JSLHR

Research Note

Motor-Induced Suppression of the N100 Event-Related Potential During Motor Imagery Control of a Speech Synthesizer Brain–Computer Interface

Jonathan S. Brumberg[a] and Kevin M. Pitt[a]

Purpose: Speech motor control relies on neural processes for generating sensory expectations using an efference copy mechanism to maintain accurate productions. The N100 auditory event-related potential (ERP) has been identified as a possible neural marker of the efference copy with a reduced amplitude during active listening while speaking when compared to passive listening. This study investigates N100 suppression while controlling a motor imagery speech synthesizer brain–computer interface (BCI) with instantaneous auditory feedback to determine whether similar mechanisms are used for monitoring BCI-based speech output that may both support BCI learning through existing speech motor networks and be used as a clinical marker for the speech network integrity in individuals without severe speech and physical impairments.
Method: The motor-induced N100 suppression is examined based on data from 10 participants who controlled a BCI speech synthesizer using limb motor imagery. We considered listening to auditory target stimuli (without motor imagery) in the BCI study as passive listening and listening to BCI-controlled speech output (with motor imagery) as active listening since audio output depends on imagined movements. The resulting ERP was assessed for statistical significance using a mixed-effects general linear model.
Results: Statistically significant N100 ERP amplitude differences were observed between active and passive listening during the BCI task. Post hoc analyses confirm the N100 amplitude was suppressed during active listening.
Conclusion: Observation of the N100 suppression suggests motor planning brain networks are active as participants control the BCI synthesizer, which may aid speech BCI mastery.

Speech motor control relies on neural processes to monitor self-generated speech (Curio, Neuloh, Numminen, Jousmäki, & Hari, 2000; Flinker et al., 2010; Houde, Nagarajan, Sekihara, & Merzenich, 2002; Numminen & Curio, 1999) in order to focus on and maintain one's own speech. The motor efference copy is often discussed as a mechanism the brain uses to compare incoming sensory information to expectations or goals (including auditory) based on motor system consequences (Eliades & Wang, 2003; Golfinopoulos, Tourville, & Guenther, 2010; Guenther, Ghosh, & Tourville, 2006; Houde & Nagarajan, 2011; Niziolek, Nagarajan, & Houde, 2013; von Holst & Mittelstaedt, 1950). In two computational accounts of speech motor control, errors in sensory feedback compared to sensorimotor expectations are used to provide corrective motor commands to improve and maintain online speech productions (Golfinopoulos et al., 2010; Guenther et al., 2006; Houde & Nagarajan, 2011). Behavioral evidence for the efference copy has been well demonstrated using speech auditory feedback perturbation paradigms (Bauer, Mittal, Larson, & Hain, 2006; Niziolek & Guenther, 2013; Tourville, Reilly, & Guenther, 2008; Villacorta, Perkell, & Guenther, 2007) in which auditory information, such as fundamental frequency, is shifted either up or down during speech production and induces a compensation in the opposite direction. In these cases, the speech motor efference copy is hypothesized to provide the brain with information about expected auditory information (e.g., anticipated pitch and formants) so that

[a]Department of Speech-Language-Hearing: Sciences & Disorders, The University of Kansas, Lawrence

Correspondence to Jonathan S. Brumberg: brumberg@ku.edu

Editor-in-Chief: Michael Hammer
Editor: Torrey Loucks

Received May 18, 2018
Revision received December 16, 2018
Accepted April 16, 2019
https://doi.org/10.1044/2019_JSLHR-S-MSC18-18-0198

Publisher Note: This article is part of the Forum: Selected Papers From the 2018 Conference on Motor Speech—Basic Science and Clinical Innovation.

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

Downloaded from: https://pubs.asha.org Lucas Steuber on 06/08/2020, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

deviations (e.g., perturbations) can be detected and used to form corrective motor commands to compensate.

Feedback perturbation paradigms have also been used to uncover neurological evidence for the speech motor efference copy mechanism, with findings converging on the N100 event-related potential (ERP)/M100 event-related field as a potential marker. The N100 is an auditory ERP with a negative polarity near 100 ms after presentation of an auditory stimulus (e.g., tone, speech), and the M100 is the equivalent magnetoencephalography response. During active listening while speaking, the N100 amplitude is suppressed relative to passive listening (Curio et al., 2000; Houde et al., 2002; Numminen & Curio, 1999) and in response to altered speech feedback (voice transformation: Heinks-Maldonado, Mathalon, Gray, & Ford, 2005; noise masking: Houde et al., 2002; and pitch shifting: Behroozmand & Larson, 2011; Heinks-Maldonado et al., 2005). In speech perturbation paradigms, there is minimal N100 suppression when individuals passively listen to non–speech sounds, maximally suppressed (i.e., minimal negativity, or closer to zero) when producing and hearing one's own speech, and in the middle for other manipulations (e.g., passively listening to one's own speech, listening to shifted auditory feedback; Behroozmand, Karvelis, Liu, & Larson, 2009; Behroozmand & Larson, 2011; Heinks-Maldonado et al., 2005; Houde et al., 2002; Martikainen, Kaneko, & Hari, 2004). In these paradigms, graded N100 suppression represents the extent to which perception of self-produced speech reflects speech motor expectations. The N100 amplitude can even be elicited when pressing a button to generate speech sounds (Martikainen et al., 2004), suggesting suppression occurs when individuals believe they are in control of the sound production and have associated a sensory outcome with a volitional motor action (Behroozmand et al., 2009).

