Brain-Computer Interfaces

ISSN: 2326-263X (Print) 2326-2621 (Online) Journal homepage: http://www.tandfonline.com/loi/tbci20

Neurolinguistic and machine-learning perspectives on direct speech BCIs for restoration of naturalistic communication

Olga Iljina, Johanna Derix, Robin Tibor Schirrmeister, Andreas Schulze-Bonhage, Peter Auer, Ad Aertsen & Tonio Ball

To cite this article: Olga Iljina, Johanna Derix, Robin Tibor Schirrmeister, Andreas Schulze-Bonhage, Peter Auer, Ad Aertsen & Tonio Ball (2017): Neurolinguistic and machine-learning perspectives on direct speech BCIs for restoration of naturalistic communication, Brain-Computer Interfaces, DOI: 10.1080/2326263X.2017.1330611

To link to this article: http://dx.doi.org/10.1080/2326263X.2017.1330611

Published online: 21 Jun 2017.

BRAIN-COMPUTER INTERFACES, 2017
https://doi.org/10.1080/2326263X.2017.1330611

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Neurolinguistic and machine-learning perspectives on direct speech BCIs for restoration of naturalistic communication

Olga Iljina[a,b,c,e,g,h], Johanna Derix[e,h], Robin Tibor Schirrmeister[e,h], Andreas Schulze-Bonhage[d,e], Peter Auer[a,b,c,f], Ad Aertsen[g,i] and Tonio Ball[e,h]

[a]GRK 1624 'Frequency effects in language', University of Freiburg, Freiburg, Germany; [b]Department of German Linguistics, University of Freiburg, Freiburg, Germany; [c]Hermann Paul School of Linguistics, University of Freiburg, Germany; [d]Epilepsy Center, Department of Neurosurgery, University Medical Center Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany; [e]BrainLinks-BrainTools, University of Freiburg, Freiburg, Germany; [f]Freiburg Institute for Advanced Studies (FRIAS), University of Freiburg, Freiburg, Germany; [g]Neurobiology and Biophysics, Faculty of Biology, University of Freiburg, Freiburg, Germany; [h]Translational Neurotechnology Lab, Department of Neurosurgery, University Medical Center Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany; [i]Bernstein Center Freiburg, University of Freiburg, Freiburg, Germany

ABSTRACT

The ultimate goal of brain-computer-interface (BCI) research on speech restoration is to develop devices which will be able to reconstruct spontaneous, naturally spoken language from the underlying neuronal signals. From this it follows that thorough understanding of brain activity and its functional dynamics during real-world speech will be required. Here, we review current developments in intracranial neurolinguistic and BCI research on speech production under increasingly naturalistic conditions. With an example of neurolinguistic data from our ongoing research, we illustrate the plausibility of neurolinguistic investigations in non-experimental, out-of-the-lab conditions of speech production. We argue that interdisciplinary endeavors at the interface of neuroscience and linguistics can provide valuable insight into the functional significance of speech-related neuronal data. Finally, we anticipate that work with neurolinguistic corpora composed of real-world language samples and simultaneous neuronal recordings, together with machine-learning methodology accounting for the specifics of the neurolinguistic material, will improve the functionality of speech BCIs.

ARTICLE HISTORY
Received 2 August 2016
Accepted 11 May 2017

KEYWORDS
BCI; ECoG; speech production; neurolinguistics; real-world communication

1. Introduction

Establishing approaches to enable communication in severely paralyzed patients is a major aim of brain-computer-interface (BCI) research [1]. One increasingly pursued method is to infer the message the patient desires to convey from the underlying neuronal activity in speech-related brain areas and to externalize the message using an assistive technical device. A speech BCI based on this method can be referred to as 'direct', as it relies on a one-to-one correspondence between the neuronal activity and the behavioral output [2]. For instance, if a paralyzed patient wants to say, 'I'd like to have a cup of strong black coffee, please', a direct speech BCI, also referred to as a 'brain-to-text' system [3,4], will try to reconstruct this utterance as precisely as possible from the underlying neuronal signal and convert it into, for example, a text message or into voiced synthetic speech via a text-to-speech device.

Speech reconstruction from brain activity is not only an exciting but also a challenging endeavor, and different recording methods, both hemodynamic and electrophysiological, have been used to attempt it. Each of them has its own drawbacks and advantages. Popular hemodynamic methods are functional magnetic resonance imaging (fMRI) and functional near-infrared spectroscopy (fNIRS). They detect brain activity by non-invasive measurements of changes in the level of cerebral oxygenation. These are correlated with changes in neuronal activation [5] although this correlation can be loose [6] and dependent on the investigated anatomical area [7]. fMRI has the advantage of whole-head coverage, and it allows studying the entire network involved in task-related processing. In comparison with fMRI, fNIRS can only measure hemodynamic responses from the outer cortical layers, but it is associated with the advantages of being silent, portable, and cheaper to acquire [8,9].

CONTACT Olga Iljina olga.iljina@uniklinik-freiburg.de

Recent BCI studies have shown that fNIRS can be used to decode intended simple communication via 'yes' or 'no' responses in paralyzed subjects [10,11], suggesting that fNIRS can aid restoration of communicative abilities in complete paralysis [12]. Hemodynamic methods of measurement are non-invasive, and they possess good spatial resolution. Nevertheless, their main demerit for speech-related research is that the temporal resolution of the recorded signals only allows detecting changes on the temporal scale of seconds, corresponding to the slow speed of metabolic processes inherent to these methods.

To be able to capture dynamic speech-related changes in the time range of milliseconds, electrophysiological recordings [3] or combinations of electrophysiological and hemodynamic methods [10] may be helpful. Electroencephalography (EEG) and magnetoencephalography (MEG) are popular non-invasive methods to study speech-related neuronal processing. These methods can record electrophysiological signals from the cortical surface, and they possess very high temporal resolution and spatial resolution comparable with fNIRS and fMRI, respectively. One demerit of these methods, however, is that they can only record bone- and scalp-filtered signals and therefore offer inferior signal-to-noise ratios compared with intracranial electrophysiology [13]. An important methodological drawback of both hemodynamic and non-invasive electrophysiological methods for studies on expressive language is that they are prone to distortion of the brain signal by myographic activity. It accompanies expressive speech, during covert production as well [14], and may thus obscure biologically relevant neuronal responses. Due to this drawback, perception-based paradigms are often preferred when studying linguistic phenomena with the help of non-invasive methods [15], and research on linguistic functions in conditions of speech production is underrepresented in comparison [16].

Electrocorticography (ECoG) obtained in neurological patients is an attractive alternative to non-invasive data, especially for studies on expressive language. Such recordings have a high signal-to-noise ratio, they possess very high temporal and high spatial resolution under the area covered by electrodes [13,17], and their invasive nature keeps the impact of myographic activity on the recorded signal moderate (a comparison of simultaneously recorded ECoG and EEG over the same cortical region is shown in Fig. 1 of Derix et al. [18]; also see Fiederer et al. [19] for a characterization of myographic impacts on ECoG signals). Due to its invasiveness, this method of measurement can only be implemented for neurolinguistic research in consented neurological patients in whom ECoG electrodes are implanted for diagnostic procedures. Accordingly, ECoG is available for research only at large neurosurgical centers, and the placement and the number of electrodes can differ between subjects due to their individual clinical needs. Also, a researcher may often have to wait for a suitable patient, in whom the main areas of interest are sufficiently covered by electrodes and whose clinical picture allows for studying the areas of interest (e.g. one may want to make sure the seizure onset zone in patients with epilepsy lies outside the respective areas) [18,20,21]. Important advantages of using this method for speech-related research are that neuronal activity can be recorded directly from the cortical surface, and that recordings, which are usually obtained over an extended period of one to three weeks of pre-neurosurgical diagnostics, can be used to study naturalistic, self-initiated human behavior [21,22] and real-world communication [18,20,23].

Another invasive method used in previous direct speech BCI research is intracortical microelectrodes, which can record extracellular potentials from small multi-unit populations. Like ECoG, this recording method possesses high signal quality and robustness against ocular and other movement-related artifacts, and it has been used in long-term clinical trials due to the comparatively small area of implantation and hence a reduced risk of infection or mechanical damage to the neuronal tissue [24,25]. A major limitation of this approach, however, is that it allows recording from small cortical regions, which may be a risk factor for long-term stability of the recorded neuronal signal.

