PERSPECTIVE

Brain-to-speech decoding will require linguistic and pragmatic data

IOP Publishing Journal of Neural Engineering
J. Neural Eng. **15** (2018) 063001 (6pp) https://doi.org/10.1088/1741-2552/aae466

Perspective

Brain-to-speech decoding will require linguistic and pragmatic data

Leon Li[1,3] **and Serban Negoita**[2]

[1] Department of Psychology and Neuroscience, Duke University, Durham, NC, United States of America
[2] University of Maryland School of Medicine, Baltimore, MD, United States of America

E-mail: leon.inbox@gmail.com (L Li)

Received 6 May 2018, revised 24 September 2018
Accepted for publication 26 September 2018
Published 23 October 2018

Abstract

Objective. Advances in electrophysiological methods such as electrocorticography (ECoG) have enabled researchers to decode phonemes, syllables, and words from brain activity. The ultimate aspiration underlying these efforts is the development of a brain-machine interface (BMI) that will enable speakers to produce real-time, naturalistic speech. In the effort to create such a device, researchers have typically followed a bottom-up approach whereby low-level units of language (e.g. phonemes, syllables, or letters) are decoded from articulation areas (e.g. premotor cortex) with the aim of assembling these low-level units into words and sentences. *Approach*. In this paper, we recommend that researchers supplement the existing bottom-up approach with a novel top-down approach. According to the top-down proposal, initial decoding of top-down information may facilitate the subsequent decoding of downstream representations by constraining the hypothesis space from which low-level units are selected. *Main results*. We identify types and sources of top-down information that may crucially inform BMI decoding ecosystems: communicative intentions (e.g. speech acts), situational pragmatics (e.g. recurrent communicative pressures), and formal linguistic data (e.g. syntactic rules and constructions, lexical collocations, speakers' individual speech histories). *Significance*. Given the inherently interactive nature of communication, we further propose that BMIs be entrained on neural responses associated with interactive dialogue tasks, as opposed to the typical practice of entraining BMIs with non-interactive presentations of language stimuli.

Keywords: brain-to-speech, brain-machine interface, neurolinguistics, neuropragmatics

To read a person's mind may be impossible, but to decode a person's brain may be the next best thing. To this end, researchers have designed brain-machine interfaces (BMIs) that can decode binary responses, phonemes, letters, and even words from brain activity. The typical strategy is to follow what may be described as a 'bottom-up' approach. Namely, researchers decode low-level units of language (e.g. phonemes, syllables or letters) from motor articulation areas (e.g. premotor cortex), with the aim of potentially assembling these low-level linguistic units into higher-level units such as phrases and sentences (Brumberg *et al* 2011, Pei *et al* 2011). However, efforts enacting this bottom-up strategy have yet to achieve sufficient accuracy to enable real-time decoding of naturalistic speech. This raises the question of whether current neuroimaging modalities have sufficient spatial and temporal precision to enable the bottom-up strategy to succeed in the foreseeable future.

To overcome this challenge, a potential solution may be to integrate additional sources of information, which we call 'top-down' information, into BMI decoding ecosystems. The intuition here is that top-down information may help BMI decoders with the technical problem of constraining the hypothesis space of options from which the decoder selects.

[3] Author to whom any correspondence should be addressed.
Department of Psychology and Neuroscience, Duke University, Durham, NC 27707, United States of America.

This strategy of incorporating top-down information into BMI decoding ecosystems has proven to be beneficial to BMI functionality when utilizing a phonemic language model (Moses *et al* 2016) or gaze information (Zander *et al* 2010). Other relevant types of top-down information, as we will describe, include speakers' communicative intentions (e.g. their speech acts), the prevailing communicative pressures in a speaker's communicative context (i.e. the pragmatics of the situation), and the formal linguistic properties of the speaker's language. In this paper, we argue that a top-down approach is feasible and warranted. Of course, the top-down approach proposed here is not intended to replace the bottom-up approach, but rather to complement the existing technologies in this field. As such, we propose conceptual recommendations for how different types of top-down information may improve BMI functionality. We also review literature that suggests that the functional neuroanatomy that underlies communicative intentions is accessible to current neuroimaging modalities. Finally, we describe the need for dialogue-based entraining paradigms.

