Published in final edited form as:

J Neural Eng. 2018 December ; 15(6): 066025. doi:10.1088/1741-2552/aae329.

The influence of prior pronunciations on sensorimotor cortex activity patterns during vowel production

E. Salari, Z.V. Freudenburg, M.J. Vansteensel, and N.F. Ramsey

Brain Center Rudolf Magnus, Department of Neurology and Neurosurgery, University Medical Center Utrecht, Utrecht, The Netherlands

Abstract

Objective—In recent years, brain-computer interface (BCI) systems have been investigated for their potential as a communication device to assist people with severe paralysis. Decoding speech sensorimotor cortex activity is a promising avenue for the generation of BCI control signals, but is complicated by variability in neural patterns, leading to suboptimal decoding. We investigated whether neural pattern variability associated with sound pronunciation can be explained by prior pronunciations and determined to what extent prior speech affects BCI decoding accuracy.

Approach—Neural patterns in speech motor areas were evaluated with electrocorticography in five epilepsy patients, who performed a simple speech task that involved pronunciation of the /i/ sound, preceded by either silence, the /a/ sound or the /u/ sound.

Main results—The neural pattern related to the /i/ sound depends on previous sounds and is therefore associated with multiple distinct sensorimotor patterns, which is likely to reflect differences in the movements towards this sound. We also show that these patterns still contain a commonality that is distinct from the other vowel sounds (/a/ and /u/). Classification accuracies for the decoding of different sounds do increase, however, when the multiple patterns for the /i/ sound are taken into account. Simply including multiple forms of the /i/ vowel in the training set for the creation of a single /i/ model performs as well as training individual models for each /i/ variation.

Significance—Our results are of interest for the development of BCIs that aim to decode speech sounds from the sensorimotor cortex, since they argue that a multitude of cortical activity patterns associated with speech movements can be reduced to a basis set of models which reflect meaningful language units (vowels), yet it is important to account for the variety of neural patterns associated with a single sound in the training process.

Keywords

Coarticulation; sensorimotor cortex; speech; brain-computer interface; ECoG

Corresponding author: N.F. Ramsey, Phone: +31 (0)88 755 6862, Fax: +31 (0)30 254 2100, n.f.ramsey@umcutrecht.nl, Address: Heidelberglaan 100, 3508 GA, Utrecht, The Netherlands. Room G03 1.24, P.O. Box 85500.

ORCID:

E.Salari: 0000-0002-7467-3297

Z.Freudenburg: 0000-0002-2790-0020

M.Vansteensel: 0000-0002-9252-5116

N.Ramsey: 0000-0002-7136-259X

Intro

People who suffer from severe paralysis can sometimes lose the ability to speak or communicate while retaining adequate levels of cognitive functioning [1–3]. This condition is called locked-in syndrome (LIS). People with LIS often rely on assistive communication technology to interact with their family and caretakers. An example of such technology is a brain-computer interface (BCI) system [4]. BCIs record brain signals and translate specific features of the recorded signal into a control signal for a computer or other effector. As such, these systems aim to provide people with paralysis the ability to, for instance, write text on a screen or produce synthesized speech, without the involvement of any muscle activity (see for instance [5] for an example of a fully implantable BCI system for home use).

Recent years have seen an increasing interest in decoding brain activity related to speech movements for the control of BCIs. Several studies using implanted electrodes have shown, for instance, that sounds (isolated or within words) can be classified from the sensorimotor cortex (accuracy for 4 sounds: $71.9 \pm 8.8\%$ [6]; accuracy for a complete set of American English phonemes $20.4 \pm 9.8\%$ [7]). Also, when speech is not actually executed but *attempted* by someone with paralysis, classification of 38 American English phonemes was possible with an average accuracy of 21% [8]. Also in natural spoken speech, it is possible to reconstruct what people said, based on brain signals alone (word classification up to 75% in a set of 10 words [9]). However, these systems are far from being a usable application for people from the target population since the accuracy levels and degrees of freedom of speech motor BCIs remain suboptimal.

A possible explanation for the suboptimal performance of current speech motor BCIs is that most decoders do not explicitly take into account that there may be variations in the brain activity patterns related to sound pronunciation and generally base classification of each individual sound on a single activity pattern. However, for speech recognition applications that are based on sound recordings [10] or facial muscle movements [11], it has been shown that taking context (previous and future phonemes) into consideration, substantially improves speech recognition. The reason for this might be that articulator placement for a particular sound depends on the sound(s) pronounced before and/or after the spoken sound, a phenomenon called coarticulation [12,13]. Similarly, the movements the articulators make to create a particular sound are likely to depend on prior sounds, since the articulators move from different starting positions. Because articulator movement is largely controlled by the sensorimotor cortex of the brain [14–17], variations in articulator positions for the same sound, or in their movements towards a certain position, are likely to be associated with variations in sensorimotor brain activity patterns. Indeed, there is evidence for a representation of coarticulation in the sensorimotor cortex in that, for single electrodes, the high-frequency band (HFB; >50 Hz) power amplitude varies with different articulator positions for the same sound [18,19]. Also, we have previously shown that repeatedly pronouncing the same sound is not associated with equal neural activity in the sensorimotor cortex if the sounds are spaced a second or less apart, suggesting that previous sounds influence sensorimotor cortex activity [20]. It is currently unclear, however, how coarticulation and the movement from one sound to another, affect our ability to distinguish a certain sound from other sounds, using neural signals. Also, it remains to be determined if

the incorporation of multiple brain patterns for a single sound in a decoder can improve classification accuracies. Clearly, answers to these questions are of great interest for the development of motor speech BCIs since they may improve the classification performance and thereby daily life usability of such systems.