In the following, we describe the latest developments in the field of ECoG research on speech production using increasingly naturalistic approaches (section 1), illustrate the plausibility of studying speech production-related neuronal activity in conditions of real-world communication using a visualized presentation of such activity from our ongoing research (section 2), and address the challenge of understanding the functional significance of brain activity in its temporal dynamics during continuous, non-experimental communication (section 3). Then, we review the latest developments in the area of direct speech BCI research with ECoG from linguistic (section 4) and machine-learning (section 6) perspectives. We argue in these sections that direct speech BCIs can profit from linguistically grounded approaches embracing multiple levels of linguistic abstraction and from the usage of decoding methodologies accounting for the hierarchical and probabilistic nature of the linguistic material. In section 5, we propose that methodology from corpus-based linguistic research may be useful to refine principles of natural speech decoding. Finally, we provide an outlook for the field of invasive direct speech BCIs in section 7.

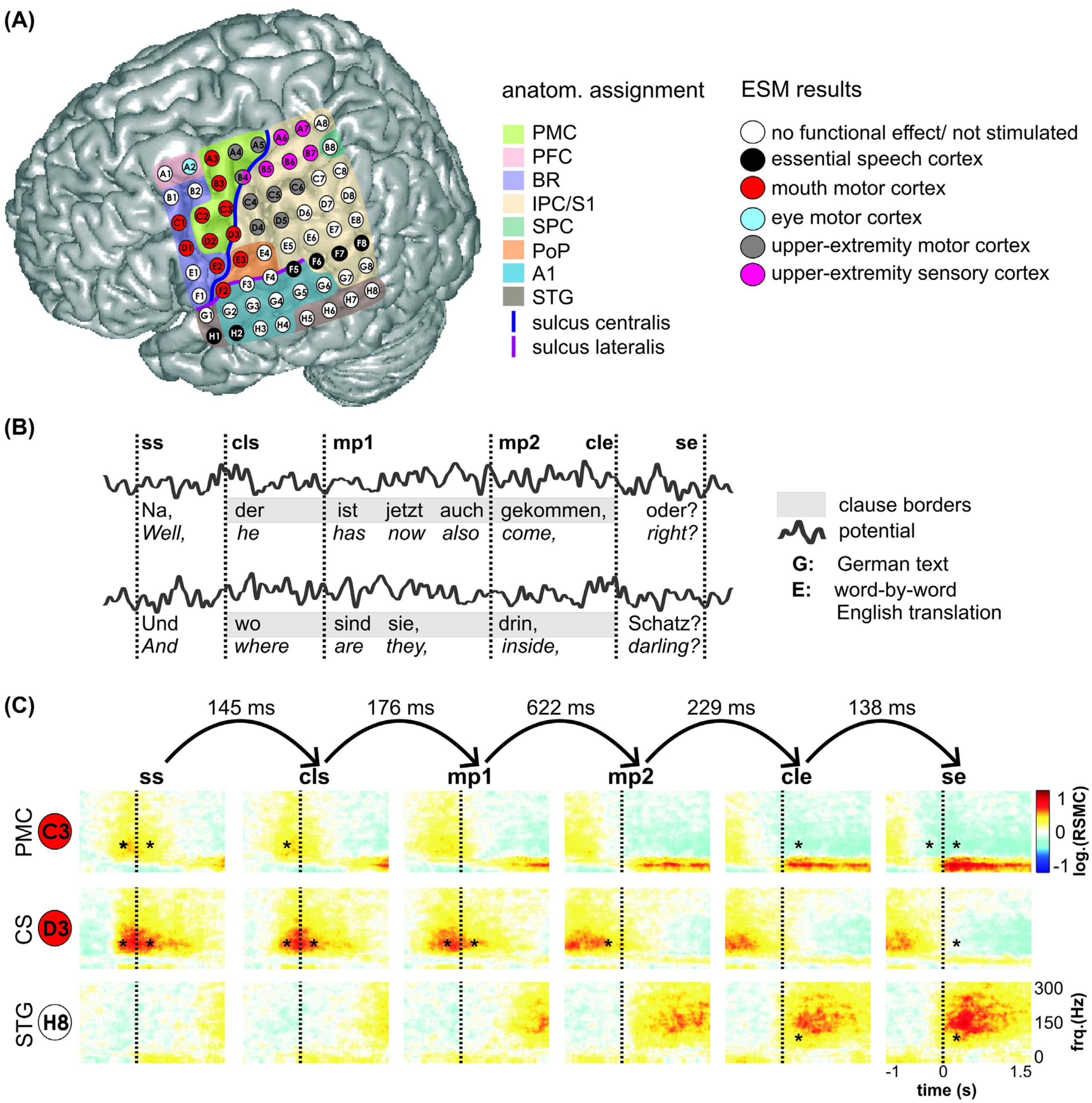

Figure 1. Spatially and temporally specific ECoG responses can be observed at different temporal stages of non-experimental, real-life production of simple clauses. (A) Individual electrode positions of one subject are visualized on the standard brain from SPM5 based on their MNI coordinates. Anatomical assignment (procedure described in detail in [21]) and the results of electrocortical stimulation mapping (ESM) are color-coded (see legend). Electrode labels (e.g. C3) are included for ease of reference. PMC, premotor cortex; PFC, prefrontal cortex; BR, Broca's area; SPC/IPC, superior/inferior parietal cortex; PoP, parietal operculum; A1, primary auditory cortex; S1, primary somatosensory cortex; STG, superior temporal gyrus without assignment to a particular Brodmann area. (B) A schematic illustrating linguistic compositions of the simple clauses. Dotted lines with marker names above them ('ss', speech start; 'cls', clause start; 'cle', clause end; 'se', speech end) indicate the positions of the corresponding linguistic events (also referred to as 'conditions') in the simultaneously recorded raw ECoG signal. A schematic of the raw potential at one electrode is visualized as a black solid line. Vertical dashed lines: onset of the respective condition. (C) Examples of relative spectral magnitude changes (RSMC) in speech-relevant anatomical areas of the respective subject during non-experimental, real-world production of 232 simple clauses. '0' (white color) indicates no change in spectral magnitude relative to the baseline period, yellow and red colors depict increases and blue colors decreases relative to the baseline. The time information above the bullets shows trial-averaged (median) temporal differences between conditions in the course of the speech production epoch. Please note that the short temporal differences between 'ss' and 'cls', 'cle' and 'se', 'cls' and 'mp1', and 'mp2' and 'cle' may be due to the fact that these trial categories could, in some instances, coincide in time due to their linguistic definition. Abbreviations: frq., frequency; log., natural logarithm. Asterisks before the dotted line, significant effects in the early time window; asterisks after the dotted line, significant effects in the later time window; other conventions as in (B).

2. Developments in ECoG research on speech production using increasingly naturalistic approaches

ECoG is more and more frequently being used in neurolinguistic experiments [26,27], and it has proven well suited to the study of the neuronal signatures of expressive language. Previous ECoG research has shown that spatially and temporally specific neuronal effects can be obtained in the fronto-temporo-parietal cortex during overt speech production [28]. They align well with speech-relevant areas identified by means of electrocortical stimulation mapping [29] and occur in a spatially [23,30] as well as temporally [31,32] specific manner. Until recently, ECoG investigations on speech production have most frequently addressed the general properties of the neuronal activity regardless of linguistic parameters of the spoken material, and they have classically been conducted to evaluate the usefulness of such data for localizing eloquent language cortex in pre-neurosurgical diagnostics [21,29,33,34].

In recent years, ECoG has increasingly been used to study speech-related neuronal processing from a linguistic perspective. Phonological properties of the spoken language material, such as the place of phoneme articulation [4,35,36], the manner of articulation, the voicing, and the distinction between vowels and consonants [4], have been studied. Using a syntactically informed approach to segmentation of continuous speech, Derix et al. [20] investigated the neuronal differences between ECoG recordings obtained during the production of simple sentences with compared to without memory content. In a sociolinguistic perspective on speech-ECoG data, Derix et al. [18] studied dialog partner-specific differences in the neuronal activity during spontaneous communication. Chen et al. [37] investigated differences between semantic categories of words during overt naming, and Fedorenko et al. [38] investigated the neural correlates of how the meaning of a sentence is constructed.

Linguistically grounded ECoG studies, however, are still a handful, and one main challenge in current research is 'to break down language function into computational primitives suitable for biology' (D. Poeppel in [26]). This is, in our opinion, a particularly urgent call for studies on speech production. Compared to the extensively studied side of speech perception, little evidence is available on the spatiotemporal representation of speech production and its subfunctions in the human brain [16]. Such evidence is essential to make linguistically informed choices of, for example, optimal (and possibly spatially dispersed) implantation sites and spatial scales for direct speech BCI applications.

The ultimate goal of direct speech BCI research is to develop devices which will be able to reconstruct spontaneous, naturally spoken language from the underlying neuronal signal. From this it follows that a thorough understanding of brain activity during connected, real-world speech will be required. Current knowledge on the neuronal processes involved in expressive language, however, has largely been gathered using decontextualized and disconnected linguistic output. It is thus conceivable that natural speech production may differ from speech elicited in simplified experiments, e.g. with regard to a different amount of working memory, behavioral restrictedness or associated attentional resources [18,39]. Non-experimental research based on speech produced during continuous, real-world communication can enable comparisons with previous experimental observations.

Brain activity during spontaneous, non-experimental communication, however, is a largely unexplored phenomenon, which has been referred to as 'the dark matter' of cognitive neuroscience [40]. To bridge this gap, several labs worldwide have started investigating the neuronal activity underlying language in extraoperatively recorded ECoG in increasingly naturalistic experiments [3,4,41–44] as well as in conditions of non-experimental, real-world communication [18,20,21,23,45–49]. Their reported effects align with previous experimental findings (cf. Crone et al. [28]) and reflect the somatotopic arrangement of the human sensorimotor cortex [21]. This research provides proof of principle for the non-experimental approach and allows elucidation of the largely unexplored neuronal signatures of authentic, uninstructed communication.

3. The plausibility of the non-experimental approach to study real-world speech production

To further extend this non-experimental line of research, we are currently constructing a multimodal neurolinguistic corpus. It contains retrospectively documented uninstructed, real-world, spontaneous speech production with concurrent ECoG recordings. The speech data have been acquired based on simultaneous audio and video materials. These multimodal recordings have been obtained at the University Medical Center Freiburg for the purpose of pre-neurosurgical diagnostics of epilepsy, and they were donated for research by consented neurological patients. All speakers included in the corpus (data sets from a total amount of eight subjects at different stages of annotation) gave written informed consent that these multimodal recordings would be made available for scientific investigation, and the Ethics Committee of the University Medical Center Freiburg approved the recruitment procedure [20]. The linguistic data in our corpus consist of continuous transcriptions of the subjects' speech production based on the audiovisual signals. Transcriptions were generated by trained linguists according to GAT-2 conventions for the 'basic' transcript [50] established in applied linguistics. A crucial advantage of using this method is that it provides detailed information about the linguistic material, including accents, pitch contours, intensity, pause duration, and several paralinguistic features. Simple clauses constitute basic units of speech production in our corpus. We identify them in continuous transcriptions of the subjects' speech according to structural linguistic criteria [51], tag the outer borders of the clauses as well as several other phenomena relevant to clause description (Figure 1(C)) in ECoG data using the Coherence EEG/PSG System software by Deltamed (Paris, France), analyze the clauses according to linguistic parameters (parts of speech, syntactic constituents, and dependency relations, etc.), and study the associated neuronal effects.

The main advantage of such an integrative approach combining expertise in linguistics and invasive electrophysiology is that it offers a unique opportunity to elucidate the neuronal correlates of linguistic processing in conditions of non-experimental, real-world human communication. The aforementioned non-experimental ECoG studies have demonstrated with the example of speech onset-related data that spatially and temporally meaningful neuronal effects can be observed during real-world speech production. Here, we would like to provide an additional illustration of the plausibility of the non-experimental approach by visualizing neuronal effects at distinctive temporal stages of real-world speech production, including the starts and ends of speech and clause production epochs (Figure 1).

Figure 1 shows typical neuronal effects in the non-experimentally-obtained data from our multimodal corpus (previously unpublished material). This figure contains examples of most prominent ECoG responses which were observed in one native speaker of German (male, 41 years old at the moment of implantation, left-lateralized language areas) at distinctive temporal stages of real-world production of simple clauses. These effects took place in the premotor cortex, on the central sulcus and on the superior temporal gyrus (classic areas implicated in expressive language [16,52]). The three electrodes visualized in Figure 1(C) showed maximum relative spectral magnitude changes in their respective anatomical region (Figure 1(A)), and they were significant at at least one of the six investigated time points (further also referred to as 'conditions'; Figure 1(B)) in the course of speech production. Below follows a description of the procedures we undertook for data analysis including the statistics.

The simple clauses were extracted from 11 hours of continuously transcribed expressive language material from the subject. Different amounts of speech production were present in these hours, depending on the length of stay of the patient's conversation partners (private visitors or medical personnel) and on the wakefulness of the patient. The patient was fully conscious and alert in the analyzed epochs of conversation, and all of these conversation periods lay at least 30 minutes away from epileptic seizures to minimize the risk of contamination of physiologically relevant responses by inter-ictal activity.

The analyzed clauses consisted of the main verb (i.e. full verb, copula verb, or a transitive modal verb in the absence of a full verb [53]) plus all its syntactically governed lexical elements. Time periods between speech start and speech end (vertical dashed lines with markers 'ss' and 'se' above them in Figure 1, respectively) represent epochs of continuous speech production with no pauses exceeding 200 ms, in which clauses corresponding to time periods from clause start (i.e. the start of the first word in a clause, marker 'cls' in Figure 1) to clause end (i.e. the end of the last word in a clause, marker 'cle' in Figure 1) were embedded. The 200-ms threshold for definition of speech production epochs was chosen based on evidence from conversation linguistics indicating that pauses shorter that 200 ms are not perceived by interlocutors as 'pauses' in a conversation [54]. Within each clause, two additional temporal stages of speech production were identified according to linguistic criteria: 'mp1' corresponds to the start of the word in the left sentence bracket: a finite verb in main clauses and a subordinate conjunction in conjunction-introduced subordinate clauses [51]; 'mp2' is the start of the word in the right sentence bracket (usually a non-finite verb in a main clause and a non-finite verb in a subordinate clause) or the right-most word in the middle field of a main clause (often an adverb or a noun), depending on whether or not the right sentence bracket was present. The linguistic positions 'mp1' and 'mp2' within the clause were defined based on the topological sentence model in German [51,53]. The resulting temporal precedence of 'ss' → 'cls' → 'mp1' → 'mp2' → 'cle' → 'se' in the trial-averaged data in Figure 1(C) therefore shows the progression of the neuronal response over the course of non-experimental, real-world speech production.

As in our previous studies [18,20,21], spectral magnitude changes were calculated based on common average reference (CAR)-filtered data using a fast Fourier transformation. Grid electrodes lying within the seizure onset area, identified by epileptologists in ongoing ECoG recordings (C2, D6, E3 in Figure 1(A)), were excluded from all analyses to reduce the impact of strong epileptic spiking activity at these electrodes on our observations. Additional details on data pre-processing are available in [20]. Here, we used a sliding window of 200 ms with a time step of 20 ms and five Slepian tapers. Since the sampling frequency of recordings was 1024 Hz, this analysis resulted in a 5-Hz frequency resolution. The absolute spectral magnitudes for each trial and electrode in the entire visualized time period of −1 o 1.5 s relative to the onset of the respective condition ('ss', 'cls', 'mp1', 'mp2', 'cle', 'se') and in each time-frequency bin were divided by the same baseline activity. This baseline was calculated by median-averaging the absolute spectral magnitude in each frequency bin over the time bins corresponding to the time period −2 to −1.5 s relative to the onset of each trial of the respective condition, median-averaging it over all trials from this condition, and by mean-averaging the obtained value over the six conditions described above. The baseline-corrected spectral magnitudes in individual trials were then averaged over trials in the respective condition.

Similarly to [20], we evaluated the statistical significance of neuronal responses in the high gamma (70–150 Hz) frequency band in two time windows (−1 to 0 and 0 to 1 s relative to the onset of the respective condition) using a sign-test (FDR-corrected over the entire number of CAR-rereferenced grid electrodes (61/64) and the two analyzed time windows at a significance threshold of q = .001). The examples of most prominent increases in high gamma activity shown in Figure 1(C) were significant in both time windows for the conditions speech start ('ss') and clause start ('cls') in the speech motor region (electrodes C3, D3), and the increase in high gamma activity observed in the superior temporal cortex (electrode H8) was significant in the late time window for the conditions 'sentence end' ('cle') and 'speech production end' ('se'). These effects occurred in speech-relevant areas identified by clinicians in the course of pre-neurosurgical diagnostics (Figure 1(A)).

With only about a dozen recent exceptions [18,20,21,23,45–49], neurolinguistic and direct speech BCI studies to date are based on experiments. Investigation of non-experimental, real-word language is a novel, largely unexplored approach. For this reason, we considered that its plausibility merits a detailed illustration (Figure 1(C)). The effects visualized in Figure 1 align in their location with results of previous experimental [23,28] and non-experimental [20,21] findings, and they enable illustration of temporal progression of neuronal activity at several temporal stages of speech production. These temporally meaningful and spatially reproducible neuronal responses lend support to our notion that spontaneous, non-experimental speech is worth investigating.

4. The temporal dynamics of neuronal activity as a challenge for research on connected speech

The classic dual-stream model by Hickok and Poeppel [55] can be helpful in interpreting the temporal dynamics of neuronal responses underlying non-experimental, real-world speech production (Figure 1). For instance, one may assume that activity in the temporal cortex enables mapping of the acoustic properties of speech onto conceptual and semantic representations, such as for the purpose of self-monitoring via state feedback control [56]. Activation in the motor cortex, in contrast, may reflect forward predictions of the intended articulatory output in the dorsal language stream, responsible for sound-to-articulation mapping.

Many open questions, however, remain as to how these dynamics are informative about the temporal precedence of linguistic functions deployed in the stream of speech. For instance, when do semantic, phonological, or syntactic processes occur? Do they take place at separate or (partly) overlapping temporal stages, e.g. in the production of words at the positions 'mp1' and 'mp2' in Figure 1(C)? Different ways of interpreting such dynamics are conceivable. For instance, the model by Levelt et al. [57] suggests that word production evolves in a sequence of temporally distinctive stages of linguistic processing. First, the conceptual and semantic information is extracted. After this, the lemma, i.e. the basic word form, is retrieved from the mental lexicon. Next, it is integrated into the morphosyntactic and phonological context, and then the resulting word is articulated. Some electrophysiological studies agree with this notion of sequential processing [58,59]. Other linguistic models, however, suggest that a parallel architecture of processing phonological, conceptual, and syntactic information is conceivable [60]. More research may be needed to evaluate the physiological plausibility of these different possibilities.

To understand the dynamics of neuronal responses underlying continuous speech, taking neurolinguistic studies beyond the level of single words and towards larger speech units of real-life-like complexity will be essential. Also on the level of multi-word sequences, more research is needed to understand the exact functional significance of brain activity at different time points of speech production. For instance, whenever we produce an utterance, how much of it do we pre-determine prior to articulation: do we produce a syntactic plan of the entire sentence before we articulate it [61] or do we plan the structure of the sentence during its production in an incremental manner online [62], such as by putting together salient combinations of words ('chunks') stored in the mental lexicon as single entities as a result of the speaker's experience with language [63]? A related and hitherto unexplored question is, what are the exact basic 'building blocks' of spontaneous, natural speech and how can their borders be detected in neuronal activity (e.g. Chafe [64], Frazier and Fodor [65])? The validity of these different linguistic accounts of how we produce continuous language sequences remains to be evaluated neurolinguistically. To this end, implementation of approaches similar to those used in experimental ECoG research on the spatiotemporal dynamics of cortical activity during speech production [66] and neurocomputational simulations of the speaking brain [67] can be helpful.

We believe that neurolinguistic research in conditions of spontaneous, non-experimental communication can shed light on such questions, and that an interdisciplinary approach bringing together neuroscientists and linguists can play an important role in understanding the exact functional dynamics of brain activity to support natural language. For instance, usage-based linguistics, which offers a broad fund of theoretical and empirical knowledge on spontaneous speech and provides detailed descriptions of its structural, temporal, and distributional properties [62,63,68], may be helpful. Linguistically informed accounts of how information is encoded in the dynamics of brain activity can not only contribute to a better understanding of the linguistic functioning of the human brain. They can also aid derivation of adequate decoding models for speech reconstruction with direct BCIs.

5. Developments in the area of direct speech BCIs from a linguistic perspective

Beyond their utility in clinically oriented and basic research, invasive recordings of cortical activity have proven valuable in direct BCI approaches to speech reconstruction. Here, we overview the advances of research on speech production with the help of direct ECoG-based BCIs from a linguistic perspective (Table 1). Please note that the present paper focuses on speech production. Studies conducted with perception-based paradigms are beyond the scope of the present paper, and information about the progress in this related field of research is available elsewhere [39,69,70].

Decoding studies have shown that speech production can reliably be distinguished from non-speech behavior based on ECoG data from the fronto-temporo-parietal region [4,22,71]. Differences between phonetic features of speech can also be decoded. Robust categorization of phonemes, for instance, has been achieved based on neuronal spiking activity obtained with neurotrophic electrodes in the speech motor cortex of locked-in individuals [24,25,72] as well as using spectral magnitude or power changes in the fronto-temporo-parietal cortex of epilepsy patients [3,4,73–78]. Recent ECoG studies report successful classification of phoneme categories and of articulatory features of speech on a subphonemic level. Mugler et al. [77] compared the classification performance for articulatory gestures (i.e. movements of the different articulators) vs. phonemes. These authors could achieve higher decoding accuracies when decoding articulatory gestures from postcentral areas, whereas frontal contacts proved more informative when decoding phonemes. Lotte et al. [4] decoded classes of phonemes based on the manner of articulation, the voicing, and the assignment to the category of vowels vs. consonants. They observed that ECoG sites which were most informative of these differences lay in spatially segregated areas of the fronto-temporo-parietal cortex. Several direct BCI studies on speech production have decoded auditory features, such as vowel formants [74] and spectrograms of the acoustic signal [79]. These studies speak for the suitability of sensorimotor [74] as well as fronto-opercular and temporal regions [79] to reconstruct overt expressive speech from the auditory signal.

Other studies with similar implantations have performed classification of individual overtly [80] and also covertly produced words [81], overtly produced sentences [82], of the semantic category the spoken word [83] or sentence [20] belongs to, and the identity of the conversation partner [18]. These studies could achieve successful classification in most of these linguistic scenarios. Martin et al. [84] also asked whether a decoding model trained on ECoG recorded during overt speech production would elicit above-chance classification when applied to data obtained during a covert speech production condition. This was indeed the case, with the best decoding performance yielded in the superior temporal and in the pericentral cortex. All in all, this research speaks for the plausibility of deciphering expressive language from underlying neuronal activity and supports the translatability of findings from overt to silent speech production.

Direct speech BCI is a young branch of research (as far as we are concerned, the first study was conducted using ECoG by Blakely et al. in 2008 [73]; also see Chaudhary et al. [12] for an overview of the history of BCI technologies). In spite of recent achievements, more work is needed to take such speech-restoration devices into the daily life of paralyzed patients. One challenge in direct speech BCI research is that decoding approaches operate on a limited, pre-selected set of linguistic categories, whereas human language is complex and rich in combinatorics. A currently popular way of dealing with linguistic variability is to reduce the number of decoding categories to a basic set of features, such as by using the phonemic inventory or the entire set of articulatory gestures speech is composed of [3,4,77]. An advantage of these approaches is that they can be applied to different items of speech regardless of their length and combinatorial properties. The latter is especially the case with articulatory gesture-based approaches, as they allow on e to account for co-articulation effects (i.e. differences in articulation between instances of the same phoneme depending on the immediate phonological context) [85]. Decoding approaches using the phonetic level

Table 1. An overview of direct ECoG-BCI studies on speech production from linguistic and machine-learning perspectives. N.s., total number of analyzed subjects; impl., implantation site; f, frontal; t, temporal; p, parietal; conv., conventional; frq., frequency; inv., investigated; C, consonant; V, vowel; i.a., inter altera; SVM, support vector machine; PCA, principal-component analysis. Frequency information in the penultimate column is spectral magnitude- or power-based.

