Published in final edited form as:
Neuroimage. 2018 October 15; 180(Pt A): 301–311. doi:10.1016/j.neuroimage.2017.10.011.

Decoding spoken phonemes from sensorimotor cortex with High-Density ECoG grids

NF Ramsey[1], **E Salari**[1], **EJ Aarnoutse**[1], **MJ Vansteensel**[1], **MG Bleichner**[2], and **ZV Freudenburg**[1]

[1]Brain Center Rudolf Magnus, University Medical Center Utrecht, Dept Neurology and Neurosurgery, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands [2]Neuropsychology Lab, Department of Psychology, European Medical School, Cluster of Excellence Hearing4all, University of Oldenburg, Ammerländer Heerstrasse 114-118, 26129 Oldenburg, Germany

Abstract

For people who cannot communicate due to severe paralysis or involuntary movements, technology that decodes intended speech from the brain may offer an alternative means of communication. If decoding proves to be feasible, intracranial Brain-Computer Interface systems can be developed which are designed to translate decoded speech into computer generated speech or to instructions for controlling assistive devices. Recent advances suggest that such decoding may be feasible from sensorimotor cortex, but it is not clear how this challenge can be approached best. One approach is to identify and discriminate elements of spoken language, such as phonemes. We investigated feasibility of decoding four spoken phonemes from the sensorimotor face area, using electrocorticographic signals obtained with high-density electrode grids. Several decoding algorithms including spatiotemporal matched filters, spatial matched filters and support vector machines were compared. Phonemes could be classified correctly at a level of over 75% with spatiotemporal matched filters. Support Vector machine analysis reached a similar level, but spatial matched filters yielded significantly lower scores. The most informative electrodes were clustered along the central sulcus. Highest scores were achieved from time windows centered around voice onset time, but a 500 ms window before onset time could also be classified significantly. The results suggest that phoneme production involves a sequence of robust and reproducible activity patterns on the cortical surface. Importantly, decoding requires inclusion of temporal information to capture the rapid shifts of robust patterns associated with articulator muscle group contraction during production of a phoneme. The high classification scores are likely to be enabled by the use of high density grids, and by the use of discrete phonemes. Implications for use in Brain-Computer Interfaces are discussed.

Keywords

ECoG; language; phonemes; decoding; Brain-Computer Interface

Corresponding Author: Nick F Ramsey, Brain Center Rudolf Magnus, University Medical Center Utrecht, Dept Neurology and Neurosurgery, Heidelberglaan 100, Room G03 1.22, 3584 CX Utrecht, The Netherlands, T +31 (0)88 755 6863, F +31 (0)30 254 2100, n.f.ramsey@umcutrecht.nl.

1 Introduction

To function in life, it is critical to be able to communicate. Spoken and written language, as well as non-verbal expressions, allow people to interact socially. Expression of language in particular is crucial for communication of ones needs, ideas and opinions. People who are completely unable to express themselves are essentially excluded from society at every level (Bruno et al., 2011; Laureys et al., 2005; Rousseau et al., 2015). Although their numbers may be small, their predicament warrants research into ways to restore communication abilities (Chaudhary et al., 2016; Wolpaw et al., 2002). Disorders leading to severe communication disability include afflictions leading to total paralysis resulting from trauma, stroke and neurodegenerative diseases (Locked-In Syndrome)(Lulé et al., 2009), and loss of muscle coordination due to trauma or developmental disorders such as Cerebral Palsy. When some muscle control is preserved (however minimal), Assistive Technologies (AT) are available to maximally utilize intentional movements. When no control is preserved, there are no technologies available to meet the patients need for communication. In recent years attempts to achieve communication by means of a Brain-Computer Interface have increased, leading to promising avenues (Farwell and Donchin, 1988; Gallegos-Ayala et al., 2014; Kennedy and Bakay, 1998; McCane et al., 2015; Sellers et al., 2010, 2014) but not yet to standard treatment for communication loss. Recently, however, a first case was presented where a Locked-In, late-state ALS patient could successfully use a Brain-Computer Interface to communicate in daily life without requirement for presence of an expert (Vansteensel et al., 2016). The system was fully implanted, and allowed the patient to generate signals, obtained from electrodes directly on the motor cortex, to select items in spelling software. Non-invasive BCI solutions, using scalp EEG and the 'P300 speller', have also resulted in encouraging results (Farwell and Donchin, 1988; Kleih et al., 2011; McCane et al., 2015; Sellers et al., 2010), but these require considerable skill from caregivers to attach the scalp electrodes and initiate the system. The systems that currently work in select patients provide a coarse, but reliable, means to communicate, and do so by decoding specific events from the brain. They are, however, a far cry from restoring communication to a level where the user can interact with others in real-time. Nevertheless, a first step has been made on the road to restoring communication by extracting information from the cerebral cortex, encouraging further development

Application of decoding algorithms, if conducted appropriately, can also reveal the mechanism by which the human brain translates neuronal activity to perceptions and actions (Brunner et al., 2015; Sadtler et al., 2014). As such, the fact that many of the associated cortical regions exhibit a topographical representation encourages the notion that different percepts or actions are associated with different topographical distributions of activity. This has been investigated notably in primary cortices (V1, A1, S1 and, to a lesser degree, M1), and has yielded successful identification of stimulus features by means of classifying the stimulus-induced cortical activity patterns (Bleichner et al., 2016; Branco et al., 2016; Formisano et al., 2008; Kay et al., 2008; Polimeni et al., 2010). The fact that cortical activity patterns map onto specific stimulus features supports the notion that topography reflects an orderly distribution of specific functions along the cortex, with each function being associated with one or more specific neuronal ensembles (or cortical columns) (Hubel and

Wiesel, 1959; Markram, 2008; Mountcastle, 1997). Although such ensembles can be modulated in terms of response amplitude by selective attention and/or predictive mechanisms (Andersson et al., 2013; Brefczynski and DeYoe, 1999; Miall and Wolpert, 1996), and can be subject to an attention-driven shift in the exact mapping onto sensory space (Klein et al., 2014), the fact that activity patterns identify stimulus features reproducibly and robustly, indicates a certain degree of segregation of neuronal ensembles and the sensory space they code for.

Several approaches have been adopted in attempts to decode cortical activity to restore a means of communication. For EEG signals, detection of brain states has been utilized to select icons on a computer screen, by identifying a specific sensory input sequence emanating from that particular icon (visual or auditory pulse sequences which differ for each icon)(Fazel-Rezai et al., 2012). The recorded neural response to the sequence (which constitutes an amplified representation thereof) reveals which icon the person is attending to. Decoding is then tightly coupled to deliberate sensory input. Decoding *internally* generated actions is currently best feasible from sensorimotor cortex. With EEG the decline in amplitude of the mu rhythm (8-12 Hz, event-related desynchronization) that accompanies attempted or actual movement (McFarland et al., 2000; Pfurtscheller and Neuper, 1997), can be used also as a brain-state detector of an intentional act. Detection is here often translated to selection of an icon during a sequential icon scanning scheme (‘switch scanning’) or a unidirectional cursor movement. Neither EEG method is of much use for exploiting the fine topographical organization of the cortex. With intracranial EEG, or electrocorticography (ECoG), topographical patterns can be probed (Crone et al., 1998; Jacobs and Kahana, 2010; Miller et al., 2012). ECoG decoding approaches utilize the distribution of functionally coherent regions as is the case in the motor cortex (Bleichner et al., 2016; Bouchard and Chang, 2014; Miller et al., 2009; Schalk and Leuthardt, 2011) or visual cortex (Andersson et al., 2011). Language regions and networks may not provide adequate points of reference for decoding elements of speech since they do not exhibit a coherent topographical map (Kellis et al., 2010; Pei et al., 2011b), as seems to be the case for associative cortex in general (although some topography has been reported such as in (Harvey et al., 2013). Decoding (attempted) language production, however, is not constrained to language regions. The final stage of language production heavily depends on the sensorimotor cortex, which generates the motor commands for speaking and, for that matter, sign language (Bleichner et al., 2016, 2015; Crone et al., 2001). Given that both motor (Bleichner et al., 2016; Kellis et al., 2010; Siero et al., 2014) and somatosensory cortex (Branco et al., 2016; Sanchez-Panchuelo et al., 2012) exhibit quite detailed topographies, and that speaking involves rapid sequential patterns of muscle contractions in the face and vocal tract, the sensorimotor cortex should conceptually provide rich and coherent spatial and temporal information about what a person wants to say (Bouchard et al., 2013). Interestingly, and crucial for BCI research, research has shown that the sensorimotor activity patterns that are generated by complex hand gestures (representing letters of the American sign language alphabet for deaf people), are also generated by *attempts* to make these gestures in arm amputees (Lotze et al., 2001; Raffin et al., 2012; Roux et al., 2003). This finding suggests that actual and attempted motor acts may yield equally decodable cortical information, and that therefore research on cortical representations of speech is directly relevant for application in BCI technology for paralyzed people.

In this study, we tested the hypothesis that even the smallest elements of speech, phonemes, should provide decodable information from sensorimotor cortex for classification. This hypothesis relies on two assumptions. First, the cortical topographical representation of speech utterances such as phonemes maps onto the constellation of muscles or muscle groups that is required to produce the sound. Second, since speech involves rapid sequential schemes of muscle contractions even for phonemes, the contribution of time in the decoding algorithms should provide a significant contribution to phoneme classification (Bouchard et al., 2013; Jiang et al., 2016).

We report on a study on decoding of phoneme production from sensorimotor cortex in five patients implanted with high-density electrocorticography (ECoG) electrode grids. All patients had grids implanted for source localization of their seizures for subsequent surgical treatment of medically intractable epilepsy. In three patients, these grids were part of the clinical grid implantation plan, and in two patients the grid was placed as an addition to the clinical plan, for research purposes. All procedures were approved by the Medical Ethical Board of the hospital, and were in accordance with the Declaration of Helsinki of 2013. The ECoG grids over the sensorimotor face area had a high density of electrodes (3-4 mm center to center), allowing for detailed investigation of topographical representation of phoneme production. For decoding we focused on high-frequency broadband signal power (HFB, 65-125 Hz) (Crone et al., 1998) since this feature of the electrophysiological signal contains the most detailed and neuronal firing rate-related information (Bleichner et al., 2016; Miller et al., 2009; Siero et al., 2014). It is thought to most accurately reflect activity of neuronal ensembles, compared to other signal features (Manning et al., 2009; Miller et al., 2009; Ray and Maunsell, 2011). The density has been shown to produce independent signals between adjacent electrodes for the HFB and thus provide rich information about underlying cortical topography (Muller et al., 2016; Siero et al., 2014)

2 Methods

2.1 Subjects & Data Acquisition

ECoG signal was collected from five intractable epilepsy patients (Table 1) who had grids implanted subdurally over the inferior sensorimotor cortex on their right (subjects R1 and R2) or left (subjects L1, L2, and L3) hemisphere (depending on the probable location of the source of seizures). We refer to these grids as high density (HD) ECoG grids due to their high electrode density (3-4 mm center-to-center). Grids were obtained from Ad-tech Medical and PMT Corporation. Electrodes had an exposed diameter of 1 or 1.2 mm. For comparison: standard clinical grids have 1 cm center-to-center and a 2.4 mm electrode diameter. Written informed consent for participation in this study was given in accordance with the Declaration of Helsinki, 2013 and the study was approved by the Medical Ethical Committee of the Utrecht University Medical Center.