In the current study, we confirm that sensorimotor cortex neural activity patterns that are related to the pronunciation of one specific sound, and likely the movements towards that sound (the /i/ vowel in this study), depends on the preceding sound. In addition, we investigate the consequences of these different /i/-related activity patterns for the decoding of different sounds based on sensorimotor cortex neural signals. We use electrocorticography (ECoG) to record neural signals since this technique encompasses a combination of high temporal resolution and high spatial specificity [21], both particularly relevant for speech related research [6,22]. We focus on the 75-135 Hz power changes in the sensorimotor cortex, which are known to be closely related to speech and articulator movements, especially in the ventral areas of the sensorimotor cortex [15,17,23], and which have been associated with underlying neural firing [23–25].

Method

Participants

Five patients (age 14-41; 3 females) participated in this study, while they were treated for epilepsy in the University Medical Center Utrecht. Subdural ECoG electrodes were placed for clinical reasons. In two participants, the coverage of the clinical grids included the area of interest, the ventral sensorimotor cortex (vSMC; left hemisphere in one participant, right in the other participant). The other three participants (all adults) gave informed consent for the placement of an extra high-density (HD) electrode grid over this area for research purposes only (left hemisphere in 2 participants, right in 1 participant). Electrodes of the clinical grids were 2.3 mm in diameter and were spaced 10 mm apart. HD-grid electrodes were 1 or 1.17 mm in diameter and spaced 3 or 4 mm apart. For the participants with HD-grid coverage we only used the HD electrodes for the current analysis.

This study is in accordance with the Declaration of Helsinki (2013) and was approved by the ethics committee of the University Medical Center Utrecht. All participants gave written informed consent before participation.

Task

Participants were visually cued to produce one of two vowels (the /u/ or /a/), or to remain silent, for 1000 ms, and then make a smooth transition from the current sound (or silence) into a new sound (the /i/ vowel), without a break between the two sounds, and hold the /i/ for 1000 ms. We chose these vowels since they are phonetically well described and are known to involve different articulator placements [26]. Visual cues were presented on a computer screen, which was placed approximately 1 m from the subject. A trial started with the visual presentation of either the letters 'oe' or 'aa', which, in Dutch, correspond to the vowels /u/ and /a/, respectively, or with a '-' sign, which indicated to remain silent. This was followed by the presentation of the letters 'ie', corresponding to the /i/ sound. Subjects were

instructed to produce the sound that was presented on the screen for as long as it was visible. Trials were interleaved with inter-trial intervals of 3000 ms, during which a fixation cross was presented on the screen. Conditions (/u/ followed by /i/, /a/ followed by /i/, and silence followed by /i/) were randomly repeated 30 times, divided over two sessions (except for subject B, who completed only 1 session and therefore performed 15 trials per condition). We also included 20 rest trials of 2000 ms (10 for subject B), interleaved with the other trials, in which the subjects remained silent.

Data Acquisition

The recorded sampling frequency of the ECoG data was either 512 Hz, 2048 Hz (both Micromed, Treviso, Italy), or 2000Hz (Blackrock Microsystems LLC, Salt Lake City, USA), similar as in [27]. For the subjects who performed two runs, we concatenated the data of the two runs. Audio was recorded by microphones in the patient's room and synchronized with the brain signals. Unless mentioned otherwise, data analysis was performed using Matlab software (The Mathworks, Inc., Natick, MA, USA).

Data preprocessing

Electrodes over the sensorimotor cortex were identified by visual inspection of a 3D rendering of the subjects preoperative MRI scan that was coregistered with a post-operative CT scan [28,29]. Electrodes with a flat or noisy signal, or with signal that was considered epileptic by a neurologist, were removed from further analysis (see [27] for details). For the remaining electrodes ($n=25, 18, 36, 114$ and 122 for subjects A-E, respectively), line noise removal and common-average re-referencing was applied. All mention of electrodes below pertain to only the remaining electrodes. Subsequently, for each of the electrodes the HFB (75-135 Hz) power trace was calculated with a Gabor wavelet [30], similar as in [27], and this signal was subsequently smoothed for noise reduction with a moving average window of 0.1 seconds, centered around each sample. We choose this window since it is within an optimal range for phoneme classification [31], while at the same time preserving neural dynamics [20]. We normalized the signals over runs by taking, per run, the average HFB amplitude (over all electrode time series), subtracting this from all electrode time series and dividing the result by the standard deviation. The voice on- and offsets were first automatically determined from the acoustic signal using an automated vowel detection algorithm [32] and, where necessary, corrected by auditory and visual inspection using PRAAT software [33]. In addition, we determined the voice transition from one sound to another by visual and auditory inspection of the acoustic signal.