Recently, we investigated the performance of a brain–computer interface (BCI) for controlling a formant frequency speech synthesizer with continuous, real-time audio feedback (Brumberg, Pitt, & Burnison, 2018). BCIs are devices that provide a direct link between an individual and a computer device through brain activity alone, without requiring any overt movement or behavior (Brumberg, Pitt, Mantie-Kozlowski, & Burnison, 2018). Most often, BCIs are designed as a technique accessing augmentative and alternative communication systems for individuals with severe speech and physical impairments due to paralysis and other neurological disorders (e.g., amyotrophic lateral sclerosis and brainstem stroke) and focus on letter or symbol spelling using discrete item selection or continuous cursor control (see Brumberg, Pitt, Mantie-Kozlowski, et al., 2018, for a review). Alternative BCI designs propose to directly decode speech from neurological recordings either invasively from the brain itself (e.g., Brumberg, Wright, Andreasen, Guenther, & Kennedy, 2011; Conant, Bouchard, Leonard, & Chang, 2018; Herff et al., 2015; Kellis et al., 2010; Mugler et al., 2014; Ramsey et al., 2018; see Chakrabarti, Sandberg, Brumberg, & Krusienski, 2015, for a review) or noninvasively from the scalp using electroencephalography (EEG; Brigham & Kumar, 2010; Brumberg, Pitt, & Burnison, 2018; DaSalla, Kambara, Sato, & Koike, 2009; Suppes, Lu, & Han, 1997). In this study, we wanted to explore whether individuals who learned to control the formant frequency speech synthesizer BCI (Brumberg, Pitt, & Burnison, 2018) demonstrated N100 suppression patterns similar to those observed for overt speech production, since speech sound output was the consequence of a volitional but imagined motor action.

In our prior work (Brumberg, Pitt, & Burnison, 2018), 16 participants (14 female, M_{age} = 27.5 years) learned to control the synthesizer BCI over three sessions (approximately 2 hr/session). During an offline training phase, participants listened (3-s duration) to synthesized vowel stimuli (/u/, /a/, or /i/) while imagining a specific movement of the hands (left: /u/, right: /a/) and feet (/i/). The resulting data were used to train a neural decoding algorithm that associated motor imagery–related changes in the EEG sensorimotor rhythm to the first two formants of the three vowels. Then, in an online testing phase (four blocks of 30 trials per vowel, per session), participants were presented with the audio and/or visual stimulus of the target vowel and instructed to first passively listen (listen without any motor imagery, 1.5-s duration) then to actively listen by using motor imagery to modulate the sensorimotor rhythm from which the BCI algorithm decoded instantaneous (< 50-ms delay) predictions of the first two formants. All decoded formants were provided as audio (synthesized) and/or visual (two-dimensional formant plane) feedback, and participants were separated into groups receiving unimodal (visual or auditory) or multimodal (audiovisual) feedback. A key difference in the two types of listening, passive versus active, in the online test phase is that, during passive listening, auditory and/or visual feedback is experimentally controlled and there is no associated motor imagery. In active listening, however, auditory and/or visual feedback is controlled directly by the participants' motor imagery. Overall, participants who received both auditory and visual feedback performed the best in the four-alternative task of producing the vowels /u/, /a/, and /i/ while avoiding the vowel /ae/ (68.3% average accuracy), followed by those who received unimodal auditory feedback (50.1% average accuracy) and unimodal visual feedback (47.2% average accuracy). Only differences between audiovisual feedback and each of the two unimodal feedback conditions were statistically significant with no differences between the unimodal groups (Brumberg, Pitt, & Burnison, 2018). There were also no statistically significant effects of session number, indicating participants did not change their performance over time. Based on past reports from other studies, we conclude this effect may change with additional training sessions (Brumberg, Pitt, & Burnison, 2018). In order to be successful in the BCI task, participants needed to coordinate their imagined movements with the real-time auditory feedback of vowel sounds. This process of BCI-based audio–motor coordination shares many similarities with conventional speech motor control, particularly motor execution with perceptual feedback monitoring and corrective motor commands (Guenther et al., 2006), though in a limb motor/acoustic domain. As a result, it is possible that motor imagery involved in BCI synthesizer control may activate

Downloaded from: https://pubs.asha.org Lucas Steuber on 06/08/2020, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

motor efference copy mechanisms that provide auditory cortical areas with information on expected auditory feedback and result in motor-induced suppression of the N100 response.