1. Bottom-up and top-down directions in the lexical stream

The consensus model of language production posits a downstream flow of information from conceptual to lexical to phonological levels of representation (Ferreira and Griffin 2003, Hickok 2012, Li and Slevc 2017). Communicative intentions at the conceptual level generate morpho-syntactic lexical units known as lemmas, which in turn generate the phonological lexemes that inform motor articulations. Language BMIs have typically followed a bottom-up strategy of attempting to decode low-level units of language such as letters or syllables. The advantage of this bottom-up approach is that low-level units are combinatorial and easier to isolate than higher-level units of discourse. Finding the unique neural correspondences for 26 English letters, for example, seems more feasible than finding the unique neural correspondences for tens of thousands of English words.

Fundamentally, accuracy is expected to be higher when the number of options in the hypothesis space is lower. Indeed, this pattern has been observed empirically. Accuracy is typically high, for example, when the hypothesis space is binary (i.e. when the decoder only selects between two options). When making a binary prediction about which of two story segments a person was reading, a decoding scheme using fMRI achieved an accuracy of 74% (Wehbe *et al* 2014). Another fMRI study achieved 90% accuracy for a binary decoding of whether participants were selectively attending to 'yes' or 'no' in a stream of sounds (Naci *et al* 2013). As well, a decoding scheme using electrocorticography (ECoG) achieved accuracies over 90% when attempting to determine which of ten sentences a participant was hearing (Moses *et al* 2018). Conversely, when using a similar ECoG decoding scheme to decode among 39 phonemes (including a silence phoneme), accuracy was only around 30% (Moses *et al* 2016). As well, another study that used microelectrodes positioned near primary motor cortex to decode among 38 phonemes only achieved accuracies of around 20% (Brumberg *et al* 2011). In comparison, a recent ECoG study in which the hypothesis space only included 4 phonemes achieved accuracy over 70% (Ramsey *et al* 2018). Thus, it would seem that a robust and intuitive way to improve decoding accuracy would be to constrain the hypothesis space of options from which a BMI decoder selects. The effectiveness of constraining the hypothesis space has already been partially demonstrated by the application of a phonemic language model (i.e. a model of expected frequencies of phoneme sequences based on a corpus analysis) for decoding phonemes from brain activity (Moses *et al* 2016).

A major disadvantage to the bottom-up approach is that low-level units are underdetermined. A given set of letters does not necessarily correspond to one and only one particular word (as anyone who has played Scrabble is aware), much less one and only one semantic meaning. The problem of ambiguity is exacerbated as the number of letters increases (e.g. if a speaker intends to produce a sentence as opposed to one individual word). It may be exceedingly difficult to isolate the unique neural signatures of the individual letters in a sentence, given the close temporal proximity and potential temporal overlap between letter representations (Indefrey 2011). In fact, a recent study that evaluated the effect of time position on phoneme classification revealed that decoding accuracy decreased as the time position of a phoneme within a sequence increased (Moses *et al* 2016). For instance, the fourth phoneme in a sequence would be less likely to be predicted correctly than the first. One approach towards circumventing this challenge was demonstrated in a recent ECoG based study, where a spatio-temporal matched filter algorithm achieved greater phoneme classification accuracy than a spatial matched filter, but comparable accuracy to a support vector machine based decoding algorithm (Ramsey *et al* 2018). These results support the value of including temporal variables in classification schemes, particularly in the idealized scenario where signals are temporally distinct. However, the 1.3 s spaces between discrete phonemes (consisting of 0.8 s average reaction times and 0.5 s average response lengths per phoneme) that were utilized in the previously described study may not be attainable when decoding naturalistic speech. A more temporally distinct signal may thus be necessary. Thus, phoneme decoding may be aided not only by the inclusion of more temporal information, but also by the inclusion of other types of information that are temporally independent of the time course of the speech stream.