Electrode selection—For every sensorimotor electrode, we determined whether it was significantly related to the pronunciation of one or more of the sounds, by computing per sound production (the three different /i/ productions and the /u/ and /a/ production) the mean HFB response amplitude in a window of 0.3 seconds around sound onset. The HFB response amplitude during sound production was contrasted with the mean HFB amplitude during the rest trials. If the HFB response of an electrode was significantly correlated to the pronunciation of one or more of the sounds ($\alpha=0.05$, false discovery rate corrected for the number of retained electrodes), it was included in the subsequent analysis.

Determining peak timing—It is known that the HFB response peak related to the pronunciation of a sound does not always occur exactly at voice onset and can be different for different locations in the sensorimotor cortex [15,17,34,35]. Yet, for the current study, we chose to analyze (and classify, as described below) the same point in time for all electrodes, and initially we analyzed the neural activity pattern for each sound only at one point in time (for other points in time see below). That is, one time-point for the first sound and one time-point for the second sound. These were the time points where the HFB amplitude was the highest in the average signal of all included electrodes. This was done because for fast decoding of sounds in (future) real-time BCI applications, each incoming time point (or small window) needs to be classified and the easiest way to do this is to compare each incoming time point's brain activity pattern to a single time window for each of the to be classified states. Thus, for the current analysis we determined the time point of maximal HFB activity related to the pronunciation of the first sound and the second sound, in the average signal of all included electrodes. This was done by epoching, for all included electrodes, the active trials twice: once aligned to the first sound (/u/ or /a/) onset, and once to the transition to the /i/ sound. We then took the average activity over all electrodes and trials and determined the timing of the mean HFB response peak closest to the sound onset and the mean HFB peak closest to the sound transition, respectively. Since we know from previous research [20] that repeated isolated sounds are accompanied by a peak neural response on or just after voice onset, we assume that the same should be true for voice transitions.

Sensorimotor cortex activity pattern analysis—We determined for all trials (aligned to the sound transition) and per electrode, the HFB amplitude related to the /i/ pronunciation, by computing per included electrode and for each trial, the average HFB amplitude in a window of 0.3 seconds centered around the previously determined HFB response peak timing. Then, we determined whether there was a significant difference in the HFB spatial patterns between the three different conditions (/i/ preceded by silence, /i/ preceded by /u/ or /i/ preceded by /a/). We reasoned that if the HFB patterns for the different /i/ conditions would be significantly different, it should be feasible to classify which (if any) sound came before the /i/, based on the HFB pattern related to the /i/ sound. If the HFB pattern for /i/ would be unaffected by prior pronunciations, however, classification results of the three /i/ conditions should be at chance level. Therefore, we classified for each /i/ pronunciation the preceding sound (or silence), based on the HFB amplitude pattern associated with the /i/ sound, using a leave-one-trial out, template matching classification procedure. We calculated for each trial the correlation with three templates (one for each /i/ condition) and assigned each trial to the template (class) with the highest correlation. We chose correlation as a distance measure since it has previously been used to successfully classify different sounds and hand gestures based on sensorimotor cortex neural activity [6,36,37]. Templates of the HFB spatial pattern were made for each /i/ condition which was done by calculating the average (per electrode) HFB activity of all trials per class (/i/ condition), excluding the test trial. The averaged activities of all electrodes formed the spatial pattern template. Note that the template of the test trial's condition was based on one trial less than the other templates since for that condition, by consequence, there was one trial less available for training. However, we chose this leave-one-trial out classification procedure to maximize the number

of training trials, given the relatively small number of trials in total. To determine chance level and classification accuracy significance, we performed a Monte Carlo procedure in which we shuffled the labels and repeated the classification a thousand times. Chance level is the mean of the Monte Carlo accuracies and significance level (or p-value) was determined by counting the number of Monte Carlo accuracies higher than the actual accuracy, divided by the number of iterations (1000).

Anatomical localization of differences—We investigated which electrodes showed differences in HFB amplitude between the three /i/ conditions, to see specifically which sensorimotor cortex areas show differences in HFB amplitude and whether or not there was an anatomical organization in the effects. For that, we used, per electrode, an ANOVA ($\alpha = 0.05$, FDR corrected), in which the dependent variable was the HFB amplitude associated with the pronunciation of an /i/ sound, and the independent variable was the experimental condition.