To test whether motor efference copy mechanisms, defined as observation of a motor imagery–induced suppression of the N100, were present in our prior study, we conducted a second analysis of our BCI data. Specifically, we examined participants' EEG recordings for evidence of suppressed N100 responses during speech synthesizer BCI control focusing on time intervals during passive listening to target vowel stimuli (e.g., no motor imagery with experimentally generated feedback) versus intervals of active listening (e.g., feedback that depends on active motor imagery control). We hypothesize if the N100 is suppressed during active listening relative to passive listening, then BCI control likely uses a motor efference copy, or similar neural mechanisms, to aid motor imagery control of the BCI formant synthesizer, which may facilitate BCI learning.

Method

Participant Data

Data from 10 participants (nine female, M_{age} = 27.5 years, range: 21–36 years) who took part in the Brumberg, Pitt, and Burnison (2018) BCI speech synthesizer study and who received continuous auditory feedback (i.e., not including the unimodal visual feedback group) were used to analyze the auditory N100 response. One group of five participants received multimodal audiovisual feedback, and the other group received unimodal audio feedback in the BCI paradigm. All participants completed four blocks of 30 trials per vowel, on each of three sessions within a 2-week time period. During the BCI experiment, no trials were rejected due to electrical artifacts. Instead, BCI processing was halted upon detection of electrical artifacts with a voltage of ±150 μV. All participants had normal hearing, normal/corrected vision, and no known neurological disorders.

Data Acquisition and Processing

EEG was obtained using a g.HIamp (g.tec) acquisition system from 62 active electrodes at 256 Hz according to the 10-10 standard placement, with a forehead ground and left earlobe reference. Since there were no effects of session on BCI performance in our previous study (Brumberg, Pitt, & Burnison, 2018), we grouped all sessions together for the present analysis. Instead, our major comparison was between N100 ERPs during passive listening (feedback without motor imagery) and active listening (motor imagery–controlled BCI speech synthesis).

To examine the N100, raw EEG signals from the BCI experiment were reprocessed in MATLAB (Mathworks, Inc.) by first high-pass filtering at 1 Hz and removing eyeblink artifacts using independent components analysis. Any trials following artifact removal with EEG amplitudes over ±150 μV were rejected from the analysis (0.97%), but no BCI trials with incorrect productions were initially rejected. In a separate analysis, we compare N100 responses for the full data set against those for correct BCI trials only. The resultant signals were then low-pass filtered at 30 Hz, windowed from −100 ms to 500 ms relative to the audio onset of either the target vowel stimulus (passive) or the motor imagery–controlled speech auditory feedback (active), and baseline corrected from −100 ms to 0 ms. Next, we identified the N100 as the first negativity near 100 ms by visually inspecting grand-average ERPs (over condition and group) for each electrode. We confirmed visual inspections using a one-sample, left-tailed *t* test of per-participant ERP averages (over condition and group) to verify putative N100 negativities were statistically less than zero (*t* tests used Bonferroni correction for multiple comparisons of the number of time points in the ERP window).

Statistical Analysis

We further examined the N100 amplitude by first choosing peak negativities in a 60-ms window centered at 100 ms for each participant average ERP for each electrode in the two conditions, passive and active listening. We then used a mixed-effects general linear model in R with the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) to examine the between-subjects factor Group (audiovisual or audio-only feedback) and within-subjects factors Condition (passive or active listening) and Electrode (29 locations; see Results section and Figure 1); participant was used as a random factor. Finally, we repeated our statistical analyses (linear mixed-effects model of N100 amplitude for the within-subjects factors Condition, Electrode, and Accuracy; random factor of participant) using only correct BCI trials to determine if there were any relationships between N100 suppression and BCI success.

