



Modulation of Auditory Responses to Speech vs. Nonspeech Stimuli during Speech Movement Planning

Ayoub Daliri^{1,2*} and Ludo Max^{2,3}

¹ Speech Lab, Department of Speech, Language and Hearing Sciences, Boston University, Boston, MA, USA, ² Laboratory for Speech Physiology and Motor Control, Department of Speech and Hearing Sciences, University of Washington, Seattle, WA, USA, ³ Haskins Laboratories, New Haven, CT, USA

Previously, we showed that the N100 amplitude in long latency auditory evoked potentials (LLAEPs) elicited by pure tone probe stimuli is modulated when the stimuli are delivered during speech movement planning as compared with no-speaking control conditions. Given that we probed the auditory system only with pure tones, it remained unknown whether the nature and magnitude of this pre-speech auditory modulation depends on the type of auditory stimulus. Thus, here, we asked whether the effect of speech movement planning on auditory processing varies depending on the type of auditory stimulus. In an experiment with nine adult subjects, we recorded LLAEPs that were elicited by either pure tones or speech syllables when these stimuli were presented prior to speech onset in a delayed-response speaking condition vs. a silent reading control condition. Results showed no statistically significant difference in pre-speech modulation of the N100 amplitude (early stages of auditory processing) for the speech stimuli as compared with the nonspeech stimuli. However, the amplitude of the P200 component (later stages of auditory processing) showed a statistically significant pre-speech modulation that was specific to the speech stimuli only. Hence, the overall results from this study indicate that, immediately prior to speech onset, modulation of the auditory system has a general effect on early processing stages but a speech-specific effect on later processing stages. This finding is consistent with the hypothesis that pre-speech auditory modulation may play a role in priming the auditory system for its role in monitoring auditory feedback during speech production.

OPEN ACCESS

Edited by:

Srikantan S. Nagarajan,
University of California,
San Francisco, USA

Reviewed by:

Erich Schröger,
University of Leipzig, Germany
Zarinah Karim Agnew,
University of California San Francisco,
USA

*Correspondence:

Ayoub Daliri
ayoub.daliri@gmail.com

Received: 30 November 2015

Accepted: 04 May 2016

Published: 18 May 2016

Citation:

Daliri A and Max L (2016) Modulation of Auditory Responses to Speech vs. Nonspeech Stimuli during Speech Movement Planning. *Front. Hum. Neurosci.* 10:234. doi: 10.3389/fnhum.2016.00234

Keywords: speech, speech planning, auditory modulation, auditory evoked potentials, EEG/ERP

INTRODUCTION

The central nervous system (CNS) modulates its response to sensory inputs that are consequences of self-produced movements. Studies have used behavioral and neurophysiological measures to examine this modulation in different sensory modalities (Waszak et al., 2012; Schröger et al., 2015b). Behavioral studies, for example, have shown that we perceive the loudness of a sound that is a consequence of our own action as less intense than the loudness of a sound produced by others (Sato, 2008, 2009; Weiss et al., 2011; Desantis et al., 2012). Using neurophysiological techniques—such as electroencephalography (EEG), magnetoencephalography (MEG), electrocorticography (ECoG), and single unit recordings—studies have shown that

cortical responses evoked by self-produced speech sounds are modulated in comparison with those evoked by hearing a played-back version of the same speech sounds¹ (EEG: Ford et al., 2001; Liotti et al., 2010; MEG: Curio et al., 2000; Houde et al., 2002; Beal et al., 2010; ECoG: Towle et al., 2008; Greenlee et al., 2011; single unit recordings: Creutzfeldt et al., 1989). The mechanism underlying this modulation is precise and specific: experimentally implemented or naturally occurring deviations in the feedback signal result in a reduction of the modulation magnitude (Heinks-Maldonado et al., 2006; Chang et al., 2013; Niziolek et al., 2013). In addition, animal studies have further confirmed this phenomenon in the auditory-motor system of monkeys, rodents, bats, and crickets (Suga and Schlegel, 1972; Suga and Shimozawa, 1974; Muller-Preuss and Ploog, 1981; Poulet and Hedwig, 2002, 2007; Eliades and Wang, 2003; Nelson et al., 2013; Schneider et al., 2014).