The proposed 'top-down' recommendation for improving BMIs is to make use of information that may constrain the hypothesis space for the subsequent decoding of downstream representations such as phonemes or syllables. An especially important type of top-down information is the speaker's communicative intention. Information that is relevant to a speaker's communicative intention can be derived from several separate sources that are temporally distinct: speech acts (or in this case, their neural signatures), situational pragmatics, lexical and syntactic constraints, and individual speech histories.

2. Speech acts and communicative intentions

According to influential theories of speech acts (e.g. Searle (2001)), a speaker's utterance may be categorized as a 'speech act' according to the communicative function of the utterance (e.g. assertive speech acts convey information to the listener, whereas directive speech acts request behavior from the listener). A developing body of research under the title of neuropragmatics has aimed to investigate the neural signatures of speech acts. Research in this area has provided evidence that communicative intentions can be inferred from neural activity.

Using fMRI, for example, Egorova *et al* (2016) examined neural responses that arose when participants observed assertive and directive speech acts, which the authors respectively referred to as 'naming' and 'requesting'. The same words were used for the two speech act conditions. The conditions differed, however, in how the words were framed by a preceding sentence and what action followed the word. Requesting speech acts generated more neural activation than naming speech acts in the left inferior frontal gyrus (LIFG), bilateral premotor cortex, right posterior superior temporal sulcus (pSTS), and left anterior inferior parietal cortex; in contrast, naming speech acts generated more activation than request speech acts in the left angular gyrus (Egorova *et al* 2016). The right temporoparietal junction (right TPJ), an area associated with theory of mind, was active during both requests and naming, which attests to the communicative nature of the gestures (Egorova *et al* 2016). Another relevant study by Committeri *et al* (2015) examined the neural substrates of the production, not only the observation, of communicative gestures. Participants observed or produced declarative or imperative points while inside a fMRI scanner. Cortical regions associated with pointing production included bilateral ventral premotor cortex, anterior midcingulate cortex (aMCC), middle insula, and right presupplementary motor area (Committeri *et al* 2015). Consistent with Egorova *et al* (2016), the production of both declarative and imperative points was associated with activation in the right TPJ, which again confirms that the gestures were communicative actions, not only motor actions (Committeri *et al* 2015). Importantly, these results suggest that speech acts are associated with neural substrates that are cortical in location. As such, a variety of BMIs (including those using ECoG) may potentially be able to measure neural signals from several of these regions of interest (e.g. the right TPJ) in order to differentiate between different types of speech acts. Speech act information could potentially then be combined with phonemic language models to adaptively give greater weight when decoding phonemes to the particular phonemes that are relevant to particular speech acts. If a 'requesting' speech act was decoded, for instance, the model could adaptively give more weight to detecting phonemes associated with request-related words such as 'can' or 'would'.

Other research has broadly investigated the neural signatures of the presence or absence of intentions, including the intention to communicate. In particular, investigations have identified a variety of regions associated with the presence of motor and communicative intentions, such as the right pSTS, medial prefrontal cortex, dorsal prefrontal cortex, and intraparietal areas (Sassa *et al* 2007, Noordzij *et al* 2009, Carota *et al* 2010, Andersen *et al* 2014). These regions may potentially serve as candidate recording sites for decoding communicative intentions.

