-frequency oscillations (Tort et al., 2008). It has been found that θ–β/γ CFC increased significantly across a range of human cortical regions during various cognitive tasks, including language-related tasks, such as active/passive listening to phonemes and words, word production, visual reading, and so on (Canolty et al., 2006). The phenomenon of θ–β/γ CFC increase has been interpreted in other studies as the neural mechanism for memory processing, including encoding and retrieval of long-term memory and working memory maintenance in both non-human mammals (Tort et al., 2008, 2009; Shirvalkar et al., 2010) and human beings (Mormann et al., 2005; Sauseng et al., 2009; Axmacher et al., 2010; Friese et al., 2013; Köster et al., 2014; Kaplan et al., 2014). It is likely, therefore, that θ–β/γ CFC is related to high-level linguistic processes like phonological working memory maintenance and retrieval of lexical–semantic information, or even sentence-level processes related to memory retrieval or encoding (e.g., contextual semantic integration and syntactic processing). Furthermore, it has recently been suggested that θ–β/γ CFC supports the hierarchical binding of both long-duration (such as syllables and long phonemes, e.g., long-vowels, at θ-scale) and short-duration (such as short phonemes, e.g., consonants and short-vowels, at β/γ-scale) phonological information during speech analysis (Giraud and Poeppel, 2012; Gross et al., 2013). Besides θ–β/γ CFC, the coupling between δ and θ oscillations (δ–θ CFC) may also be important. δ–θ CFC was found to be higher when listening to forward than to backward utterances, indicating a possible role of hierarchical binding between even longer-duration information of prosody or phrases/words (at δ-scale) and the θ-scale information in speech perception (Gross et al., 2013), although one should be cautious when interpreting the δ–θ CFC effects due to the close frequency ranges between δ and θ oscillations that could cause intrinsic coupling effects mathematically.

In spite of the abundant findings on brain oscillations to describe language processing as reviewed above, few studies have examined these oscillatory indices for different linguistic hierarchical levels simultaneously. How brain oscillations index and separate processes at these levels therefore remains obscure. The current study aims at revealing oscillatory EEG indices for phonological and higher-level linguistic (semantic/syntactic) processing during listening to auditory sentences in Mandarin. We used three types of continuous utterance stimuli in Mandarin in order to dissociate the effects caused by acoustics, phonology, and the higher linguistic levels: (1) sentences consisting of meaningful disyllabic words assembled with a valid syntactic structure (real-word condition); (2) utterances with morphologically valid syllables, but no valid disyllabic words (pseudo-word condition); and (3) backward versions of both the real-word and pseudo-word utterances ('non-speech' condition). In this design, real-word and pseudo-word utterances can be distinguished by their differences in semantic content. For example, in the real-word condition, the syllable pair, '喜' and '欢', constitutes a disyllabic word, '喜欢' ('enjoy'), while in the pseudo-word condition, the two successive syllables, '书' and '实', do

Please cite this article as: Mai, G., et al., Delta, theta, beta, and gamma brain oscillations index levels of auditory sentence processing, NeuroImage (2016), http://dx.doi.org/10.1016/j.neuroimage.2016.02.064

not form a meaningful disyllabic word (i.e., a pseudo-word '书实', see more detailed examples in 'Stimuli and tasks' section). Thus, the real-word condition involves semantic integration of two successive syllables into a meaningful word compared to the pseudo-word condition. Also, as real-word utterances have a valid syntactic structure that pseudo-word utterances do not have, the real-word condition also involves syntactic processing compared to the pseudo-word condition. Since both real-word and pseudo-word utterances are composed of morphologically valid syllables, we also designate them as 'speech' conditions. The backward utterances are closely matched in terms of acoustic complexity to their respective speech utterances, thereby providing a control condition by which to dissociate the psychoacoustic processing of speech-like physical properties from speech-specific processing (Binder et al., 2000; Londei et al., 2010; Saur et al., 2010; Peña and Melloni, 2012; Gross et al., 2013). We consider that EEG indices with statistically greater magnitude in the real-word condition than in the pseudo-word condition involve semantic/syntactic processing. Meanwhile, we consider indices with greater magnitude for speech than for non-speech, but with no statistical difference between real-word and pseudo-word, to be signatures relevant to phonological rather than semantic/syntactic processing. We thus focus on two types of comparisons: (1) speech (real-word plus pseudo-word) vs. non-speech (backward) condition; and (2) real-word vs. pseudo-word condition.

Based on suggestions in previous studies, that processing of phonology and higher linguistic levels engage different anatomical and functional neural networks (Saur et al., 2010) that can be segregated into different brain oscillations (McNab et al., 2012, with stimuli of visual nouns), we hypothesize that phonological and higher-level linguistic organization in auditory sentence processing can be separately indexed by EEG signatures relevant to δ (2–4 Hz), θ (4–8 Hz), β (13–30 Hz), and γ (30–50 Hz) oscillations. Specifically, we propose the following four EEG parameters, which have been studied in previous research related to speech and language processing, as candidates of the signatures for our current hypothesis. Firstly, we propose to use power changes in brain oscillations, as has been commonly used in previous studies of phonological and semantic/syntactic functions (e.g., Pulvermüller et al., 1996; Bastiaansen et al., 2008; Peña and Melloni, 2012). Secondly, we propose to use EEG-acoustic entrainment (EAE) as an index for auditory sentence processing. As reviewed above, δ to θ brain–acoustic entrainment has been shown to co-vary with speech intelligibility (Peelle et al., 2013; Doelling et al., 2014). However, seen simply from changes in speech intelligibility, what linguistic hierarchical levels (phonological or semantic/syntactic processing) are involved and how these hierarchical levels respectively influence brain–acoustic entrainment remain unclear. Thirdly, we propose to use δ–θ and θ–β/γ cross-frequency couplings (CFC) as candidates for indexing phonological and semantic/syntactic computations and memory processes, such as working memory and long-term memory retrieval in phonological and semantic/syntactic processing, and hierarchical binding between speech components in different timescales (prosody/phrases/words, syllables, phonemes). Finally, based on the findings of auditory-motor inter-areal connectivity embodied via β and γ synchronies during phonological processing (Alho et al., 2014) and large-scale connectivity brain networks for lexical–semantic retrieval (Patterson et al., 2007; Pulvermüller, 2013), we propose to use renormalized partial directed coherence (rPDC) in the β/γ range as an index to describe neural connectivity during phonological and/or semantic processing, where rPDC is a quantitative method for calculating the extent of directed connectivity between brain regions (Schelter et al., 2009).

Methods

Subjects

21 subjects (8 males and 13 females, aged 19–25 years old), all undergraduate or postgraduate students of the Chinese University of Hong Kong, consented to participate in this study. All subjects were native Mandarin speakers from mainland China with normal-hearing, as confirmed by a monaural pre-test on both ears. Data for one subject (female) were not used for further analysis due to the excessive percentage of trial rejection (>60% target trials were rejected) due to eye artifacts (while <30% trials were rejected in every other subject). Of the 20 subjects whose data were used for analyses, 18 were right-handed (all with handedness indices (HI) > 40) and 2 were ambidextrous (both with HI of 33.3) according to the Edinburgh Handedness Inventory (Oldfield, 1971).

Stimuli and tasks

Stimuli were auditory utterances comprising three conditions: (1) real-word utterances; (2) pseudo-word utterances; and (3) backward versions of (1) and (2). Real-word and pseudo-word utterances were designated as 'speech' and were naturally produced by an adult, male native Mandarin speaker with syllable rate approximately 4 Hz (i.e., about 250 ms per syllable) recorded at the sampling rate of 22,050 Hz in a quiet environment using the software PRAAT (Boersma and Weenink, University of Amsterdam). The real-word utterances consisted of semantically unpredictable sentences (SUSs), which are syntactically acceptable but semantically anomalous sentences (Benoit et al., 1996). Each SUS included four valid disyllabic (2-character) words, e.g., a sample SUS is '网络喜欢坚强的空气', in which the disyllabic words are '网络' ('Internet'), '喜欢' ('enjoy'), '坚强' ('tough'), and '空气' ('air'); the other syllable, '的', is a grammatical particle without substantive meaning. All SUSs had the syntactic structure 'Subject + Verb + Attribute + 的 + Object'. The purpose of using SUSs was to prevent the subjects from inferring the sentence content via contextual information and to guarantee that they attended to the entire utterance. Pseudo-word utterances were sentences consisting of morphologically valid syllables, but with no two adjacent syllables forming a valid word, e.g., '书实生宇树飞的视身' with the particle '的' at the same position as in the real-word condition (i.e., the seventh syllable in each utterance). In this example, the first two adjacent syllables '书' and '实' failed to form a meaningful word. Backward utterances were the time-reversed versions of the 'speech' utterances and were designated as 'non-speech'. Previous studies (Binder et al., 2000; Londei et al., 2010; Saur et al., 2010; Peña and Melloni, 2012; Gross et al., 2013) have also used backward speech as the non-speech baseline in order to retain similar acoustic complexity (e.g., speech-like temporal fluctuations, formant distributions, and harmonic structures) as the speech versions in spite of the substantial phonological distortion. Most subjects reported that backward utterances sound like utterances of an unknown foreign language (e.g., Saur et al., 2010; Peña and Melloni, 2012). There were 80 real-word and 80 pseudo-word utterances with different contents and without repetition of any same utterance during each experiment session. Half of the real-word and half of the pseudo-word utterances were time-reversed to generate the backward utterances, hence 80 backward utterances in total. The stimuli all had a similar duration (2200–2300 ms), contained 9 syllables (in the forward 'speech' condition), and were adjusted to the same average intensity (~70 dB SPL). During the experiments, all subjects listened to the same sets of real-word, pseudo-word, and backward stimuli.