Classification of different sounds—Next, we investigated if the /i/ sound patterns could be distinguished reliably from the patterns of the other two sounds (i.e. /a/ and /u/). In other words, we investigated whether or not the (different) patterns for the /i/ sound contained a unique signature and were more similar to each other than to the patterns of other sounds. To that purpose, we calculated the HFB amplitude patterns for the other two sounds in a similar way as explained for the /i/ sound, but now with trials aligned to the voice onset instead of sound transition. Subsequently we performed a leave-one trial-out, template matching classification of the three different sounds (/i/, /a/, and /u/) based on the HFB amplitude patterns. If the (different) patterns for the three /i/ conditions contained a signature that is distinct from the /u/ and /a/ sound, all /i/ productions (no matter what sound came before it) should be classifiable using a template based on the data of only one of the /i/-combinations (for instance only on the isolated /i/ productions). However, this may not necessarily be the most optimal method for classification since the variation in patterns for the same sound is then not taken into account. Therefore, an individual template for each /i/ combination might be more optimal. Considering the number of sound combinations in natural speech, a separate template for each sound combination may, however, not be practical for a BCI. It is therefore interesting to see if a single template, derived from the average of sound combinations, can be used for accurate classification, as this takes into account the variations in patterns, yet only uses one template per sound. To investigate these questions, we used three different methods to create the template(s) for the /i/ sound patterns. In the first method (the ‘Isolated /i/ method’), we used only the trials of the isolated /i/ (i.e. the one preceded by silence) to create the /i/ template. For the creation of the other two templates (for /u/ and /a/), the data during the production of the /u/ and the /a/ were used. Note that in this case, even though the template for the /i/ sound was based only on isolated /i/ productions, all /i/-combinations (/i/ preceded by silence, by /u/ or by /a/) were included in the classification of the three different sound productions (/i/, /u/ and /a/), meaning that the number of /i/ productions was three times larger than that of each of the other sounds. A correction for this difference in number of trials was applied, as described below. In the second method (the ‘Multiple /i/ method’), we used five different templates in total: three different templates for the /i/ sound (one for each of the conditions) and one for

each of the other two sounds. In this case, if the test trial had the highest correlation with one of the three /i/ templates, the test trial was classified as belonging to the /i/ sound. In the third method (the 'Mean /i/ method'), we took the average /i/ sound pattern across all three /i/ conditions as the /i/ template. By using these different methods for creating the /i/ template, we could determine if and how multiple activity patterns for one sound need to be incorporated in a BCI.

We compared the classification performance using the three different methods for creating the /i/ template with three different T-tests (comparing method 1 vs 2, 1 vs 3 and 2 vs 3, $\alpha=0.05$, FDR corrected). Since the /i/ sound has three times the number of productions as the other two sounds, we corrected the classification accuracy for the different number of productions per sound. This was done by taking first the accuracy of every class (the number of correctly classified trials of one class divided by the total number of trials within that class) and then taking the average of that as the overall accuracy (as opposed to taking the total number of correctly classified trials divided by the total number of trials). We determined the chance level with a Monte Carlo procedure, shuffling the labels 1000 times.

Finally, to see if all patterns (the three different /i/ patterns and the patterns for the /u/ and /a/) could be differentiated from each other, we also performed a classification of all five patterns, see supplementary material Figure S1.

Classification of the different conditions over time—In the above analyses, sound templates were based on the moment of peak neural activity, which is, likely, closely related to movements (articulator transitions), rather than to steady state pronunciation. To get some insights, however, in how long after the transition the different /i/ condition neural activity patterns stayed different we performed a classification over time between the three conditions while at the same time moving the templates in time (from 2 seconds before transition cue to 2 seconds after, in steps of 0.05 seconds). This analysis enabled us to see if the within-sound classification (of the three /i/ conditions) was based mostly on the transition phase or was also possible at later points in time (e.g. the steady state phase). To be able to compare results over trials, we epoched all trials, and corrected differences in voice onset, transition and offset in the brain recordings by down sampling and interpolation, similar as in [27], effectively aligning all voice onsets, transitions and offsets to their respective mean. The significance level for classification was set to the 95% confidence interval on the binominal distribution above chance level.

Results

Task performance and electrode selection

Participants performed the task adequately. Only one trial for one subject (C) was excluded due to a missed response. On average, subjects produced the /u/ for 0.64 (SD=0.13) seconds, the /a/ for 0.65 (SD=0.13) seconds and the /i/ for 0.67 (SD=0.18) seconds. Three subjects showed a significant difference between the durations of the different sounds (A: $F(2,147)=8.26$, $p<0.001$, C: $F(2,145)=27.65$, $p<0.001$ & E: $F(2,147)=29.85$, $p<0.001$). For the three different /i/ conditions, subjects produced the /i/ preceded with silence for 0.75 (SD=0.19) seconds, preceded by /u/ for 0.62 (SD=0.13) seconds and preceded by /a/ for 0.64

(SD=0.13) seconds. Two subjects showed a significant difference between the speech durations of the /i/ conditions (B: $F(2,42)=29.49$, $p<0.001$ & E: $F(2,87)=14.88$, $p<0.001$).

For subjects A-E respectively, 72% (18/25), 61% (11/18), 47% (17/36), 48% (55/114) and 89% (108/122) of the electrodes over the sensorimotor cortex showed a significant correlation with the task. Only these were used for subsequent analyses.

Classification of the different /i/ combinations

Figure 1 shows that for most subjects, there are clearly two peaks of activity (one per sound) and it also suggests that there are on average (across electrodes) differences between conditions. These differences are not equal over subjects though. Classification of the HFB patterns of the three /i/ sound conditions showed that, indeed, for all participants there were significant differences between these HFB patterns (Table 1 and Figure 1), indicating that prior sound production influences the brain activity pattern for the /i/ sound.

Investigation of whether or not differences between conditions occurred in all electrodes revealed that, of all electrodes, 22% (4/18), 27% (3/11), 29% (5/17), 45% (25/55) and 13% (14/108) showed a significant difference in HFB amplitudes between the three different /i/ production conditions (Figure 2). Notably, these electrodes appeared to be localized in clusters, which is especially clear in the HD-grid participants.

Classification of /a/, /u/ and /i/ sound production

Investigating the differences of the /i/ sound pattern(s) with that of the other sounds revealed that all subjects show significant classification accuracies for all three template methods (Figure 3 and 4 and Table 2). In other words, despite the fact that /i/ production is associated with different HFB patterns that depend on prior utterance, the /i/ sound can be distinguished reliably from the /a/ and the /u/, indicating that the different neural patterns for the /i/ sound are more similar to each other than to the other sounds. However, for the 'Isolated /i/ method', significantly lower classification accuracies were obtained ($M=63.05\%$, $SD=11.17\%$) than for the 'Multiple /i/ method' ($M=77.82\%$, $SD=13.19\%$, $T=-5.02$, $p=0.01$), or for the 'Mean /i/ method' ($M=76.64\%$, $SD=11.41\%$, $T=-6.33$, $p<0.001$). There was no significant difference in classification accuracy between the latter two methods (Figure 4).

Classification at different points in time

To see if the classification of the three /i/ conditions was based mostly on the transition phase or was also possible at later points in time (e.g. the steady state phase) we did a classification over time between the three conditions using different time points to generate the templates (as opposed to using the peak neural activity). Results showed that, using this approach, the classification accuracy of the three /i/ conditions was highest at the neural peak activity (which was usually very close to the voice transition) and that it dropped rapidly after the peak, suggesting that the difference of neural activity between the /i/ conditions is mostly the result of differences in the movement of the articulators toward the sound (Figure 5). However, the different /i/ conditions stayed significantly separable between 0.28 - 1.01 seconds after the transition.

Discussion

In this study, using a simple speech task we confirmed that sensorimotor neural signals related to sound production are affected by previous sound productions.

Electrocorticography measurements in five epilepsy patients revealed that the HFB amplitude pattern in the sensorimotor cortex associated with production of a certain sound may differ, depending on the presence and nature of a preceding sound production. These data indicate that production of the same sound is associated with multiple HFB amplitude patterns in the (ventral) sensorimotor cortex, likely due to differences in the movements towards a sound, which is of direct relevance for the development of speech BCI. This phenomenon was not only observed in the averaged signal of all electrodes, but also at single electrode level. The anatomical localization of electrodes that showed a significant difference in HFB amplitude during the different productions of the /i/ sound seemed to be somewhat clustered for most subjects, but given the limited number of participants, it cannot be concluded whether or not there is a typical anatomical distribution for the effects we found. Importantly, the differences in HFB amplitude between the three different /i/ conditions were most apparent just after voice transition (close to the moment of peak neural activity) and classification accuracy decreased rapidly after the transition. This suggests that the difference in neural activity between the /i/ conditions is mostly due to the articulator movements toward the sound, and that, once the articulators are in place, pronunciation of the /i/ sound is less affected by prior utterances. For most subjects, however, the different /i/ conditions stayed significantly distinguishable for 0.28 – 1.01 seconds after the transition, suggesting that there might be some differences in the neural activity that are not only due to the transition.

Interestingly, despite the different response patterns associated with /i/ pronunciation, we found that the different sounds (/i/, /u/, /a/) can be classified accurately from one another, irrespective of whether the /i/ was preceded by another sound or not. This means that the representation of /i/ production in the sensorimotor cortex contains a distinct signature that can be distinguished from the representation of /a/ and /u/ production, and would seem to be independent of preceding sound production. However, we also show that classification accuracies for the different sounds increase significantly when the multiple HFB amplitude patterns for the /i/ sound are taken into account. These results suggest that, for the purpose of speech BCI, the training of a classifier benefits from taking into account the differences in sound representation that are caused by preceding speech production, since this will lead to higher classification accuracies.

Neural underpinnings of multiple HFB amplitude patterns

The different HFB amplitude patterns related to the production of a single vowel may be attributed to coarticulation. Coarticulation can occur in two directions: the pronunciation of a previous sound can influence the articulator position of an upcoming sound (carryover) or vice versa (anticipatory) [12,13]. Literature from linguistic experiments have shown for instance that the tongue position during the pronunciation of the /k/ in the word 'seek' is higher than in the word 'arc', since the /i/ is a fronted vowel (with the tongue front high) and the /a/ has a low tongue position and the tongue position during the vowel extends to the

upcoming sound(s) (example taken from [13]). An example of anticipatory coarticulation can be seen during the pronunciation of the words 'ski' and 'cool' in which the /k/ is more fronted (tongue front is high) in 'ski' due to an anticipation of the fronted vowel /i/ and less fronted in 'cool', in anticipation of the /u/ which has a more backward tongue position [13]. Recent investigations have shown that articulator position differences due to coarticulation can lead to different levels of activity for the same sound in some parts of the sensorimotor cortex [18,19]. Here we extend those results by showing where these differences may be anatomically localized for one sound and how these effects can influence decoding accuracy for speech BCI. In addition, we show that, even though coarticulation effects are present in the sensorimotor cortex, the different patterns for one particular sound have something in common that distinguishes them from the patterns of other sounds.