An especially promising candidate area for decoding communicative intentions is the middle temporal gyrus, as models of speech production have suggested that activity in this region emerges early in the temporal time course of the production stream, prior to the neural markers of phonetic articulation in the superior temporal gyrus, anterior cingulate, LIFG, left precentral gyrus, left thalamus, and cerebellum (Indefrey 2011, Tankus *et al* 2012). The sequence of this time course allows for the possibility that neural data about communicative intentions, once decoded, could be rapidly employed to constrain the hypothesis space for immediately subsequent decoding of motor regions. Notably, the rostral anterior cingulate and the medial orbitofrontal cortex may be candidate recording sites (in addition to the typical premotor areas) for decoding phonetic information, as research has suggested that these regions are highly tuned to particular vowel representations (Tankus *et al* 2012).

Another important consideration for decoding phonetic information is the calibration of the time window around the onset of the articulation. Research has suggested that the most informative time window occurs close to the onset of articulation. For example, Ramsey *et al* (2018) found that a time window of 400 to 500 ms centered around voice onset was highly effective. Similarly, Mugler *et al* (2014) found that 'features spanning 200 ms before to 200 ms after phoneme onset accounted for 88.1% of peak performance' (p. 6). Even more specifically, Mugler *et al* (2014) stated that the 'most informative time bin occurred right at phoneme onset (0–50 ms) across subjects' (p. 6). These recommendations will be highly useful for researchers seeking to incorporate temporal information into their decoding ecosystems. A further proposal is that speakers' communicative intentions could, in theory, also be decoded with respect to arousal or valence (e.g. positive or negative valence), which may be associated with activity in the amygdala, limbic regions, and autonomic nervous system. Valence information may be important to decode because the propositional content of a speech act often depends on whether the speaker wishes to say something positive or negative about a referent (e.g. even the very same word may convey different meanings when spoken in a warm versus a stern tone).

3. Situational and pragmatic data

In addition to the speaker's communicative intention, the communicative situation itself is a highly enriched source of information about what the speaker is likely to say. There exist robust cultural expectations about the communicative pressures and communicative intentions that are recurrent in particular situations (Tomasello 2003). Indeed, situational data alone may be sufficient to infer speakers' words in some cases. In studies using the human simulation paradigm (HSP), adults who view muted clips of parent-child interactions were able to

estimate, with relatively high accuracy, what word the parent said (e.g. Cartmill *et al* (2013)). As such, this situational information could be beneficial to a BMI's decoding ecosystem. Incorporating situational information into the BMI decoding ecosystem may involve using a combination of eye-tracking and object recognition software to detect what object a participant is attending to. This information, in combination with phonemic language models, could enable the BMI to predict phonemes and words that are relevant to the attended objects.

In previous integrations of BMIs and eye-tracking, BMIs have served as the 'clicking mechanism' to signal the user's intention to 'select' what they are viewing. In a study by Zander *et al* (2010), participants using BMI were able to signal a click by visualizing wringing a towel. Zander *et al* (2010) found that the use of the BMI as the clicking mechanism was, in some cases, more accurate and less frustrating than the use of dwell times as the clicking mechanism. This ability of a BMI to serve as a clicking mechanism demonstrates that eye-tracking in combination with BMI input can be used to indicate what object a speaker wishes to draw attention to. With appropriate object recognition software, the attended object could be classified in terms of its linguistic features, such as its phonetic or lexical properties. Similarly, an object-detector could enable the BMI decoding ecosystem to consider the communicative expectations associated with the objects in the speaker's vicinity. Information about the communicative situation may also be derived from location-tracking technologies such as GPS (e.g. if the BMI device detects that a speaker is at a grocery store, it may adaptively expect that the speaker will talk about objects that are typically found in grocery stores).

Importantly, a challenge to integrating contextual information via the top-down approach is that speakers may sometimes wish to talk about topics that are not localized to the current situation (e.g. a speaker at a grocery store may wish to discuss politics rather than groceries). A brain-to-speech BMI that assumes that speakers only wish to talk about situationally proximal topics would be far too restrictive for naturalistic speech. In order to address this concern, it will be important for brain-to-speech BMIs to somehow optimally give weight to situationally proximal topics without excluding the possibility that speakers may wish to discuss topics that are far removed from the context. Future research will be needed to discover how to optimize the relative weights of situationally proximal and situationally distal influences on a speaker's intended discourse. In summary, although decoded brain data may be highly informative about people's mental states (Haynes and Rees 2006), additional sources of relevant contextual information may also prove useful to the decoding ecosystems of BMIs.