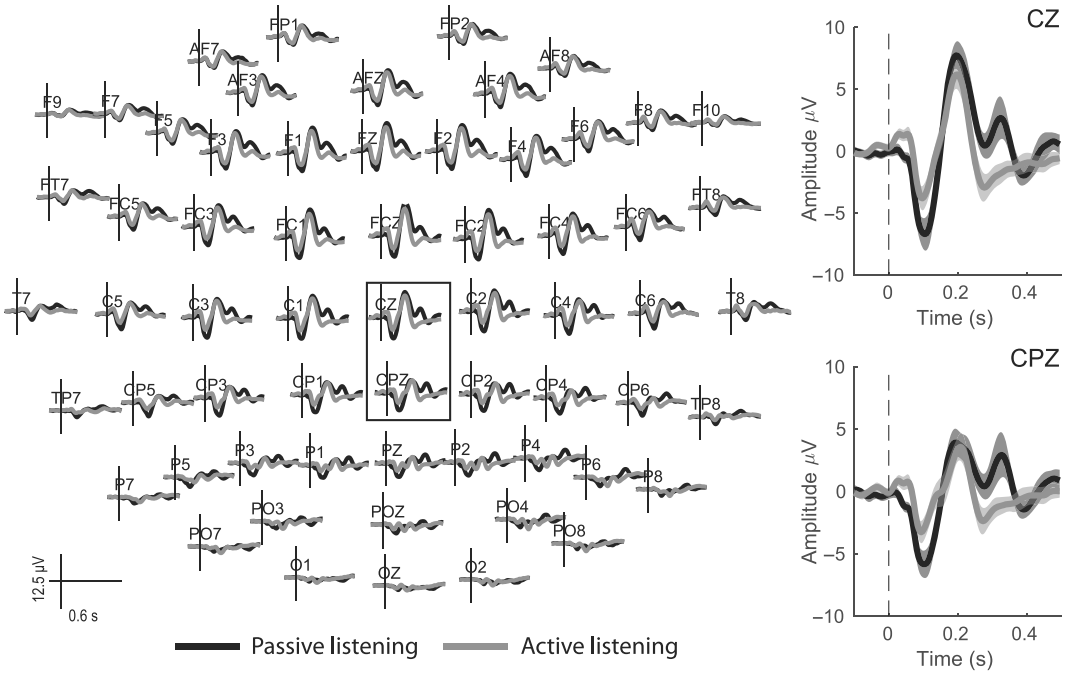
Results

N100 Interval and Spatial Properties

The average peak N100 amplitude was centered at 111 ms (98–121 ms over all electrodes) and was found for each electrode through a combination of visual inspection and one-sample, left-tailed *t* tests (with Bonferroni correction for all time points in the ERP window). Electrodes either without a participant average negativity around 100 ms or negativities that were not statistically significantly less than zero (one-sample, left-tailed, Bonferroni-corrected *t* test) were not included in subsequent analyses, leaving 29 of 62 electrodes for additional study (anterior sites: $FP_{1,2}$, $AF_{z,3,4,7}$, $F_{1,3,5,7}$; temporal sites: FT_7, T_7; and central sites: $FC_{z,1,2,3,5}$, $C_{z,1,2,3,4,5,}$, $CP_{z,1,2,3,4}$). Figure 1 also clearly shows a P200 for both active and passive listening as well as a P300 for passive listening; however, our analysis was not intended to explore effects of ERPs other than the N100 and its motor-induced suppression. Therefore, the P200 and P300

Downloaded from: https://pubs.asha.org Lucas Steuber on 06/08/2020, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Figure 1. Grand-average event-related potentials (ERPs) at each scalp location labeled according to the 10-10 standard for passive listening (black) and active listening (gray) conditions. The vertical line indicates time 0 ms and ERP windows span –100 ms to 500 ms and its height spans –5 μV to 7.5 μV from bottom to top. Example ERPs from electrodes CZ and CPZ (outlined in black) are shown in greater detail to the right with 95% confidence intervals (shaded).



components are not analyzed and will be the subject of future work.

N100 Suppression

The linear mixed-effects model analysis of N100 peak amplitude differences at each electrode with an N100 response ($N = 29$) for all BCI trials revealed statistically significant main effects of Condition (passive vs. active listening, Wald test: $\chi^2(1) = 575.1$, $p < .001$) and Electrode (Wald test: $\chi^2(28) = 1155.8$, $p < .001$) with no main effect of Group, as well as statistically significant interaction effects of Condition × Electrode (Wald test: $\chi^2(28) = 137.4$, $p < .001$) and Group × Electrode (Wald test: $\chi^2(28) = 85.7$, $p < .001$). A simple effects analysis of the Condition × Electrode interaction (linear mixed-effects model of amplitude for each level of Electrode) revealed 22 electrodes (of the 29 with a verified N100 component) with statistically significantly different N100 amplitudes (Bonferroni correction applied for the number of electrodes, $p < .05$) in central ($C_{Z,1,2,3,4,5}$, $CP_{Z,1,2,3,4}$, $FC_{Z,1,2,3,4,5}$), temporal (FT_7, T_7), and anterior ($F_{1,3}$, AF_Z) scalp locations. For each electrode, a Tukey's post hoc test found N100 amplitudes had greater negativity during passive listening than during active listening (all comparisons, $p_{Tukey} < .05$). A simple effects analysis of the Group × Electrode interaction revealed no statistically significant differences in feedback type (unimodal audio or multimodal audiovisual), which suggests all effects were for cross terms and are not the focus of this study. The normalized N100 suppression index (Behroozmand & Larson, 2011):

$$\frac{\left|\mathrm{N100}_{\mathrm{passive}} - \mathrm{N100}_{\mathrm{active}}\right|}{\left|\mathrm{N100}_{\mathrm{passive}}\right|} \times 100 \qquad (1)$$