Typically, these results have been interpreted in the context of theoretical frameworks of motor control involving efference copy and forward internal models (Waszak et al., 2012; Horvath et al., 2015; Schröger et al., 2015b). Specifically, it has been suggested that the CNS uses an efference copy of issued motor commands and forward internal models to predict the auditory consequences of self-produced movements. The CNS then compares this prediction with the actual auditory feedback, and it attenuates its response to auditory feedback that matches the prediction (Houde et al., 2002; Heinks-Maldonado et al., 2006; Behroozmand and Larson, 2011; Chang et al., 2013; Niziolek et al., 2013).

In addition to such sensory modulation during movement execution, a growing body of evidence suggests that the CNS already modulates sensory processing during movement planning (Creutzfeldt et al., 1989; Max et al., 2008; Mock et al., 2011, 2015; Daliri and Max, 2015a,b). In a recent speech study (Daliri and Max, 2015b), we recorded long latency auditory evoked potentials (LLAEPs) in response to probe tones that were played either prior to speaking (i.e., during speech movement planning) or during no-speaking control conditions (silent reading or seeing nonlinguistic symbols). Results for a group of participants with typical speech showed that N100 amplitude in the speaking condition was attenuated in comparison with both control conditions. We suggested that, during speech planning, the CNS uses an efference copy of planned control signals to prime the auditory system for its role in processing the upcoming auditory feedback resulting from execution of those control signals. In our previous studies (Daliri and Max, 2015a,b), we examined pre-speech auditory modulation only by means of pure tone probe stimuli. Thus, it has remained unknown whether the overall nature and magnitude of this auditory modulation vary for different types of stimuli. To elucidate the phenomenon's potential role in the monitoring of auditory feedback during speech production, however, it is necessary to first understand whether speech movement planning differentially affects the auditory system's

processing of speech stimuli as compared with nonspeech stimuli. Hence, in the present study, we aimed to investigate whether the effect of speech movement planning on auditory processing varies depending on the general characteristics of the auditory probe stimulus. Specifically, we studied pre-speech modulation of LLAEPs that were elicited by either pure tones or speech syllables. For both types of stimuli, we quantified pre-speech auditory modulation by comparing the N100 and P200 components' amplitudes in a speaking condition vs. a silent reading condition.

N100 is the largest negative peak in the electrical cortical response to a transient auditory stimulus, occurring approximately 100 ms after onset of the stimulus. The N100 component is primarily generated by neural populations located in the primary auditory cortex (Näätänen and Picton, 1987; Zouridakis et al., 1998; Godey et al., 2001) and reflects processes involved in detecting acoustic change in the environment (e.g., Hyde, 1997). P200 is the largest positive peak that follows N100, approximately 180 ms after onset of the auditory stimulus. The neural generators of P200 are less well understood. It has been suggested that primary neural generators of P200 are located in multiple auditory areas, including primary and secondary auditory cortices (Hari et al., 1987; Scherg et al., 1989; Baumann et al., 1990; Mäkelä and Hari, 1990; Godey et al., 2001; Steinschneider and Dunn, 2002). However, it has also been suggested that P200 may have additional neural generators separate from the auditory areas (Crowley and Colrain, 2004). The functional role of the P200 component is not entirely clear (Crowley and Colrain, 2004). For speech stimuli, however, the available data suggest that whereas the N100 component is involved in early stages of auditory processing such as encoding basic physical aspects of the auditory input, the P200 component is involved in later stages of auditory processing such as speech-specific, higher-level analysis (Näätänen and Picton, 1987; Crowley and Colrain, 2004; Tremblay et al., 2009; Pratt and Lightfoot, 2012). Therefore, in the current study, we hypothesized that if pre-speech modulation plays a role in priming the auditory system for its role in monitoring auditory feedback during speech production, then the N100 and P200 components might be differentially affected when the stimuli used to probe the auditory system during speech planning are pure tones vs. speech syllables.