4. Lexical and syntactic constraints

Formal linguistic information will be necessary for producing naturalistic sentence outputs, as opposed to isolated words. Even if the neural signatures of individual letters and words, as well as the neural signatures of communicative intentions, could be decoded from the brain, there would still remain the problem of assembly. Communicative intentions are underdetermined with respect to syntactic constructions because the same message may be said in multiple ways (e.g. *dog bites man* versus *man was bitten by dog*). The danger, in particular, is that slight alterations in syntactic features such as morphological marking or word order (e.g. compare *man bites dog* to *dog bites man*) may drastically distort the semantic meaning of an utterance (Gertner *et al* 2006). A naturalistic BMI will need to be able to assemble decoded words in the correct, intended order. The solution to this ordering problem will be to provide the decoding ecosystem with information about the formal grammar of the target language (Chomsky 1967, Yang *et al* 2017) as well as information about high-frequency sentence constructions that speakers in the language community are likely to employ (Tomasello 2003). For example, the BMI's syntactic assembly algorithm may be aided by information about regularities in verb arguments. The transitive verb 'eat', for instance, may take as a direct object a noun that is edible (Ferguson *et al* 2014). Thus, when a linguistically enriched BMI decodes the word 'eat', it may be able to adaptively expect the subsequent occurrence of a noun representing an edible object (with implications for downstream phoneme predictions). Relatedly, a BMI's lexical selections may be aided by information about collocations (i.e. regularities in the co-occurrences of words), which have been extensively described in linguistic corpora studies (e.g. Heylen *et al* (2015)). If a BMI has decoded the word 'cat' with high confidence, for instance, it may then expect that related words such as 'pet' or 'house' will likely occur in the upcoming speech stream as well.

Although it is advisable to equip brain-to-speech BMIs with syntactic expectations, it is also important to acknowledge that naturalistic speech itself often does not conform to syntactic expectations. Naturalistic speech is full of speech errors, incomplete sentence fragments, and ambiguities (Tomasello 2003, Slevc and Ferreira 2006). A brain-to-speech BMI that only expects to find syntactically coherent speech outputs when decoding neural activity may thereby fail to decode naturalistic speech that is genuinely ungrammatical. One solution to this challenge is to leave the assumption of syntactic coherence intact, in which case speakers could adapt to this assumption by learning to only produce syntactically coherent outputs. This solution may prove to be cumbersome and tiring for speakers, however. Another solution that may result in a more user-friendly BMI will be to relax the expectation that speech outputs are always syntactically coherent. If so, the degree to which syntactic coherence is expected may be optimized for ease of use and efficiency (perhaps even at the level of the individual speaker). As with the case of optimizing the predictive contribution of situationally pragmatic information, the question of how to best optimize the predictive contribution of a top-down category of information (in this case, syntactic information) without making the top-down expectation too restrictive will be an important direction for future investigation and theory.

5. Individual speech histories

Another relevant type of information is the speaker's individual history of speech, given that speakers are generally likely to repeat the words, constructions, and messages that they have communicated before (Tomasello 2000). Accordingly, an individual speaker's personal BMI could be calibrated to expect the production of the words and constructions that the speaker tends to use. Information about a speaker's individual speech history may also be used in combination with information about the communicative context (e.g. the BMI device may observe that a speaker tends to discuss certain topics at certain locations). Analysis of the distributional, lexical, and syntactic properties of decoded letters and words will require conversion of the neural data into text as an intermediary step, as opposed to direct conversion into acoustic output (Herff *et al* 2015). This intermediate step is needed because textual information lends itself to linguistic analysis, whereas purely acoustic information does not. In sum, the convergence of pragmatic information from linguistic, situational, and speaker history data may greatly assist BMIs in the task of sentence assembly.