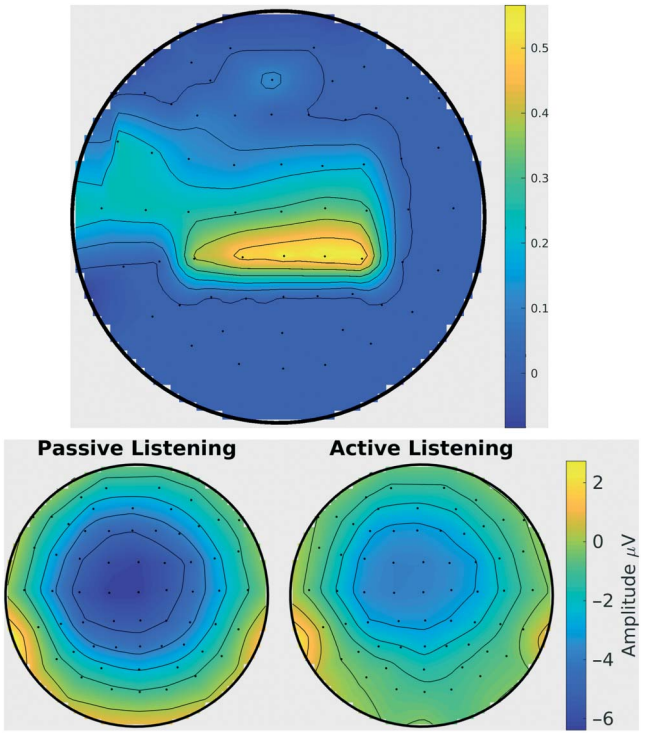
was computed for all electrodes with a statistically significant N100 suppression and ranged from 7% to 57%, with the highest amount of suppression over central electrodes $CP_{z,1-4}$ (41%–57%) and $C_{z,1-5}$ (25%–32%), followed by temporal site electrodes FT_7/T_7 (24%–26%), fronto-central $FC_{z,1-5}$ (16%–22%), and anterior/frontal sites AF_z, $F_{1,3}$ (7%–14%). The spatial topography of the normalized N100 suppression index for electrodes with a statistically significant N100 suppression is represented in Figure 2; electrodes without suppression were set to zero.

Relationship Between N100 Suppression and BCI Accuracy

Our second analysis focused on the relationship between BCI accuracy and observed suppression of the N100 ERP component for electrodes with a verified grand-average N100 response. The mixed-effects analysis of N100 peak amplitude differences at each electrode with an N100 ($N = 29$) for only correct BCI trials revealed statistically significant main effects of Condition (passive vs. active listening, Wald

Downloaded from: https://pubs.asha.org Lucas Steuber on 06/08/2020, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Figure 2. A two-dimensional topographical scalp representation of the normalized N100 suppression index when considering all brain–computer interface trials. The top of the circle is anterior, and the bottom is posterior. Blue colors are close to or equal to 0% suppression (note: scalp locations without a verified N100 response were entered with a 0% suppression), and warm colors indicate greater suppression. The N100 scalp topography is shown for both passive listening (bottom left) and active listening (bottom right), with blue colors indicating greater negativity and red colors indicating positivity.



test: $\chi^2(1) = 316.4$, $p < .001$) and Electrode (Wald test: $\chi^2(28) = 457.6$, $p < .001$) and their interaction (Wald test: $\chi^2(28) = 62.8$, $p < .001$). A simple effects analysis of the Condition × Electrode interaction (linear mixed-effects model of amplitude for each level of Electrode) revealed 18 electrodes (of the 29 with a verified N100 component) with statistically significantly different N100 amplitudes (Bonferroni correction for the total number of electrodes, $p < .05$) in central ($C_{Z,1,2,3,4,5}$; $CP_{Z,1,2,3,4}$; $FC_{Z1,2,3,4,5}$) and temporal (T_7) locations. For each electrode, a Tukey's post hoc test revealed greater N100 negativities during passive listening than during active listening (all comparisons, $p_{Tukey} < .05$), similar to the analysis of all BCI trials, though over a smaller number of electrode sites.

Discussion

N100 Suppression and Motor Efference Copy

In this research note, we provide evidence that an auditory N100 response to BCI-controlled synthesized speech feedback was smaller in amplitude compared to the N100 generated while passively listening to synthesized speech. Past work has suggested that motor-induced suppression of auditory cortical responses is largely due to neurological mechanisms that mediate the motor efference copy (Curio et al., 2000; Flinker et al., 2010). Furthermore, the efference copy mechanism is hypothesized to provide the sensory system an estimate of expected consequences of motor behavior in order to verify successful action completion and to generate corrective motor commands in the presence of motor error (Eliades & Wang, 2003; Niziolek et al., 2013). In speech, the motor efference copy is hypothesized to provide auditory cortex with the information needed to determine if incoming perception of self-produced utterances are correct and if any orofacial corrective motor commands are needed (Guenther et al., 2006; Houde & Nagarajan, 2011; Niziolek et al., 2013).

In this study, we apply a definition of the motor-induced N100 suppression during active perception of self-produced speech sounds as a representation or consequence of speech motor efference copy mechanisms. Therefore, the motor (imagery) induced N100 suppression observed while participants actively controlled the speech synthesizer BCI and listened to the online feedback provides evidence that a motor efference copy mechanism was involved and possibly used by the auditory cortex to monitor BCI-produced speech sounds. More specifically, the presence of a suppressed N100 response during active listening suggests that participants may utilize motor planning and feedback neural pathways similar to those used during speech motor learning (e.g., forward models, feedback learning; Guenther et al., 2006; Houde & Nagarajan, 2011) for speech synthesizer BCI learning. From one perspective, our BCI study is a form of voice perturbation or transformation; participants learn to control a device that sounds human like, but is not their own. In this way, the results of our N100 study analysis confirm past work on N100 suppression in response to feedback modifications such as voice transformation (e.g., alien voice; Heinks-Maldonado et al., 2005). Similarly, our BCI used limb motor imagery, rather than speech motor imagery, making our N100 suppression results, and the neuromotor control mechanism, more similar to past N100 studies in which manual button pressing led to the production of speech sounds (Martikainen et al., 2004). In addition, the BCI control strategy involved motor imagery, or covert movement, and supports past studies on covert movement-induced N100 suppression (Numminen & Curio, 1999). However, this study goes further than these past findings by combining covert limb motor actions (imagery) and transformed voice output into a single experiment (i.e., producing speech output using a non–speech motor strategy that results in synthesized audio feedback).