MATERIALS AND METHODS

Participants

Participants were nine right-handed adults (3 females; age range: 22–30 years, $M = 25.44$ years, $SD = 2.45$) who were naive to the purpose of the study. All participants were native speakers of American English who self-reported no current or prior neurological, psychological, or communication disorders. Only participants with normal binaural hearing (≤ 20 dB hearing level (HL) at octave frequencies 250–8000 Hz) were included. Prior to the experiment, written informed consent was obtained from all subjects. All procedures were approved by the Institutional

¹Similar results have been obtained when participants press a button that causes a tone to be played vs. passively listening to tones (for a review see Schröger et al., 2015b).

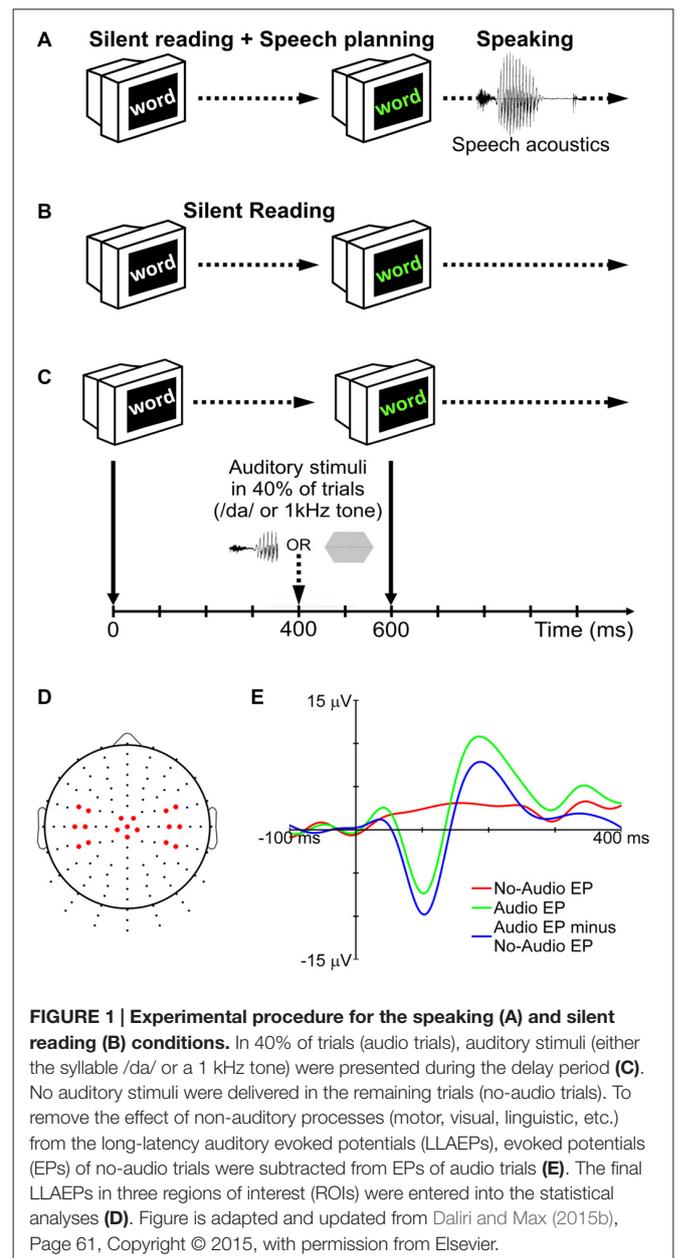
Review Boards of the University of Washington (where the data were collected and analyzed).

Instrumentation and Procedure

The design of this study was based on our previously published studies (Daliri and Max, 2015a,b). During the test session, each participant was seated inside a sound-attenuated room in front of a 23-inch LCD monitor. The participant's speech output was transduced with a wireless microphone (WL185, Shure Incorporated, Niles, IL, USA) placed approximately 15 cm from the mouth, amplified with a microphone amplifier (DPS II, ART ProAudio, Niagara Falls, NY, USA) and headphone amplifier (S.phone, Samson Technologies Corp., Syosset, NY, USA), and played back to the participant through insert earphones (ER-3A, Etymotic Research Inc., Grove Village, IL, USA). Prior to each experiment, the overall amplification level was calibrated such that a 75 dB SPL speech signal at the microphone produced an output of 73 dB sound pressure level (SPL) in the insert earphones. The earphones' output level was measured with a 2 cc coupler (Type 4946, Bruel and Kjaer Inc., Norcross, GA, USA) connected to a sound level meter (Type 2250A Hand Held Analyzer with Type 4947 $\frac{1}{2}$ " Pressure Field Microphone, Bruel and Kjaer Inc., Norcross, GA, USA).