Another explanation for the current results could be the fact that the movements of the articulators to a particular sound differ depending on the prior sound. For instance, it is known for the Dutch language that the tongue position when producing an /a/ is mostly low (somewhat similar as the rest position), while for the /u/ sound the back of the tongue moves up the most and for the /i/ sound the front of the tongue moves up the most [35]. This means that the movement from an /a/ to an /i/ sound may be more similar to producing solely the /i/ sound, than to that of producing an /u/ sound before the /i/ sound. However, even though these movement differences are likely to be quite similar across subjects, there was no convincing evidence in our study for a consistently higher similarity of some of the brain activity patterns (e.g. those of the isolated /i/ and the /i/ with the /a/ before it), compared to others (e.g. the /i/ with the /u/ before it). Differences in electrode positions and coverage over subjects, as well as the sparse spatial sampling of ECoG, may have influenced the current results, however, since only when similar parts of the cortex are recorded it is possible to compare patterns over subjects.

Implications for BCI

The current data contribute to a better understanding of the activity properties of the sensorimotor cortex and this may be of interest for the development of BCIs that aim to decode speech sounds based on sensorimotor cortex activity. We show that multiple sensorimotor cortex HFB amplitude patterns exist for pronunciation of a single sound, which are likely related to the movements toward that sound, and that classification of different sounds can benefit from taking these different patterns into consideration. This does not mean however, that one should employ separate templates for all combinations of sounds. Our results indicate that an average template of the different patterns will suffice for accurate classification. Previous studies that have trained a classifier for phoneme classification not based on isolated phonemes but on phonemes within a word or sentence (e.g. [7,9]), implicitly took some context for sound production in consideration. Our results suggest that explicit and careful consideration of different contexts is essential for accurate sound classification.

Limitations and future work

A higher number of participants and a more extensive electrode coverage would have strengthened the current results. In addition, in the current experiment we tested only a

limited number of sound combinations. Further research is needed to investigate the effects of prior sound production on other phonemes than the /i/ and the influence of this on BCI classification accuracy. Another limitation is that, although we tried to stay as constant as possible with respect to marking the transitions, finding the exact transition onset is not always clear and might therefore have caused some variability in transition onset marking. Finally, as mentioned above, coarticulation effects can also **occur** in an anticipatory fashion, meaning that the sounds that follow a particular sound influence the articulator position of the prior sound. More research on the consequences of such an anticipatory coarticulation on BCI classification seems warranted.

Conclusion

We demonstrate that the production of one sound is accompanied with multiple sensorimotor cortex activity patterns. This is of importance for BCIs that aim to decode speech from the sensorimotor cortex since we show that multiple movements related to the same sound can be classified into a single vowel sound class. Classification is, however, further improved when the multiple patterns for one sound are taken into account by the classifier (either by training a classifier for the separate patterns of the same sound or by training on the averaging of the multiple patterns for one sound).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported by the European Union (ERC-Advanced ‘iConnect’ grant 320708). All authors and co-authors declare that they have no conflict(s) of interest. Our thanks go out to the participants for their contribution and the clinical neurophysiology department staff and the neurosurgeons for their collaboration and cooperation.

References

- [1]. American Congress of Rehabilitation Medicine. Recommendations for use of uniform nomenclature pertinent to patients with severe alterations in consciousness. *Arch Phys Med Rehabil.* 1995; 76:205–9. [PubMed: 7848080]
- [2]. Smith E, Delargy M. Locked-in syndrome. *BMJ.* 2005; 330:406–9. [PubMed: 15718541]
- [3]. Posner JB, Plum F, Saper CB, Schiff N. Plum and Posner’s Diagnosis of Stupor and Coma. Oxford University Press; USA: 2007.
- [4]. Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain-computer interfaces for communication and control. *Clin Neurophysiol.* 2002; 113:767–91. [PubMed: 12048038]
- [5]. 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; 375:2060–6. [PubMed: 27959736]
- [6]. Ramsey NF, Salari E, Aarnoutse EJ, Vansteensel MJ, Bleichner MB, Freudenburg ZV. Decoding spoken phonemes from sensorimotor cortex with high-density ECoG grids. *NeuroImage.* 2017; 180:301–11. [PubMed: 28993231]
- [7]. 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; 11 035015.

- [8]. 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. [PubMed: 21629876]
- [9]. Herff C, Heger D, de Pestiers A, Telaar D, Brunner P, Schalk G, Schultz T. Brain-to-text: decoding spoken phrases from phone representations in the brain. *Neural Technol.* 2015; 9:217.
- [10]. Lee K. Context-independent phonetic hidden Markov models for speaker-independent continuous speech recognition. *IEEE Trans Acoust Speech Signal Process.* 1990; 38:599–609.
- [11]. Schultz T, Wand M. Modeling coarticulation in EMG-based continuous speech recognition. *Speech Commun.* 2010; 52:341–53.
- [12]. Rietveld A, van Heuven V. *Algemene Fonetiek*. Bussum: Coutinho; 2001.
- [13]. Davis J. *Phonetics: fundamentals, potential applications and role in communicative disorders*. New York: Nova Publishers; 2015.
- [14]. Penfield W, Boldrey E. Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain J Neurol.* 1937; 60:389–443.
- [15]. 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.* 1998; 121:2301–15. [PubMed: 9874481]
- [16]. Towle VL, Yoon H-A, Castelle M, Edgar JC, Biassou NM, Frim DM, Spire J-P, Kohrman MH. ECoG gamma activity during a language task: differentiating expressive and receptive speech areas. *Brain.* 2008; 131:2013–27. [PubMed: 18669510]
- [17]. Bouchard KE, Mesgarani N, Johnson K, Chang EF. Functional organization of human sensorimotor cortex for speech articulation. *Nature.* 2013; 495:327–32. [PubMed: 23426266]
- [18]. Chartier J, Anumanchipalli GK, Johnson K, Chang EF. Encoding of Articulatory Kinematic Trajectories in Human Speech Sensorimotor Cortex. *Neuron.* 2018; 98:1042–54. [PubMed: 29779940]
- [19]. Bouchard KE, Chang EF. Control of Spoken Vowel Acoustics and the Influence of Phonetic Context in Human Speech Sensorimotor Cortex. *J Neurosci.* 2014; 34:12662–77. [PubMed: 25232105]
- [20]. Salari E, Freudenburg ZV, Vansteensel MJ, Ramsey NF. Repeated Vowel Production Affects Features Of Neural Activity In Sensorimotor Cortex. *Brain Topogr.* 2018
- [21]. Siero JC, Hermes D, Hoogduin H, Luijten PR, Ramsey NF, Petridou N. BOLD matches neuronal activity at the mm scale: A combined 7 T fMRI and ECoG study in human sensorimotor cortex. *NeuroImage.* 2014; 101:177–84. [PubMed: 25026157]
- [22]. 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. [PubMed: 26138146]
- [23]. Miller KJ, Leuthardt EC, Schalk G, Rao RPN, Anderson NR, Moran DW, Miller JW, Ojemann JG. Spectral Changes in Cortical Surface Potentials during Motor Movement. *J Neurosci.* 2007; 27:2424–32. [PubMed: 17329441]
- [24]. 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.
- [25]. Ray S, Maunsell JHR. Different Origins of Gamma Rhythm and High-Gamma Activity in Macaque Visual Cortex. *PLOS Biol.* 2011; 9:e1000610.
- [26]. Booij G. *The Phonology of Dutch*. Clarendon Press; 1999.
- [27]. Salari E, Freudenburg ZV, Vansteensel MJ, Ramsey NF. Spatial-Temporal Dynamics of the Sensorimotor Cortex: Sustained and Transient Activity. *IEEE Trans Neural Syst Rehabil Eng.* 2018; 26:1084–92. [PubMed: 29752244]
- [28]. Branco MP, Gaglianese A, Glen DR, Hermes D, Saad ZS, Petridou N, Ramsey NF. ALICE: A tool for automatic localization of intra-cranial electrodes for clinical and high-density grids. *J Neurosci Methods.* 2018; 301:43–51. [PubMed: 29100838]
- [29]. Hermes D, Miller KJ, Noordmans HJ, Vansteensel MJ, Ramsey NF. Automated electrocorticographic electrode localization on individually rendered brain surfaces. *J Neurosci Methods.* 2010; 185:293–8. [PubMed: 19836416]

- [30]. Bruns A. Fourier-, Hilbert- and wavelet-based signal analysis: are they really different approaches? *J Neurosci Methods*. 2004; 137:321–32. [PubMed: 15262077]
- [31]. Branco MP, Freudenburg ZV, Aarnoutse EJ, Vansteensel MJ, Ramsey NF. Optimization of sampling rate and smoothing improves classification of high frequency power in electrocorticographic brain signals. *Biomed Phys Eng Express*. 2018; 4 045012.
- [32]. Hermes DJ. Vowel-onset detection. *J Acoust Soc Am*. 1990; 87:866–73. [PubMed: 2307780]
- [33]. Boersma P. Praat, a system for doing phonetics by computer. *Glott Int*. 2002; 5
- [34]. Hermes D, Siero JCW, Aarnoutse EJ, Leijten FSS, Petridou N, Ramsey NF. Dissociation between Neuronal Activity in Sensorimotor Cortex and Hand Movement Revealed as a Function of Movement Rate. *J Neurosci*. 2012; 32:9736–44. [PubMed: 22787059]
- [35]. Coudé G, Ferrari PF, Rodà F, Maranesi M, Borelli E, Veroni V, Monti F, Rozzi S, Fogassi L. Neurons Controlling Voluntary Vocalization in the Macaque Ventral Premotor Cortex. *PLOS ONE*. 2011; 6 e26822.
- [36]. Branco MP, Freudenburg ZV, Aarnoutse EJ, Bleichner MG, Vansteensel MJ, Ramsey NF. Decoding hand gestures from primary somatosensory cortex using high-density ECoG. *NeuroImage*. 2017; 147:130–42. [PubMed: 27926827]
- [37]. 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; 221:203–16. [PubMed: 25273279]

Figure 1. Confusion matrices and average response for the different /i/ productions.