6. Entraining decoders on dialogue

To maximize the yield of the aforementioned strategies, it may be of benefit to entrain BMI decoding ecosystems on tasks involving dialogue, as opposed to the static presentation of context-independent stimuli. Language is used, after all, for communication and interaction (Pickering and Garrod 2004, Tomasello 2008). This approach would be consistent with the emerging consensus in the field of social cognitive neuroscience (Schilbach 2010, Gallotti and Frith 2013) that researchers ought to investigate the neural signatures of actual social interactions (e.g. in the 'second-party stance' or the 'we-mode'), not just the neural signatures associated with observations of others' social interactions (e.g. in the 'third-party stance'). Interaction differs from observation in several ways. Participants are more motivated to be responsive, are more emotionally engaged as opposed to detached, and are able to influence the environment, not just observe it (Pfeiffer *et al* 2013). Since these are all features of real communication, it will be important for BMIs to take the neural signatures of these features into account. To date, much of the entraining that has been enacted in the BMI literature has not used interactive contexts due to a variety of logistical and technical challenges. In the training phase of Herff *et al* (2015), for instance, participants were asked to read aloud text that was presented to them on a screen. The neural responses recorded during this reading phase were then used to decode test material. However, given that the neural responses in this training phase were associated with no real communicative interactions, and therefore, presumably, no real communicative intentions, it is uncertain whether this decoding scheme could have been applied to decode naturalistic speech reflecting actual communicative intentions.

In order to entrain the BMI decoder on patterns of brain activity associated with dialogue, the speaker may participate in a training phase that involves a simulated dialogue. In other words, the speaker may engage in a conversation with an interlocutor who follows a script that highly constrains what the speaker is likely to say (e.g. the interlocutor asks leading questions). This would allow the BMI to decode the neural activity that is associated with the conceptual preparation of a message (while controlling for the content of the message, given that the content will be predictably constrained by the dialogue task). It is likely that the neural data associated with the comprehension of an interlocutor's message will also be informative to BMIs, especially if this information could be used in combination with the neural data associated with what the speaker herself intends to say (Pickering and Garrod 2007).

7. Conclusion

To conclude, we have offered a set of strategic recommendations for the effort to design a naturalistic brain-to-speech BMI. Researchers have made tremendous progress in characterizing the neural substrates of the motor execution of speech (Tourville and Guenther 2011) and the neural language network more broadly (Friederici 2002, Hagoort 2005, Xiang *et al* 2010, Catani and Bambini 2014, Fedorenko and Thompson-Schill 2014). Researchers have also made progress in using BMIs to decode phonemes, syllables, and words from neural data (Kellis *et al* 2010, Pei *et al* 2011, Tankus *et al* 2012). However, further advances in BMI design will likely require locating and leveraging additional kinds of data. These data may include communicative intentions, situational communicative pragmatics, and linguistic content (e.g. syntactic rules and constructions, lexical collocations and frequencies, speakers' individual speech histories). We propose, in summary, that these kinds of top-down information will increase the accuracies of BMI decoders by constraining the hypothesis space of options from which a decoder selects. The search is on for the development of a natural, intuitive, and accurate brain-to-speech decoding device that will draw upon both neural and external inputs.

Acknowledgments

Both authors express gratitude to Dr Susan Courtney for formative research training at Johns Hopkins University. The first author is thankful to Dr L Robert Slevc, Dr Elika Bergelson, and Dr Michael Tomasello for helpful discussions on language. Useful feedback on this paper was provided by Rohan Ahuja, Gagan Tunuguntla, Jared Vasil, and the Duke University Writing Studio. The authors also express thanks to Prigel Family Creamery and Loch Raven Reservoir for providing tranquil settings for the conversations that motivated this paper. The authors declare that there are no financial conflicts of interest regarding this paper.

ORCID iDs

Leon Li https://orcid.org/0000-0002-7289-5101
Serban Negoita https://orcid.org/0000-0003-2477-0379

References

Andersen R A, Kellis S, Klaes C and Aflalo T 2014 Toward more versatile and intuitive cortical brain-machine interfaces *Curr. Biol.* **24** R885–97
Brumberg J S, Wright E J, Andreasen D S, Guenther F H and Kennedy P R 2011 Classification of intended phoneme production from chronic intracortical microelectrode recordings in speech-motor cortex *Frontiers Neurosci.* **5** 65
Carota F, Posada A, Harquel S, Delpuech C, Bertrand O and Sirigu A 2010 Neural dynamics of the intention to speak *Cereb. Cortex* **20** 1891–7
Cartmill E A, Armstrong B F, III, Gleitman L R, Goldin-Meadow S, Medina T N and Trueswell J C 2013 Quality of early parent input predicts child vocabulary 3 years later *Proc. Natl Acad. Sci.* **110** 11278–83
Catani M and Bambini V 2014 A model for social communication and language evolution and development (SCALED) *Curr. Opin. Neurobiol.* **28** 165–71
Chomsky N 1967 Recent contributions to the theory of innate ideas *Synthese* **17** 2–11
Committeri G, Cirillo S, Costantini M, Galati G, Romani G L and Aureli T 2015 Brain activity modulation during the production of imperative and declarative pointing *NeuroImage* **109** 449–57
Egorova N, Shtyrov Y and Pulvermüller F 2016 Brain basis of communicative actions in language *NeuroImage* **125** 857–67
Fedorenko E and Thompson-Schill S L 2014 Reworking the language network *Trends Cogn. Sci.* **18** 120–6
Ferguson B, Graf E and Waxman S R 2014 Infants use known verbs to learn novel nouns: evidence from 15- and 19-month-olds *Cognition* **131** 139–46
Ferreira V S and Griffin Z M 2003 Phonological influences on lexical (mis)selection *Psychol. Sci.* **14** 86–90
Friederici A D 2002 Towards a neural basis of auditory sentence processing *Trends Cogn. Sci.* **6** 78–84
Gallotti M and Frith C D 2013 Social cognition in the we-mode *Trends Cogn. Sci.* **17** 160–5
Gertner Y, Fisher C and Eisengart J 2006 Learning words and rules: abstract knowledge of word order in early sentence comprehension *Psychol. Sci.* **17** 684–91
Hagoort P 2005 On Broca, brain, and binding: a new framework *Trends Cogn. Sci.* **9** 416–23
Haynes J-D and Rees G 2006 Decoding mental states from brain activity in humans *Nat. Rev. Neurosci.* **7** 523–34
Herff C, Heger D, de Pesters A, Telaar D, Brunner P, Schalk G and Schultz T 2015 Brain-to-text: decoding spoken phrases from phone representations in the brain *Frontiers Neurosci.* **9** 217
Heylen K, Wielfaert T, Speelman D and Geeraerts D 2015 Monitoring polysemy: word space models as a tool for large-scale lexical semantic analysis *Lingua* **157** 153–72
Hickok G 2012 Computational neuroanatomy of speech production *Nat. Rev. Neurosci.* **13** 135–45
Indefrey P 2011 The spatial and temporal signatures of word production components: a critical update *Frontiers Psychol.* **2** 255
Kellis S, Miller K, Thomson K, Brown R, House P and Greger B 2010 Decoding spoken words using local field potentials recorded from the cortical surface *J. Neural Eng.* **7** 056007
Li L and Slevc L R 2017 Of papers and pens: polysemes and homophones in lexical (mis)selection *Cogn. Sci.* **41** 1532–48
Moses D A, Leonard M K and Chang E F 2018 Real-time classification of auditory sentences using evoked cortical activity in humans *J. Neural Eng.* **15** 036005