Participants completed five blocks of trials for a speaking condition and five blocks of trials for a silent reading condition (Figures 1A,B). The order of the 10 blocks was randomized for each participant. Prior to the start of each block, participants were informed about the condition to be completed in that block (note that each block contained trials for only one condition). Each block consisted of 90 trials. In each trial, a word in white characters was presented against a black background on the computer monitor (Figure 1C). The color of the word changed to green after 600 ms. In the speaking condition, this change of color signaled the go cue for the participant to say the word aloud. Participants ignored the characters' change of color in the silent reading condition. During 40% of the trials in each block (audio trials), auditory stimuli were delivered binaurally through the insert earphones. In the remaining trials (no-audio trials), no auditory stimuli were presented.

The auditory stimulus was either a pure tone (1 kHz; 40 ms duration; 10 ms rise/fall time; 75 dB SPL) or a truncated recording of the syllable /da/ (40 ms duration; 75 dB SPL) spoken by the participant at the beginning of the session. Praat 5.3 (Boersma and Weenink, 2016) was used to record the audio signal from the production of the syllable, and to truncate it to the first 40 ms. The rationale for truncation was to keep the length of the two stimuli the same (tone and syllable), and to maintain consistency with our previous studies (Daliri and Max, 2015a,b). As shown in Figure 1C, auditory stimuli were presented 400 ms after appearance of the target word in white characters (for a discussion on the time-point of stimulus delivery, see Max et al., 2008). The trial ended when the word disappeared from the screen 500 ms after the go signal. The temporal interval between two successive trials was randomly selected from a set of possible intervals (1500, 2000, 2500, 3000, or 3500 ms). The words to be spoken or read in each block of trials were randomly selected from a list containing 300 monosyllabic consonant-



vowel-consonant (CVC) words with no consonant clusters. All words were 3–5 letters long. Psyscope X B53 was used: (a) to present words on the screen; (b) to deliver auditory stimuli to participants; and (c) to send external triggers for the EEG system (see below).

Electroencephalographic Recordings

Using a Biosemi active-electrode EEG system with Ag/AgCl electrodes (Active Two, Biosemi Inc., Amsterdam, Netherlands), EEG signals were recorded from 128 standard sites on the scalp (Figure 1D) according to an extension of the international 10–10 electrode system (Gilmore, 1994; Oostenveld and Praamstra, 2001). Electrooculogram (EOG) signals related to blinking and eye movements were recorded using two electrodes placed on

the outer canthus and below the left eye. Electromyogram (EMG) signals related to orofacial muscle activity were recorded using four electrodes placed on the skin overlying right-side orofacial muscles (masseter: jaw elevation; anterior belly of the digastric: jaw depression; orbicularis oris: lower and upper lip displacement/rounding). Serving as reference electrodes, two additional electrodes were placed on the left and right mastoids. The signals from all electrodes (EEG, EOG, EMG, references) and from an additional microphone placed 15 cm away from the participant's mouth (SM58, Shure, Niles, IL, USA) were continuously recorded at a sampling rate of 1024 Hz.

Data Analysis

The EEGLAB toolbox (Delorme and Makeig, 2004) and custom-written MATLAB scripts (The MathWorks, Inc., Natick, MA, USA) were used for offline data analysis. Signals from the two mastoid electrodes were mathematically averaged to reconstruct a reference signal. All EEG signals were re-referenced to this average mastoid signal and low-pass filtered with a cut-off frequency of 50 Hz using a finite impulse response (FIR) filter (Kaiser windowed sinc FIR filter; deviation: 0.005; transition bandwidth: 1 Hz). The continuous data were then segmented into epochs from 100 ms before the auditory stimulus to 400 ms after the auditory stimulus in audio trials or the equivalent time interval in no-audio trials. To adjust for baseline differences across epochs, the average amplitude of the 100 ms pre-stimulus period for each epoch was subtracted from the whole epoch. Epochs were inspected to reject those with: (a) EEG amplitudes exceeding $\pm 100 \mu\text{V}$; (b) large EOG signals associated with eye movements and blinking; and (c) EMG activity before the *go* signal. Each participant's remaining epochs for the audio trials and no-audio trials from a given condition (either speaking or silent reading) were then averaged separately.