For each participant (A-E) the confusion matrix for classification of the different /i/ HFB patterns is shown on the left. The numbers indicate the percentage (%) of trials within the class (indicated by the rows) that were classified as the class indicated by the columns. On the right, the average HFB response (over electrodes and trials) is shown for each participant and each condition, with time on the x-axis and the HFB-power (in arbitrary units) on the y-axis. Voice onset and voice transition are shown by vertical red lines and the corresponding response peak times are shown by vertical black lines. Trials were aligned to voice transition

and the voice onset standard deviation is indicated with a red horizontal bar. Note that since all trials are aligned to voice transition, there is no standard deviation for this. The sound before the /i/ is indicated without slashes to highlight the fact that classification specifically focused on the /i/ sound here.

Figure 2. Sensorimotor cortex electrodes that show differences in HFB amplitude between the different /i/ sound conditions.

The brain surface rendering of each subject (A-E) is shown. The sensorimotor cortex is highlighted in white and the central sulcus in red. For each electrode, an ANOVA was performed to test for differences between the HFB amplitudes during the different /i/ productions and the inverted p-values (1-p) thereof are shown here by the color shading of each electrode. Lighter colors represent statistically stronger differences between different /i/ sound conditions.

Figure 3. Average confusion matrices for classification of the three different sounds using the different /i/ template methods.

The average (over subjects) confusion matrices are shown for each of the three different methods used to create /i/ templates. The numbers indicate the percentage (%) of trials within the class that is indicated by the rows that were classified as the class indicated by the columns.

Figure 4. Average and individual classification accuracies for classification of the different sounds using the three different /i/ template methods.

In panel (a) the average (over subjects) classification accuracies are shown for each of the three different /i/ template methods. In panel (b) the same is shown for each individual separately. In panel (a), an asterisk indicates a significant difference in accuracy between methods. The horizontal red line indicates the average chance level (over subjects and methods).

Figure 5. The classification accuracy between the three conditions over time, using templates generated at moving time points.

The classification accuracy (y-axis) is shown at each time point (x-axis) between the three different conditions (/i/, /u/-/i/ & /a/-/i/), for each subject. Time point 0 is 2 seconds before transition cue onset. Templates were based on the same point in time as the time point of classification (rather than only at the peak neural activity, as used earlier). Voice onset, transition and offset are indicated by vertical red lines. The point at which originally, the /i/ templates were based (the peak in neural activity) is indicated by a dashed black line. Note, that for most participants the most separable moment was around the peak of neural activity.

For subject E, the peak of neural activity was rather wide (see figure 1) which may explain the time difference between the accuracy peak and the neural response peak. Chance level (33.33%) is indicated by a thin horizontal line and the thick horizontal line indicates the accuracy level at which classification is significantly above chance level. Note that, at the onset of a trial, classification should be at chance level, since at this moment all three templates were based on rest and should not be able to predict the future, upcoming, pronunciation. Between voice onset and voice transition, the classification is between rest, the /u/ and the /a/ sound. Between voice transition and voice offset the classification is between the three different /i/ conditions (the isolated /i/, the /i/ with an /u/ before and the /i/ with an /a/ before), basically predicting if and what sound was said in the past (just before the /i/ sound).

Table 1
Classification of the different /i/ patterns.

For all subjects (A-E), the classification accuracy, chance level and p-value (as determined by a Monte Carlo procedure) are shown, for the classification of the different /i/ productions. Significant classification of these patterns indicates that there are overall differences between HFB patterns associated with /i/ sound productions, indicating these patterns depend on the previous sound.

Participant	Accuracy (%)	Chance level (%)	p
A	48.89	33.11	0.018
B	64.44	33.12	0.003
C	58.20	33.14	0.001
D	83.33	33.08	<0.001
E	75.56	33.23	<0.001

Table 2
Statistics for the different sound patterns classification, using three different /i/ template methods.

For all subjects (A-E) the classification accuracy is indicated for the classifications of the different sounds per /i/ template method. Also, the average accuracy (over subjects) is shown. Chance level, as determined by a Monte Carlo procedure is shown between brackets. Asterisks indicate a classification accuracy significantly above chance (* $p < 0.01$, ** $p < 0.001$).

Participant	Isolated /i/	Multiple /i/	Mean /i/
A	50.74% (33.14%)*	61.85% (33.31%)**	64.44% (33.38%)**
B	60.00% (33.05%)*	74.81% (33.45%)**	71.11% (33.22%)**
C	63.74% (33.19%)**	71.34% (33.40%)**	72.47% (34.53%)**
D	81.11% (33.15%)**	96.30% (33.23%)**	94.07% (33.72%)**
E	59.63% (33.29%)**	84.81% (33.15%)**	81.11% (32.97%)**
Mean	63.05% (33.16%)	77.82% (33.31%)	76.64% (33.57%)