N100 Suppression Relationship to BCI Accuracy

In our second analysis, we examined the interaction of BCI control and N100 response for accurate BCI trials only, as opposed to aggregating both accurate and failure trials. In both analyses, we found subsets of electrodes containing ERP responses that differed between active and

Downloaded from: https://pubs.asha.org Lucas Steuber on 06/08/2020, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

passive listening (22 in the full analysis, 18 in the BCI accurate-only analysis). Despite the difference in the number of electrodes, the same overall pattern emerged from both analyses, namely, that the N100 response was reduced during active listening relative to passive listening. It is interesting that the spatial topography of N100 suppression was greatest over sensorimotor regions (FC, C, and CP electrodes) with a slight left lateralization (more so for the BCI accurate-only analysis), though these effects were not studied in detail and future study will be needed to fully account for differences in the N100 response over electrode locations. Finally, though there were similar response patterns between the analysis of BCI accurate-only and all BCI trials, it remains that more pronounced differences in N100 suppression may emerge as BCI accuracy improves over longer training periods than initially used in the BCI paradigm (more than three sessions; Nijboer, Birbaumer, & Kubler, 2010).

P200 and P300 Responses

In addition to the N100, we also found a large P200 response in our paradigm for both active and passive listening and a P300 response during passive listening only. These ERP components are not hypothesized to be involved in processing the motor efference copy and were not subject to statistical analysis in this study. However, their presence requires some discussion. There is some evidence that the P200 is involved in active monitoring of vocalizations (Behroozmand & Larson, 2011) and for processing auditory stimulus properties such as intensity and pitch (see Crowley & Colrain, 2004, for a review). Certainly, the BCI study required participants to maintain active attention on the continuously changing formant frequencies of the BCI synthesizer, and future BCI studies should be designed to specifically investigate the effects of BCI synthesizer control on the P200. Similarly, we observed a P300 response during passive listening, but not during active listening, which may be due to an "oddball" response in which the passive listening period is a "surprise" after the random silence interval between trials. In contrast, the active listening/motor imagery period can be reasonably anticipated as it always follows passive listening, possibly reducing any P300 response. The role of both the P200 and P300, since observed in this study, must be studied further for their relevance either to screening participants for potential use of the BCI synthesizer or as neurological markers for objective assessment of BCI learning.

Limitations

Notably, the data used for the present analysis were derived from a limb motor task that resulted in speech-related feedback; therefore, it is not possible to conclude that a speech-specific motor efference copy mechanism was used to control the synthesizer BCI. A confirmatory study is needed in which speech motor imagery is used to control the BCI synthesizer. However, our results do lend some evidence toward a general, multimodal motor efference copy mechanism that can be shared between motor modalities. It is also possible that, as participants become more proficient, their mental strategies for BCI synthesizer control (i.e., limb motor imagery) and the desired BCI output (i.e., vowel production) may merge into a single goal-directed behavior. If so, subsequent analyses would need to determine whether the motor efference copy becomes speech specific or retains general motor relationships (e.g., test speech vs. limb output using the limb-based sensorimotor rhythm BCI). Since the main experiment by Brumberg, Pitt, and Burnison (2018) did not find any effects of session on BCI performance, it is not possible to determine whether there are level effects of progressive BCI mastery on N100 suppression. Further study of the BCI synthesizer over longer training periods is needed to determine whether training-related improvements in BCI performance are associated with changes in N100 suppression.

Future Work

Technical Innovation

The method for eliciting the auditory evoked N100 suppression response using a continuous speech synthesizer BCI is a novel approach. The N100 suppression effect shown in this study was obtained in a speech production task using a BCI that relied only on motor imagery without actual muscular contractions associated with natural vocalizations. In addition, all audio feedback was provided from a speech synthesizer, and participants did not experience any bone-conducted audio feedback. Therefore, our method does not suffer from any artifacts common to N100-style experiments, including those due to auditory feedback intensity (needed to mask bone-conducted responses) or orofacial electromyography (possible motor contamination during speech). Furthermore, our study reaffirms motor-induced suppression as a result of covert or imagined movement (Numminen & Curio, 1999), which supports future study of the motor efference copy using imagined or inner speech (Schultz et al., 2017). Confirmation of the suppressed N100 also opens the door for future experiments using BCI-based speech synthesis to further examine the role of the speech motor efference copy using established methods, including pitch and formant perturbation, in the absence of electromyographic artifacts and competing, unaltered audio feedback.