Moses D A, Mesgarani N, Leonard M K and Chang E F 2016 Neural speech recognition: continuous phoneme decoding using spatiotemporal representations of human cortical activity *J. Neural Eng.* **13** 056004
Mugler E M, Patton J L, Flint R D, Wright Z A, Schuele S U, Rosenow J, Shih J J, Krusienski D J and Slutzky M W 2014 Direct classification of all American English phonemes using signals from functional speech motor cortex *J. Neural Eng.* **11** 035015
Naci L, Cusack R, Jia V Z and Owen A M 2013 The brain's silent messenger: using selective attention to decode human thought for brain-based communication *J. Neurosci.* **33** 9385–93
Noordzij M L, Newman-Norlund S E, de Ruiter J P, Hagoort P, Levinson S C and Toni I 2009 Brain mechanisms underlying human communication *Frontiers Hum. Neurosci.* **3** 14
Pei X, Barbour D L, Leuthardt E C and Schalk G 2011 Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans *J. Neural Eng.* **8** 046028
Pfeiffer U J, Timmermans B, Vogeley K, Frith C D and Schilbach L 2013 Towards a neuroscience of social interaction *Frontiers Hum. Neurosci.* **7** 22
Pickering M J and Garrod S 2004 Toward a mechanistic psychology of dialogue *Behav. Brain Sci.* **27** 169–225
Pickering M J and Garrod S 2007 Do people use language production to make predictions during comprehension? *Trends Cogn. Sci.* **11** 105–10
Ramsey N F, Salari E, Aarnoutse E J, Vansteensel M J, Bleichner M G and Freudenburg Z V 2018 Decoding spoken phonemes from sensorimotor cortex with high-density ECoG grids *NeuroImage* **180** 301–11
Sassa Y, Sugiura M, Jeong H, Horie K, Sato S and Kawashima R 2007 Cortical mechanism of communicative speech production *NeuroImage* **37** 985–92
Schilbach L 2010 A second-person approach to other minds *Nat. Rev. Neurosci.* **11** 449
Searle J R 2001 *Rationality in Action* (Cambridge, MA: MIT Press)
Slevc L R and Ferreira V S 2006 Halting in single word production: A test of the perceptual loop theory of speech monitoring *J. Mem. Lang.* **54** 515–40
Tankus A, Fried I and Shoham S 2012 Structured neuronal encoding and decoding of human speech features *Nat. Commun.* **3** 1015
Tomasello M 2000 Do young children have adult syntactic competence? *Cognition* **74** 209–53
Tomasello M 2003 *Constructing a Language: a Usage-Based Theory of Language Acquisition* (Cambridge, MA: Harvard University Press)
Tomasello M 2008 *Origins of Human Communication* (Cambridge, MA: MIT Press)
Tourville J A and Guenther F H 2011 The DIVA model: a neural theory of speech acquisition and production *Lang. Cogn. Process.* **26** 952–81
Wehbe L, Murphy B, Talukdar P, Fyshe A, Ramdas A and Mitchell T 2014 Simultaneously uncovering the patterns of brain regions involved in different story reading subprocesses *PLoS One* **9** e112575
Xiang H-D, Fonteijn H M, Norris D G and Hagoort P 2010 Topographical functional connectivity pattern in the perisylvian language networks *Cereb. Cortex* **20** 549–60
Yang C, Crain S, Berwick R C, Chomsky N and Bolhuis J J 2017 The growth of language: universal Grammar, experience, and principles of computation *Neurosci. Biobehav. Rev.* **81** 103–19
Zander T O, Gaertner M, Kothe C and Vilimek R 2010 Combining eye gaze input with a brain-computer interface for touchless human-computer interaction *Int. J. Hum. Comput. Interact.* **27** 38–51