A participant's averaged response for audio trials (tone or syllable) reflected brain activity related to both auditory and non-auditory processing (e.g., activity associated with motor, linguistic, cognitive, and visual processes necessary to complete the task). The averaged response for no-audio trials, on the other hand, reflected only brain activity related to the non-auditory processes. Therefore, to derive LLAEPs that best estimated the actual auditory response, each participant's averaged response for no-audio trials was subtracted from her or his averaged response for audio trials (Martikainen et al., 2005; Bäess et al., 2008, 2011; Luck, 2014; Daliri and Max, 2015a,b). **Figure 1E** illustrates this procedure which was used to derive LLAEPs for all individual electrodes.

As a last step, directly motivated by the results from our prior work (Daliri and Max, 2015a,b), the data from selected electrodes located in three regions of interest (ROIs) were averaged. As illustrated in **Figure 1D**, these ROIs included electrodes over the left hemisphere (Left ROI: electrodes D11, D12, D19 [equivalent to C3], D20, D27, D28), the central region (Central ROI: electrodes A1 [equivalent to Cz], A2, B1, C1, D1, D15), and the right hemisphere (Right ROI: electrodes B17, B18, B22 [equivalent to C4], B23, B30, B31). Our previous work (Daliri and Max, 2015b) indicated that auditory modulation is larger at electrodes located over the central region and over the left

hemisphere than over the right hemisphere. Thus, we used the same ROIs for the present study as this allows an examination of the consistency of such ROI effects, if any, on the auditory modulation phenomenon. The evoked response obtained for each ROI was further low-pass filtered (cut-off frequency 15 Hz; Kaiser windowed sinc FIR filter; deviation: 0.005; transition bandwidth: 1 Hz) before the peak amplitude and peak latency of the N100 and P200 components were extracted. N100 was defined as the largest negative peak between 70 and 130 ms, and P200 was defined as the largest positive peak between 150 and 250 ms. We used a custom written MATLAB script together with visual verification to detect N100 and P200 peaks for each individual participant in each of the conditions (Luck, 2014).

Statistical Analyses

IBM SPSS Statistics 19 (IBM, Armonk, NY, USA) was used to conduct the statistical analysis. For each dependent variable, analysis of variance (ANOVA) for repeated measures was used with Condition (speaking and silent reading), Stimulus (tone and syllable), and ROI (left ROI, central ROI, and right ROI) as the repeated measures. To account for potential violations of the sphericity assumption, degrees of freedom were adjusted using the Huynh-Feldt correction (Max and Onghena, 1999). As appropriate, repeated measures of ANOVA were followed up by *post hoc* analyses using *t*-tests with Bonferroni corrections for multiple comparisons.

RESULTS

Figure 2 shows grand average LLAEP (averaged over all subjects) waveforms from the central ROI when tones (A) or syllables (B) were presented during speech planning and during silent reading. As illustrated, the LLAEP amplitudes for both tones and syllables are reduced in the speaking condition as compared with the silent reading condition. Statistical results for measures of N100 and P200 amplitude and latency are described in the following sections.

N100 Amplitude

Figure 2C shows N100 amplitudes in the speaking and reading conditions for both tones and syllables. The magnitude of N100 amplitude modulation (i.e., amplitude in the reading condition minus amplitude in the speaking condition) for tones vs. syllables is shown in **Figure 2D**.

We found a statistically significant main effect of Stimulus, $F_{(1,8)} = 13.927$, $p = 0.006$, with a larger N100 amplitude in response to syllables vs. tones. We also found statistically significant main effects of Condition, $F_{(1,8)} = 35.601$, $p < 0.001$, and ROI, $F_{(1.562,12.498)} = 57.199$, $p < 0.001$, as well as a Condition \times ROI interaction, $F_{(2,16)} = 9.602$, $p = 0.002$. This interaction occurred because the magnitude of N100 modulation (i.e., the difference in N100 amplitude between the speaking and reading conditions) was larger in the central ROI than in the left ROI, $t_{(8)} = -3.401$, $p = 0.009$, or the right ROI, $t_{(8)} = -4.148$, $p = 0.003$. Most relevant to the hypothesis under investigation, however, is that the Stimulus \times Condition

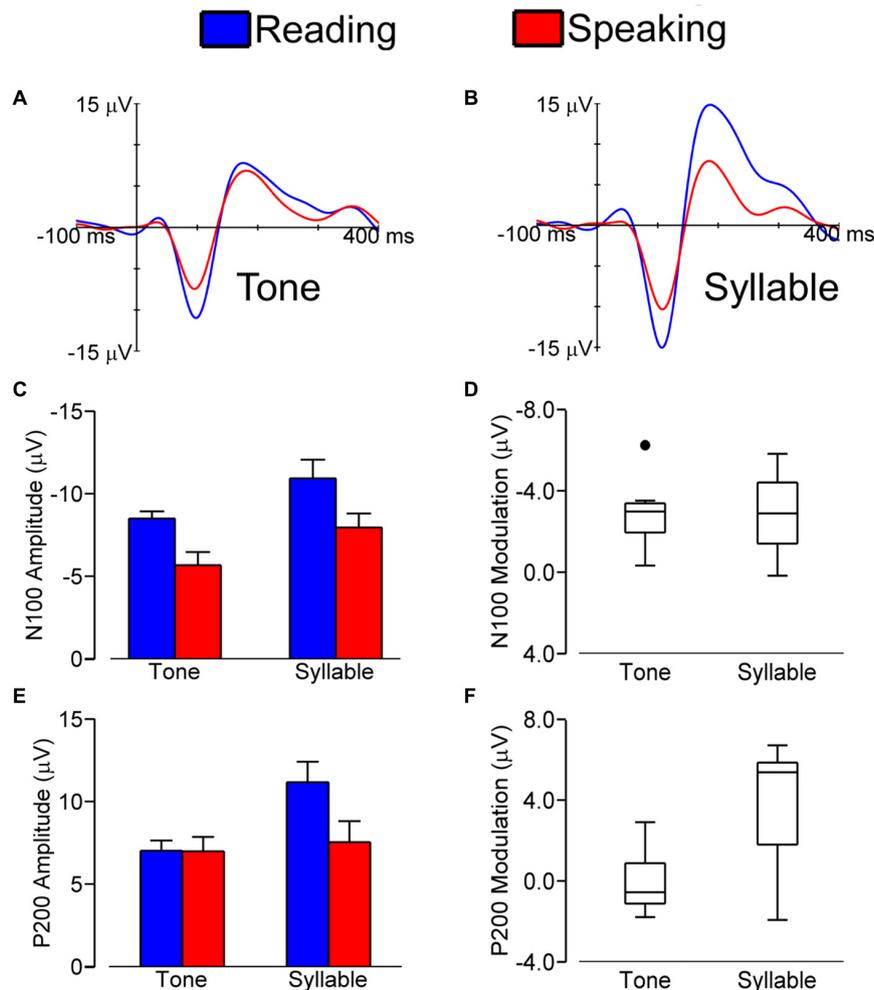


FIGURE 2 | Grand average (over all subjects) LLAEPs in response to tones (A) and syllables (B) for the central ROI in the speaking (red) and the silent reading (blue) conditions. Bar graphs show the group average N100 amplitude (C) and P200 amplitude (E), with error bars indicating standard errors. Each bar represents data averaged across participants and ROIs. Box plots show the distribution of N100 modulation (D) and P200 modulation (F). We found similar N100 amplitude modulation for tones and syllables. For P200 amplitude on the other hand, statistically significant modulation was found for responses elicited by syllables and not for responses elicited by tones.

interaction was not statistically significant ($p = 0.841$). Thus, with regard to N100 amplitude, pre-speech auditory modulation did not differ for speech vs. non-speech stimuli. Lastly, we also found no statistically significant interactions of Stimulus \times ROI ($p = 0.074$) or Condition \times Stimulus \times ROI ($p = 0.097$).