BCI and Clinical Implications

We hypothesize that engagement of the speech motor network will be beneficial for learning to control BCIs with continuously synthesized speech output (Brumberg, Nieto-Castanon, Kennedy, & Guenther, 2010; Brumberg, Pitt, & Burnison, 2018), which are modeled after the biological vocal mechanism, and other direct speech decoding approaches (e.g., Brumberg et al., 2011; Herff et al., 2015; Mugler et al., 2014). In one sense, BCIs bridge damaged or disordered biological functions, and for speech, BCIs are intended to replace impaired vocal output as a result of neurological disease or injury. Since successful speech production relies

Downloaded from: https://pubs.asha.org Lucas Steuber on 06/08/2020, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

on a complex network of brain regions, it is likely important for speech BCIs to similarly engage the full speech motor brain network. That is, BCI learning may be enhanced when motor signals that are used for decoding intended speech articulations generate audio output that is processed by brain regions involved in monitoring one's productions and generating/sending corrective feedback commands back to the motor cortex for tuning BCI control. N100 suppression has properties that may be useful for an objective, neural marker of speech motor processing that may also be used to confirm the presence of functional neural mechanisms needed for BCI synthesizer control. Confirmation of N100 suppression during active listening suggests the BCI synthesizer motor network (planning, production, and feedback), perhaps similar to the speech motor network, is involved in our BCI task and implies that individuals with neuromotor impairments (e.g., severe dysarthria due to amyotrophic lateral sclerosis, cerebral palsy; anarthria due to brainstem stroke) who use speech synthesizer BCIs may benefit from existing neural pathways for speech motor learning (e.g., forward models, feedback learning; Guenther et al., 2006; Houde & Nagarajan, 2011) to achieve BCI control. In particular, N100 suppression could be used either as a potential screening tool to identify whether speech motor networks are still active in individuals with progressive neurological impairments (e.g., amyotrophic lateral sclerosis) or as a marker for skill learning and neural rehabilitation through emergence and strengthening of the N100 suppression as a function of speech motor recovery or BCI control. Future studies are needed to focus on the clinical implications of N100 suppression, especially with respect to continuous speech synthesizer BCI control.

Conclusions

We explored whether motor imagery control of a BCI with continuously synthesized speech output elicited electrophysiological markers hypothesized to represent speech motor efference copy processing. Presence of the N100 suppression in this study suggests a motor efference copy related to expected synthesized speech is generated in our speech output BCI task, which may be beneficial for learning BCI control. In addition, it is possible the N100 may be helpful for identifying whether speech motor brain networks are intact in individuals with acquired neurological impairments who may use BCI for communication. Future study of the N100 and other observed ERP components will help clarify the role of the motor efference copy for speech output BCIs as a function of training and mastery and as a potential clinical marker for speech motor network health.

Acknowledgments

This work was supported in part by the National Institute on Deafness and Other Communication Disorders Grant R03-DC011304, awarded to principal investigator J. Brumberg.

References

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Bauer, J. J., Mittal, J., Larson, C. R., & Hain, T. C. (2006). Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voice amplitude. *The Journal of the Acoustical Society of America, 119*(4), 2363–2371.

Behroozmand, R., Karvelis, L., Liu, H., & Larson, C. R. (2009). Vocalization-induced enhancement of the auditory cortex responsiveness during voice F0 feedback perturbation. *Clinical Neurophysiology, 120*(7), 1303–1312.

Behroozmand, R., & Larson, C. R. (2011). Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. *BMC Neuroscience, 12*(1), 54.

Brigham, K., & Kumar, B. V. K. V. (2010). Imagined speech classification with EEG signals for silent communication: A preliminary investigation into synthetic telepathy. *2010 4th International Conference on Bioinformatics and Biomedical Engineering,* 1–4.

Brumberg, J. S., Nieto-Castanon, A., Kennedy, P. R., & Guenther, F. H. (2010). Brain–computer interfaces for speech communication. *Speech Communication, 52*(4), 367–379.

Brumberg, J. S., Pitt, K. M., & Burnison, J. D. (2018). A noninvasive brain–computer interface for real-time speech synthesis: The importance of multimodal feedback. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 26*(4), 874–881.

Brumberg, J. S., Pitt, K. M., Mantie-Kozlowski, A., & Burnison, J. D. (2018). Brain–computer interfaces for augmentative and alternative communication: A tutorial. *American Journal of Speech-Language Pathology, 27*(1), 1–12.

Brumberg, J. S., Wright, E. J., Andreasen, D. S., Guenther, F. H., & Kennedy, P. R. (2011). Classification of intended phoneme production from chronic intracortical microelectrode recordings in speech-motor cortex. *Frontiers in Neuroscience, 5*(65). https://doi.org/10.3389/fnins.2011.00065

Chakrabarti, S., Sandberg, H. M., Brumberg, J. S., & Krusienski, D. J. (2015). Progress in speech decoding from the electrocorticogram. *Biomedical Engineering Letters, 5*(1), 10–21.