authors	year	N. s.	impl.	ECoG type	study paradigm	classification / regression targets	frq. range(s) of Inv. neuronal features	decoding approach / classifier
Blakely et al.	2008	1	f-t	Micro-ECoG	Arbitrary overt production of phonemes	One phoneme vs. another (two phoneme pairs, binary classification)	Six frequency ranges (7–12, 10–13,14–25, 26–35, 36–70, 70–150 Hz)	Linear SVM
Kellis et al.	2010	1	f-t	Micro-ECoG	Overt visually cued production of words	Ten words	2–Hz bands (0.3–500 Hz)	Custom PCA- and clustering-based classifier
Pei et al.	2011	8	f-t-p	Conv. ECoG	Visually cued overt or covert repetition of words corresponding to CVC syllables	Four vowels, nine consonant pairs	Three frequency ranges (8–12 Hz, 18–26 Hz, 70–170 Hz)	Naive Bayes classifier
Leuthardt et al.	2011	4	f-t-p	Conv. ECoG	Visually cued overt and covert production of phonemes	Overt vs. covert phoneme, imagined phoneme vs. rest	2–Hz bands (0–550 Hz)	Custom linear decoder
Wang et al.	2011	4	f-t-p	Conv. ECoG	Visually cued overt speech production (picture naming, property identification, naming the semantically closest object)	Different semantic categories of objects	High gamma (60–120 Hz)	Naive Bayes classifier, linear SVM
Zhang et al.	2012	1	f-t	Conv. ECoG	Visually cued repetition of 8-character sentences (famous proverbs) upon aural presentation	Two sentences	High gamma (60–90 Hz)	Linear SVM
Derix et al.	2012	3	f-t-p	Conv. ECoG	Non-experimental, real-world conversation epochs with different communication partners (treating physician vs. life partner) leveled in the duration of speech perception and production	The identity of the conversation partner (doctor vs. partner)	Theta (3–5 Hz), alpha (8–12 Hz) and their combination, high gamma (70–150 Hz)	Regularized LDA
Kanas et al.	2014	1	f-t-p	Conv. ECoG	Visually and aurally cued repetition of syllables	Speech vs. non –speech	1–Hz bands (0–256 Hz)	Radial-basis-function SVM (i.a.)
Mugler et al.	2014	2	f-p, f	Conv. and micro-ECoG	Visually cued overt production of words	Thirteen articulatory gestures and 39 phonemes	High to very high gamma (70–300 Hz)	LDA classifier
Ikeda et al.	2014	4	f-t-p	Conv. ECoG	Covert articulation of three visually presented vowels	Three vowels	High gamma (70–110 Hz)	Linear SVM
Martin et al.	2014	7	f-t-p	Conv. ECoG	Overt and covert production of excerpts from historical political speeches or a children's story upon visual presentation	Reconstruction of auditory signals	High gamma (70–150 Hz)	Linear regression-based decoder for spectrogram features
Derix et al.	2014	3	f-t-p	Conv. ECoG	Non-experimental, real-world production of simple clauses with differences in mnemonic content	Sentence semantics (sentences with vs. without memory content)	Theta (3–5 Hz), alpha (8–12 Hz) and their combination, high gamma (70–150 Hz)	Regularized LDA
Bouchard et al.	2014	2	f-t-p	Micro-ECoG	Overt reading of CV syllables	Vowel formants	High gamma (70–150 Hz)	Principal-component linear regression with a custom two-staged optimization procedure
Lotte et al.	2015	7	f-t-p	Conv. ECoG	Overt reading of well-known political speeches	Place of articulation, manner of articulation, voicing status and phonological category of consonant or vowel, active speaking vs. silence	High gamma (70–170 Hz)	LDA classifier
Herff et al.	2015	7	f-t-p	Conv. ECoG	Overt production of excerpts from political speeches, fan fiction and children rhymes upon visual presentation	Words (from 23 phonemes decoded by ECoG)	High gamma (70–170 Hz)	Custom decoder with per-phoneme Gaussian feature distributions and an in-built bigram-based probabilistic language model
Wang et al.	2016	6	f-t-p	Conv. ECoG	Non-experimental, real-world behavior	Unsupervised, movement/speech/rest clusters were discovered	Multiple bands (1–53 Hz)	Hierarchical *k*-means clustering
Martin et al.	2016	5	f-t-p	Conv. ECoG	Overt and covert repetition of aurally presented words	All pairs from six words (binary classification)	High gamma (70–150 Hz)	Dynamic time-warping-based multiple-kernel SVM
Herff et al.	2016	1	f-t	Conv. ECoG	Repetition of a visually and aurally presented sentence	Reconstruction of auditory signals	High gamma (70–170)	Linear regression-based decoder for spectrogram features

of analysis in these recent studies proved to elicit relatively high accuracies of decoding. Nevertheless, exploration of other levels of linguistic description may help to further enhance the performance of direct speech BCIs.

Recent work shows that neighborhood probability estimates can be useful to improve speech reconstruction by evaluating the decoded speech unit with regard to its statistical probability in the linguistic context (e.g. a personal pronoun ('he') in English is more likely to precede a finite verb ('reads') than vice versa). It has been shown that predictive methods from linguistics, which rely on probabilities of co-occurrence of language units, can enhance the speed of spelling in P300 interfaces in healthy [86] and in paralyzed [87] subjects. A recent development in ECoG-BCI studies is that probabilistic *n*-gram-based models have been applied in production [3] and perception [88] studies to constrain the number of meaningful choices for speech decoding. Together with a recent review on the integration of language models into BCI classifiers [89], these publications support the notion that the use of predictive models of language, which can be derived from linguistic research, is a promising strategy for future direct speech BCI studies.

Language is a highly complex system, which possesses multiple levels of abstraction ranging from articulatory primitives to complex and abstract syntactic and semantic structures. Multi-level approaches with in-built models of language can be useful to improve the accuracy of speech reconstruction with direct speech BCIs. Information on several levels of linguistic abstraction may be helpful, e.g. when decoding results on these different levels are mutually incompatible. For instance, if the first estimate of the phoneme-to-phoneme approach is \' sō\ ('so'), the second is \'sȯ\ ('saw'), and the best estimate for the part-of-speech category decoding is 'noun' followed by the category 'verb', the mutually compatible solution ('saw', 'noun') would be selected. Such linguistically multi-level approaches can be expected to harness more information which can be used for decoding and to restrict the number of meaningful choices given the linguistic context [89].

6. Relevance of corpus-based linguistic methodology for direct speech BCI research

As proposed above, the performance of direct BCIs may benefit from accounting for multiple levels of linguistic abstraction. Nevertheless, a challenge when implementing such an integrative approach is that neurolinguistic evidence of neuronal distinctions between linguistic entities on many of these levels still needs to be generated. Consider, for instance, the linguistic category 'word.' In a BCI study on speech decoding by Kellis et al. [80], classification of 10 individual words from the neuronal signal was evaluated. Compared with this number, The International Corpus of English (ICE) [90] contains one million words per variety of the English language. Wouldn't it be interesting to extend the approach by Kellis et al. [80] to any word in the whole register of ICE to ensure better coverage of natural speech by BCI applications? This is, however, hardly feasible (or at least methodologically difficult without breaking down the words into a limited number of primitives) for the following reason. For a decoding algorithm to recognize differences in brain activity, training data with multiple examples of neuronal recordings underlying each word would be required. Thus, one would have to ask the subject to repeat the entire ICE corpus multiple times to obtain such data, which would be extremely time-consuming, burdensome to the subject, and most likely impossible in ECoG studies which rely on data recorded over relatively short time periods of about one to three weeks prior to surgical resection.

We anticipate that such ample data can be generated by adopting methodology from linguistics, similar to the approach we have illustrated in Figure 1. A central assumption in corpus-based linguistics is that the statistical properties of a language generalize across users of the respective language, and that these statistics are similarly reflected in the (para-)linguistic behavior of different speakers [63]. For this reason, for example, word frequency extracted from a retrospective corpus can be associated with comparable response latencies between speakers of the same language [91]. Neurolinguistic evidence on the functional organization of language areas is also based on the principle of generalizability: in spite of inter-individual variability of language-relevant anatomical [92] and functional areas [52], reproducible spatial [16] and temporal [57] patterns of functional organization can be observed. Following the same principle, one could expect that the neuronal signal components which are most informative of linguistic distinctions will, to some extent, generalize between speakers. It is an interesting open question for speech BCI research, whether and under what conditions a decoder trained on data from other subject(s) can yield successful decoding from neuronal recordings of a different individual [93].

An important advantage of generating and studying neurolinguistic corpora may be that they would allow for classification of linguistic phenomena which have a large number of realizations in natural language, as is the case with the above-mentioned linguistic category 'word'. Such neurolinguistic evidence can be generated similarly to the way linguistic corpora are generated, that is, by recording continuous speech of multiple speakers over extensive time periods, segmenting it in linguistic units (e.g. words) and annotating those units with regard to various linguistic categories (similarly to what we have done in the example in Figure 1). Many linguistic corpora are currently available online, which greatly promotes quantitative linguistic research. To our knowledge, however, no published neurolinguistic corpus bringing together spontaneously spoken speech and concurrent neuronal recordings is currently available, and the joint effort of linguists and neuroscientists is needed to create such data and make them available to interested researchers.

7. Developments in the area of direct speech BCIs from a machine-learning perspective

Direct BCI studies on speech production have taken different machine-learning approaches, which are summarized in Table 1. These can be classified by their usage of high-bias vs. high-variance classifiers and usage of standard machine-learning algorithms vs. non-standard, customized algorithms tailored to the specific decoding problem. As in the case of recoding techniques, each of these approaches has its own advantages and disadvantages [94].

Most studies summarized in Table 1 have used high-bias classifiers, i.e. classifiers that make strong assumptions about the mathematical relationships between input features and decoding targets. For instance, linear decoders assume a linear relationship. Examples of high-bias classifiers include linear support vector machines (SVM [73,75,82]), linear discriminant analysis (LDA [4,18,20,77]), linear regression-based [79,84] or custom linear decoders [76], principal-component linear regression [74], and a naive Bayes classifier [78]. Such classifiers have the advantages of being fairly robust against overfitting. Their strong assumptions can prevent them from learning 'false' relationships, which only exist due to noise in the particular data they are trained on. Also, they are usually fast to train and fast to apply compared with more complex high-variance classifiers. The demerit of this robustness is inflexibility: if the imposed assumptions are far from the true mathematical relationships between input features and decoding targets, the classifier will yield suboptimal decoding performance. High-variance classifiers impose less strong assumptions on the mathematical relationships in the data, and thus they can learn a larger range of relationships, while being more vulnerable to overfitting. High-variance approaches used in previous studies (Table 1) include radial-basis-function SVM [71], dynamic time-warping-based multiple-kernel SVM [81], hierarchical *k*-means clustering [22], and custom non-linear clustering-based classifiers [3,80]. Some authors have also evaluated multiple classifiers, thereby showing which classifiers are better suited for their decoding problem (e.g. [71]). All high-bias and high-variance classifiers summarized in this paragraph are associated with the demerit of manual feature extraction based on a priori assumptions as to what signal features are most informative for classification.

The decoders used in previous direct speech BCI research can be divided into standard algorithms using standard classifiers (linear regression [79,84], linear [73,75,82] and radial-basis-function [71] SVMs, LDA [4,18,20,77], a naive Bayes classifier [78], hierarchical *k*-means clustering [22]) and customized, non-standard methods [3,74,76,80]. As can be seen from Table 1, about two-thirds of previous direct speech-ECoG-BCI studies have opted for standard methods. These methods have the advantage that the decoding algorithms are well known and thus more transparent for the reader to understand, and that differences in decoding performance elicited with the same standard classifier can be more easily attributed to the other steps of the decoding pipeline such as data pre-processing and feature extraction. Custom methods make such comparisons difficult. However, they can achieve a substantially better performance than standard decoders by virtue of being adapted to the requirements of the specific decoding problem. An example is the usage of an in-built probabilistic language model to alleviate the multi-class decoding problem in the study by Herff et al. [3].

Another fundamental difference is between supervised and unsupervised decoding approaches. Supervised decoding approaches require known decoding targets such as the words produced by the subject during a recording. Unsupervised approaches can be used to discover meaningful structure in the data even if one does not have decoding targets. One example is the study by Wang et al. [22], in which the authors recorded the subjects' brain signals during spontaneous activities. Through unsupervised clustering methods, they could detect clusters of similar brain activity for different behaviors including movement, speaking, and resting, even though they did not use a priori labels indicating which behaviors took place. Unsupervised approaches therefore open the door for discovering meaningful insights from ample unlabeled recordings of brain signals.

Another class of decoding algorithm relevant to direct speech BCI research is deep learning, which is currently gaining attention in BCI research [95–97]. Deep learning has most prominently been used for recognition of graphical patterns in computer vision, and it has rapidly penetrated multiple other areas due to its general applicability. Schmidhuber [98] provides a detailed historical review of the evolution of deep learning; also see LeCun et al. [99] for an introductory review. Deep learning approaches, such as convolutional neural networks or recurrent neural networks, are of interest to the BCI community. They can handle complex, hierarchical non-linear relationships in the data and use the entire signal without a priori limitations on the amount of potentially informative features. Disadvantages of deep learning are that models take longer to train and that they are harder to interpret compared with those yielded with other methods. A crucial advantage, however, is that deep learning approaches can handle large numbers of classes [100,101]. This makes them particularly attractive to direct speech BCI research, which has to face the complexity and variability of natural human language. To our knowledge, no published direct speech BCI study so far has used these methods, and they bear hitherto unexplored opportunities for research to come.

8. Future prospects of invasive BCI for speech restoration

BCI technology is now developing towards autonomous wireless implants, and, although several technical challenges need to be surmounted [39], long-term recordings with such implants are conceivable in the near future. A companion paper in this special issue [102] provides a detailed discussion of the current trends and future perspectives of implant technology for human application. We believe that ECoG-based invasive BCI is a promising strategy to restore communication in paralyzed patients. It is nevertheless difficult to estimate when this will be possible in clinical practice. This will depend on the newest and conceivable future advances in implant, decoding, and neurolinguistic methodologies. Speech-related neuronal signals obtained continuously over months or perhaps even years during daily-life experiences of implanted neurological patients will most likely be available soon. Such 'big data' can provide valuable content for neurolinguistic corpora. Decoding approaches suitable for long-term recordings of brain activity, such as that implemented in [22], and also deep learning [95–97], together with automated or semi-automated approaches to account for linguistic information [3,88], may advance current understanding of the dynamic processes underlying speech production and open up new possibilities for research on direct speech BCIs.

Acknowledgements

We thank GRK 1624 'Frequency effects in language' (University of Freiburg) and BrainLinks-BrainTools (DfG grant EXC1086) for financial support.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

[1] Birbaumer N, Gallegos-Ayala G, Wildgruber M, et al. Direct brain control and communication in paralysis. Brain Topogr. 2014;27(1):4–11.
[2] Thinnes-Elker F, Iljina O, Apostolides JK, et al. Intention concepts and brain-machine interfacing. Front Psychol. 2012;3:1–10. Article id 455.
[3] Herff C, Heger D, de Pesters A, et al. Brain-to-text: decoding spoken phrases from phone representations in the brain. Front Neurosci. 2015;9:1–11. Article id 217.
[4] Lotte F, Brumberg JS, Brunner P, et al. Electrocorticographic representations of segmental features in continuous speech. Front Hum Neurosci. 2015;9:1–13. Article id 97.
[5] Logothetis NK. What we can do and what we cannot do with fMRI. Nature. 2008;453(7197):869–878.
[6] Ojemann GA, Ojemann J, Ramsey NF. Relation between functional magnetic resonance imaging (fMRI) and single neuron, local field potential (LFP) and electrocorticography (ECoG) activity in human cortex. Front Hum Neurosci. 2013;7:1–9. Article id 34.
[7] Conner CR, Ellmore TM, Pieters TA, et al. Variability of the relationship between electrophysiology and bold-fMRI across cortical regions in humans. J Neurosci. 2011;31(36):12855–12865.
[8] Hong K-S, Santosa H. Decoding four different sound-categories in the auditory cortex using functional near-infrared spectroscopy. Hear Res. 2016;333:157–166.
[9] Santosa H, Hong MJ, Hong K-S. Lateralization of music processing with noises in the auditory cortex: an fNIRS study. Front Behav Neurosci. 2014;8:1–9. Article id 418.
[10] Chaudhary U, Xia B, Silvoni S, et al. Brain-computer interface-based communication in the completely locked-in state. PLoS Biol. 2017;15(1):e1002593.
[11] Gallegos-Ayala G, Furdea A, Takano K, et al. Brain communication in a completely locked-in patient using bedside near-infrared spectroscopy. Neurology. 2014;82(21):1930–1932.
[12] Chaudhary U, Birbaumer N, Ramos-Murguialday A. Brain-computer interfaces for communication and rehabilitation. Nat Rev Neurol. 2016;12(9):513–525.
[13] Ball T, Kern M, Mutschler I, et al. Signal quality of simultaneously recorded invasive and non-invasive EEG. NeuroImage. 2009;46(3):708–716.
[14] McGuigan FJ. Covert oral behavior during the silent performance of language tasks. Psychol Bull. 1970;74(5):309.
[15] Ganushchak LY, Christoffels IK, Schiller NO. The use of electroencephalography in language production research: a review. Front Psychol. 2011;2:1–6. Article id 208.
[16] Price CJ. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. NeuroImage. 2012;62(2):816–847.
[17] del Millán J del R, Carmena JM. Invasive or noninvasive: understanding brain-machine interface technology. IEEE Eng Med Biol Mag. 2010;29(1):16–22.
[18] Derix J, Iljina O, Schulze-Bonhage A, et al. "Doctor" or "darling"? Decoding the communication partner from ECoG of the anterior temporal lobe during non-experimental, real-life social interaction. Front Hum Neurosci. 2012;6:1–14. Article id 251.

[19] Fiederer LDJ, Lahr J, Vorwerk J, et al. Electrical stimulation of the human cerebral cortex by extracranial muscle activity: effect quantification with intracranial EEG and FEM simulations. IEEE Trans Biomed Eng. 2016;63(12):2552–2563.
[20] Derix J, Iljina O, Weiske J, et al. From speech to thought: the neuronal basis of cognitive units in non-experimental, real-life communication investigated using ECoG. Front Hum Neurosci. 2014;8:383.
[21] Ruescher J, Iljina O, Altenmüller D-M, et al. Somatotopic mapping of natural upper- and lower-extremity movements and speech production with high gamma electrocorticography. NeuroImage. 2013;81:164–177.
[22] Wang NXR, Olson JD, Ojemann JG, et al. Unsupervised decoding of long-term, naturalistic human neural recordings with automated video and audio annotations. Front Hum Neurosci. 2016;10:165. [Epub ahead of print]. doi: 10.1016/j.bandl.2016.06.003
[23] Towle VL, Yoon H-A, Castelle M, et al. ECoG gamma activity during a language task: differentiating expressive and receptive speech areas. Brain. 2008;131(8):2013–2027.
[24] Brumberg JS, Wright EJ, Andreasen DS, et al. Classification of intended phoneme production from chronic intracortical microelectrode recordings in speech-motor cortex. Front Neurosci. 2011;5:1-12. Article id 65.
[25] Guenther FH, Brumberg JS, Wright EJ, et al. A wireless brain-machine interface for real-time speech synthesis. PLoS ONE. 2009;4(12):e8218.
[26] Ritaccio A, Matsumoto R, Morrell M, et al. Proceedings of the seventh international workshop on advances in electrocorticography. Epilepsy Behav. 2015;51: 312–320.
[27] Towle VL, Dai Z, Zheng W, et al. Mapping cortical function with event-related electrocorticography. In: RW Byrne, Editor. Functional mapping of the cerebral cortex. Springer International Publishing; 2016. p. 91–104. doi: 10.1007/978-3-319-23383-3_6
[28] Crone NE, Hao L, Hart J Jr, et al. Electrocorticographic gamma activity during word production in spoken and sign language. Neurology. 2001;57(11):2045–2053.
[29] Sinai A, Bowers CW, Crainiceanu CM, et al. Electrocorticographic high gamma activity versus electrical cortical stimulation mapping of naming. Brain. 2005;128(7):1556–1570.
[30] Fukuda M, Rothermel R, Juhász C, et al. Cortical gamma-oscillations modulated by listening and overt repetition of phonemes. NeuroImage. 2010;49(3):2735–2745.
[31] Edwards E, Nagarajan SS, Dalal SS, et al. Spatiotemporal imaging of cortical activation during verb generation and picture naming. NeuroImage. 2010 Mar;50(1):291–301.
[32] Kunii N, Kamada K, Ota T, et al. The dynamics of language-related high-gamma activity assessed on a spatially-normalized brain. Clin Neurophysiol. 2013;124(1):91–100.
[33] Leuthardt EC, Miller K, Anderson NR, et al. Electrocorticographic frequency alteration mapping: a clinical technique for mapping the motor cortex. Neurosurgery. 2007;60(4 Suppl 2):260–271.
[34] Taplin AM, de Pesters A, Brunner P, et al. Intraoperative mapping of expressive language cortex using passive real-time electrocorticography. Epilepsy Behav Case Rep. 2016;5:46–51.
[35] Bouchard KE, Mesgarani N, Johnson K, et al. Functional organization of human sensorimotor cortex for speech articulation. Nature. 2013 Mar 21;495(7441):327–332.
[36] Toyoda G, Brown EC, Matsuzaki N, et al. Electrocorticographic correlates of overt articulation of 44 English phonemes: intracranial recording in children with focal epilepsy. Clin Neurophysiol. 2013;125(6):1129–1137.
[37] Chen Y, Shimotake A, Matsumoto R, et al. The “when” and “where” of semantic coding in the anterior temporal lobe: temporal representational similarity analysis of electrocorticogram data. Cortex. 2016;79:1–13.
[38] Fedorenko E, Scott TL, Brunner P, et al. Neural correlate of the construction of sentence meaning. Proc Nat Acad Sci. 2016;113(41):E6256–E6262.
[39] Martin S, Millán JDR, Knight RT, et al. The use of intracranial recordings to decode human language: challenges and opportunities. Brain Lang. 2016. [Epub ahead of print]. doi: 10.1016/j.bandl.2016.06.003
[40] Przyrembel M, Smallwood J, Pauen M, et al. Illuminating the dark matter of social neuroscience: Considering the problem of social interaction from philosophical, psychological, and neuroscientific perspectives. Front Hum Neurosci. 2012;6:1-15. Article id 190.
[41] Brown EC, Rothermel R, Nishida M, et al. In vivo animation of auditory-language-induced gamma-oscillations in children with intractable focal epilepsy. NeuroImage. 2008;41(3):1120–1131.
[42] Kojima K, Brown EC, Rothermel R, et al. Multimodality language mapping in patients with left-hemispheric language dominance on Wada test. Clin Neurophysiol. 2012;123(10):1917–1917.
[43] Genetti M, Tyrand R, Grouiller F, et al. Comparison of high gamma electrocorticography and fMRI with electrocortical stimulation for localization of somatosensory and language cortex. Clin Neurophysiol. 2015;126(1):121–130.
[44] Nourski KV, Steinschneider M, Rhone AE. Electrocorticographic activation within human auditory cortex during dialog-based language and cognitive testing. Front Hum Neurosci. 2016;10:1–15. Article id 202.
[45] Wray CD, Blakely TM, Poliachik SL, et al. Multimodality localization of the sensorimotor cortex in pediatric patients undergoing epilepsy surgery. J Neurosurg Pediatr. 2012;10(1):1–6.
[46] Cho-Hisamoto Y, Kojima K, Brown EC, et al. Cooing- and babbling-related gamma-oscillations during infancy: Intracranial recording. Epilepsy Behav. 2012;23(4):494–496.
[47] Vansteensel MJ, Bleichner MG, Dintzner LT, et al. Task-free electrocorticography frequency mapping of the motor cortex. Clin Neurophysiol. 2013;124:1169–1174.
[48] Arya R, Wilson JA, Vannest J, et al. Electrocorticographic language mapping in children by high-gamma synchronization during spontaneous conversation: comparison with conventional electrical cortical stimulation. Epilepsy Res. 2014;110:78–87.

[49] VanGilder JL. Electrocorticographic analysis of spontaneous conversation to localize receptive and expressive language areas [master thesis]. Tempe (AZ): Arizona State University; 2013.
[50] Selting M, Auer P, Barth-Weingarten D, et al. Gesprächsanalytisches Transkriptionssystem 2 (GAT 2) [Conversational transcription system 2 (CTS 2)]. Gesprächsforschung: Online-Zeitschrift zur verbalen Interaktion; 2009.
[51] Musan R. Satzgliedanalyse. Heidelberg: Winter; 2009.
[52] Penfield W, Boldrey E. Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. Brain. 1937;60(4):389–443.
[53] Eisenberg P. Grundriss der deutschen Grammatik: Band 2: Der Satz. (überarb. und aktualisierte Aufl.) [Foundations of the German grammar: volume 2: the sentence (modified and revised edition)]. Stuttgart: Metzler; 2004.
[54] Walker MB, Trimboli C. Smooth transitions in conversational interactions. J Physiol. 1982;117(2):305–306.
[55] Hickok G, Poeppel D. The cortical organization of speech processing. Nat Rev Neurosci. 2007;8(5):393–402.
[56] Hickok G. Computational neuroanatomy of speech production. Nat Rev Neurosci. 2012;13(2):135–145.
[57] Levelt WJ, Roelofs A, Meyer AS. A theory of lexical access in speech production. Behav Brain Sci. 1999;22(1):1–38.
[58] Levelt WJ, Schriefers H, Vorberg D, et al. The time course of lexical access in speech production: a study of picture naming. Psychol Rev. 1991;98(1):1–8. Article id 122.
[59] Sahin NT, Pinker S, Cash SS, et al. Sequential processing of lexical, grammatical, and phonological information within Broca's area. Science. 2009;326(5951):445–449.
[60] Jackendoff R. A parallel architecture perspective on language processing. Brain Res. 2007;1146:2–22.
[61] Garrett MF. Production of speech: observations from normal and pathological language use. In: AW Ellis, editor. Normality and pathology in cognitive functions. London: Academic; 1982. p. 19–76.
[62] Auer P. On-line syntax: thoughts on the temporality of spoken language. Lang Sci. 2009;31:1–13.
[63] Bybee J. Language, usage and cognition. Cambridge: Cambridge University Press; 2010.
[64] Chafe W. Discourse, consciousness, and time: the flow and displacement of conscious experience in speaking and writing. Chicago, IL: University of Chicago Press; 1994.
[65] Frazier L, Fodor JD. The sausage machine: a new two-stage parsing model. Cognition. 1978;6(4):291–325.
[66] Brumberg JS, Krusienski DJ, Chakrabarti S, et al. Spatio-Temporal progression of cortical activity related to continuous overt and covert speech production in a reading task. PLoS ONE. 2016;11(11):e0166872.
[67] Schomers MR, Garagnani M, Pulvermüller F. Neurocomputational consequences of evolutionary connectivity changes in perisylvian language cortex. J Neurosci. 2017;37(11):3045–3055.
[68] Pfänder S, Behrens H, Auer P. Erfahrung zählt. Frequenzeffekte in der Sprache: ein Werkstattbericht. Zeitschrift für Literaturwissenschaft und Linguistik [Experience matters. Frequency effects in language: a lab report]. 2013;43(1):7–32.
[69] Chakrabarti S, Sandberg HM, Brmberg JS, et al. Progress in speech decoding from the electrocorticogram. Biomed Eng Lett. 2015;5(1):10–21.
[70] Herff C, Schultz T. Automatic speech recognition from neural signals: a focused review. Front Neurosci. 2016;10:1–7. Article id 429.
[71] Kanas VG, Mporas I, Benz HL, et al. Joint spatial-spectral feature space clustering for speech activity detection from ECoG signals. IEEE Trans Biomed Eng. 2014;61(4):1241–1250.
[72] Kennedy PR, Bakay RA. Restoration of neural output from a paralyzed patient by a direct brain connection. NeuroReport. 1998;9(8):1707–1711.
[73] Blakely T, Miller KJ, Rao RPN, et al. Localization and classification of phonemes using high spatial resolution electrocorticography (ECoG) grids. Conf Proc IEEE Eng Med Biol Soc. 2008;2008:4964–4967.
[74] Bouchard KE, Chang EF. Neural decoding of spoken vowels from human sensory-motor cortex with high-density electrocorticography. Conf Proc IEEE Eng Med Biol Soc. 2014;2014:6782–6785.
[75] Ikeda S, Shibata T, Nakano N, et al. Neural decoding of single vowels during covert articulation using electrocorticography. Front Hum Neurosci. 2014;8:1–8.
[76] Leuthardt EC, Gaona C, Sharma M, et al. Using the electrocorticographic speech network to control a brain-computer interface in humans. J Neural Eng. 2011;8(3):036004.
[77] Mugler EM, Patton JL, Flint RD, et al. Direct classification of all American English phonemes using signals from functional speech motor cortex. J Neural Eng. 2014;11(3):035015.
[78] Pei X, Barbour DL, Leuthardt EC, et al. Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans. J Neural Eng. 2011;8(4):046028.
[79] Herff C, Johnson G, Diener L, et al. Towards direct speech synthesis from ECoG: a pilot study. Conf Proc IEEE Eng Med Biol Soc. 2016;2016:1540–1543.
[80] Kellis S, Miller K, Thomson K, et al. Classification of spoken words using surface local field potentials. Conf Proc IEEE Eng Med Biol Soc. 2010;2010:3827–3830.
[81] Martin S, Brunner P, Iturrate I, et al. Word pair classification during imagined speech using direct brain recordings. Sci Rep. 2016;6:25803.
[82] Zhang D, Gong E, Wu W, et al. Spoken sentences decoding based on intracranial high gamma response using dynamic time warping. Conf Proc IEEE Eng Med Biol Soc. 2012;2012:3292–3295.
[83] Wang W, Degenhart AD, Sudre GP, et al. Decoding semantic information from human electrocorticographic (ECoG) signals. Conf Proc IEEE Eng Med Biol Soc. 2011;6294–6298.
[84] Martin S, Brunner P, Holdgraf C, et al. Decoding spectrotemporal features of overt and covert speech from the human cortex. Front Neuroeng. 2014;7:1–15. Article id 14.
[85] Liberman AM, Delattre PC, Cooper FS, et al. The role of consonant-vowel transitions in the perception of the stop and nasal consonants. Psych Mon. 1954;68:1–13.

[86] Ryan DB, Frye GE, Townsend G, et al. Predictive spelling with a P300-based brain-computer interface: increasing the rate of communication. Int J Hum Comput Interact. 2011;27(1):69–84.
[87] Speier W, Chandravadia N, Roberts D, et al. Online BCI typing using language model classifiers by ALS patients in their homes. Brain-Computer Interfaces. 2016;4:1–8.
[88] Moses DA, Mesgarani N, Leonard MK, et al. Neural speech recognition: continuous phoneme decoding using spatiotemporal representations of human cortical activity. J Neural Eng. 2016;13(5):056004.
[89] Speier W, Arnold C, Pouratian N. Integrating language models into classifiers for BCI communication: a review. J Neural Eng. 2016;13(3):031002.
[90] Greenbaum S, Nelson G. The International Corpus of English (ICE) project. World Englishes. 1996;15(1):3–15.
[91] Oldfield RC, Wingfield A. Response latencies in naming objects. Q J Exp Psychol. 1965;17(4):273–281.
[92] Amunts K, Schleicher A, Bürgel U, et al. Broca's region revisited: Cytoarchitecture and intersubject variability. J Comp Neurol. 1999;412(2):319–341.
[93] Jayaram V, Alamgir M, Altun Y, et al. Transfer learning in brain-computer interfaces. IEEE Comput Intell Mag. 2016;11(1):20–31.
[94] Lotte F, Congedo M, Lécuyer A, et al. A review of classification algorithms for EEG-based brain-computer interfaces. J Neural Eng. 2007;4(2):R1–R13.
[95] Hajinoroozi M, Mao Z, Jung TP, et al. EEG-based prediction of driver's cognitive performance by deep convolutional neural network. Signal Process Image. 2016;47:549–555.
[96] Manor R, Mishali L, Geva AB. Multimodal neural network for rapid serial visual presentation brain computer interface. Front Comput Neurosci. 2016;10: 1–13. Article id 130.
[97] Tabar YR, Halici U. A novel deep learning approach for classification of EEG motor imagery signals. J Neural Eng. 2017;14(1):016003.
[98] Schmidhuber J. Deep learning in neural networks: an overview. Neural Networks. 2015;61:85–117.
[99] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–444.
[100] Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542(7639):115–118.
[101] Morin F, Bengio Y. Hierarchical probabilistic neural network language model. Aistats. 2005;5:246–252.
[102] Kohler F, Gkogkidis CA, Bentler C, et al. Closed-loop interaction with the cerebral cortex: present and future of wireless implant technology. Brain Comput Interfaces. forthcoming.