Subjects were seated in front of a computer screen listening to the acoustic stimuli with EARTONE 3-A insert earphones (Etymotic Research, Illinois, USA) via a Scarlett 2i2 sound card (Focusrite Audio, Bucks, UK) and were presented with a visual paradigm controlled by E-Prime 2.0 (Psychology Software Tools, Pennsylvania, USA). The experimental procedure of each trial is shown in Fig. 1(A). At the start of each trial, there was a 1.5-s silence during which time a white cross was centered on the screen for eye fixation. Then a cue sound (a naturally produced syllable or a backward syllable) was played and it was followed by another 2-s silence after which the target utterance was played. A forward cue sound was always followed by a 'speech' utterance, while

Please cite this article as: Mai, G., et al., Delta, theta, beta, and gamma brain oscillations index levels of auditory sentence processing, NeuroImage (2016), http://dx.doi.org/10.1016/j.neuroimage.2016.02.064

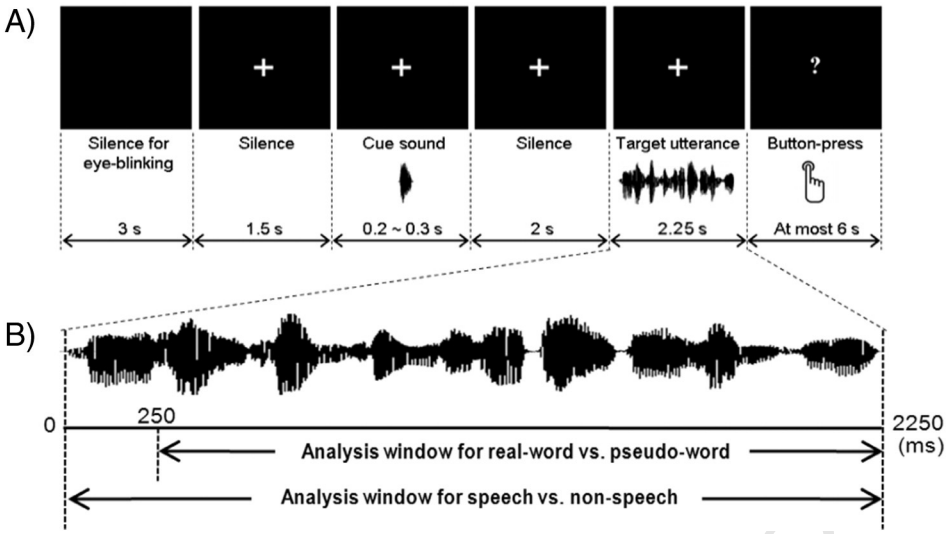

Fig. 1. *The experimental paradigm.* (A) The experimental procedure in each trial; (B) analysis windows for speech vs. non-speech (0–2250 ms) and for real-word vs. pseudo-word (250–2250 ms).

a backward cue sound was always followed by a backward utterance. The task for the subject was to make a single forced-choice judgment whether the cue sound was present in the target utterance by making a button press when a question mark was presented on-screen after each utterance. To counterbalance the possible effects of the motor responses on EEG, half of the subjects were instructed to press a specified button on the left side of the keyboard as 'yes' and a button on the right side as 'no', the buttons being reversed for the other half of the subjects. Prior to each trial, there was a 3-s silent period, during which time subjects could blink. The experiment consisted of 8 blocks, each comprising 10 trials per condition (80 real-word, 80 pseudo-word, and 80 backward utterances in total) in random order. There were only 20% of the trials in which cue sounds were actually present in the target stimuli. Average response accuracies were shown to the subjects on the screen after they finished each block, and they were all encouraged to respond as accurately as possible. The purpose of the task was to keep subjects alert during presentation of the target stimulus. To exclude the possibility that subjects did not attend to the entire utterance period and to avoid auditory repetition effects, trials in which cue sounds appeared in the target stimuli were not used in the subsequent analysis, hence 64 stimuli per condition (80% of 80 trials) prior to EEG artifact rejection.

Presentations of sample stimuli and two practice blocks preceded each formal experiment. Subjects were allowed to take breaks between blocks. Before the experiments as well as during the breaks they were reminded to sit still, to keep their eyes focused on the white fixation cross and to avoid eye-blinks and muscle movements during the target period.

Data processing and analyses

Accuracies and reaction times of subjects' behavioral responses were recorded by E-Prime 2.0 via button press. EEG data were recorded with a 32-electrode ActiveTwo system (10–20 system, BioSemi, Amsterdam, The Netherlands) sampled at 1024 Hz and downsampled to 512 Hz before subsequent analyses. The electrode configuration is shown in Fig. 2. Eye artifacts were detected through both horizontal and vertical EOG electrodes. Signal processing was conducted using Matlab 2010a (The Mathworks Inc., Massachusetts, USA) and statistical analyses were conducted using SPSS 13.0 (SPSS Inc., Illinois, USA).

Within-subject factors and guidelines for statistical analyses

In the current study, statistical analyses were conducted on various EEG signatures relevant to five frequency ranges: δ (2–4 Hz), θ (4–8 Hz), β1 (13–20 Hz), β2 (20–30 Hz), and γ (30–50 Hz). Within-subject repeated-measures ANOVAs were applied to the data obtained from the 20 subjects. We defined two within-subject factors: 'Condition' and 'Region'. Levels of 'Condition' were designated according to the two comparisons focused on in this paper (see Introduction): (1) speech (real-word and pseudo-word) vs. non-speech (backward) condition; and (2) real-word vs. pseudo-word condition. For (1), the analysis window was set from the stimulus onset to 2250 ms after onset (i.e., the approximate offset of each stimulus). For (2), the starting point of the analysis window was set to 250 ms after stimulus onset based on the assumption that there should be no lexico-semantic or sentence-level differences between the real-word and pseudo-word conditions within the first syllable period (i.e., the first 250 ms of each target stimulus) (see Fig. 1(B)). For the levels of 'Region', we grouped the 32 EEG electrodes into 8 regions, each of which contained 4 neighboring electrodes (see Fig. 2): Left Frontal, Fp1, AF3, F7, F3; Right Frontal, Fp2, AF4, F8, F4; Left Temporal, FC5, T7, C3, CP5; Right Temporal, FC6, T8, C4, CP6; Left Parieto-Occipital, P7, P3, PO3, O1; Right Parieto-Occipital, P8, P4, PO4,

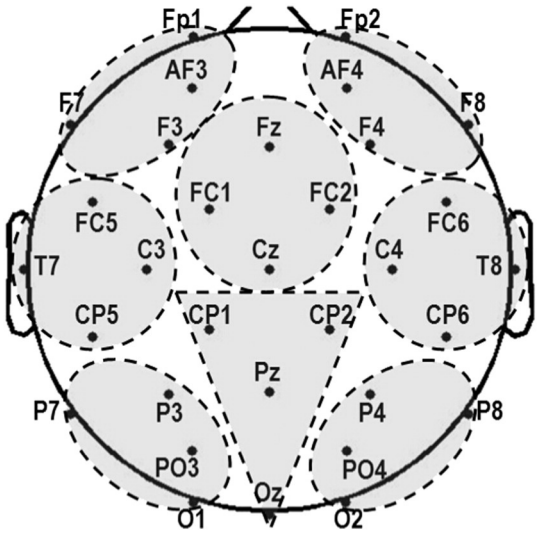

Fig. 2. *Channel configuration.* The 32 channels were divided into 8 regions during the statistical analyses as indicated by the grey ellipses/triangle.

O2; Mid Centro-Frontal, Cz, FC1, FC2, Fz; Mid Parieto-Occipital, CP1, CP2, Pz, Oz. The ANOVAs were then conducted accordingly with Greenhouse–Geisser correction.

Following guidelines for analysis in the two-factor design for repeated-measures ANOVA (Maxwell and Delaney, 2004), interaction between Condition and Region was examined first. In cases that an interaction was found to be significant, post hoc pairwise comparisons between different levels of Condition (speech vs. non-speech or real-word vs. pseudo-word) within individual regions were then conducted; otherwise, the main effects of Condition were examined.

Preprocessing

The EEG signals in each block were first re-referenced to the temporal average of the left and right mastoid electrodes and then passed through a high-pass filter at 0.5 Hz and a 1-Hz-width notch filter centered at 50 Hz (frequency of AC) and at its odd harmonics (150 and 250 Hz). The filtering was based on the least-square FIR filter design, followed by a zero-phase (all-zero) FIR filter (functions 'firls' and 'filtfilt' in the Matlab Signal Processing Toolbox). The signals were then segmented, each segmented epoch containing the target utterance (fixed to the length of 2250 ms) of a particular trial plus a 500-ms pre-stimulus baseline. Moreover, each epoch also included a 500-ms period before the pre-stimulus baseline as well as a 500-ms period after the target stimulus offset. The purpose of such inclusion was to reduce the possibility of border artifacts within the −500 to 2250 ms interval during application of filtering in the subsequent signal processing.

During artifact rejection, trials in which absolute amplitude exceeded 100 μV in the vertical EOG (top minus bottom vEOG), 50 μV in the horizontal EOG (left minus right hEOG), or 80 μV at any scalp electrode at any time point between −500 and 2250 ms were rejected. In addition, muscle movements can produce abnormally large amplitude at high frequencies (>20 Hz) (Muthukumaraswamy, 2013) that could contaminate the data. We thus further band-pass filtered the EEG signals to 20–50 Hz and extracted the maximal absolute amplitude out of all 32 scalp electrodes between −500 and 2250 ms of each trial. We then applied studentized residuals (Meyers et al., 2006) for every individual subject within each condition (real-word, pseudo-word, and non-speech, respectively) to reject trials which contained abnormally large amplitudes that were possibly generated by muscle movements (see *S1 Supplementary Methods* in the *Supplementary Materials* for details).

EEG power

The EEG signal in each segmented period was band-pass filtered into δ, θ, β1, β2, and γ bands. The filter type was the same as in the section Preprocessing, and the filter order was determined by the lower-cutoff frequency, which was three times the ratio of the sampling frequency to the lower-cutoff frequency (c.f. Tort et al., 2008, 2009, 2010; Scheffer-Teixeira et al., 2012; Yanovsky et al., 2014). However, because the filter implementations in Matlab require that the filter order should not exceed 1/3 of the length of the filtering period, the filter order was then set to 1/3 of the length of the filter period once the previous setting did not fulfill such requirement. The power in each band was calculated via the Hilbert transform (square of Hilbert envelope) and then normalized,

$$\sigma_k = (S_k - \mu)/\Delta,$$

where S_k and μ refer to the power during the target period at the kth sampling point and the average baseline power, respectively, and Δ refers to the standard deviation across the baseline period. The normalized power was then averaged across the analysis window for each trial. In the statistical analyses, within-subject ANOVAs with factors of Condition and Region were conducted separately for the δ, θ, β1, β2, and γ bands.

EEG–acoustic entrainment (EAE)

EEG–acoustic entrainment (EAE) was quantified as the temporal correlations between the EEG signals and the envelope profile of the acoustic stimulus in each trial. Each acoustic stimulus was restricted to 100–4000 Hz before the envelope profile was extracted. The correlation values for the δ, θ, β1, β2, and γ bands were calculated separately.

For the low-frequency components of δ and θ, the EEG waveforms were band-passed (same filtering method as described in the section EEG power) at the range of 2–8 Hz in 0.5-Hz steps with 2-Hz bandwidths. The acoustic stimulus envelopes were extracted via Hilbert transform and decimated to the same sampling rate as the EEG, and then band-passed in the same way. The entrainment between the EEG waveforms and the stimulus envelope at the same rates were then measured via Pearson correlation for each frequency step. Kong et al. (2014) showed that the greatest EEG–speech correlations occur at time lags of <200 ms when actively attending to speech stimuli. Considering this possible time lag of EEG tracking of the acoustic stimulus, we calculated the correlation coefficient for each lag (in 10-ms steps) ranging from 20 to 220 ms and then chose the maximal coefficient as the EAE value (Ahissar et al., 2001; Kong et al., 2014; also see *Fig. S1* in *S1 Supplementary Methods* in the *Supplementary Materials*). Furthermore, all EAE values were Fisher-transformed in every trial (Silver and Dunlap, 1987). In the subsequent statistical analyses, δ EAE and θ EAE were obtained by averaging the EAE values across the frequency steps within the δ range (2–4 Hz) and θ range (4–8 Hz), respectively.

For the high-frequency components of β and γ, unlike the δ and θ bands, phases of which have been shown to be entrained to speech envelopes (Peelle et al., 2013; Doelling et al., 2014), no β or γ phase-locking effects were found in speech comprehension (Luo and Poeppel, 2007). On the other hand, however, time series of high-frequency power have been found to be involved in speech envelope tracking (Nourski et al., 2009; Golumbic et al., 2013; Kubanek et al., 2013; Fontolan et al., 2014). We therefore quantified the β1, β2, and γ EAE by correlating the power profiles of EEG with the stimulus acoustic envelope profiles. The time series of EEG power profiles were firstly generated by extracting the RMS energy of the band-passed EEG waveforms at β1, β2, and γ, respectively, with a sliding window of 80 ms in 10-ms steps. The time series of the stimulus envelope profiles were generated by extracting the RMS energy of each stimulus waveform also with an 80-ms sliding window in 10-ms steps and then correlated with the corresponding β1, β2, and γ EEG power profiles, respectively (cf., Kubanek et al., 2013). The 80-ms window length was chosen to be both short enough to capture the general variations of the stimulus envelopes and long enough so that each window contained at least one cycle of the EEG signal (one cycle of β1 EEG < 80 ms) for estimating EEG power profiles. Furthermore, as moving average using 80-ms sliding windows is equivalent to applying a low-pass filter with cutoff <6 Hz (at −3 dB rolloff) (Smith, 1999), this manipulation thus retained slow-varying components of the β and γ EEG power information. Like for δ and θ EAE, the maximal Pearson correlation value was chosen among the values obtained at different time lags ranging from 20 to 220 ms (in 10-ms steps) and then Fisher-transformed for each trial.

Furthermore, following Peelle et al. (2013) in calculating cerebro-acoustic coherence, we also generated a random-level correlation as a baseline for each condition. A 32-'electrode' random matrix was created for each trial which consisted of 32 zero-mean, Gaussian pseudorandom signals (using the Matlab function 'randn') with length equal to that of the analysis period in real EEG. The entrainment between random signals and the stimulus acoustic envelope was calculated for the δ, θ, β1, β2, and γ bands, respectively, following the same procedure as calculating EAE. We then obtained the baseline via averaging all 32 resultant random-level correlations for each trial. During the statistical analyses, Within-subject ANOVAs were conducted with the factors of Condition ((speech − $\text{random}_{\text{speech}}$) vs. (non-speech − $\text{random}_{\text{non-speech}}$), or (real-word − $\text{random}_{\text{real-word}}$) vs. (pseudo-word − $\text{random}_{\text{pseudo-word}}$)) and Region.

Please cite this article as: Mai, G., et al., Delta, theta, beta, and gamma brain oscillations index levels of auditory sentence processing, NeuroImage (2016), http://dx.doi.org/10.1016/j.neuroimage.2016.02.064

Cross-frequency coupling (CFC)

CFC was quantified as the values of modulation index (MI) based on the normalized entropy of the high-frequency power according to the low-frequency phase (cf. Tort et al., 2008). Low- and high-frequency signals were first obtained via band-pass filtering within each segmented epoch (same filtering method as in the section EEG power). The time series of both the low-frequency phase and high-frequency power were obtained using the Hilbert transform. The low-frequency phases were then split uniformly into 18 phase bins (each of 20 degrees) across the target period in each trial. Next, the mean high-frequency power within each phase bin was measured and normalized to the scale from 0 to 1. The entropy of the high-frequency power according to the low-frequency phase was calculated to obtain the final MI value (for complete descriptions of procedure and formulae, see *S1 Supplementary Methods* in the *Supplementary Materials*; also see Tort et al., 2008). Matrices, in which elements represent the MI values for all low- and high-frequency pairs, were then generated for every trial. For calculating θ–β/γ CFCs, the low-frequency (frequency for phase) θ band ranged from 4 to 8 Hz in 0.5-Hz steps with 2-Hz bandwidth, and the high-frequency (frequency for power) β and γ bands ranged from 13 to 50 Hz in 1-Hz steps with 4-Hz bandwidth. For calculating δ–θ CFCs, the low-frequency δ band ranged from 2 to 3.5 Hz in 0.5-Hz steps with 1-Hz bandwidth, and the high-frequency θ band ranged from 4 to 8 Hz in 1-Hz steps with 2-Hz bandwidth. CFCs in each trial were finally calculated by averaging the MI values across the elements in the MI matrix that encompassed the corresponding ranges. Within-subject ANOVAs with the factors of Condition and Region were conducted in the statistical analyses for θ–β1, θ–β2, θ–γ, and δ–θ CFC, respectively.

Renormalized partial directed coherence (rPDC)

PDC is a promising method in neuroscience capable of detecting the directional connectivity between different brain regions/EEG electrodes based on Granger causality of forecasting outflow signal from one region (or electrode) to other region(s) (or electrode(s)) (Schelter et al., 2005). Although PDC can detect significant connectivity, it does not allow conclusions on the absolute strength of the connectivity, and thus is not suitable for comparing the connectivity strength of different variables and conditions, which weakens its interpretability (Schelter et al., 2009). In order to overcome this drawback, Renormalized PDC (rPDC) was developed (Schelter et al., 2009), which has been applied to compare connectivity strengths in several recent EEG studies (e.g., Elshoff et al., 2013; Michels et al., 2013; Japaridze et al., 2013; Maksimow et al., 2014).

rPDC is based on the framework of the vector autoregressive (VAR) model,

$$X(t) = \sum_{r=1}^{p} a(r)X(t-r) + \varepsilon(t)$$

where $X(t)$ is the N dimensional vector, $(X_1(t), X_2(t), \ldots X_N(t))^{\mathrm{T}}$ (T denotes the transpose) at time point t, where $X_i(t)$ is the zero-mean serial of ith electrode, $a(r)$ is the coefficient matrix of the VAR model at delayed time step r, in which each element $a_{ij}(r)$ represents the contribution of the jth electrode (delayed at r), $X_j(t-r)$, to the current ith electrode, $X_i(t)$, $\varepsilon(t)$ denotes an N dimensional vector of Gaussian noise, and p denotes the VAR model order which determines the spectral resolution (e.g., if $p = f_s \times M$, where f_s is the EEG sampling frequency and M is the delayed period in time, then the spectral resolution is defined as $1/M$). In order to avoid over-parameterization, the number of variables ($p \times N^2$) should not exceed the total number of data points ($N \times T$, where T is the number of sampling points in the target period)—a balance between the required spectral resolution (determined by p) and number of electrodes N should be carefully considered. In the current study, we selected the spectral resolution to be 10 Hz (i.e., $p = f_s \times 100\ ms$), allowing us to examine rPDC at the β–γ range (i.e., 20, 30, and 40 Hz). To avoid over-parameterization, we selected 16 electrodes for the rPDC analysis (Fp1, Fp2, F3, Fz, F4, T7, C3, Cz, C4, T8, P3, Pz, P4, O1, Oz, and O2 according to the configuration of the 16-electrode ActiveTwo system layout, corresponding to 2 channels for each of the 8 regions, see Fig.2), rather than all 32 electrodes. The VAR model coefficient, $a(r)$, and noise, $\varepsilon(t)$, were then estimated by multivariate least squares (MLS). rPDC values between the 16 selected electrodes were then calculated based on the estimated $a(r)$ and $\varepsilon(t)$. Complete procedures and formulae for calculating the rPDC values are described in *S1 Supplementary Methods* in the *Supplementary Materials* (also see Schelter et al., 2009).

Furthermore, a random-level baseline of rPDC was generated by creating 200 random trials, each of which consisted of 16 zero-mean, Gaussian pseudorandom signals with length equal to that of the analysis period in real EEG, similar in the section EEG–acoustic entrainment. The rPDC values for each random trial were calculated via the same procedure as in real EEG and were averaged across all 'electrode' pairs (16×15 pairs in total). The final random-level rPDC was obtained by averaging values across all 200 trials. During the statistical analysis, above-random-level rPDCs were extracted for each condition (speech, non-speech, real-word, and pseudo-word, respectively) by comparing rPDCs in these 4 conditions with the random-level baseline using within-subject pairwise T-tests (2-tailed). We defined the directed coherence from the jth electrode to the ith electrode, $rPDC_{ij}$, as above-random-level if its value was significantly higher than the random-level rPDC at the significance level $p < 0.01$ (with Bonferroni correction according to the total number of electrode pairs, i.e., 16×15). To refine the range of analysis, we only included those rPDC pairs for further analysis, each of which was classified as above-random-level in either of speech and non-speech for speech vs. non-speech (in either of real-word and pseudo-word for real-word vs. pseudo-word). Within-subject ANOVAs with the factor of Condition were then conducted for rPDCs at 20, 30, and 40 Hz, respectively.

Results

After EEG artifact rejection, over 70% of the trials were retained for further analysis in all 20 subjects. The average number of retained trials were 56.4 (SE: 1.2), 56.0 (SE: 1.2), and 56.8 (SE: 1.2) for the real-word, pseudo-word, and non-speech conditions, respectively.

Behavioral results

The response accuracies and reaction times (RTs) of the behavioral judgments were compared across conditions via within-subject T-tests. RT in each trial was calculated as the interval between visual onset of the question mark and the button-press (see Fig. 1). Trials that were not used in further EEG analyses were excluded for the behavioral calculation. The results are shown as Fig. S2 in the *Supplementary Materials*. For the response accuracy (Fig. S2(A)), results for both the real-word and pseudo-word conditions were significantly higher than that for the backward (non-speech) condition (both $p < 10^{-8}$), while no significant difference of accuracy was found between the real-word and pseudo-word conditions ($p > 0.2$). Accuracies in all conditions were significantly higher than the 50% chance-level ($p < 10^{-27}$ for real-word, $p < 10^{-25}$ for pseudo-word, and $p < 10^{-8}$ for non-speech). For RT (Fig. S2(B)), significant longer RT was found for non-speech than for the real-word ($p < 0.001$) and pseudo-word ($p < 10^{-4}$) conditions, while there was no significant difference of RT between real-word and pseudo-word ($p > 0.1$).

EEG band power

Results for δ, θ, β1, β2, and γ power are illustrated in Fig. 3.

For speech versus non-speech (Fig. 3(A)), a significant Condition × Region interaction was found for δ power ($p < 0.004$). Post hoc pairwise

Please cite this article as: Mai, G., et al., Delta, theta, beta, and gamma brain oscillations index levels of auditory sentence processing, NeuroImage (2016), http://dx.doi.org/10.1016/j.neuroimage.2016.02.064

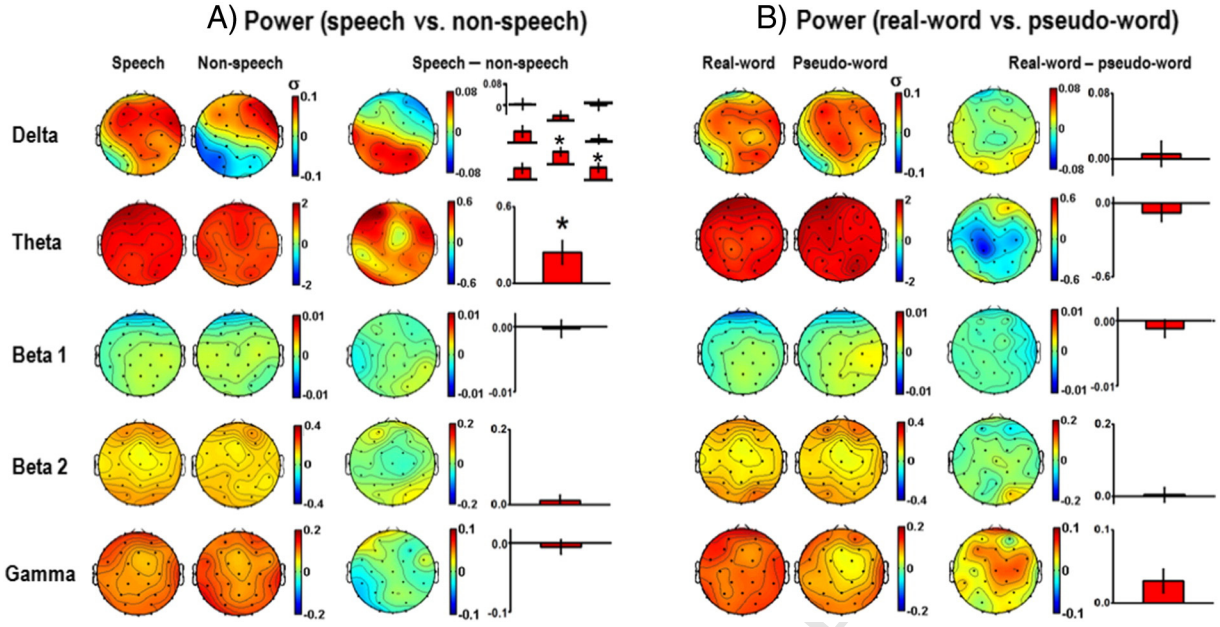

Fig. 3. *Results for δ, θ, β1, β2, and γ EEG power.* Because of the significant Condition × Region interaction in δ power for speech versus non-speech, post hoc pairwise comparisons were undertaken in all 8 regions. Asterisks denote the significance levels $p < 0.05$ (Bonferroni corrections have been conducted during post hoc pairwise comparisons in individual regions).

comparisons were therefore conducted in individual regions, showing that δ power was higher for speech than for non-speech in the Right Parieto-Occipital ($p < 0.006$) and Mid Parieto-Occipital ($p < 0.003$) regions (significant at the level of 0.05 after Bonferroni correction across the 8 regions). For θ power, there was no significant Condition × Region interaction ($p > 0.6$), but a significant main effect of Condition was observed with higher θ power for speech than for non-speech ($p < 0.02$). For β1, β2, and γ power, neither Condition × Region interactions (β1 ($p > 0.5$), β2 ($p > 0.2$), γ ($p > 0.2$)) nor main effects of Condition (β1 ($p > 0.8$), β2 ($p > 0.4$), γ ($p > 0.5$)) were found.

For the real-word versus pseudo-word comparison (Fig. 3(B)), neither Condition × Region interactions (δ ($p > 0.6$), θ ($p > 0.5$), β1 ($p > 0.6$), β2 ($p > 0.4$), γ ($p > 0.6$)) nor main effects of Condition (δ ($p > 0.6$), θ ($p > 0.2$), β1 ($p > 0.3$), β2 ($p > 0.9$), γ ($p > 0.07$)) were found at any frequency range.

Furthermore, we also conducted an additional power analysis comparing the real-word condition with the pseudo-word condition within the analysis window of 300–500 ms after onset of the second syllable of each disyllabic word (except for the final disyllabic word of each utterance due to contaminations of button-press), rather than the analysis window of 250–2250 ms of the utterance period. The reason for this analysis is that we predicted that semantic differences between real-word and pseudo-word should be maximal at approximately 400 ms after onset of the second syllable of each disyllabic word. The window of 300–500 ms is in accordance with the classical N400 component for lexical retrieval and contextual semantic integration (Lau et al., 2008). The results are shown in Fig. 4, in which only the result for γ power was shown. The result showed no significant Condition × Region interactions at any frequency range (δ, θ, β1 β2, or γ, all $p > 0.2$), and no significant main effects of Condition for δ, θ, β1, or β2 power (all $p > 0.1$). There was a significant main effect of Condition showing that γ power was higher for real-word than for pseudo-word ($p < 0.02$).

EEG–acoustic entrainment (EAE)

For detailed statistical illustrations comparing EAE values in each condition (speech, non-speech, real-word, and pseudo-word) with the respective random-level baselines, see Fig. S3 in the *Supplementary Materials*. Fig. 5 shows the results comparing speech with non-speech (Fig. 5(A)) and real-word with pseudo-word (Fig. 5(B)). The two topmost panels illustrate the EAE values relative to random baselines in each condition within the δ and θ range (2–8 Hz) averaged across all the 32 electrodes. The lower panels illustrate the statistical results.

For speech versus non-speech (Fig. 5(A)), a significant Condition × Region interaction was found for β2 EAE ($p < 0.001$), but not for δ ($p > 0.08$), θ ($p > 0.2$), β1 ($p > 0.1$), or γ ($p > 0.5$) EAE. The post hoc pairwise comparisons showed that β2 EAE was higher for (speech − $\text{random}_{\text{speech}}$) than for (non-speech − $\text{random}_{\text{non-speech}}$) in the Right Parieto-Occipital ($p < 10^{-4}$), Mid Centro-Frontal ($p < 0.001$), and Mid Parieto-Occipital ($p < 0.001$) regions (significant at the level of 0.05 after Bonferroni correction across the 8 regions). Main effects of Condition were found with higher δ ($p < 10^{-4}$) and θ ($p < 0.02$) EAE for (speech − $\text{random}_{\text{speech}}$) than for (non-speech − $\text{random}_{\text{non-speech}}$). No main effect of Condition was found for β1 ($p > 0.2$) or γ ($p > 0.3$) EAE.

For real-word versus pseudo-word (Fig. 5(B)), no significant Condition × Region interactions were found for δ ($p > 0.5$), θ ($p > 0.6$), β1 ($p > 0.1$), β2 ($p > 0.5$), or γ ($p > 0.5$) EAE. There was a significant main effect of Condition showing lower δ EAE for (real-word − $\text{random}_{\text{real-word}}$) than for (pseudo-word − $\text{random}_{\text{pseudo-word}}$) ($p < 0.003$). No main effects of Condition were found for θ ($p > 0.1$), β1 ($p > 0.6$), β2 ($p > 0.7$), or γ ($p > 0.4$) EAE.

In addition, we noticed that for the comparison in which β2 EAE was significantly higher for (speech − $\text{random}_{\text{speech}}$) than for (non-speech − $\text{random}_{\text{non-speech}}$), β2 EAE was found to be significantly higher

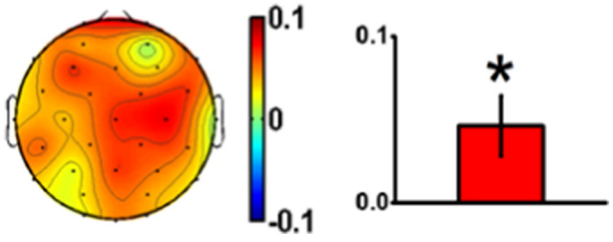

Fig. 4. *Additional result for γ EEG power.* An additional γ power result was obtained when comparing real-word with pseudo-word within the analysis window of 300–500 ms after onset of the second syllable of each disyllabic word. The asterisk denotes the significance levels $p < 0.05$.

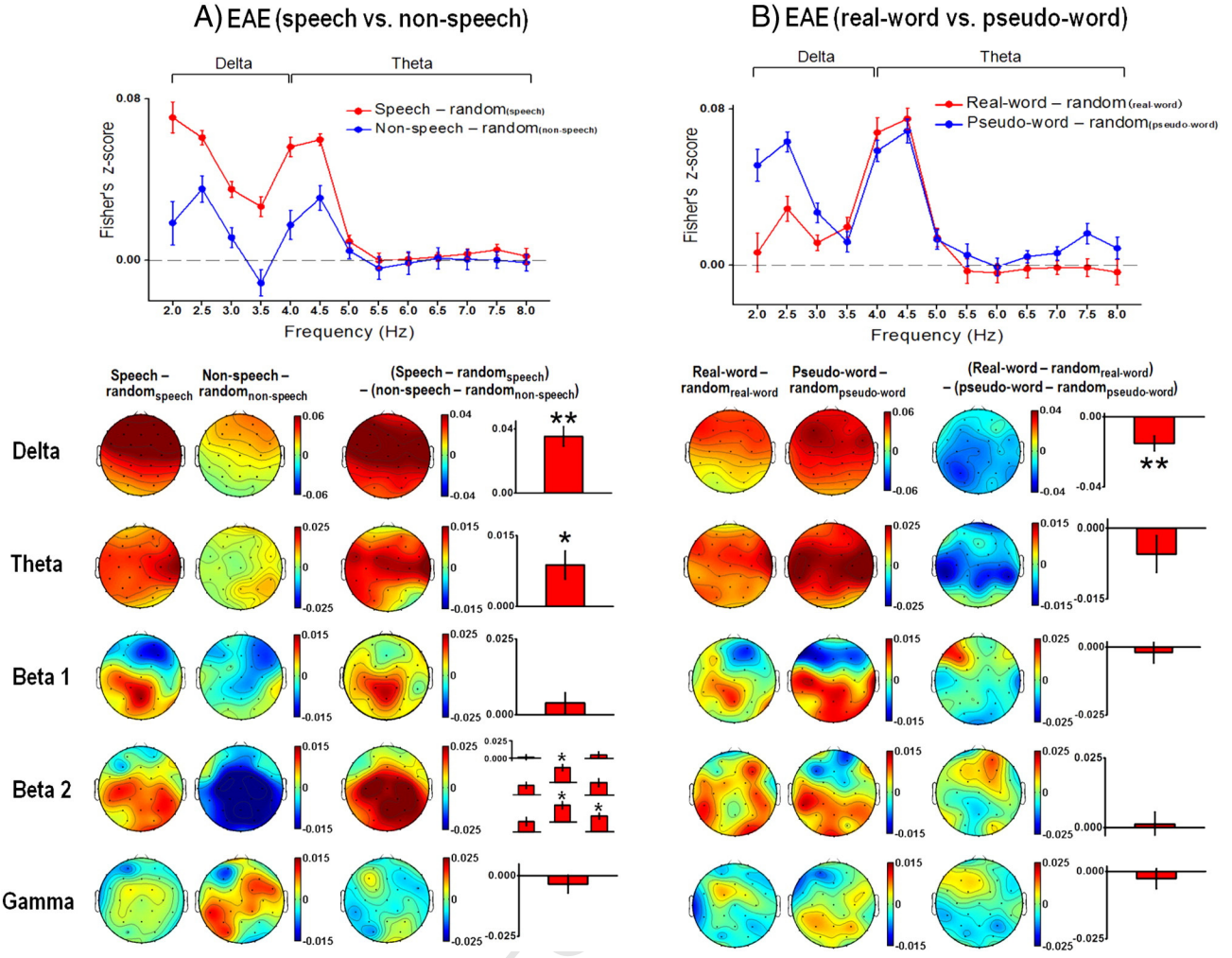

Fig. 5. *Results for δ, θ, β1, β2, and γ EAE.* 'Random' refers to the random-level EAE in each respective condition. The top panels illustrate the EAE value as a function of frequency of δ and θ (2–8 Hz) averaging across all the 32 electrodes. The lower panels show the statistical results. Because of the significant Condition × Region interaction in β2 EAE for speech versus non-speech, post hoc pairwise comparisons were undertaken in all 8 regions. Single asterisks denote the significance levels $p < 0.05$ (Bonferroni corrections have been conducted during post hoc pairwise comparisons in individual regions). Double asterisks denote significance level $p < 0.01$. Error bars denote the SEM.

for speech than for $\text{random}_{\text{speech}}$ but lower for non-speech than for $\text{random}_{\text{non-speech}}$ at various central, parietal, and occipital electrodes (see Fig. S3 for details). This indicates that the significant effect of β2 EAE for speech vs. non-speech may be partly attributed to this below-random effect in non-speech. This may imply suppression of acoustic envelope entrainment in non-speech, which is further discussed in the Discussion section.

Cross-frequency couplings (CFC)

Results for θ–β/γ CFCs are illustrated in Fig. 6, in which the top panels show the phase-power representations ((speech − non-speech) and (real-word − pseudo-word), respectively) for each of the 8 regions, and the lower panels show statistical results for θ–β1, θ–β2, and θ–γ CFCs. For δ–θ CFCs, see Fig. S4 in the *Supplementary Materials*.

For speech versus non-speech, neither Condition × Region interaction ($p > 0.4$) nor main effect of Condition ($p > 0.3$) was found for δ–θ CFC (Fig. S4(A)). For the θ–β/γ CFCs (Fig. 6(A)), no significant Condition × Region interactions were found for θ–β1 ($p > 0.1$), θ–β2 ($p > 0.1$), or θ–γ ($p > 0.5$) CFC. We found main effects of Condition for θ–β1 and θ–γ CFCs with significantly higher θ–β1 CFC for speech than for non-speech ($p < 0.02$) and significantly lower θ–γ CFC for speech than for non-speech ($p < 0.05$). No significant main effect of Condition was found for θ–β2 CFC ($p > 0.8$).

For the real-word versus pseudo-word comparison, neither Condition × Region interaction ($p > 0.5$) nor main effect of Condition ($p > 0.3$) was found for δ–θ CFC (Fig. S4(B)). For the θ–β/γ CFCs (Fig. 6(B)), neither Condition × Region interactions nor main effects of Condition were found for θ–β1 (interaction: $p > 0.4$; main effect: $p > 0.2$), θ–β2 (interaction: $p > 0.5$; main effect: $p > 0.6$), or θ–γ (interaction: $p > 0.08$; main effect: $p > 0.2$) CFC.

Renormalized partial directed coherence (rPDC)

Due to the tradeoff between over-parameterization and spectral resolution, we only considered rPDCs in the β to γ range (20, 30, and 40 Hz). For statistical results comparing rPDCs of different conditions (speech, non-speech, real-word, and pseudo-word) with the random-level rPDC, see Fig. S5 in the *Supplementary Materials*. Fig. 7 shows the results based on rPDC pairs, each of which was at above-random-level in both or in either of speech and non-speech (above-random-level in both or in either of real-word and pseudo-word for real-word vs. pseudo-word). Red arrows represent above-random-level rPDC pairs in speech but not in non-speech, or in the real-condition but not the pseudo-word condition, while blue arrows represent above-random-level rPDC pairs in non-speech but not in speech, or in pseudo-word but not in real-word.

Please cite this article as: Mai, G., et al., Delta, theta, beta, and gamma brain oscillations index levels of auditory sentence processing, NeuroImage (2016), http://dx.doi.org/10.1016/j.neuroimage.2016.02.064

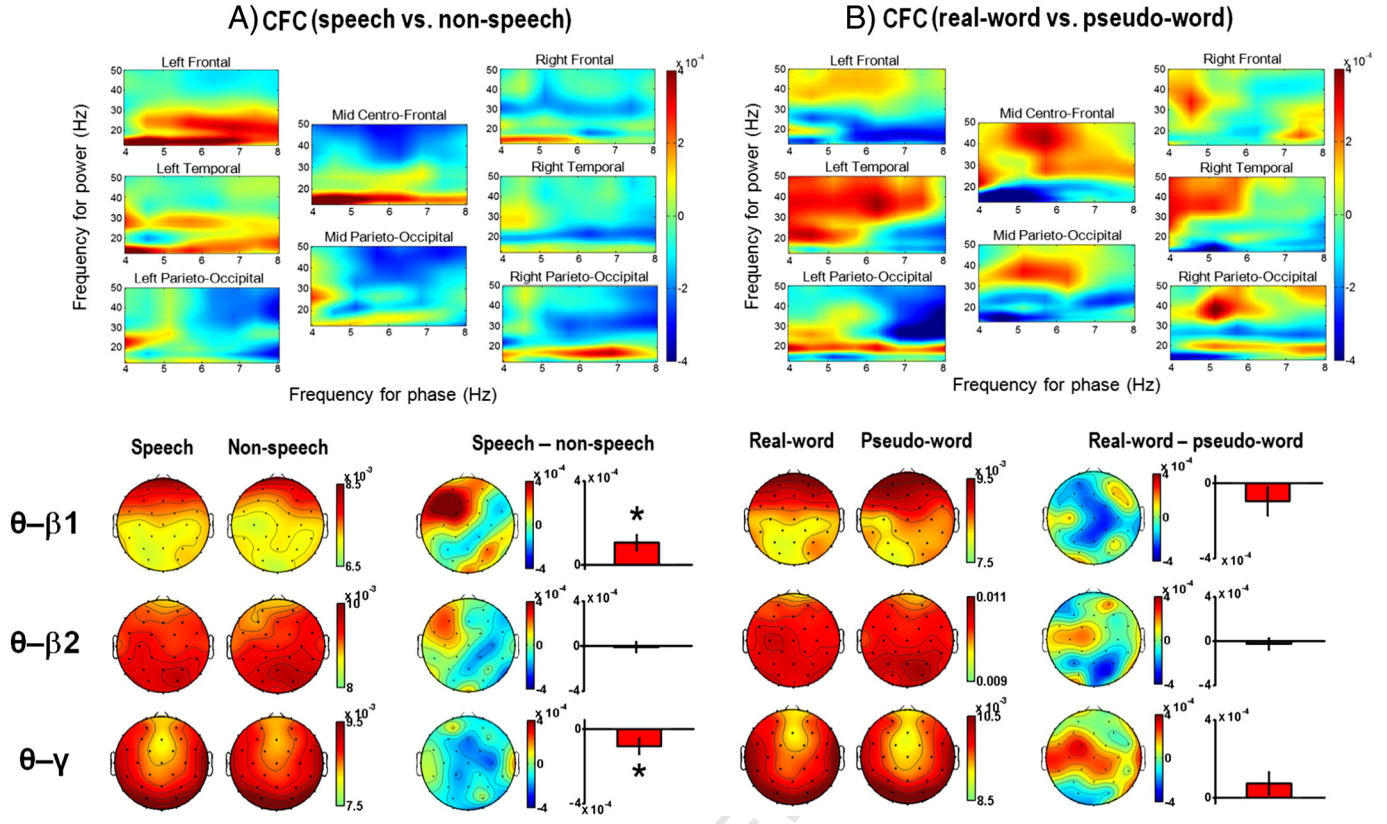

Fig. 6. *Results for θ–β1, θ–β2, and θ–γ CFCs.* Top panels illustrate MI value as a function of frequency for phase (4–8 Hz) and frequency for power (13–50 Hz) in all 8 regions. Lower panels show the statistical results. Asterisks denote the significance levels $p < 0.05$. Error bars denote the SEM.

For speech versus non-speech (Fig. 7(A)), after excluding the random-level pairs that simultaneously existed in both conditions, there was a significant main effect of Condition for 20-Hz rPDC (speech > non-speech, $p < 0.05$), but not for 30-Hz ($p > 0.5$) or 40-Hz ($p > 0.8$) rPDCs.

For real-word versus pseudo-word (Fig. 7(B)), after excluding the random-level pairs that simultaneously existed in both conditions, no significant main effects of Condition were found for 20-Hz ($p > 0.6$), 30-Hz ($p > 0.4$), or 40-Hz ($p > 0.4$) rPDCs.

Result summary

For the low-frequency EEG components of δ and θ, the results show that power and EAE of δ and θ were significantly higher for the speech than the non-speech condition; however, no such effects were found when comparing real-words with pseudo-words. Interestingly, we further found a significant effect with lower δ EAE for real-word than for pseudo-word. For the high-frequency components of β and γ, on one hand, all β-related significant effects (β2 EAE, θ–β1 CFCs, and 20-Hz

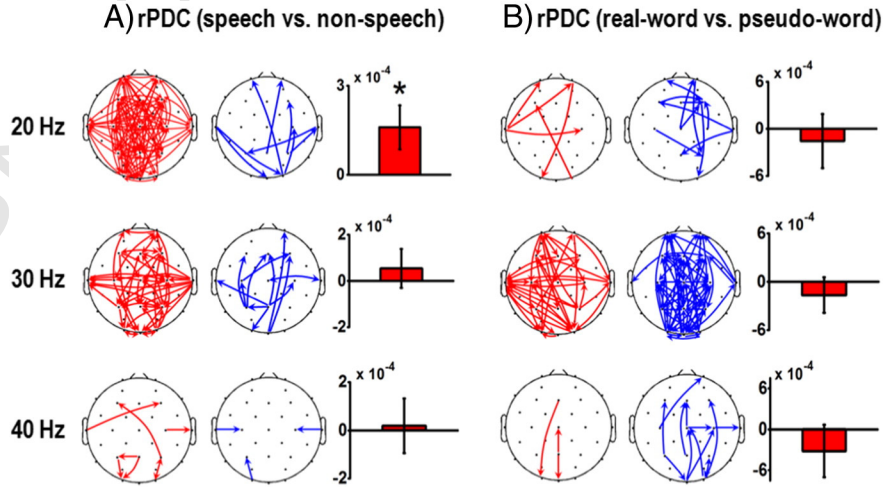

Fig. 7. *Results for 20-, 30-, and 40-Hz rPDCs.* Red arrows represent above-random-level pairs in speech but not non-speech, or in real-word but not in pseudo-word; blue arrows represent above-random-level pairs in non-speech but not in speech, or in pseudo-word but not in real-word (N.B., above-random-level pairs that existed in both speech and non-speech, or in both real-word and pseudo-word were not shown). Bar graphs show comparisons of mean rPDC values, each pair of which was at above-random level either of speech and non-speech (or either of real-word and pseudo-word). The single asterisk denotes the significance level $p < 0.05$. Error bars denote the SEM.

Please cite this article as: Mai, G., et al., Delta, theta, beta, and gamma brain oscillations index levels of auditory sentence processing, NeuroImage (2016), http://dx.doi.org/10.1016/j.neuroimage.2016.02.064

rPDC) were obtained only for speech versus non-speech but not for real-word versus pseudo-word; on the other hand, significant higher γ power was obtained for real-word than for pseudo-word within the time window of 300–500 ms after onset of the second syllable of the disyllabic words.

Discussion

Attention control and working memory load across conditions

In the current experiment, subjects were instructed to pay attention to the target stimuli and to perform a sound-matching task. Results of response accuracies indicate that the task was easier for speech (>95% in both real-word and pseudo-word) than for non-speech (<75%). However, the observation that accuracies were significantly higher than chance level (50%) even for non-speech, confirms that subjects complied with the instruction to pay active attention to the target stimuli in all conditions. In addition, reaction time (RT) was significantly longer for non-speech than for speech, which is consistent with the higher task difficulty in non-speech. We argue that, the higher task difficulty indicates that the working memory (WM) load was greater for non-speech than for speech, as also suggested in the previous study (Peña and Melloni, 2012).

Roles of power and EAE at δ and θ bands for phonological processing

δ and θ EEG power

Results of the normalized EEG power in the current study show that the δ and θ power was significantly higher for speech than for non-speech, but no δ or θ power differences were found between the real-word and pseudo-word conditions. The θ power effect is thus consistent with previous studies (Peña and Melloni, 2012; Ding et al., 2015), which showed higher θ EEG/MEG power when listening to forward sentences than to the backward versions, regardless of the listeners being native or non-native speakers. The authors' interpretation was that the forward utterances were perceived more as syllable strings than the backward utterances, supporting the role of θ activities in tracking θ-rate syllabic patterns (Peña and Melloni, 2012). In the current study, besides higher θ power, higher δ power was also observed for speech than for non-speech, implicating a possible role of δ activities in tracking δ-rate phonetic features, such as supra-syllabic patterns. However, we did not obtain higher δ power for real-word than pseudo-word and this is not consistent with the study by Ding et al. (2015) which also used Mandarin sentence stimuli with similar acoustic property of 4-Hz syllable rate as in the current study. This study obtained a higher 1 Hz (rate of 4-syllable phrases) and 2 Hz (rate of disyllabic words) compared to pseudo-words, arguing for the role of δ activities for processing at higher linguistic hierarchical levels (i.e., phrase and word levels) (Ding et al., 2015). We suggest that the inconsistency is likely because we focused on the predefined δ oscillations at 2–4 Hz range rather than 1 or 2 Hz and a sound matching task was used in the current study that may thus result in more observable lower-level neurophysioloigcal effects.

Furthermore, considering that backward utterances, which served as the non-speech baseline, preserve properties closely matched to the acoustic complexity (such as speech-like temporal fluctuations, formant distributions, and harmonic structures) but cause serious phonological distortions (Binder et al., 2000; Saur et al., 2010; Peña and Melloni, 2012; Gross et al., 2013), we argue that higher degree of phonological processing may thus result in the observed higher δ and θ power.

δ and θ EAE

Previous behavioral studies have shown that low-frequency (especially <10 Hz) acoustic envelopes are essential for human speech recognition (e.g., Shannon et al., 1995; Arai et al., 1999; Xu et al., 2005). Furthermore, recent neurophysiological studies have highlighted the role of entrainment of neural oscillations to acoustic envelopes at δ and θ frequencies in speech comprehension (Luo and Poeppel, 2007; Peelle et al., 2013; Doelling et al., 2014; Ding and Simon, 2014; Ding et al., 2015). However, which linguistic hierarchical levels are involved in such entrainment and how these hierarchical levels influence the entrainment remain unclear. Here, we have shown significantly higher δ and θ EAE for speech than for non-speech, but no higher δ or θ EAE was observed for real-word than for pseudo-word. Brain entrainment to the speech envelope reflects how cortical activity interacts with sensory input (i.e., inputs of acoustic envelope via auditory systems) and is modulated by speech intelligibility, therefore implying top-down control of the entrainment by high-level linguistic processing (Peelle et al., 2013; Ding and Simon, 2014). Thus, we argue that the current results reflect a top-down linguistic effect enhancing low-frequency brain–acoustic entrainment driven by phonological processing.

In addition, we have shown a significantly lower δ EAE for the real-word condition than for the pseudo-word condition. A plausible explanation for this result is that, as the subjects were required to perform a sound matching task, the richer semantic/syntactic contents of real-word utterances may assist in the recognition of their phonological contents, reducing the phonological processing demands as indexed by δ EAE, thereby giving rise to lower EAE for real-word than for pseudo-word. This is also consistent with previous studies showing the importance of δ-range acoustic envelopes in human recognition of semantically meaningless syllables (Arai et al., 1996, 1999) and a recent study arguing that δ neural-envelope entrainment reflects an increased listening effort under more attention-demanding speech recognition conditions (Ding et al., 2014).

Roles of β for phonological and γ for semantic/syntactic processing

β2 EAE

The higher β2 EAE observed for speech than for non-speech, but not for real-words compared to pseudo-words, implies that β2 EAE is involved in phonological processing. The analyses that compared the β2 EAE for speech and non-speech to their respective random baselines showed that β2 EAE was significantly higher for speech than for $random_{speech}$, but lower for non-speech than for $random_{non\text{-}speech}$ at various central and parietal and occipital electrodes (Fig. S3). For the speech condition, similar to δ and θ EAE, we argue that β2 EAE was enhanced via top-down control driven by phonological processing. For non-speech, we interpret that the effect of β2 EAE below the random level may be related to another top-down control mechanism. The behavioral results, which showed significantly lower response accuracy and longer reaction time for non-speech than for speech, indicate a higher working memory (WM) demand in non-speech. Previous studies have shown that WM maintenance of target information can cause top-down suppression of attention to irrelevant information in order to prevent overloading the limited WM capacity (Gazzaley et al., 2005; Zanto and Gazzaley, 2009). We suspect that in non-speech, top-down processing was driven to suppress the attention required for β2 EAE in order to optimize WM maintenance of the past acoustic information, which therefore caused the effect on β2 EAE to be lower than the random-level baseline.

θ–β1 and θ–γ CFCs

θ–β/γ CFC has been observed in human cortices during various cognitive tasks, including language-related tasks, such as active/passive listening to phonemes and words, word production, and visual reading (Canolty et al., 2006). The underlying neurophysiological mechanisms of CFC are still not fully understood, but it has been suggested that, because the phase of low-frequency (θ) oscillations represents the time course of membrane potential fluctuations and the power of high-frequency (β and γ) oscillations reflects local neural excitability (e.g., Chapman and Lacaille, 1999; Niessing et al., 2005), θ–β/γ CFC is therefore a representation of activities regulating neural excitability at

Please cite this article as: Mai, G., et al., Delta, theta, beta, and gamma brain oscillations index levels of auditory sentence processing, NeuroImage (2016), http://dx.doi.org/10.1016/j.neuroimage.2016.02.064

critical time points controlled by θ phase (Schroeder and Lakatos, 2008). Due to the regularity between the θ phase and β/γ power, the pattern produced by strong θ–β/γ CFC is similar to what is generated by experimental protocols to produce the long-term synaptic potentiation (LTP) (Canolty and Knight, 2010). CFC is thus (at least some of the time) related to the potentiation of synaptic activities that entail high-level cognitive functioning, especially learning and memory processes (Canolty and Knight, 2010). Furthermore, as previously introduced (see Introduction), θ–β/γ CFC was experimentally found to be involved in encoding and retrieval of long-term memory (LTM), and WM maintenance in both non-human mammals (Tort et al., 2008, 2009; Shirvalkar et al., 2010) and human beings (Mormann et al., 2005; Sauseng et al., 2009; Axmacher et al., 2010; Friese et al., 2013; Köster et al., 2014; Kaplan et al., 2014).

Higher θ–β1 CFC was found for speech than for non-speech, but not for real-word than for pseudo-word. This may be related to the retrieval of long-term stored phonological information and phonological WM maintenance in the left parietal and inferior frontal regions reported by previous studies (Nixon et al., 2004; Zaehle et al., 2008; Strand et al., 2008; Liebenthal et al., 2013). Alternatively, this effect may be explained by the hypothesis that θ–β/γ CFC is responsible for the hierarchical binding between long-duration (e.g., syllables and long-vowels) and short-duration phonetic information (consonants and short-vowels) during phonological analysis (Giraud and Poeppel, 2012; Gross et al., 2013). This latter claim, however, does not seem to be well supported according to our data. The hierarchical binding theory states that both the low- and high-frequency neural components are entrained to the acoustic envelopes and they lie in a nesting relation (Giraud and Poeppel, 2012). We would thus predict that, if the hierarchical binding theory were correct, the current θ–β1 CFC effect would be coherent with the effects of θ and β1 EAE that we reported earlier. However, while we found significantly higher θ–β1 CFC for speech than for non-speech, we did not obtain a significant effect of β1 EAE for speech compared to non-speech (see section EEG–acoustic entrainment).

We also found a significantly lower θ–γ CFC effect for speech than for non-speech. We infer that this may reflect the higher psychoacoustic WM demands for non-speech than for speech due to the higher difficulty of the sound matching tasks in non-speech, as previously discussed. We therefore speculate that the two significant CFC effects found in the current study, one being θ–β1 CFC reflects retrieval of long-term stored phonological information and phonological WM processing, while the other being θ–γ CFC reflects lower-level psychoacoustic WM processing. We suggest that in future studies, the θ–γ/β CFC effects involved in different linguistic hierarchical levels could be better tested with more cautious control of the task difficulties across different conditions.

20-Hz rPDC

We proposed the rPDC that can quantify the extent of directed connectivity between brain regions as an index for phonological and semantic processing based on two observations in previous studies. First, a recent MEG study has revealed that during identification of semantically meaningless syllables, connectivity between auditory regions (in the temporal lobe) and motor and premotor regions (in the frontal lobe) was activated via inter-areal phase synchronies at β and γ frequencies, reflecting the role of inter-areal interactions for phonological processing (Alho et al., 2014). Second, during semantic processing, lexical-semantic retrieval requires a large-scale connectivity network throughout the brain (Patterson et al., 2007; Pulvermüller, 2013).

We examined the rPDCs at β to γ frequencies (20, 30, and 40 Hz) and found a significantly higher 20-Hz rPDC for speech than for non-speech, but not for real-word compared to pseudo-word. Also, no significant effects were found for 30- or 40-Hz rPDC. This implies that 20-Hz rPDC reflects phonological processing rather than semantic/syntactic processing. This finding is compatible with the results of Alho et al. (2014), which show that inter-areal β phase synchronies have significantly positive correlations with the syllable identification accuracies, thereby indicating the importance of β oscillations during phonological processing. That study also highlighted the recruitment of the auditory-motor network during speech perception, as was also reported by a number of previous studies (Wilson et al., 2004; Meister et al., 2007; Londei et al., 2010; Liebenthal et al., 2013). Furthermore, synchronies in β activities have been found to be associated with functions of sensorimotor interaction (see a review by Siegel et al., 2012). In connection with these previous studies, we propose that the current 20-Hz rPDC result may imply the role of β oscillations in auditory-motor interaction for phonological processing.

γ power

Higher γ was found for real-word than pseudo-word within the 300–500 ms window after onset of the second syllable of the disyllabic words. This window was chosen as we predicted that semantic differences between real-word and pseudo-word should be maximal at approximately 400 ms after onset of the second syllable of each disyllabic word, corresponding to the classical N400 component for lexical retrieval and contextual semantic integration (Lau et al., 2008). This is in accordance with previous studies showing the involvement of γ power in lexico-semantic retrieval (Lutzenberger et al., 1994; Pulvermüller et al., 1996, 1999; Mainy et al., 2008). In addition, although in the current study, the real-word utterances were SUSs that were semantically anomalous, they were syntactically valid in which all disyllabic words kept correct grammatical word categories. There was a previous study that found higher γ power for sentences with correct grammatical word category than for sentences with violated word category (Bastiaansen et al., 2010). Therefore, it is possible that, besides lexical retrieval, γ power found in the current study is also relevant to syntactic processing in terms of perception of word category.

Different roles of β and γ oscillations

The current data show that β and γ oscillations play different roles during auditory sentence processing. β-related effects comprise the 20-Hz rPDC, β2 EAE, and θ–β1 CFC for phonological processing, while the γ-related effect comprises the γ power for semantic/syntactic processing. Because β oscillations are closely associated with sensorimotor interaction (Siegel et al., 2012; Alho et al., 2014), the current results may therefore imply the recruitment of auditory-motor networks during phonological processing. On the other hand, γ oscillations were shown to be involved in higher-level linguistic processing.

Furthermore, previous neurophysiological studies have shown that β and γ oscillations are generated by different underlying neuronal mechanisms. Firstly, the functional connections between excitatory neurons and inhibitory interneurons within circuits that generate β and γ oscillations were shown to be different (Kopell et al., 2000; Olufsen et al., 2003). Secondly, β and γ oscillations are expressed with different strengths across cortical lamina: β oscillations originate primarily from the deep laminar layers, whereas γ oscillations originate primarily from the superficial layers (Roopun et al., 2010; Kramer et al., 2008; Maier et al., 2008; Buffalo et al., 2011). These backgrounds thus provide hints for the interpretation of our current data that phonological processing and semantic/syntactic processing may engage different functional and anatomical cortical networks, as supported by a previous neuroimaging study (Saur et al., 2010). In this study by Saur et al. (2010), stimulus design was similar to the present study (comprising real-word, pseudo-word, and backward utterances) and cortical networks supporting phonological and semantic processing were examined using combined functional and anatomical connectivity approaches in MRI. Distinct functional and anatomical temporal–frontal connectivity was identified between phonological and semantic processing (Saur et al., 2010). Our current data therefore echo the results

Please cite this article as: Mai, G., et al., Delta, theta, beta, and gamma brain oscillations index levels of auditory sentence processing, NeuroImage (2016), http://dx.doi.org/10.1016/j.neuroimage.2016.02.064

of this study, showing distinct networks associated with processing at different linguistic hierarchical levels.

Summary

The present study investigated the δ, θ, β, and γ EEG oscillations during auditory sentence processing. We hypothesized that the phonological and higher-level semantic/syntactic processing can be separately indexed. First, we observed significant effects of band power and EEG-acoustic entrainment of the δ and θ oscillations elicited during phonological processing, but found no such effects during semantic/syntactic processing. This thus indicates the tracking of phonetic patterns by δ and θ oscillations during phonological processing. Second, we observed significant β-related and γ-related effects during phonological and semantic/syntactic processing, respectively. By relating these observations to previous findings of anatomical and functional differences between β and γ oscillations, we infer that the current β- and γ-related effects may originate from different cortical networks. Taken together, our data demonstrate that phonological and higher-level linguistic (semantic/syntactic) processes during auditory sentence processing can be indexed by distinct EEG signatures relevant to δ, θ, β, and γ oscillations. The results are therefore compatible with previous neuroimaging evidence showing that phonological processing and higher-level linguistic processing engage distinct neural networks.

Uncited references

Q12

Cooke et al., 2006
Humphries et al., 2007
Rogalsky and Hickok, 2009

Acknowledgments

This research was supported in part by grant 455911 made by the Research Grants Council of Hong Kong SAR to Prof. William S-Y. Wang at The Chinese University of Hong Kong. We would also like to appreciate the reviewers for their insightful comments and suggestions during the paper revision.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2016.02.064.

References

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., Merzenich, M.M., 2001. Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proc. Natl. Acad. Sci. 98 (23), 13367–13372.
Alho, J., Lin, F.-H., Sato, M., Tiitinen, H., Sams, M., Jääskeläinen, I.P., 2014. Enhanced neural synchrony between left auditory and premotor cortex is associated with successful phonetic categorization. Front. Psychol. 5 (394), 1–10.
Arai, T., Pavel, M., Hermansky, H., Avendano, C., 1999. Syllable intelligibility for temporally filtered LPC cepstral trajectories. J. Acoust. Soc. Am. 105, 2783–2791.
Arai, T., Hermansky, H., Pavel, M., Avendano, C., 1996. Intelligibility of speech with filtered time trajectories of spectral envelopes. Proceedings of ICSLP'96, Philadelphia, USA, pp. 2490–2493.
Axmacher, N., Henseler, M.M., Jensen, O., Weinreich, I., Elger, C.E., Fell, J., 2010. Cross-frequency coupling supports multi-item working memory in the human hippocampus. Proc. Natl. Acad. Sci. 107 (7), 3228–3233.
Bastiaansen, M.C.M., Magyari, L., Hagoort, P., 2010. Syntactic unification operations are reflected in oscillatory dynamics during on-line sentence comprehension. J. Cogn. Neurosci. 22, 1333–1347.
Bastiaansen, M.C.M., Oostenveld, R., Jensen, O., Hagoort, P., 2008. I see what you mean: theta power increases are involved in the retrieval of lexical semantic information. Brain Lang. 106 (1), 15–28.
Bastiaansen, M.C.M., van Berkum, J.J.A., Hagoort, P., 2002. Syntactic processing modulates the theta rhythm of human EEG. NeuroImage 17, 1479–1492.
Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S.F., Springer, J.A., Kaufman, J.N., Possing, E.T., 2000. Human temporal lobe activation by speech and nonspeech sounds. Cereb. Cortex 10, 512–528.
Benoit, C., Grice, M., Hazan, V., 1996. The SUS test: a method for the assessment of text to speech synthesis intelligibility using semantically unpredictable sentences. Speech Comm. 18, 381–392.
Buffalo, E.A., Fries, P., Landman, R., Buschman, T.J., Desimone, R., 2011. Laminar differences in gamma and alpha coherence in the ventral stream. Proc. Natl. Acad. Sci. 108, 11262–11267.
Canolty, R.T., Edwards, E., Dalal, S.S., Soltani, M., Nagarajan, S.S., Kirsch, H.E., Berger, M.S., Barbaro, N.M., Knight, R.T., 2006. High gamma power is phase-locked to theta oscillations in human neocortex. Science 313, 1626–1628.
Canolty, R.T., Knight, R.T., 2010. The functional role of cross-frequency coupling. Trends Cogn. Sci. 14 (11), 507–515.
Chapman, C.A., Lacaille, J.-C., 1999. Intrinsic theta-frequency membrane potential oscillations in hippocampal CA1 interneurons of stratum lacunosum-moleculare. J. Neurophysiol. 81 (3), 1296–1307.
Cogan, G.B., Poeppel, D., 2011. A mutual information analysis of neural coding of speech by low-freuqency MEG phase information. J. Neurophysiol. 106, 554–563.
Cooke, A., Grossman, M., DeVita, C., Gonzalez-Atavales, J., Moore, P., Chen, W., Gee, J., Detre, J., 2006. Large-scale neural network for sentence processing. Brain Lang. 96 (1), 14–36.
Ding, N., Chatterjee, M., Simon, J.Z., 2014. Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. NeuroImage 88, 41–46.
Ding, N., Melloni, L., Zhang, H., Tian, X., Poeppel, D., 2015. Cortical tracking of hierarchical linguistic structures in connected speech. Nat. Neurosci. 19 (1), 158–164.
Ding, N., Simon, J.Z., 2014. Cortical entrainment to continuous speech: functional roles and interpretations. Front. Hum. Neurosci. 8 (311), 1–7.
Doelling, K.B., Arnal, L.H., Ghitza, O., Poeppel, D., 2014. Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. NeuroImage 85, 761–768.
Elshoff, L., Muthuraman, M., Anwar, A.R., Deuschl, G., Stephani, U., Raethjen, J., Siniatchkin, M., 2013. Dynamic imaging of coherent sources reveals different network connectivity underlying the generation and perpetuation of epileptic seizures. PLoS One 8 (10), 1–11.
Fell, J., Axmacher, N., 2011. The role of phase synchronization in memory processes. Nat. Rev. Neurosci. 12, 105–118.
Fontolan, L., Morillon, B., Liegeois-Chauvel, C., Giraud, A.-L., 2014. The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. Nat. Commun. 5 (4694), 1–10.
Friese, U., Köster, M., Hassler, U., Martens, U., Trujillo-Barreto, N., Gruber, T., 2013. Successful memory encoding is associated with increased cross-frequency coupling between frontal theta and posterior gamma oscillations in human scalp-recorded EEG. NeuroImage 66, 642–647.
Gazzaley, A., Cooney, J.W., Rissman, J., D'Esposito, M., 2005. Top-down suppression deficit underlies working memory impairment in normal aging. Nat. Neurosci. 8, 1298–1300.
Giraud, A.-L., Poeppel, D., 2012. Cortical oscillations and speech processing: emerging computational principles and operations. Nat. Neurosci. 15 (4), 511–517.
Golumbic, E.M.Z., Ding, N., Bickel, S., Lakatos, P., Schevon, C.A., McKhann, G.M., Goodman, R.R., Emerson, R., Mehta, A.D., Simon, J.Z., Peoppel, D., Schroeder, C.E., 2013. Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". Neuron 77, 980–991.
Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., Garrod, S., 2013. Speech rhythms and multiplexed oscillatory sensory coding in the human brain. PLoS Biol. 11 (12), 1–14.
Humphries, C., Binder, J.R., Medler, D.A., Liebenthal, E., 2007. Time course of semantic processes during sentence comprehension: an fMRI study. NeuroImage 36 (3), 924–932.
Japaridze, N., Muthuraman, M., Moeller, F., Boor, R., Anwar, A.R., Deuschl, G., Stephani, U., Raethjen, J., Siniatchkin, M., 2013. Neuronal networks in west syndrome as revealed by source analysis and renormalized partial directed coherence. Brain Topogr. 26 (1), 157–170.
Jensen, O., Kaiser, J., Lachaux, J.P., 2007. Human gamma-frequency oscillations associated with attention and memory. Trends Neurosci. 30 (7), 317–324.
Kaplan, R., Bush, D., Bonnefond, M., Bandettini, P.A., Barnes, G.R., Doeller, C.F., Burgess, N., 2014. Medial prefrontal theta phase coupling during spatial memory retrieval. Hippocampus 24, 656–665.
Klimesch, W., 2012. Alpha-band oscillations, attention, and controlled access to stored information. Trends Cogn. Sci. 16 (12), 606–617.
Kong, Y.-Y., Mullangi, A., Ding, N., 2014. Differential modulation of auditory responses to attended and unattended speech in different listening conditions. Hear. Res. 316, 73–81.
Kopell, N., Ermentrout, G.B., Whittington, M.A., Traub, R.D., 2000. Gamma rhythms and beta rhythms have different synchronization properties. Proc. Natl. Acad. Sci. 97 (4), 1867–1872.
Köster, M., Friese, U., Schöne, B., Trujillo-Barreto, N., Gruber, T., 2014. Theta-Gamma Coupling during Episodic Retrieval in the Human EEG Brain Research. in press.
Kramer, M.A., Roopun, A.K., Carracedo, L.M., Traub, R.D., Whittington, M.A., Kopell, N.J., 2008. Rhythm generation through period concatenation in rat somatosensory cortex. PLoS Comput. Biol. 4 (9), 1–16.
Kubanek, J., Brunner, P., Gunduz, A., Poeppel, D., Schalk, G., 2013. The tracking of speech envelope in the human cortex. PLoS One 8 (1), 1–9.
Lau, E.F., Phillips, C., Poeppel, D., 2008. A cortical network for semantics: (de)constructing the N400. Nat. Rev. Neurosci. 9, 920–933.
Lewis, A.G., Bastiaansen, M., 2015. A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. Cortex 68, 155–168.
Lewis, A.G., Wang, L., Bastiaansen, M., 2015. Fast oscillatory dynamics during language comprehension: unification versus maintenance and prediction? Brain Lang. 148, 51–63.