Conant, D. F., Bouchard, K. E., Leonard, M. K., & Chang, E. F. (2018). Human sensorimotor cortex control of directly-measured vocal tract movements during vowel production. *The Journal of Neuroscience, 38*(12), 2955–2966.

Crowley, K. E., & Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: Age, sleep and modality. *Clinical Neurophysiology, 115*(4), 732–744.

Curio, G., Neuloh, G., Numminen, J., Jousmäki, V., & Hari, R. (2000). Speaking modifies voice-evoked activity in the human auditory cortex. *Human Brain Mapping, 9*(4), 183–191.

DaSalla, C. S., Kambara, H., Sato, M., & Koike, Y. (2009). Single-trial classification of vowel speech imagery using common spatial patterns. *Neural Networks, 22*(9), 1334–1339.

Eliades, S. J., & Wang, X. (2003). Sensory–motor interaction in the primate auditory cortex during self-initiated vocalizations. *Journal of Neurophysiology, 89,* 2194–2207.

Flinker, A., Chang, E. F., Kirsch, H. E., Barbaro, N. M., Crone, N. E., & Knight, R. T. (2010). Single-trial speech suppression of auditory cortex activity in humans. *The Journal of Neuroscience, 30*(49), 16643–16650.

Golfinopoulos, E., Tourville, J. A., & Guenther, F. H. (2010). The integration of large-scale neural network modeling and functional

Downloaded from: https://pubs.asha.org Lucas Steuber on 06/08/2020, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

brain imaging in speech motor control. *NeuroImage, 52*(3), 862–874.

Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language, 96*(3), 280–301.

Heinks-Maldonado, T. H., Mathalon, D. H., Gray, M., & Ford, J. M. (2005). Fine-tuning of auditory cortex during speech production. *Psychophysiology, 42*(2), 180–190.

Herff, C., Heger, D., de Pesters, A., Telaar, D., Brunner, P., Schalk, G., & Schultz, T. (2015). Brain-to-text: Decoding spoken phrases from phone representations in the brain. *Frontiers in Neuroscience, 9,* 217.

Houde, J. F., & Nagarajan, S. S. (2011). Speech production as state feedback control. *Frontiers in Human Neuroscience, 5,* 82.

Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: An MEG study. *Journal of Cognitive Neuroscience, 14*(8), 1125–1138.

Kellis, S., Miller, K., Thomson, K., Brown, R., House, P., & Greger, B. (2010). Decoding spoken words using local field potentials recorded from the cortical surface. *Journal of Neural Engineering, 7*(5), 056007.

Martikainen, M. H., Kaneko, K., & Hari, R. (2004). Suppressed responses to self-triggered sounds in the human auditory cortex. *Cerebral Cortex, 15*(3), 299–302.

Mugler, E. M., Patton, J. L., Flint, R. D., Wright, Z. A., Schuele, S. U., Rosenow, J. M., . . . Slutzky, M. W. (2014). Direct classification of all American English phonemes using signals from functional speech motor cortex. *Journal of Neural Engineering, 11*(3), 035015.

Nijboer, F., Birbaumer, N., & Kubler, A. (2010). The influence of psychological state and motivation on brain-computer interface performance in patients with amyotrophic lateral sclerosis—A longitudinal study. *Frontiers in Neuroscience, 4*. https://doi.org/10.3389/fnins.2010.00055

Niziolek, C. A., & Guenther, F. H. (2013). Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations. *The Journal of Neuroscience, 33*(29), 12090–12098.

Niziolek, C. A., Nagarajan, S. S., & Houde, J. F. (2013). What does motor efference copy represent? Evidence from speech production. *The Journal of Neuroscience, 33*(41), 16110–16116.

Numminen, J., & Curio, G. (1999). Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex. *Neuroscience Letters, 272*(1), 29–32.

Ramsey, N. F., Salari, E., Aarnoutse, E. J., Vansteensel, M. J., Bleichner, M. G., & Freudenburg, Z. V. (2018). Decoding spoken phonemes from sensorimotor cortex with high-density ECoG grids. *NeuroImage, 180,* 301–311.

Schultz, T., Wand, M., Hueber, T., Krusienski, D., Christian, H., & Brumberg, J. (2017). Biosignal-based spoken communication: A survey. *IEEE Transactions on Audio, Speech and Language Processing, 25*(17), 2257–2271.

Suppes, P., Lu, Z., & Han, B. (1997). Brain wave recognition of words. *Proceedings of the National Academy of Sciences of the United States of America, 94*(26), 14965–14969.

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage, 39*(3), 1429–1443.

Villacorta, V. M., Perkell, J. S., & Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *The Journal of the Acoustical Society of America, 122*(4), 2306–2319.

von Holst, E., & Mittelstaedt, H. (1950). The reafference principle: Interaction between the central nervous system and the periphery. *Die Naturwissenschaften, 37,* 464–476.

Downloaded from: https://pubs.asha.org Lucas Steuber on 06/08/2020, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions