Temporal dynamics of the speech readiness potential, and its use in a neural decoder of speech-motor intention.

Jonathan S. Brumberg\textsuperscript{1,3}, Nichol Castro\textsuperscript{2}, Akshatha Rao\textsuperscript{3}

\textsuperscript{1}Department of Speech-Language-Hearing, University of Kansas, USA
\textsuperscript{2}Department of Psychology, University of Kansas, USA
\textsuperscript{3}Department of Electrical Engineering and Computer Science, University of Kansas, USA

brumberg@ku.edu, ncastro@ku.edu, akshatharao@ku.edu

Abstract
We investigated the temporal dynamics of brain activity related to speech and motor preparation. Previous electroencephalography (EEG) studies of speech production have identified a slow wave negativity that occurs as early as 2 seconds prior to articulation, known as the readiness potential (RP). This EEG potential is commonly described as having two phases, an early slow component and a late fast component. In our study, we collected RPs from 15 subjects and fit a linear spline with four fixed knots and their locations as free parameters. The result was a piece-wise linear approximation to the subject RPs from which it was possible to identify the onset and termination of both the slow and fast components, as well as their slopes. In addition, we used the collected RPs to train an adaptive filter neural decoding algorithm to predict occurrences of RPs from trial-based epochs of imagined speech and motor movements for use in a brain-computer interface for speech communication. The initial spline analysis will help to determine the contribution of each RP phase to the representations of intended speech-motor behavior, reducing the complexity of the adaptive filter for more efficient use in real-time.

Index Terms: speech production, electroencephalography, brain-computer interface

1. Introduction
Brain-computer interfaces (BCIs) for communication have been designed based on a number of different neurophysiological phenomena including modulation of motor cortical potentials through noninvasive electroencephalography (EEG) \cite{1, 2, 3} and intracortical microelectrodes \cite{4}. Motor-based BCIs use endogenous neural activity, meaning they are generated solely through voluntary changes to neurological state and do not depend on external stimuli. However, this freedom from external markers makes detecting true motor cortical events difficult if an individual has no voluntary motor behavior, as is the cases of quadriplegia with anarthria. This problem, known as the “midas touch” problem, is a major obstacle to fluent communication using motor-based BCIs, and is especially so for systems that directly control speech synthesizers for real-time speech acoustic output \cite{4, 5}. As an example, imagine a speaker was able to move their tongue, lips and jaw, but unable to control the onset and termination of voicing. In this scenario, the speaker would likely be frustrated by either attempting to produce a sound, but with no acoustic output, or find themselves vocalizing when a speech sound was unintended.

One solution is to continuously monitor neurological activity for signs that an error has occurred (error potential, ErrP) \cite{6}. Unfortunately, this technique suffers two major disadvantages for a speech output BCI: 1) it is retrospective, so in the case of false voicing onset, the acoustic signal has already been generated before the ErrP is detected, and 2) the ErrP is best elicited in event-related paradigms rather than continuous decoding. A better solution lies in developing a method to predict when intentional speech and motor behaviors will occur as a preemptive confirmation for any subsequent motor-based BCI activity.

Speech production is the result of a complex coordination of neural, muscular and perceptual processes. At minimum, the brain must: 1) prepare to speak by converting an abstract message into speech motor commands, 2) instruct the vocal tract to execute specific movements, 3) listen to the acoustic consequences of self productions and 4) make any necessary online adjustments in the face of speech errors. The neurological activity related to speech preparation is known as the readiness potential (RP) \cite{7}, and is characterized as a slow wave negative deflection that begins as early as 2 s prior to voluntary motor behaviors when observed using electroencephalography (EEG). Early researchers recognized speech production as just another motor task, albeit complex, and confirmed that the RP is present prior to speech and orofacial tasks \cite{8, 9, 10}. The RP, therefore, is an excellent candidate for decoding intention of upcoming speech and motor behavior in a motor-based BCI application. Though consistently observed, the RP has significant inter- and intra-subject variability and is typically only revealed after averaging dozens of trials. A speech output BCI must be able to respond in a continuous manner, or single trial at the very least. In addition, duration (2 s) and low frequency of the RP may result in a significant delay in processing when attempting to predict upcoming intended speech and motor behavior.

In this paper, we describe efforts to decode the RP in a continuous fashion using an adaptive filter technique (e.g., Wiener filter), and to better characterize the RP temporal dynamics to help focus our decoding algorithm on an analysis window with maximal relevance to upcoming intended speech and motor behavior. The temporal dynamics of the RP are known, but poorly described. It is commonly defined as having at least two preparatory components, a very slow, low amplitude deflection at the beginning of the RP wave followed by a relatively fast and higher amplitude deflection immediately prior to movement and a late, post-onset potential \cite{7, 11, 12}. Using a combination of statistical analysis, neural decoding and linear polynomial fitting, we comprehensively describe the onset, duration and slope magnitude of each RP component for a range of motor behaviors including finger extension, lip puckering and production of words involving a bilabial stop (/p/).
2. Methods

2.1. Subjects
Fifteen native speakers of American English (7 female) participated in our experiment, with a mean age of 24.9 years (range from 21-32). Two participants reported left hand dominance. All study procedures were approved by the Human Subjects Committee at the University of Kansas, Lawrence, and all participants provided informed consent before engaging in experimental activities.

2.2. Biophysiological data acquisition
Electromyography (EMG), electroencephalography, electrooculography (EOG) and respiratory signals were collected from all participants. EMG was recorded from six Ag/AgCl electrodes with pairs placed on the left and right extensor digitorum muscle and a third pair over the orbicularis oris inferior muscle approximately 1 cm apart. Each pair was recorded using a bipolar reference and a common ground was shared with the EEG electrodes. All EMG recording sites were first cleaned with alcohol and abraded to reduce skin impedance.

EEG was recorded using a 62 active Ag/AgCl (g.LadyBird, g.tec, Graz, AT) electrodes placed in a cap (g.GammaCap2, g.tec, Graz, AT) at locations according to the modified 10-10 standard [13, 14]. The ground was placed on the forehead at location AFz and a common reference electrode was clipped to the left ear. EOG signals were obtained from electrodes FP1 and AF8, which are located centrally above and lateral to the left eye. Acoustic responses were captured using a head-worn condenser microphone (AKG C520) and voice detected (g.TRIGBox, g.tec, Graz, AT). All biophysiological measures were obtained synchronously using a multipurpose data acquisition system (g.HIAmp, g.tec, Graz, AT) at 512 Hz sampling rate and notch filter from 58–62 Hz to remove the 60 Hz powerline artifact.

2.3. Tasks
Participants completed four tasks: left and right index finger extensions, lip puckers, and production of /p/ words. For the finger extensions, participants were asked to hold their hand in a loose fist and forcefully point their index finger. For lip puckers, participants were asked to push their lips forward to form “fish lips” or a “kissing” shape. For the production of /p/ words, participants were asked to say any word that began with a /p/ sound of their choosing, and were asked to avoid using the same words consistently.

All tasks were performed voluntarily at the participants own pace (i.e., no cues were given to signal action onset). Participants were asked to remain as still as possible for the entire session, which lasted approximately 1.5 hours, to minimize EEG artifacts. Each participant completed 10 blocks of 10 trials for each of the four tasks (400 trials per session). An additional instruction was given to minimize movements, including eye blinks, for approximately 3 sec before performing a task. A real-time oscilloscope visualization of the EMG, EOG and respiratory signals was provided to the participants to help minimize their movements prior to each task-related movement. They were also asked to hold the position of the motor action act for approximately 3 sec and to allow another 5 sec to pass in which they could blink if needed. For example, a right finger extension was held for approximately 3 sec before returning to a loose fist.

2.4. Data analysis

2.4.1. Obtaining the RP
Raw EMG was first downsampled to 128 Hz, full-wave rectified and smoothed with a 10 Hz lowpass FIR filter. The onset of task movements was then manually identified from the rectified EMG signal and stored for later use. The raw EEG signals were first bandpass filtered from 0.1 to 10 Hz then segmented into 2 s windows aligned to EMG onset markers. The epoch windows extended 1.5 s prior to EMG onset and 0.5 s following EMG onset, and the average from the interval (-1.5 s, -1.0 s) was removed from each epoch. Eye movement artifacts were removed from the EEG data using a semiautomated procedure in EEGLAB [15]. First, independent components analysis was performed, and components were manually rejected if they contained scalp patterns that reflect EOG contamination. Then, an automatic thresholding procedure found and rejected all trials with EEG voltage above ±150 µV. Less than 10% of trials were eliminated.

The readiness potential was obtained by averaging the epoch EEG voltages per subject, electrode and condition; a grand average was taken as an average of the individual subject RPs. One-sample t-tests were used to confirm the interval in which the RP becomes increasingly negative prior to movement onset. A single t-test per time point over all subject RPs was obtained and p-values were corrected for multiple comparisons using the false discovery rate (α=0.05) [16].

2.4.2. Linear polynomial analysis
The statistical analysis of RP negativity provides a gross estimate of its temporal structure (onset & termination), but does not identify intra-waveform dynamics. We fit a linear spline with four knots to the grand-average RP in MATLAB (Mathworks, Natick, MA, US), and left the knot location as a free parameter. The analysis resulted in three values per spline segment: slope, intercept and knot location (time point in the RP waveform). Subject specific spline fits are currently in progress.

2.4.3. Adaptive filtering & neural decoding
The neural decoding algorithm to predict occurrences of intended voluntary movement uses a Wiener filter to first denoise the data relative to an indicator variable that is one when the participant should be planning to move, and zero otherwise. The results are passed through a logistic regression filter and finally a first-order Markov model to make a final classification of movement intention state. The logistic regression filter is used to transform the output of the Wiener filter into the binary states used to describe movement preparation.

First, the epoched data from a montage of six electrodes (Fz, FCz, F1, F2, FC1, and FC2) are concatenated and averaged across electrodes to form a single data set to train the Wiener filter parameters. The averaging procedure reduces variability with minimal loss of information since the spatial resolution of EEG is very low. We additionally confirmed that the voltages were not statistically different between the six electrodes. A two-fold cross-validation procedure was used to train the Wiener filter and to obtain estimates of the optimal filter order.

The Wiener filter output is used to train the logistic regression filter which uses a step-size of 0.01 and 10 regression coefficients. Rather than using a hard decision threshold of 0.5, a first order Markov model is used to determine the optimal transition probability from baseline values to the occurrence of an
RP (e.g., transitioning into an intentional movement preparation period). The combined results define a filtering pipeline that is capable of finding onsets of the RP in single trial data.

3. Results

3.1. Identification of RP interval

An example of the grand average RP is shown for electrode Cz in Figure 1 for each of the experimental conditions: left finger extension (green), right finger extension (yellow), lip pucker (red) and /p/ word production (blue). The RPs are aligned to EMG onset represented by the black line at time 0 s. The result of the FDR corrected t-tests of RP amplitude \( p_{\text{FDR}} < 0.05 \) is shown as a shaded region from -1.14 s to 0.5 s; the post-movement period (0.0 s, 0.5 s) is also statistically less than zero, but we are most interested in the preparatory period.

3.2. Spline fitting of RP waveform

The linear spline fit of the grand average RP at electrode Cz is shown in Figure 2 for the four tasks: left finger extension (green), right finger extension (yellow), lip pucker (red) and /p/ word production (blue). The intervals defined by the knots, and their associated slopes are summarized in Table 1. The first interval for all four conditions largely corresponds to the non-significant region in the statistical analysis of the RP waveform, ending between -1.19 s and -0.93 s. The slopes of these intervals are also close to zero, further associating them with the baseline period. The following two intervals represent the slow and fast components of the RP while the final interval represents the return to baseline post-movement. The slow phase component ends between -0.45 s and -0.37 s with slopes ranging from -1.46 \( \mu \text{V/s} \) to -2.94 \( \mu \text{V/s} \). The fast phase component typically ends after movement onset at the peak negativity, and has larger slopes than the slow phase, ranging from -3.61 \( \mu \text{V/s} \) to -4.86 \( \mu \text{V/s} \). These results provide a new quantitative measurement of the slow and fast phases of the readiness potential.

3.3. Adaptive filtering & neural decoding

An example of the RP decoding results is shown in Figure 3. In this image, a single trial from one subject is plotted in blue and the shaded region represents the prediction of an RP occurrence.

The study described within this report provides one of the most detailed quantitative descriptions of the readiness potential temporal dynamics. The results of our statistical analysis of the RP waveform fits with the general morphology explained by the RP spline fit. Specifically, the portion of the RP not statistically less than zero was fit by a spline segment of the same approximate duration and near zero slope. The subsequent spline segments objectively describe the slow and fast phases based on duration.

<table>
<thead>
<tr>
<th>Task</th>
<th>Knot points interval (s)</th>
<th>Slope (( \mu \text{V/s} ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left Extensions</td>
<td>-1.50 to -1.19</td>
<td>-0.97</td>
</tr>
<tr>
<td></td>
<td>-1.19 to -0.37</td>
<td>-1.95</td>
</tr>
<tr>
<td></td>
<td>-0.37 to 0.17</td>
<td>-4.28</td>
</tr>
<tr>
<td></td>
<td>0.17 to 0.50</td>
<td>7.15</td>
</tr>
<tr>
<td>Right Extensions</td>
<td>-1.50 to -1.29</td>
<td>-0.56</td>
</tr>
<tr>
<td></td>
<td>-1.29 to -0.39</td>
<td>-1.97</td>
</tr>
<tr>
<td></td>
<td>-0.39 to 0.18</td>
<td>-4.11</td>
</tr>
<tr>
<td></td>
<td>0.18 to 0.50</td>
<td>7.14</td>
</tr>
<tr>
<td>Puckers</td>
<td>-1.50 to -0.93</td>
<td>-0.96</td>
</tr>
<tr>
<td></td>
<td>-0.93 to -0.45</td>
<td>-2.94</td>
</tr>
<tr>
<td></td>
<td>-0.45 to 0.08</td>
<td>-3.61</td>
</tr>
<tr>
<td></td>
<td>0.08 to 0.50</td>
<td>6.27</td>
</tr>
<tr>
<td>/p/ Words</td>
<td>-1.50 to -1.17</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>-1.17 to -0.41</td>
<td>-1.46</td>
</tr>
<tr>
<td></td>
<td>-0.41 to 0.18</td>
<td>-4.86</td>
</tr>
<tr>
<td></td>
<td>0.18 to 0.50</td>
<td>8.93</td>
</tr>
</tbody>
</table>
between knot points, and their related changes in amplitude by the segment slopes.

Our future analysis will examine these spline fits on the subject average RPs per condition in addition to the grand average analysis reported here. With the additional estimates of knots and slopes, we will be able to perform statistical tests of knot location and degree of negativity across the four conditions. Additionally, this analysis will be useful for examining any effects of laterality, especially when considering contralateral hemisphere for finger movements and language-related laterality for /p/ word productions. Most studies to date have not found compelling evidence for RP lateralization prior to speech production, which may be attributed to the wide inter- and intra-subject variability. The spline analysis seeks to reduce this variability and permit evaluation of the spline parameters as a proxy for the actual RP waveform.

To our knowledge, this is the first report of successful single trial neural decoding of the speech-related readiness potential, and predictions of motor intention. Our decoder has immediate application for all motor-based BCIs since the RP is observed prior to all voluntary motor behavior. Current motor-based BCIs are not capable of discerning intended from erroneous motor activations, which may result in poor performance. The consequence of these errors are quite significant for a BCI for controlling real-time speech synthesized output [4, 5]. Online prediction of motor intentions reduces the possibility of erroneous neural decoding and can lead to more natural and fluent use of the BCI-based communication device. Our future work on this decoder will attempt to decode RP events from continuously streaming EEG, which is a more realistic usage, as opposed to the data epochs reported here. We will also combine the RP decoder with our EEG-based speech synthesizer BCI [5] for a new hybrid approach to allow users to voluntarily control both voicing and synthesizer parameters.

5. Conclusions

We reported on a novel analysis of the speech and motor readiness potential, using a combination of statistical analysis and linear polynomial curve fitting to describe its temporal dynamics. Specifically, we quantified the durations and amplitudes of each of the major RP phases, the so-called slow and fast phases. This data-driven parameterization of the RP has potential use in future experiments that target processing in each phase as well as improving analysis of laterality in limb and speech movement preparation. We also developed and tested a neural decoding algorithm capable of predicting RP occurrences from single trial EEG data. The goal of this algorithm is to improve control of other motor-based BCIs by preemptively confirming that upcoming changes neural activity are related to intended movements (or imagined movements) rather than erroneous neurophysiological fluctuations. The benefits of such a system are potentially very large for speech output BCIs, which require precise control over voicing and phonation for fluent speech output.

6. Acknowledgments

This research was supported in part by the NIDCD (R03-DC011304 and T32-DC000052).

7. References