The placement of the grids was targeted at the sensorimotor face area. Exact coverage and electrode grid depended on patient-specific surgical considerations, and is shown in Figure

1. All HD grids covered at least part of the face area (see also **Figure 6**). After implantation, the locations of electrodes on the cortex were determined using cortical surface reconstructions from pre-implantation anatomical MRI and post-implantation CT scans according to the method described by (Branco et al., 2016; Hermes et al., 2010). Grid positions shown in Figure 1 are based on this method.

A 128-channel Micromed recording system (Treviso, Italy) was used for continuous recording of the HD-ECoG signal at a sampling frequency of 512 Hz (22 bits, band pass filter 0.15-134.4Hz) for subjects R1, R2, L1, and L2. A 256-channel Blackrock Neuroport system was used for recording in subject L3, at 2000 Hz (16 bits, high pass filter 0.5 Hz). Video recordings saved with the electrophysiological recordings were used to extract audio signals produced during the phoneme production tasks. All electrode signals were evaluated by trained epilepsy neurologists for signs of epileptic activity (interictal spikes), and were discarded if such activity was observed. Electrodes with poor signal quality (poor contact) were also discarded. This led to 10 electrodes from subject L2 being excluded from analysis. In addition, the corner electrodes of the grids for subjects R2 and L1 (designed to be upward skull facing for alternative rereferencing) and the top 7 rows of electrodes from subject L3's grid (located at or above the anatomical hand knob) were excluded (Figure 1 and Table 1).

2.2 Overt phoneme production Task

Patients, all of whom were native Dutch speakers, were visually cued using the Presentation® (Neurobehavioral Systems Inc) software package to pronounce the phonemes /p/, /k/, /u/, and /a:/ or remain silent and fixate on an asterisk on the screen (Figure 2). A microphone recording of their voice was extracted from the synchronously recorded clinical video system (or as an additional channel in the Blackrock system for subject L3) to evaluate task performance. Only trials in which the cued phoneme could be correctly acoustically identified by the experimenter (or rest trials in which no audible sound was produced) were included. A degree of variance in the exact acoustic features, and hence the corresponding mouth positioning and movement was tolerated as long as the utterance could be identified as the correct phoneme. In addition, voice onset times (VOTs) were marked for trial alignment using the annotation program Praat (version 5.2.29, www.praat.org). The clinical setting and the subjects' ability to perform the task led to variability in the number of task runs and trials per run. Subjects R1, R2, L1, L2 and L3 performed a total of 174, 70, 153, 119 and 114 correctly spoken phonemes, plus 63, 16, 43, 30 and 30 rest trials respectively. Frequencies of each of the 5 classes ranged between18 and 22% for R2, L1,2 and 3, and between 10 and 25% for R1. The average spoken response reaction time was 0.8s and the responses lasted 0.5s on average.

For three datasets (R1, R2 and L1) there was a systematic offset between the ECoG recordings and the video from which the auditory signal was extracted for the VOT. This offset was corrected by determining the exact time mismatch between the stimulus marker in the ECoG data file and the onset of the stimulus in the video recording (the patient task screen was recorded as Picture in Picture). This then also aligned the VOT's to the ECoG data. For L2 and L3 the auditory signal was directly co-recorded with the ECoG data.

2.3 Preprocessing

The HFB was extracted from each electrode recording in a five-step process. First, signal from each electrode was notch-filtered to remove line noise (using the ‘filtfilt’ and ‘butter’ function in Matlab for the ranges 49:51Hz and 97:103Hz), and periods of poor signal quality were removed (leading to the removal of only a single correctly performed trial from subject L3). Second, each electrode signal was re-referenced to the mean signal of all electrodes in the HD grid within a subject (common average re-referencing), which has yielded good decoding results in multiple HD ECoG studies (Bleichner et al., 2016; Branco et al., 2016). Third, the spectral response was computed using a Gabor Wavelet Dictionary (Bruns, 2004) by convolving Gabor wavelets with frequencies from 65 to 125Hz (in 1 Hz increments) and a full-width at half maximum of 4 wavelengths with the HD-ECoG signal. Fourth, the absolute values of the complex convolved responses were summed over the 61 frequencies, and the log of the sum signal was z-scored for each task run (the z-score was computed using the mean and standard deviation from all electrodes and not per individual electrode to preserve inter-electrode differences in HFB signal variance). Finally, the HFB responses were smoothed with a 100ms kernel and were divided into trials based on the VOT markers or rest trial temporal mid points. The period between 0.5s before and 0.5s after each VOT marker was used to form 1s trials of HFB response.

2.4 Spatial-Temporal Cortical Activation Pattern Classification

Classification scores of the broadband signals were computed using a leave-one-out approach and spatio-temporal Matched filter (STMF) classification was applied. In addition, basic Support Vector Machine (SVM) classifiers were trained and used for comparison to the STMF results. For classification, all trials except for one were used to screen the electrodes for inclusion in the STMFs. Electrodes with a significant difference in HFB power (averaged over the trial) between rest and any of the phonemes ($p<0.05$) were included in the feature set for training (Figure 3).

For STMF analyses, the mean HFB power trace for the 1s trial period was computed for each electrode across trials for each of the five classes (/p/, /k/, /u/, /a:/, and rest). The mean traces of electrodes were then concatenated into a single vector representing the STMF for each class. The correlation between the SMTF of each class and the concatenated HFB traces of the included electrodes of the left-out trial (the test trial) was then computed. The test trial was then classified according to the STMF it correlated highest with. The analysis was repeated for 500 ms windows to evaluate feasibility of decoding shorter periods before, centered at, and after VOT.

For SVM analyses, classifiers were trained using the same leave-one-out strategy as applied for the STMF classifiers. Electrodes were included where all phonemes together differed significantly from rest. The SVM used linear kernels and were trained for each class versus all other classes. Each trial was classified by computing the distance from the trial features to each ‘1 vs all’ other class boundaries and was given the class label that maximized the distance to the boundaries. The SVM classifier was implemented using the ‘fitcsvm’ function in the Matlab 2016a statistics toolbox.

The classification accuracies were not weighted to account for differences in numbers of trials for each class. To assess the chance level of our classifications we computed a non-parametric distribution of classification scores by randomly shuffling the trial labels 120 for SVM and 500 times for all other analyses (difference in numbers is due to computation time), and applying the leave-one-out procedure to each set of shuffled trials according to the procedure described in (Maris and Oostenveld, 2007). Chance levels were then determined to be the upper 95% confidence bound of the distribution. Classification results exceeding these chance levels are significant at the level $p < 0.05$, and become rapidly more significant with distance from these levels. Furthermore, since we did leave-one-out classification, the distribution of trials over classes for each training set does not change except that the training set sometimes had one trial less in a class.

2.5 Effects of temporal information

To assess the extent to which the temporal information of the HFB power response added to the classification results, a procedure identical to the one described in the section above was used, but here matched filters were computed not only as the means over different trials, but also averaging over a trial period. This resulted in purely spatial matched filter (SMF) patterns. Electrodes were included where all phonemes together differed significantly from rest.

2.6 Assessing locations of most informative electrodes

Given that the electrodes likely covered cortex that was not involved in phoneme production, we expected that individual electrodes would contribute to the classification to different degrees. To quantify the extent to which the individual neural populations covered by each electrode contributed to the classification results, we determined the relative contribution of each electrode to the classification scores. For this, we again applied a non-parametric sampling technique by choosing 5000 random subsets of the electrodes. For each subset, a number (N_r) of electrodes between 1 and the total number of electrodes included in the analysis for a given subject was randomly generated and a set of N_r electrodes was randomly selected from all analyzed electrodes of the corresponding subject. Hence, on average half of the electrodes were included. Next the above described STMF classification procedure was applied. For each of the random subsets the resulting leave-one-out classification score was recorded, which was assigned to all electrodes included in the subset. This way, each electrode had a range of classification scores assigned to it. Finally, the mean score over the distributions represented a quantitative contribution, for each electrode, to the classification.

To evaluate where cortical activity could be expected during phoneme production, we conducted a group analysis of previously acquired 7T fMRI data. These were obtained for an earlier study (Bleichner et al., 2015), but were not published. Twelve healthy volunteers sequentially generated, in a randomized scheme, four phonemes upon visual cues, at a rate of 1 per 15.6 seconds (10 repetitions per phoneme). The fMRI scans were analyzed with SPM8, with an event-related design and taking all phonemes together (no contrast between phonemes), and the generated b-maps (representing the degree to which each voxel responds to the task) were used for group analysis. After normalization to MNI space, all individual b-maps were averaged into one group-b-map, which was finally displayed on the surface of the average anatomical scan of 12 healthy subjects. All details of data acquisition and individual subject data analysis can be found in (Bleichner et al., 2015).

3 Results

3.1 Spoken phoneme ECoG classification

The main finding of our analysis was that the 5 classes (4 spoken phoneme classes plus rest) could be classified with STMF analysis, with a mean accuracy of 75.5% (sd 6.5%), at a mean empirically determined chance level of 26.4% (Table 2, Figure 4. Given that one condition may dominate the classification of rest versus active, we also calculated classification scores for phonemes combined versus rest, and for the 4 phonemes without rest (Table 2). This revealed that active versus rest trials could be classified at 87.6% (sd 14.1%), and that 4 phonemes could be distinguished at a score of 71.9% (sd 8.8%). A mean confusion matrix for the 5-class classification (Figure 5) shows that although rest was distinguished the most, each of the phonemes was identified well above chance. Note that the chance levels were empirically determined and differed from theoretical levels due to non-equal numbers of trials per class.

For STMF, three 500 ms windows were evaluated for the 4- and 5-class datasets. All three yielded significant classification (Table 2), but for both 4- and 5-class analysis the window spanning 250 ms before to 250 after VOT, reached the same level of classification as the 1000 ms analyses. The same 500 ms optimal window was analyzed with SMF for comparison, which resulted in slightly lower scores than for the 1000 ms window for SMF. The difference between STMF and SMF, however, remained the same, with STMF yielding well over 10 % better performance in all comparisons

To compare the STMF approach to SVM analyses, classification scores were obtained with SVM for each dataset. Classification based on SVM yielded similar results as the STMF analysis, as is shown in Table 2. Hence, despite the computationally simple nature of the STMF classifier scheme it was not outperformed by the SVM approach (paired t-test, p>0.3 for 2-, 4- and 5-class analyses).

3.2 Informative electrodes

Evaluation of the distribution of most informative electrodes, based on their contribution to classification, shows that most of these electrodes were clustered along the central sulcus, essentially along the length of the inferior half, and with no particular preference for cortex located anterior versus posterior of the sulcus (Figure 6). The distribution of activity obtained with fMRI in healthy volunteers covers pre-and post-central gyri ranging from the most inferior aspect (abutting the Sylvian fissure) to below the typical hand knob (Figure 6). The vast majority of the most informative electrodes fall within this region, suggesting that indeed the sensorimotor cortex contributes to decoding. A verification was performed to ascertain that 5000 subsets were sufficient for determining the most informative electrodes. We found that after 3000 iterations the set of most informative electrodes did not change, indicating that 5000 was sufficient.

3.3 Contribution of temporal information in decoding

When temporal information was not included in the Matched Filters (SMF), classification dropped considerably to 63.5% for the 5-class analysis (2-tailed paired t-test p=0.046) and 60.3% for the 4-class analysis (2-tailed paired t-test p=0.064) (Table 2). Even for the Active versus Rest analysis, classification scores dropped by more than 10%. These findings corroborate the notion that phoneme production is accompanied by brief time-varying events across multiple electrodes, as different muscles are contracted in a specific sequence. Even so, SMF analysis did yield well-above chance classification of the phonemes (Table 2). As mentioned before, the difference between STMF and SMF analysis remained, also when comparing the optimal 500 ms window decoding (Table 2, significant for the 4- and the 5-class classification at p<=0.05).

4 Discussion

We addressed the hypothesis that elementary components of speech production, phonemes, engage the sensorimotor cortex in a decodable fashion. To this end, we conducted research in epilepsy patients with implanted HD electrode grids placed on the sensorimotor face area, and asked them to perform a phoneme production task. The cortical spatiotemporal activity patterns generated during this task proved to be highly reproducible and phoneme-specific, as evidenced by a high 5-class classification score (75.5%). When omitting the temporal information, the evolution of activity over time within a single trial, classification dropped significantly to 63.5%, corresponding to the notion that the production of phonemes involves multiple sequential, rapidly changing, combinations of transient activity across multiple small cortical foci (Bouchard et al., 2013; Jiang et al., 2016). The fact that even without temporal information classification was highly significant, suggests that each phoneme engaged a different set of cortical foci, likely related to the collection of muscles contracted for production (Kellis et al., 2010). The significant improvement with the inclusion of temporal information and visual inspection of the STMFs (see figure 4) also indicates that the set of cortical foci overlapped between phonemes.

Of note, subject R1 did not benefit from adding temporal information in the classifiers. This subject gave the highest scores, over 80%, for the SMF analysis compared to the other subjects. This high score may reflect optimal grid placement combined with density of electrodes (3 mm). One explanation for the lack of benefit from temporal information could be that the trials were less well aligned, possibly due to varying durations of phoneme utterance or less accurate VOT determinations. This would cause a loss of temporal information in the STMFs due to temporal smoothing resulting from averaging poorly aligned trials. One way to solve this would be to align trials based on the timing of the HFB response, an example of which was reported recently for decoding gestures (Branco et al., 2016). Alternatively, performance could be improved by machine learning algorithms that allow for variations in length of the HFB responses, such as advanced neural networks.

Since classification can be biased by a single class, in our case most likely the rest condition compared to any phoneme, the analyses were also conducted on only the phoneme trials (excluding rest trials), thus addressing directly the discriminability of the four phonemes. This analysis yielded only a slightly lower classification score (71.9% versus 75.5%, albeit with different chance levels, see Table 2), which also holds when decoding the 500 ms window centered on the VOT and capturing the HFB response (Figure 4) as shown in Table 2. As shown in Figure 5, rest trials are distinguished best from all phonemes. The vowels /u/ and /a:/ are least distinguishable (Figure 5), perhaps because they both lack plosives and differ only in lip positions, but this remains speculative.

The analysis with SVM yielded similar results in terms of classification scores. One would expect SVM to result in better scores given the more advanced algorithms to extract the most discriminative information, but in this study this was not the case. One explanation could be that the underlying neurophysiological spatiotemporal patterns are in effect quite robust and reproducible. In principle SVMs handle noise contributions better than template matching, suggesting that the reason for similar performance could be that classification performance was limited more by a variation in producing the phonemes than by noise sources. For aligning trials, we used the voice onset time based on the audio signal. Yet phonemes may well have been produced slightly different in terms of duration and exact pronunciation, hence capping classification performance for both STMF and SVM.

The robustness of our findings is further indicated by several results. First, a more advanced classification method, SVM, did not yield better results than the STMF technique. The fact that a straightforward averaging of the spatiotemporal patterns of same-phoneme trials provided quite discriminable features across phonemes, indicates that the neuronal activity patterns were considerably unique and reproducible across repetitions. Second, we found no clear difference in decoding performance for left versus right hemisphere. This suggests that there is no lateralization for phoneme productions, which in turn supports the notion that we decoded neuronal activity associated with bilateral muscle contraction. The comparison with only spatial patterns confirms a clear improvement when adding temporal information in the feature set. The higher density of electrodes (than the standard 1 cm spacing used in many studies) is likely to also explain the improved classification, as for each subject the most informative electrodes were clustered within spaces of a square cm around the central sulcus (Figure 6). Regarding the time window required for decoding phonemes, Jiang (Jiang et al., 2016) found that a window of 500 ms (centered around VOT) yielded the best decoding results. Mugler noted that most of the information was obtained from a 400 ms window, also centered around VOT (Mugler et al., 2014). The current results corroborate these reports, indicating that optimal decoding of phonemes from sensorimotor cortex requires 400-500 ms. Of note, the time needed for decoding a phoneme is associated with the seemingly sluggish temporal features of the HFB response, which thereby limits performance of shorter windows. The 500 ms delay in decoding would not necessarily hamper use for BCI if a sliding window would capture attempted speech, in which case the synthetic speech generator would simply have a lag of about half a second. However, our study does not provide information about the ability to decode attempted speech, and it may well be that the sluggish HFB response prohibits real-time decoding of rapid sequences of phonemes.

Decoding speech is the goal of multiple research endeavors. Roughly two approaches may be distinguished, namely discrete classification of speech elements, as in the present study, and reconstruction of speech. For speech reconstruction, brain signals are linked to articulators in software, which allow for synthetic generation of sounds similar to sounds produced with the natural human speech production system (Brumberg et al., 2010; Guenther et al., 2009). Alternatively, spatiotemporal ECoG patterns may be mapped onto acoustic features directly (Bouchard and Chang, 2014; Martin et al., 2014). Sounds generated by linear modulation of articulators may by principle of feedback be amenable to improvement, conceptually resulting in intelligible speech (Brumberg et al., 2010; Guenther et al., 2009). Discrete classification aims to provide people with a vocabulary of phonemes or words. Several studies with intracranial electrodes have shown that production of words is accompanied by activity in various regions associated with language processing (Leuthardt et al., 2012; Pei et al., 2011b). Feasibility of discriminating among words was shown in several studies where all implanted electrodes were included in the feature selection (Kellis et al., 2010), with a predominance for auditory cortex even when subjects produce words covertly (Martin et al., 2016). Others investigated decodability of phonemes embedded in monosyllabic words, either phonemes versus rest (Leuthardt et al., 2011), or between phonemes (Herff et al., 2015; Mugler et al., 2014; Pei et al., 2011a). One study examined decodability of single, imagined phoneme production, and reported classification scores per single electrode on the order of 43%, at a chance level of 33% (Ikeda et al., 2014). Informative electrodes were found across various regions including superior temporal gyrus and premotor cortex. The present study differs from previous studies in that we combined several factors. First, we specifically targeted the sensorimotor face region, (which has a closer relation to overt or attempted speech production), excluding any electrodes on other regions, such as auditory cortex, which conceivably have a higher degree of overlap with features of perceived speech which a communications BCI does not seek to decode. Second, all our participants were implanted with high-density grids on this region to maximize information content. Even higher densities have also been shown to capture decodable information, but can currently only cover a fraction of the motor face area (Kellis et al., 2016; Leuthardt et al., 2009). Third, we included a more fine-grained temporal evolution of the HFB signal in our classification than previous studies, by a simple concatenation of electrode HFB traces, capitalizing on the fact that phoneme production involves rapid activity pattern changes over time (Bouchard et al., 2013). Fourth, we employed a classification technique that does not require any parameter optimization or prior selection of parameters (other than discarding electrodes that did not respond to the task), namely a spatiotemporal matched filter. Although it is difficult to compare across studies given the wide range of tasks used, the classification results appear to exceed those of all studies we are aware of that specifically addressed decoding of generated phonemes from the sensorimotor region, being approximately 41% for 4 vowels or consonants [4 classes, chance level 25%] (Pei et al., 2011a), 73% for three vowels [3 classes, chance level 33%] (Jiang et al., 2016), 37% for all English phonemes [39 classes, chance level 11%] (Mugler et al., 2015), and 45% for 4 phonemes [4 classes, chance level 25%] (Mugler et al., 2014). Additionally, several case studies reported classification scores on the order of 70-75% [2 classes, chance level 50%] with similar grids as the ones we used (Blakely et al., 2008), and 20% [38 classes, chance level 2.6%] with neurotrophic electrodes in a quadriplegic subject (albeit including 14 un-decodable phonemes which might bias the results) (Brumberg et al., 2011). Importantly, the fact that phonemes were spoken in clear isolation likely makes decoding easier than when words or syllables are used, making a direct comparison of classification performance with any of the studies mentioned above uninformative.

Employing a paradigm where subjects speak monosyllabic words rather than discrete phonemes, may well lead to an underestimation of decodability since the HFB responses of each phoneme are convolved with those of neighboring phonemes.

The results show that different phonemes correspond to distinguishable STMFs, suggesting that we captured the underlying cortical representation of the various articulator-related muscle groups. The findings align with other similar studies, but with a seemingly higher classification score. Factors contributing to this include high-density grids, coverage of the inferior aspect of the sensorimotor cortex, and the use of discrete phonemes. Whether the results with a small set of phonemes can predict performance with the full set of phonemes is not clear. Mugler (Mugler et al., 2014) noted that similarity in the articulator sets used for similar phonemes reduced discriminability between them. Nevertheless, for BCI purposes one can imagine that a limited, optimally discriminable, set of phonemes may enable communication albeit perhaps with a consequently limited vocabulary. An interesting approach for identifying the most discriminable phonemes is one where ECoG is analyzed according to higher-level structure of phoneme articulation with categories such as obstruent/sonorant, and labial/coronal/dorsal, as reported by Lotte (Lotte et al., 2015). Their analyses included electrodes across the perisylvian regions and revealed a significant contribution of both sensorimotor cortex and auditory cortex to discrimination of categories. It would be interesting to apply their method to sensorimotor cortex with high-density grids coverage. An additional unknown is whether phonemes can be discriminated when spoken in (rapid) sequence, given that the neuronal spatiotemporal patterns are likely to overlap with those of preceding or ensuing phonemes. Herff (Herff et al., 2015) reported promising results in several cases with data obtained during reading text out loud, showing that decoding of continuous speech may be feasible. In that study, the auditory cortex contributed considerably to decoding. Decoding perceived speech has been shown by several groups to be feasible from auditory cortex, which in itself contributes to understanding how the auditory cortex is organized (Formisano et al., 2008; Mesgarani et al., 2014; Pasley et al., 2012). However, perceived speech can be seen as a possible source of undesired decoded speech events in the context of a BCI focused on decoding only speech that was intended for overt communication. As such, decoding of perceived speech is not a feasible approach for BCI purposes for communication. Imagined speaking in people without speech impediments appears to be significantly less decodable from auditory cortex than overt or heard speech (Martin et al., 2016). As described below, speech attempted by speech-disabled subjects may, however, be better decodable.

Interestingly, we found that the time window of 500 ms before VOT could be classified at a highly significant level (64% for the 5-class set), which was almost as high as the 500 ms after VOT. This suggests that decoding does not depend entirely on sensory feedback, an important issue for BCI applications for paralyzed people who can only produce phonemes covertly. Several studies have shown significant classification for imagined phonemes or words (Ikeda et al., 2014; Martin et al., 2016; Pei et al., 2011a), albeit at sometimes considerably lower levels than what was shown for overt speech (Martin et al., 2016; Mugler et al., 2014; Pei et al., 2011a). An important question is whether imagined speech is predictive of neuronal activity generated by *attempted* speech in paralyzed people. For hand movements, this is not convincingly the case, as we have reported that the primary motor cortex fails to exhibit clear activation for imagined movement in a carefully controlled fMRI and EEG experiment (Hermes et al., 2011). Attempted hand movements, on the other hand, have been shown to generate activation patterns in arm amputees that are similar to the patterns observed during actual movement in healthy volunteers without amputations, as opposed to patterns generated by imagined movements in the latter group (Lotze et al., 2001; Raffin et al., 2012; Roux et al., 2003). Thus, although it seems logical to use imagined movement as a predictor of decodability in paralyzed people, actual movements may prove to be more predictive, albeit with a certain degree of confound due to somatosensory feedback activity. Finally, an issue we could not address is the nature of the topographical organization of sensorimotor cortex in relation to phoneme production. We believe that this may need to await the ability to fully cover the sensorimotor face region (Bouchard et al., 2013)(and Figure 6) in combination with denser electrode configurations (Slutzky et al., 2010) in single subjects to enable exhaustive mapping and assessment of topographical similarities between subjects. Signals recorded by electrodes are dominated by tissue in their immediate vicinity, and leave tissue between electrodes (in our case still some 85%) unrecorded, constituting significant undersampling of contiguous cortical surface.

Some aspects of the study impose some limitations for data interpretation. For one, the number of trials was too small to separate data into a training and a test set, leading us to classify with a leave-one-out schedule. Larger numbers of trials are clearly better, but is often not feasible due to the limited time available with ECoG patients. Yet, the results are fairly robust considering that good performance was achieved by a simple averaging of trials to obtain the phoneme-specific classifiers. Second, it is not possible to predict from this study whether similar decoding performance can be achieved for BCI, where there is no somatosensory feedback. This will ultimately require research with subjects who cannot communicate by speech.

Conclusion

A set of four phonemes could be classified with an accuracy that encourages further research on decoding speech from neuronal spatiotemporal activity patterns. The findings support and build upon reports that high-density grids on sensorimotor cortex improve decoding, and that inclusion of the finegrained temporal evolution of brain signals captures the rapid sequence of articulatory muscle groups employed in phoneme production. Whether these findings translate to decoding of attempted speech in communication-challenged people ultimately requires research in this target population, although the fact that significant decoding was achieved in the 500 ms before VOT is encouraging. This study, and several other studies, contribute to building a case for conducting such research in the future.

Acknowledgements

This research was funded by the ERC-Advanced ‘iConnect’ project (grant 320708). We thank Frans Leijten, Cyrille Ferrier, Geertjan Huiskamp, and Tineke Gebbink for their help in collecting data, Peter Gosselaar and Peter van Rijen for implanting the electrodes, as well as the technicians, the staff of the clinical neurophysiology department and the subjects for their time and effort. We also thank the members of the UMC Utrecht ECoG research team (Elmar Pels, Mariana Branco) for data collection.

References

Andersson P, Pluim JPW, Siero JCW, Klein S, Viergever MA, Ramsey NF. Real-time decoding of brain responses to visuospatial attention using 7T fMRI. PloS One. 2011; 6:e27638.doi: 10.1371/journal.pone.0027638 [PubMed: 22110702]

Andersson P, Pluim JPW, Viergever MA, Ramsey NF. Navigation of a telepresence robot via covert visuospatial attention and real-time fMRI. Brain Topogr. 2013; 26:177–185. DOI: 10.1007/s10548-012-0252-z [PubMed: 22965825]

Blakely T, Miller KJ, Rao RPN, Holmes MD, Ojemann JG. Localization and classification of phonemes using high spatial resolution electrocorticography (ECoG) grids. Conf Proc Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Conf. 2008; 2008:4964–4967. DOI: 10.1109/IEMBS.2008.4650328

Bleichner MG, Freudenburg ZV, Jansma JM, Aarnoutse EJ, Vansteensel MJ, Ramsey NF. Give me a sign: decoding four complex hand gestures based on high-density ECoG. Brain Struct Funct. 2016; 221:203–216. DOI: 10.1007/s00429-014-0902-x [PubMed: 25273279]

Bleichner MG, Jansma JM, Salari E, Freudenburg ZV, Raemaekers M, Ramsey NF. Classification of mouth movements using 7 T fMRI. J Neural Eng. 2015; 12doi: 10.1088/1741-2560/12/6/066026

Bouchard, KE; Chang, EF. Neural decoding of spoken vowels from human sensory-motor cortex with high-density electrocorticography. 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Presented at the 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2014. 6782–6785.

Bouchard KE, Mesgarani N, Johnson K, Chang EF. Functional organization of human sensorimotor cortex for speech articulation. Nature. 2013; 495:327–332. DOI: 10.1038/nature11911 [PubMed: 23426266]

Branco MP, Freudenburg ZV, Aarnoutse EJ, Bleichner MG, Vansteensel MJ, Ramsey NF. Decoding hand gestures from primary somatosensory cortex using high-density ECoG. NeuroImage. 2016; 147:130–142. DOI: 10.1016/j.neuroimage.2016.12.004 [PubMed: 27926827]

Brefczynski JA, DeYoe EA. A physiological correlate of the “spotlight” of visual attention. Nat Neurosci. 1999; 2:370–374. DOI: 10.1038/7280 [PubMed: 10204545]

Brumberg JS, Nieto-Castanon A, Kennedy PR, Guenther FH. Brain–computer interfaces for speech communication. Speech Commun, Silent Speech Interfaces. 2010; 52:367–379. DOI: 10.1016/j.specom.2010.01.001

Brumberg JS, Wright EJ, Andreasen DS, Guenther FH, Kennedy PR. Classification of intended phoneme production from chronic intracortical microelectrode recordings in speech-motor cortex. Front Neurosci. 2011; 5:65.doi: 10.3389/fnins.2011.00065 [PubMed: 21629876]

Brunner C, Birbaumer N, Blankertz B, Guger C, Kübler A, Mattia D, Millán J, del R, Miralles F, Nijholt A, Opisso E, et al. BNCI Horizon 2020: towards a roadmap for the BCI community. Brain-Comput Interfaces. 2015; 2:1–10. DOI: 10.1080/2326263X.2015.1008956

Bruno M-A, Bernheim JL, Ledoux D, Pellas F, Demertzi A, Laureys S. A survey on self-assessed well-being in a cohort of chronic locked-in syndrome patients: happy majority, miserable minority. BMJ Open. 2011; 1doi: 10.1136/bmjopen-2010-000039

Bruns A. Fourier-, Hilbert- and wavelet-based signal analysis: are they really different approaches? J Neurosci Methods. 2004; 137:321–332. DOI: 10.1016/j.jneumeth.2004.03.002 [PubMed: 15262077]

Chaudhary U, Birbaumer N, Ramos-Murguialday A. Brain-computer interfaces for communication and rehabilitation. Nat Rev Neurol. 2016; 12:513–525. DOI: 10.1038/nrneurol.2016.113 [PubMed: 27539560]

Crone NE, Hao L, Hart J, Boatman D, Lesser RP, Irizarry R, Gordon B. Electrocorticographic gamma activity during word production in spoken and sign language. Neurology. 2001; 57:2045–2053. [PubMed: 11739824]

Crone NE, Miglioretti DL, Gordon B, Lesser RP. Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band. Brain J Neurol. 1998; 121(Pt 12):2301–2315.

Farwell LA, Donchin E. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. Electroencephalogr Clin Neurophysiol. 1988; 70:510–523. [PubMed: 2461285]

Fazel-Rezai R, Allison BZ, Guger C, Sellers EW, Kleih SC, Kübler A. P300 brain computer interface: current challenges and emerging trends. Front Neuroengineering. 2012; 5doi: 10.3389/fneng.2012.00014

Formisano E, De Martino F, Bonte M, Goebel R. “Who” is saying “what”? Brain-based decoding of human voice and speech. Science. 2008; 322:970–973. DOI: 10.1126/science.1164318 [PubMed: 18988858]

Gallegos-Ayala G, Furdea A, Takano K, Ruf CA, Flor H, Birbaumer N. Brain communication in a completely locked-in patient using bedside near-infrared spectroscopy. Neurology. 2014; 82:1930–1932. DOI: 10.1212/WNL.0000000000000449 [PubMed: 24789862]

Guenther FH, Brumberg JS, Wright EJ, Nieto-Castanon A, Tourville JA, Panko M, Law R, Siebert SA, Bartels JL, Andreasen DS, Ehirim P, et al. A Wireless Brain-Machine Interface for Real-Time Speech Synthesis. PLOS ONE. 2009; 4:e8218.doi: 10.1371/journal.pone.0008218 [PubMed: 20011034]

Harvey BM, Klein BP, Petridou N, Dumoulin SO. Topographic representation of numerosity in the human parietal cortex. Science. 2013; 341:1123–1126. DOI: 10.1126/science.1239052 [PubMed: 24009396]

Herff C, Heger D, de Pesters A, Telaar D, Brunner P, Schalk G, Schultz T. Brain-to-text: decoding spoken phrases from phone representations in the brain. Front Neurosci. 2015; 9doi: 10.3389/fnins.2015.00217

Hermes D, Miller KJ, Noordmans HJ, Vansteensel MJ, Ramsey NF. Automated electrocorticographic electrode localization on individually rendered brain surfaces. J Neurosci Methods. 2010; 185:293–298. DOI: 10.1016/j.jneumeth.2009.10.005 [PubMed: 19836416]

Hermes D, Vansteensel MJ, Albers AM, Bleichner MG, Benedictus MR, Mendez Orellana C, Aarnoutse EJ, Ramsey NF. Functional MRI-based identification of brain areas involved in motor imagery for implantable brain-computer interfaces. J Neural Eng. 2011; 8doi: 10.1088/1741-2560/8/2/025007

Hubel DH, Wiesel TN. Receptive fields of single neurones in the cat’s striate cortex. J Physiol. 1959; 148:574–591. [PubMed: 14403679]

Ikeda S, Shibata T, Nakano N, Okada R, Tsuyuguchi N, Ikeda K, Kato A. Neural decoding of single vowels during covert articulation using electrocorticography. Front Hum Neurosci. 2014; 8doi: 10.3389/fnhum.2014.00125

Jacobs J, Kahana MJ. Direct brain recordings fuel advances in cognitive electrophysiology. Trends Cogn Sci. 2010; 14:162–171. DOI: 10.1016/j.tics.2010.01.005 [PubMed: 20189441]

Jiang, W; Pailla, T; Dichter, B; Chang, EF; Gilja, V. Decoding speech using the timing of neural signal modulation. 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Presented at the 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2016. 1532–1535.

Kay KN, Naselaris T, Prenger RJ, Gallant JL. Identifying natural images from human brain activity. Nature. 2008; 452:352–355. DOI: 10.1038/nature06713 [PubMed: 18322462]

Kellis S, Miller K, Thomson K, Brown R, House P, Greger B. Decoding spoken words using local field potentials recorded from the cortical surface. J Neural Eng. 2010; 7doi: 10.1088/1741-2560/7/5/056007

Kellis S, Sorensen L, Darvas F, Sayres C, O’Neill K III, Brown RB, House P, Ojemann J, Greger B. Multi-scale analysis of neural activity in humans: Implications for micro-scale electrocorticography. Clin Neurophysiol. 2016; 127:591–601. DOI: 10.1016/j.clinph.2015.06.002 [PubMed: 26138146]

Kennedy PR, Bakay RA. Restoration of neural output from a paralyzed patient by a direct brain connection. Neuroreport. 1998; 9:1707–1711. [PubMed: 9665587]

Kleih SC, Kaufmann T, Zickler C, Halder S, Leotta F, Cincotti F, Aloise F, Riccio A, Herbert C, Mattia D, Kübler A. Out of the frying pan into the fire--the P300-based BCI faces real-world challenges.

Prog Brain Res. 2011; 194:27–46. DOI: 10.1016/B978-0-444-53815-4.00019-4 [PubMed: 21867792]

Klein BP, Harvey BM, Dumoulin SO. Attraction of Position Preference by Spatial Attention throughout Human Visual Cortex. Neuron. 2014; 84:227–237. DOI: 10.1016/j.neuron.2014.08.047 [PubMed: 25242220]

Laureys S, Pellas F, Van Eeckhout P, Ghorbel S, Schnakers C, Perrin F, Berré J, Faymonville M-E, Pantke K-H, Damas F, Lamy M, et al. The locked-in syndrome : what is it like to be conscious but paralyzed and voiceless? Prog Brain Res. 2005; 150:495–511. DOI: 10.1016/S0079-6123(05)50034-7 [PubMed: 16186044]

Leuthardt EC, Freudenberg Z, Bundy D, Roland J. Microscale recording from human motor cortex: implications for minimally invasive electrocorticographic brain-computer interfaces. Neurosurg Focus. 2009; 27:E10.doi: 10.3171/2009.4.FOCUS0980

Leuthardt EC, Gaona C, Sharma M, Szrama N, Roland J, Freudenberg Z, Solis J, Breshears J, Schalk G. Using the Electrocorticographic Speech Network to Control a Brain-Computer Interface in Humans. J Neural Eng. 2011; 8doi: 10.1088/1741-2560/8/3/036004

Leuthardt EC, Pei X-M, Breshears J, Gaona C, Sharma M, Freudenberg Z, Barbour D, Schalk G. Temporal evolution of gamma activity in human cortex during an overt and covert word repetition task. Front Hum Neurosci. 2012; 6doi: 10.3389/fnhum.2012.00099

Lotte F, Brumberg JS, Brunner P, Gunduz A, Ritaccio AL, Guan C, Schalk G. Electrocorticographic representations of segmental features in continuous speech. Front Hum Neurosci. 2015; 9doi: 10.3389/fnhum.2015.00097

Lotze M, Flor H, Grodd W, Larbig W, Birbaumer N. Phantom movements and pain. An fMRI study in upper limb amputees. Brain J Neurol. 2001; 124:2268–2277.

Lulé D, Zickler C, Häcker S, Bruno MA, Demertzi A, Pellas F, Laureys S, Kübler A. Life can be worth living in locked-in syndrome. Prog Brain Res. 2009; 177:339–351. DOI: 10.1016/S0079-6123(09)17723-3 [PubMed: 19818912]

Manning JR, Jacobs J, Fried I, Kahana MJ. Broadband shifts in LFP power spectra are correlated with single-neuron spiking in humans. J Neurosci Off J Soc Neurosci. 2009; 29:13613–13620. DOI: 10.1523/JNEUROSCI.2041-09.2009

Maris E, Oostenveld R. Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods. 2007; 164:177–190. DOI: 10.1016/j.jneumeth.2007.03.024 [PubMed: 17517438]

Markram H. Fixing the location and dimensions of functional neocortical columns. HFSP J. 2008; 2:132–135. DOI: 10.2976/1.2919545 [PubMed: 19404466]

Martin S, Brunner P, Holdgraf C, Heinze H-J, Crone NE, Rieger J, Schalk G, Knight RT, Pasley BN. Decoding spectrotemporal features of overt and covert speech from the human cortex. Front Neuroengineering. 2014; 7doi: 10.3389/fneng.2014.00014

Martin S, Brunner P, Iturrate I, Millán J, del R, Schalk G, Knight RT, Pasley BN. Word pair classification during imagined speech using direct brain recordings. Sci Rep. 2016; 6doi: 10.1038/srep25803

McCane LM, Heckman SM, McFarland DJ, Townsend G, Mak JN, Sellers EW, Zeitlin D, Tenteromano LM, Wolpaw JR, Vaughan TM. P300-based Brain-Computer Interface (BCI) Event-Related Potentials (ERPs): People with Amyotrophic Lateral Sclerosis (ALS) vs. Age-Matched Controls. Clin Neurophysiol Off J Int Fed Clin Neurophysiol. 2015; 126:2124–2131. DOI: 10.1016/j.clinph.2015.01.013

McFarland DJ, Miner LA, Vaughan TM, Wolpaw JR. Mu and beta rhythm topographies during motor imagery and actual movements. Brain Topogr. 2000; 12:177–186. [PubMed: 10791681]

Mesgarani N, Cheung C, Johnson K, Chang EF. Phonetic Feature Encoding in Human Superior Temporal Gyrus. Science. 2014; 343:1006–1010. DOI: 10.1126/science.1245994 [PubMed: 24482117]

Miall RC, Wolpert DM. Forward Models for Physiological Motor Control. Neural Netw, Four Major Hypotheses in Neuroscience. 1996; 9:1265–1279. DOI: 10.1016/S0893-6080(96)00035-4

Miller KJ, Hermes D, Honey CJ, Hebb AO, Ramsey NF, Knight RT, Ojemann JG, Fetz EE. Human Motor Cortical Activity Is Selectively Phase-Entrained on Underlying Rhythms. PLoS Comput Biol. 2012; 8doi: 10.1371/journal.pcbi.1002655

Miller KJ, Zanos S, Fetz EE, den Nijs M, Ojemann JG. Decoupling the Cortical Power Spectrum Reveals Real-Time Representation of Individual Finger Movements in Humans. J Neurosci. 2009; 29:3132–3137. DOI: 10.1523/JNEUROSCI.5506-08.2009 [PubMed: 19279250]

Mountcastle VB. The columnar organization of the neocortex. Brain J Neurol. 1997; 120(Pt 4):701–722.

Mugler, EM; Goldrick, M; Rosenow, JM; Tate, MC; Slutzky, MW. Decoding of articulatory gestures during word production using speech motor and premotor cortical activity. 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Presented at the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2015. 5339–5342.

Mugler EM, Patton JL, Flint RD, Wright ZA, Schuele SU, Rosenow J, Shih JJ, Krusienski DJ, Slutzky MW. Direct classification of all American English phonemes using signals from functional speech motor cortex. J Neural Eng. 2014; 11doi: 10.1088/1741-2560/11/3/035015

Muller L, Hamilton LS, Edwards E, Bouchard KE, Chang EF. Spatial resolution dependence on spectral frequency in human speech cortex electrocorticography. J Neural Eng. 2016; 13doi: 10.1088/1741-2560/13/5/056013

Pasley BN, David SV, Mesgarani N, Flinker A, Shamma SA, Crone NE, Knight RT, Chang EF. Reconstructing Speech from Human Auditory Cortex. PLoS Biol. 2012; 10doi: 10.1371/journal.pbio.1001251

Pei X, Barbour D, Leuthardt EC, Schalk G. Decoding Vowels and Consonants in Spoken and Imagined Words Using Electrocorticographic Signals in Humans. J Neural Eng. 2011a; 8doi: 10.1088/1741-2560/8/4/046028

Pei X, Leuthardt EC, Gaona CM, Brunner P, Wolpaw JR, Schalk G. Spatiotemporal Dynamics of Electrocorticographic High Gamma Activity During Overt and Covert Word Repetition. NeuroImage. 2011b; 54:2960–2972. DOI: 10.1016/j.neuroimage.2010.10.029 [PubMed: 21029784]

Pfurtscheller G, Neuper C. Motor imagery activates primary sensorimotor area in humans. Neurosci Lett. 1997; 239:65–68. DOI: 10.1016/S0304-3940(97)00889-6 [PubMed: 9469657]

Polimeni JR, Fischl B, Greve DN, Wald LL. Laminar analysis of 7 T BOLD using an imposed spatial activation pattern in human V1. NeuroImage. 2010; 52:1334–1346. DOI: 10.1016/j.neuroimage.2010.05.005 [PubMed: 20460157]

Raffin E, Mattout J, Reilly KT, Giraux P. Disentangling motor execution from motor imagery with the phantom limb. Brain J Neurol. 2012; 135:582–595. DOI: 10.1093/brain/awr337

Ray S, Maunsell JHR. Different Origins of Gamma Rhythm and High-Gamma Activity in Macaque Visual Cortex. PLoS Biol. 2011; 9doi: 10.1371/journal.pbio.1000610

Rousseau M-C, Baumstarck K, Alessandrini M, Blandin V, Billette de Villemeur T, Auquier P. Quality of life in patients with locked-in syndrome: Evolution over a 6-year period. Orphanet J Rare Dis. 2015; 10doi: 10.1186/s13023-015-0304-z

Roux F-E, Lotterie J-A, Cassol E, Lazorthes Y, Sol J-C, Berry I. Cortical areas involved in virtual movement of phantom limbs: comparison with normal subjects. Neurosurgery. 2003; 53

Sadtler PT, Quick KM, Golub MD, Chase SM, Ryu SI, Tyler-Kabara EC, Yu BM, Batista AP. Neural constraints on learning. Nature. 2014; 512:423–426. DOI: 10.1038/nature13665 [PubMed: 25164754]

Sanchez-Panchuelo RM, Besle J, Beckett A, Bowtell R, Schluppeck D, Francis S. Within-Digit Functional Parcellation of Brodmann Areas of the Human Primary Somatosensory Cortex Using Functional Magnetic Resonance Imaging at 7 Tesla. J Neurosci. 2012; 32:15815–15822. DOI: 10.1523/JNEUROSCI.2501-12.2012 [PubMed: 23136420]

Schalk G, Leuthardt EC. Brain-Computer Interfaces Using Electrocorticographic Signals. IEEE Rev Biomed Eng. 2011; 4:140–154. DOI: 10.1109/RBME.2011.2172408 [PubMed: 22273796]

Sellers EW, Ryan DB, Hauser CK. Noninvasive brain-computer interface enables communication after brainstem stroke. Sci Transl Med. 2014; 6doi: 10.1126/scitranslmed.3007801

Sellers EW, Vaughan TM, Wolpaw JR. A brain-computer interface for long-term independent home use. Amyotroph. Lateral Scler Off Publ World Fed Neurol Res Group Mot Neuron Dis. 2010; 11:449–455. DOI: 10.3109/17482961003777470

Siero JCW, Hermes D, Hoogduin H, Luijten PR, Ramsey NF, Petridou N. BOLD matches neuronal activity at the mm scale: a combined 7T fMRI and ECoG study in human sensorimotor cortex. NeuroImage. 2014; 101:177–184. DOI: 10.1016/j.neuroimage.2014.07.002 [PubMed: 25026157]

Slutzky MW, Jordan LR, Krieg T, Chen M, Mogul DJ, Miller LE. Optimal Spacing of Surface Electrode Arrays for Brain Machine Interface Applications. J Neural Eng. 2010; 7doi: 10.1088/1741-2560/7/2/026004

Vansteensel MJ, Pels EGM, Bleichner MG, Branco MP, Denison T, Freudenburg ZV, Gosselaar P, Leinders S, Ottens TH, Van Den Boom MA, Van Rijen PC, et al. Fully Implanted Brain-Computer Interface in a Locked-In Patient with ALS. N Engl J Med. 2016; 375:2060–2066. DOI: 10.1056/NEJMoa1608085 [PubMed: 27959736]

Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain-computer interfaces for communication and control. Clin Neurophysiol Off J Int Fed Clin Neurophysiol. 2002; 113:767–791.

Highlights

1) Discrete, spoken phonemes can be classified with high performance from sensorimotor cortex, even before voice onset.

2) Phoneme production is accompanied by brief sequences of robust, sub-centimeter patterns of electrical activity on the sensorimotor face area, which reflect the sequence of engaged, articulator-related, muscle groups.

3) Decoding spoken phonemes benefits from inclusion of the temporal evolution of high frequency band power, to capture the rapid sequence of activity patterns.

4) Decoding spoken phonemes benefits from sampling from the whole inferior sensorimotor region, with electrodes spaced 4 mm apart or less.

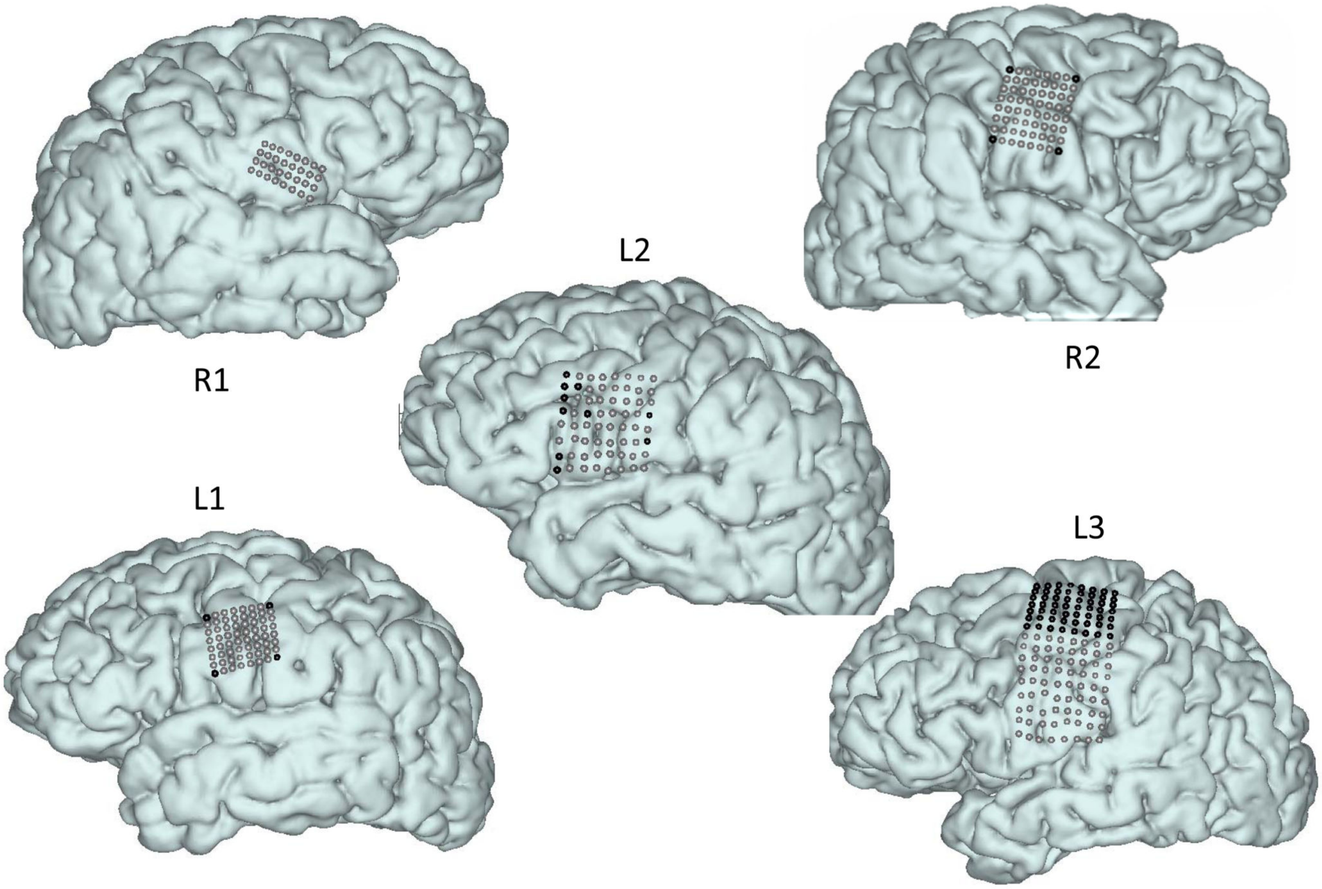

Figure 1.

Localization of HD-ECoG electrodes used for decoding. Each electrode location is indicated by a sphere on the cortical surface. Black spheres indicate electrodes that were excluded from analysis due to either orientation (facing the skull), position outside of the target region, or poor signal quality. Some electrodes seem out of line within grids, but this is due to correction for brain shift (Branco et al., 2016; Hermes et al., 2010).

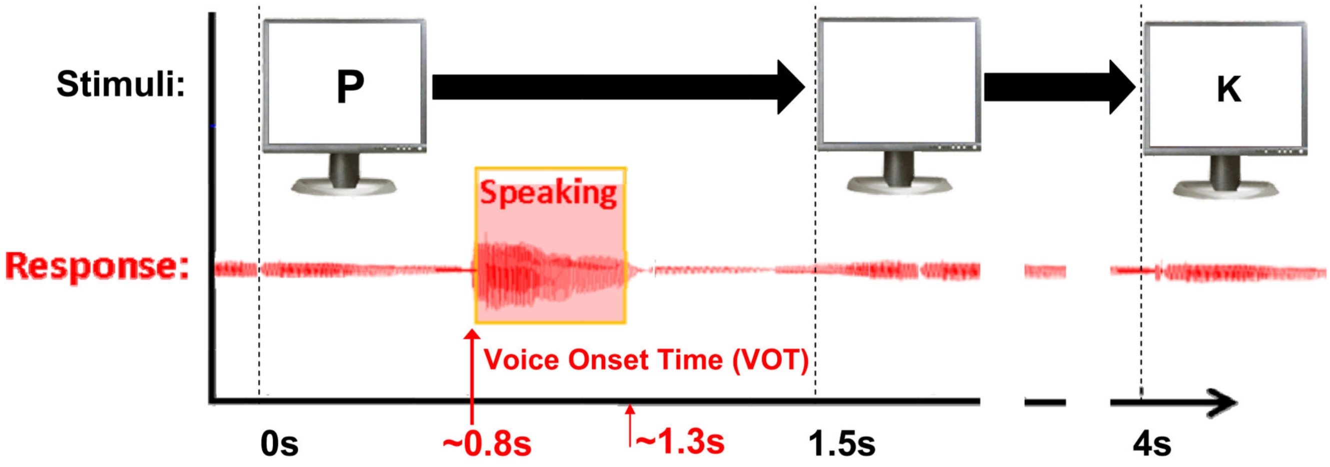

Figure 2.

The phoneme production task with timing of the stimuli. The lower part displays the aligned audio signal which was used to determine the onset of the phoneme production.

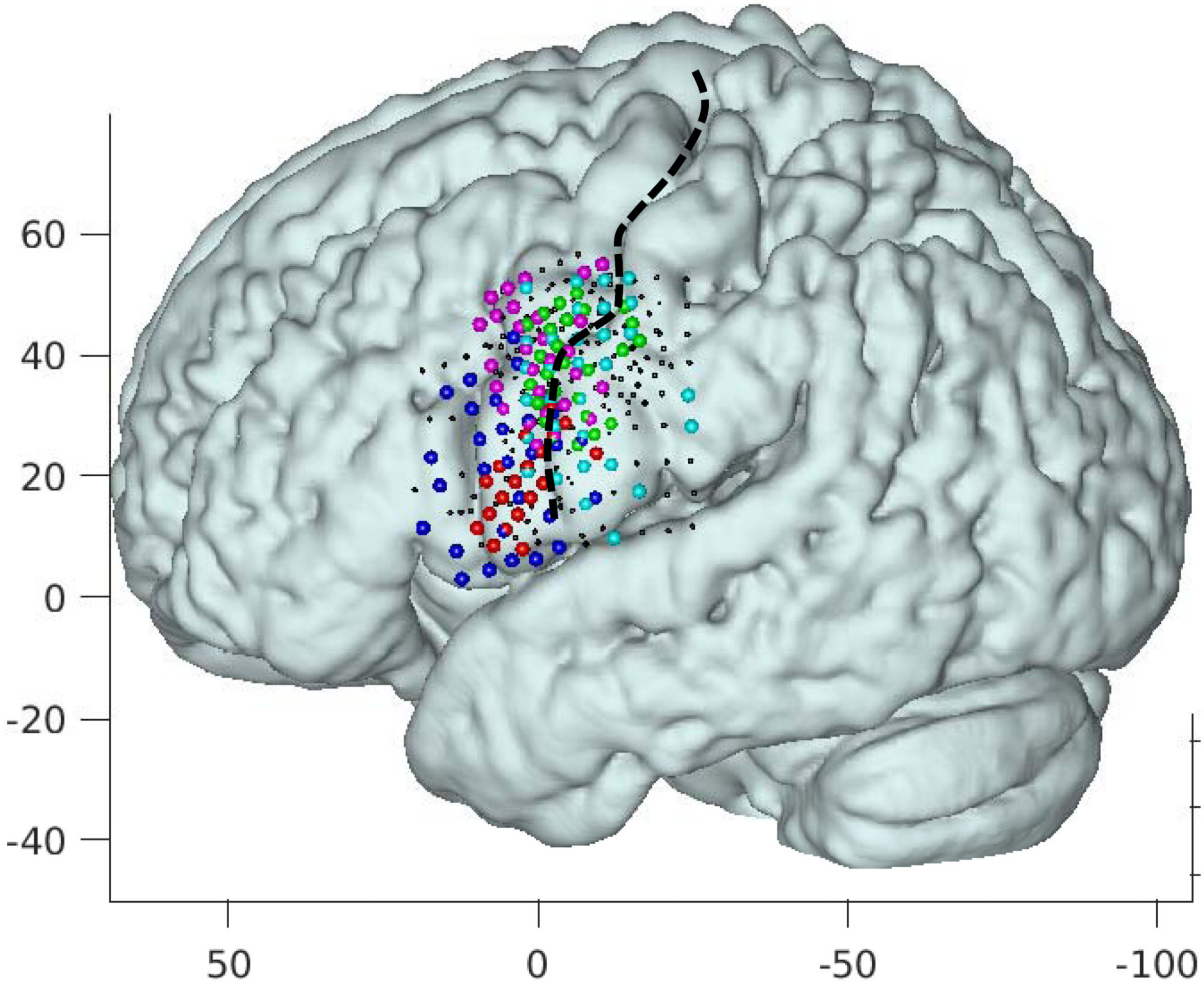

Figure 3.

All electrode grids projected onto the left hemisphere in MNI space (projected onto an average of 12 normal brains). Each color denotes the electrodes with a significant response to the task (as described in the methods for STMFs) of a different patient with red, magenta, green, blue, and cyan corresponding to subject R1, R2, L1, L2, and L3 respectively. The central sulcus is indicted with a black broken line. Axes indicate MNI coordinates

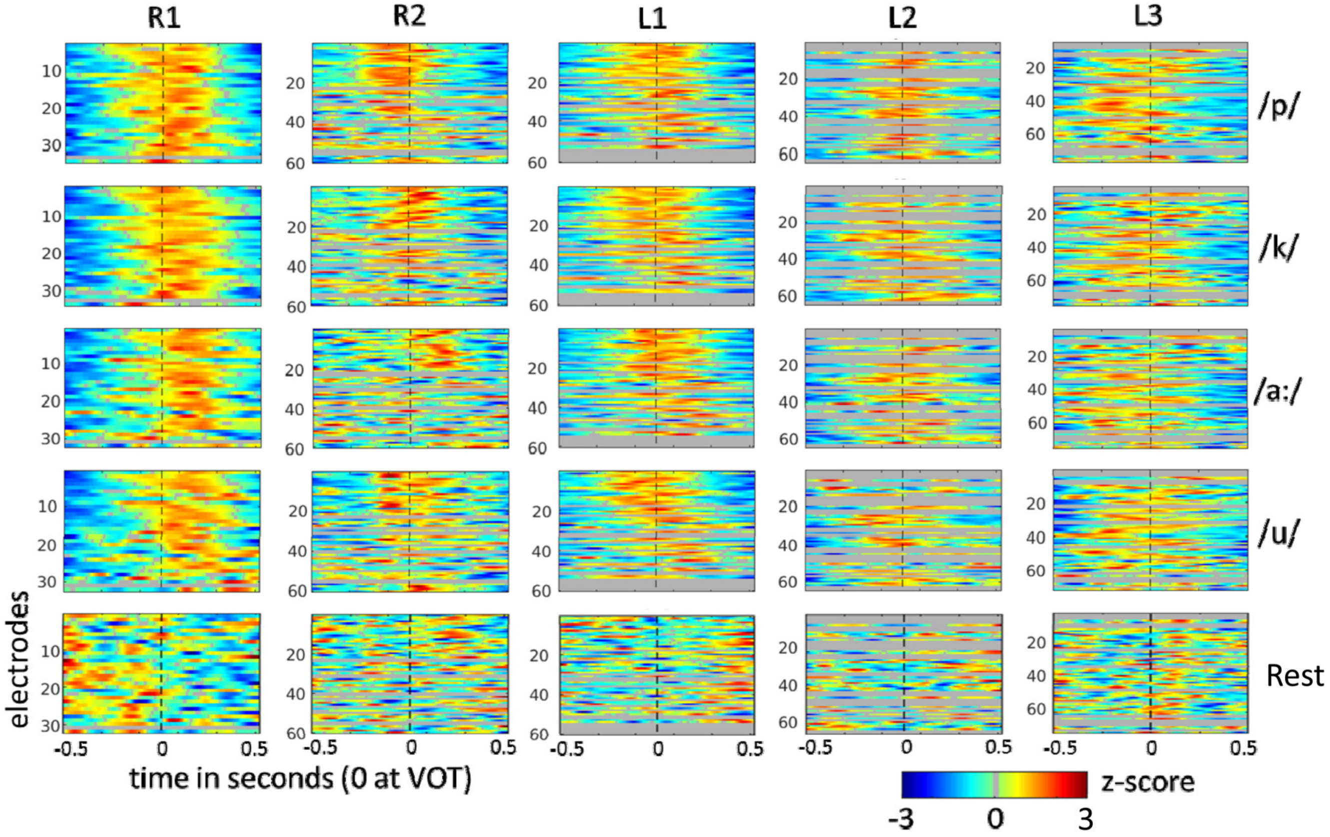

Figure 4.

All STMFs for all subjects and classes. The mean HFB responses (over trials and leave-one-out training sets) for each class, converted to z-scores, are shown, with time relative to the voice onset time (VOT) on the x-axis and electrodes on the y-axis. Grey lines represent electrodes that were excluded from classification (explained in Methods). Dotted vertical lines indicate the VOT marker.

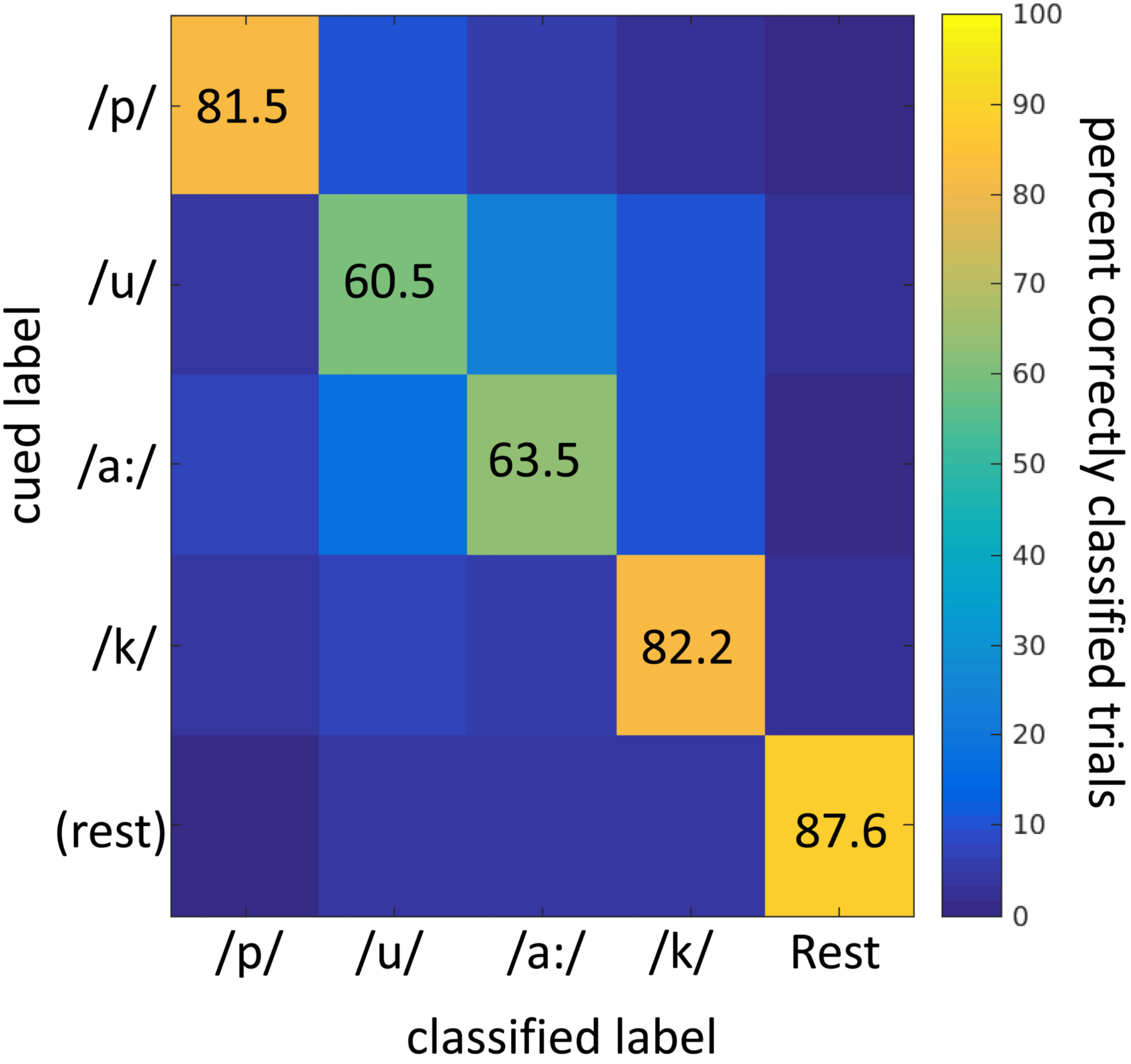

Figure 5.

STMF group-mean classification confusion matrix for 5-class classification. The y-axis indicates the presented cue, the x-axis indicates the assigned class. The scores are given as percentages of all cued trials.

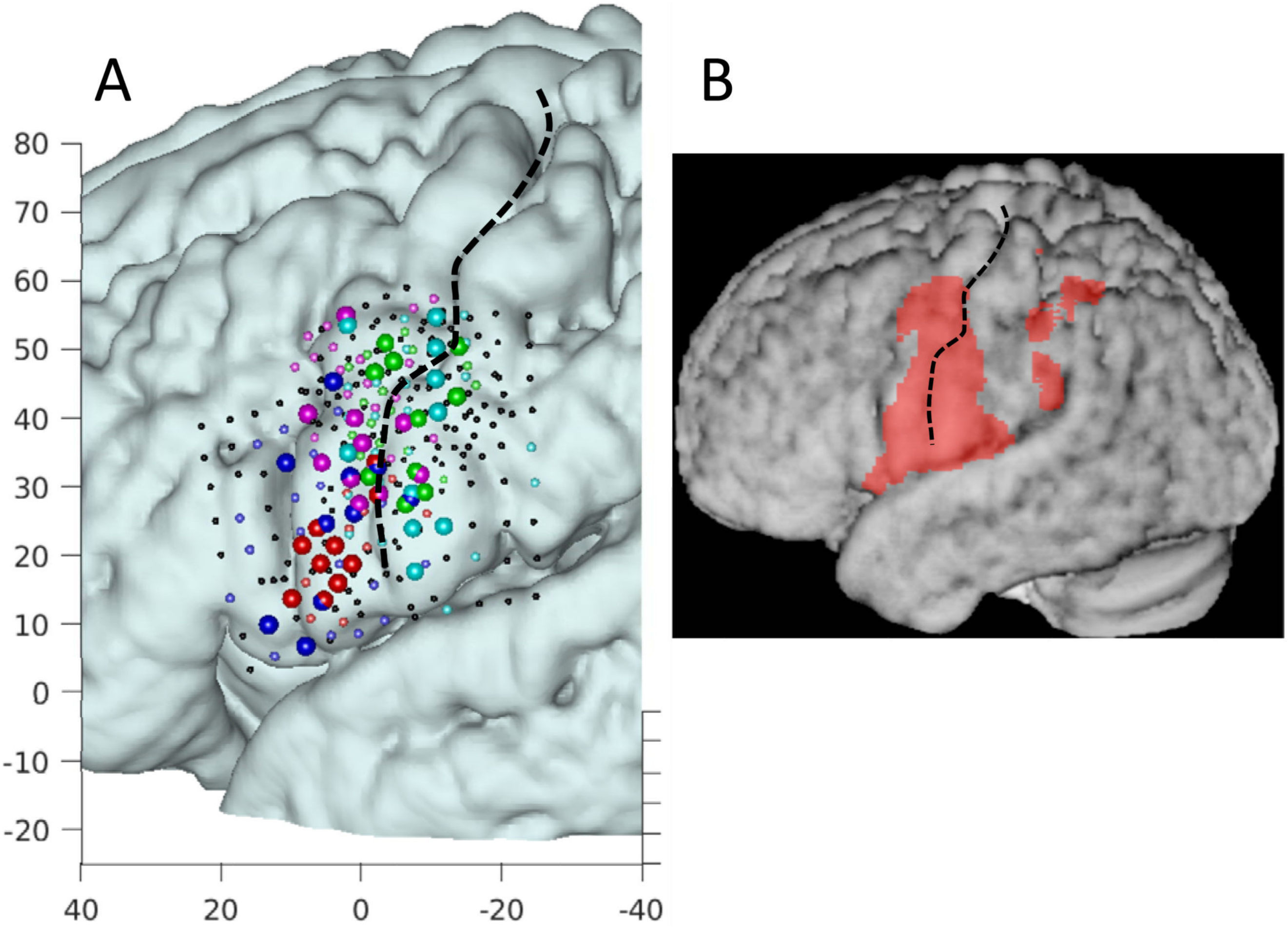

Figure 6.

A: Localizations of informative electrodes. For each patient a different color is used, with red, magenta, green, blue, and cyan corresponding to subject R1, R2, L1, L2, and L3 respectively. For display purposes the 10 most informative electrodes are shown as larger spheres and the remaining significant electrodes are show as smaller colored spheres. Axes denote MNI coordinates. B: Activity averaged across 12 healthy volunteers measured with the phoneme task with 7 Tesla fMRI (see Methods). Electrodes and fMRI activity are displayed in MNI space, projected onto an average anatomy of 12 healthy volunteers.

Table 1

Patient and electrode characteristics

patient	Number of electrodes [number analyzed]	Electrode distance (center-to-center)	Age	Gender
R1	32 [32]	3	19	F
R2	64 [60]	3	28	M
L1	64 [60]	3	18	M
L2	64 [54]	4	20	M
L3	128 [72]	4	36	F

Table 2

Classification results. All individual classification results for each subject and method (spatiotemporal matched filters STMF, spatial matched filters SMF, support vector machine SVM). Classification was performed for 2, 4 and 5 classes (phonemes versus rest, 4 phonemes only and 4 phonemes plus rest, respectively). In straight brackets the computed chance levels are given (upper 95% confidence bound of the distribution of permutations as explained in the Methods). For 4- and 5-class classification the classification scores for three 0.5 s windows are given for STMF and SMF (-0.5 − 0, - 0.25 - +0.25, 0 − 0.5). In addition, classification for the 0.5s window centered around VOT is also given for SMF.

	R1 **%correct** **[chance]**	**R2** **%correct** **[chance]**	**L1** **%correct** **[chance]**	**L2** **%correct** **[chance]**	**L3** **%correct** **[chance]**	**Mean (std)** **[chance]**
2-class: All phonemes versus rest						
STMF	100 [36.5]	81.3 [18.8]	100 [30.2]	90.0 [23.3]	66.7 [26.7]	87.6 (14.1) [27.1]
SMF	95.2 [42.9]	81.3 [37.5]	88.4 [41.9]	66.7 [46.7]	43.3 [44.3]	75.0 (20.6) [42.7]
SVM	98.4 [33.3]	93.8 [31.3]	95.3 [33.3]	90.0 [32.6]	86.7 [32.2]	92.8 (4.6) [32.7]
4-class: Phonemes only						
STMF	77.0 [24.7]	81.4 [32.9]	62.1 [28.1]	63.0 [30.3]	76.1 [31.0]	71.9 (8.8) [29.4]
SMF	78.7 [23.6]	65.7 [31.4]	58.2 [26.8]	39.5 [26.9]	59.3 [28.3]	60.3 (14.2) [27.4]
SVM	77.0 [24.1]	74.3 [28.6]	75.2 [26.9]	52.1 [24.2]	85.8 [26.2]	72.9 (12.5) [26.0]
STMF -0.5 to 0s	68.4 [25.3]	61.4 [31.4]	54.2 [26.8]	46.2 [30.3]	61.1 [30.1]	58.3 (8.4) [28.8]
STMF -0.25 to 0.25s	77.6 [24.1]	71.4 [32.9]	71.9 [26.8]	58.8 [29.4]	82.3 [30.1]	72.4 (8.8) [28.7]
STMF 0 to 0.5s	80.5 [25.9]	71.4 [32.9]	55.6 [28.1]	54.6 [29.4]	64.6 [30.1]	65.3 (10.9) [29.3]
SMF -0.25 to 0.25s	76.4 [24.1]	35.7 [31.4]	47.7 [26.8]	49.6 [26.9]	69.0 [29.2]	55.7 (16.6) [27.7]
5-class: Phonemes + rest						
STMF	83.1 [25.7]	81.4 [27.9]	70.4 [25.5]	68.5 [26.2]	74.1 [26.6]	75.5 (6.5) [26.4]
SMF	83.1 [25.7]	68.6 [29.1]	64.8 [25.0]	45.0 [25.5]	55.9 [26.6]	63.5 (14.2) [26.4]
SVM	82.7 [25.3]	77.9 [27.9]	79.6 [26.2]	59.7 [24.5]	86.0 [25.1]	77.2 (10.2) [25.3]
STMF -0.5 to 0s	75.5 [26.2]	66.3 [26.7]	63.8 [25.0]	51.7 [26.2]	62.9 [26.6]	64.0 (8.5) [26.1]
STMF -0.25 to 025s	83.5 [25.7]	70.9 [27.9]	78.1 [25.5]	62.4 [26.2]	77.6 [26.6]	74.5 (8.1) 26.4]
STMF 0 to 0.5s	85.7 [26.2]	68.6 [26.7]	63.8 [25.0]	60.4 [25.5]	64.3 [26.6]	68.6 (10.0) 26.0]
SMF -0.25 to 0.25s	81.0 [24.5]	38.4 [27.9]	58.2 [25.0]	55.7 [26.8]	65.0 [27.3]	59.7 (15.4) [26.3]