Please cite this article as: Mai, G., et al., Delta, theta, beta, and gamma brain oscillations index levels of auditory sentence processing, NeuroImage (2016), http://dx.doi.org/10.1016/j.neuroimage.2016.02.064

Liebenthal, E., Sabri, M., Beardsley, S.A., Mangalathu-Arumana, J., Desai, A., 2013. Neural dynamics of phonological processing in the dorsal auditory stream. J. Neurosci. 33 (39), 15414–15424.

Lisman, J.E., Jensen, O., 2013. The theta-gamma neural code. Neuron 77, 1002–1016.

Londei, A., D'Ausilio, A., Basso, D., Sestieri, C., Gratta, C.D., Romani, G.L., Belardinelli, M.O., 2010. Sensory-motor brain network connectivity for speech comprehension. Hum. Brain Mapp. 31 (4), 567–580.

Luo, H., Poeppel, D., 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron 54, 1001–1010.

Lutzenberger, W., Pulvermüller, F., Birbaumer, N., 1994. Words and pseudowords elicit distinct patterns of 30-Hz EEG responses in humans. Neurosci. Lett. 176 (1), 115–118.

Maier, J.X., Chandrasekaran, C., Ghazanfar, A.A., 2008. Integration of bimodal looming signals through neuronal coherence in the temporal lobe. Curr. Biol. 18, 963–968.

Mainy, N., Jung, J., Baciu, M., Kahane, P., Schoendorff, B., Minotti, L., Hoffmann, D., Bertrand, O., Lachaux, J.P., 2008. Cortical dynamics of word recognition. Hum. Brain Mapp. 29 (11), 1215–1230.

Maksimow, A., Silfverhuth, M., Langsjo, J., Kaskinoro, K., Georgiadis, S., Jaaskelainen, S., Scheinin, H., 2014. Directional connectivity between frontal and posterior brain regions is altered with increasing concentrations of propofol. PLoS One 9 (11), 1–16.

Maxwell, S.E., Delaney, H.D., 2004. Designing Experiments and Analyzing Data: A Model Comparison Perspective. second ed. Lawrence Erlbaum Associates, Mahwah, H.J.; London.

McNab, F., Hillebrand, A., Swithenby, S.J., Rippon, G., 2012. Combining temporal and spectral information with spatial mapping to identify differences between phonological and semantic networks: a magnetoencephalographic approach. Front. Psychol. 3 (273), 1–12.

Meister, I.G., Wilson, S.M., Deblieck, C., Wu, A.D., Iacoboni, M., 2007. The essential role of premotor cortex in speech perception. Curr. Biol. 17, 1692–1696.

Meyers, L.S., Gamst, G., Guarino, A.J., 2006. Applied Multivariate Research: Design and Interpretation. Sage Publications Inc., Thousand Oaks, CA.

Michels, L., Muthuraman, M., Lüchinger, R., Martin, E., Anwar, A.R., Raethjen, J., Brandeis, D., Siniatchkin, M., 2013. Developmental changes of functional and directed resting-state connectivities associated with neuronal oscillations in EEG. NeuroImage 81, 231–242.

Mormann, F., Fell, J., Axmacher, N., Weber, B., Lehnertz, K., Elger, C.E., Fernández, G., 2005. Phase/amplitude reset and theta-gamma interaction in the human medial temporal lobe during a continuous word recognition memory task. Hippocampus 15, 1–11.

Muthukumaraswamy, S.D., 2013. High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations. Front. Hum. Neurosci. 7 (138), 1–11.

Niessing, J., Ebisch, B., Schmidt, K.E., Niessing, M., Singer, W., Galuske, R.A.W., 2005. Hemodynamic signals correlate tightly with synchronized gamma oscillations. Science 309, 948–951.

Nourski, K.V., Reale, R.A., Oya, H., Kawasaki, H., Kovach, C.K., Chen, H., Howard, M.A., Brugge, J.F., 2009. Temporal envelope of time-compressed speech represented in the human auditory cortex. J. Neurosci. 29 (49), 15564–15574.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.

Olufsen, M.S., Whittington, M.A., Camperi, M., Kopell, N., 2003. New roles for the gamma rhythm: population tuning and preprocessing for the beta rhythm. J. Comput. Neurosci. 14, 22–54.

Patterson, K., Nestor, P.J., Rogers, T.T., 2007. Where do you know what you know? the representation of semantic knowledge in the human brain. Nat. Rev. Neurosci. 8, 976–988.

Peelle, J.E., Gross, J., Davis, M.H., 2013. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cereb. Cortex 23, 1378–1387.

Peña, M., Melloni, L., 2012. Brain oscillations during spoken sentence processing. J. Cogn. Neurosci. 24 (5), 1149–1164.

Pulvermüller, F., 2013. How neurons make meaning: brain mechanisms for embodied and abstract symbolic semantics. Trends Cogn. Sci. 17 (9), 458–470.

Pulvermüller, F., Lutzenberger, W., Preissl, H., 1999. Nouns and verbs in the intact brain: evidence from event-related potentials and high-frequency cortical responses. Cereb. Cortex 9 (5), 497–506.

Pulvermüller, F., Eulitz, C., Pantev, C., Mohr, B., Feige, B., Lutzenberger, W., Elbert, T., Birbaumer, N., 1996. High-frequency cortical responses reflect lexical processing: an MEG study. Electroencephalogr. Clin. Neurophysiol. 98, 76–85.

Rogalsky, C., Hickok, G., 2009. Selective attention to semantic and syntactic features modulates sentence processing networks in anterior temporal cortex. Cereb. Cortex 19, 786–796.

Roopun, A.K., LeBeau, F.E.N., Rammell, J., Cunningham, M.O., Traub, R.D., Whittington, M.A., 2010. Cholinergic neuromodulation controls directed temporal communication in neocortex in vitro. Front. Neural Circ. 4 (8), 1–10.

Saur, D., Schelter, B., Schnell, S., Kratochvil, D., Küpper, H., Kellmeyer, P., Kümmerer, D., Klöppel, S., Glauche, V., Lange, R., Mader, W., Feess, D., Timmer, J., Weiller, C., 2010. Combining functional and anatomical connectivity reveals brain networks for auditory language comprehension. NeuroImage 49, 3187–3197.

Sauseng, P., Klimesch, W., Helse, K.F., Gruber, W.R., Holz, E., Karim, A.A., Glennon, M., Gerloff, C., Birbaumer, N., Hummel, F.C., 2009. Brain oscillatory substrates of visual short-term memory capacity. Current Biology19, 1846–1852.

Scheffer-Teixeira, R., Belchior, H., Caixeta, F.V., Souza, B.C., Ribeirol, S., Tort, A.B.L., 2012. Theta phase modulates multiple layer-specific oscillations in the CA1 region. Cereb. Cortex 22 (10), 2404–2414.

Schelter, B., Winterhalder, M., Eichler, M., Peifer, M., Hellwig, B., Guschlbauer, B., Lücking, C.H., Dahlhaus, R., Timmer, J., 2005. Testing for directed influences among neural signals using partial directed coherence. J. Neurosci. Methods 152, 210–219.

Schelter, B., Timmer, J., Eichler, M., 2009. Assessing the strength of directed influences among neural signals using renormalized partial directed coherence. J. Neurosci. Methods 179, 121–130.

Schroeder, C.E., Lakatos, P., 2008. Low-frequency neuronal oscillations as instruments of sensory selection. Trends Neurosci. 32 (1), 9–18.

Shannon, R.V., Zeng, F.-G., Kamath, V., Wygonski, J., Ekelid, M., 1995. Speech recognition with primarily temporal cues. Science 270 (5234), 303–304.

Shirvalkar, P.R., Rapp, P.R., Shapiro, M.L., 2010. Bidirectional changes to hippocampal theta-gamma comodulation predict memory for recent spatial episodes. Proc. Natl. Acad. Sci. 107 (15), 7054–7059.

Siegel, M., Donner, T.H., Engel, A.K., 2012. Spectral fingerprints of large-scale neuronal interactions. Nat. Rev. Neurosci. 13, 121–134.

Silver, N.C., Dunlap, W.P., 1987. Averaging correlation coefficients: should Fisher's z transformation be used? J. Appl. Psychol. 72 (1), 146–148.

Smith, S.W., 1999. The Scientist and Engineer's Guide to Digital Signal Processing. second ed. California Technical Pub, San Diego, CA.

Tort, A.B.L., Fontanini, A., Kramer, M.A., Jones-Lush, L.M., Kopell, N.J., Katz, D.B., 2010. Cortical networks produce three distinct 7–12 Hz rhythms during single sensory responses in the awake rat. J. Neurosci. 30 (12), 4315–4324.

Tort, A.B.L., Komorowski, R.W., Manns, J.R., Kopell, N.J., Eichenbaum, H., 2009. Theta-gamma coupling increases during the learning of item-context associations. Proc. Natl. Acad. Sci. 106 (49), 20942–20947.

Tort, A.B.L., Kramer, M.A., Thorn, C., Gibson, D.J., Kubota, Y., Graybiel, A.M., Kopell, N.J., 2008. Dynamic cross-frequency couplings of local field potential oscillations in rat striatum and hippocampus during performance of a T-maze task. Proc. Natl. Acad. Sci. 105 (51), 20517–20522.

Wang, R., Perreau-Guimaraes, M., Carvalhaes, C., Suppes, P., 2012. Using phase to recognize English phonemes and their distinctive features in the brain. Proc. Natl. Acad. Sci. 109 (50), 20685–20690.

Ward, L.M., 2003. Synchronous neural oscillations and cognitive processes. Trends Cogn. Sci. 7 (12), 553–559.

Wilson, S.M., Saygin, A.P., Sereno, M.I., Iacoboni, M., 2004. Listening to speech activates motor areas involved in speech production. Nat. Neurosci. 7, 701–702.

Xu, L., Thompson, C.S., Pfingst, B.E., 2005. Relative contributions of spectral and temporal cues for phoneme recognition. J. Acoust. Soc. Am. 117, 3255–3267.

Yanovsky, Y., Ciatipis, M., Draguhn, A., Tort, A.B.L., Brankack, J., 2014. Slow oscillations in the mouse hippocampus entrained by nasal respiration. J. Neurosci. 34 (17), 5949–5964.

Zanto, T.P., Gazzaley, A., 2009. Neural suppression of irrelevant information underlies optimal working memory performance. J. Neurosci. 29 (10), 3059–3066.

Please cite this article as: Mai, G., et al., Delta, theta, beta, and gamma brain oscillations index levels of auditory sentence processing, NeuroImage (2016), http://dx.doi.org/10.1016/j.neuroimage.2016.02.064