N100 Latency

N100 latency data showed a statistically significant main effect of Stimulus, $F_{(1,8)} = 35.051$, $p < 0.001$, and a statistically significant Stimulus \times Condition interaction, $F_{(1,8)} = 5.450$, $p = 0.048$. The N100 latency was longer for responses to syllables than responses to tones (110.7 ms vs. 98.2 ms), and the difference between N100 latencies for syllables vs. tones was larger in the speaking condition than in the reading condition (15.7 ms vs. 9.3 ms). None of the other main effects (Condition and ROI) or interactions (Condition \times ROI, Stimulus \times ROI,

and Condition \times Stimulus \times ROI) were statistically significant ($p > 0.069$ in all cases).

P200 Amplitude

Figure 2E illustrates P200 amplitudes in both conditions (speaking and reading) and in response to both stimuli (tone and syllable); the magnitudes of the modulation of P200 amplitudes in response to tones and syllables are shown in Figure 2F.

P200 amplitude data showed statistically significant main effects of Condition, $F_{(1,8)} = 7.884$, $p = 0.023$, Stimulus, $F_{(1,8)} = 7.938$, $p = 0.023$, and ROI, $F_{(1,460,11.681)} = 30.675$, $p < 0.001$. These effects were modified by statistically significant interactions of Condition \times ROI, $F_{(1,572,12.218)} = 19.917$, $p < 0.001$, Stimulus \times ROI, $F_{(2,16)} = 6.684$, $p = 0.008$, Condition \times Stimulus, $F_{(1,8)} = 11.834$, $p = 0.008$, and Condition \times Stimulus \times ROI, $F_{(1,438,11.507)} = 4.579$, $p = 0.044$.

Most important for the aim of the present study, these results showed that only for the auditory responses to syllables—and not those to tones—the P200 amplitude in the speaking condition was statistically significantly smaller than the P200 amplitude in the reading condition, $t_{(8)} = 3.460$, $p = 0.008$. This difference (i.e., P200 modulation, defined as P200 amplitude in the reading condition minus P200 amplitude in the speaking condition) was larger in the central ROI than in the left ROI, $t_{(8)} = 3.816$, $p = 0.005$, and the right ROI, $t_{(8)} = 4.863$, $p = 0.001$.

P200 Latency

P200 latency data showed no statistically significant main effects of Stimulus ($p = 0.750$), Condition ($p = 0.500$), or ROI ($p = 0.169$). Furthermore, none of the two-way and three-way interactions were statistically significant ($p > 0.086$ in all cases).

DISCUSSION

We previously showed that, during speech planning, the CNS modulates auditory responses to nonspeech, pure tone probe stimuli (Daliri and Max, 2015a,b). We suggested that, during the preparation of speech movements, the CNS uses an efference copy of planned control signals to prime the auditory system for its role in processing the upcoming auditory feedback resulting from execution of those control signals (Daliri and Max, 2015a,b). However, to elucidate this auditory modulation phenomenon's potential role in the monitoring of auditory feedback, it is essential to understand whether speech movement planning differentially affects the auditory system's processing of speech stimuli as compared with nonspeech stimuli. Thus, in the present study, we investigated whether the effect of speech movement planning on auditory processing varies depending on the general characteristics of the auditory probe stimulus. We studied pre-speech modulation of LLAEPs elicited by self-produced, pre-recorded syllables vs. pure tones when these stimuli were presented during speech planning or silent reading. We hypothesized that if pre-speech modulation plays a role in priming the auditory system for its role in monitoring auditory feedback during speech production, then the N100 and P200 components might be differentially affected for pure tone probes as compared with speech probes.

With direct relevance to this main hypothesis, we report three primary findings. First, we replicated again (see also Daliri and Max, 2015a,b) a statistically significant modulation of the auditory N100 amplitude during speech movement planning in comparison with silent reading. Second, as an entirely novel result, we found that the magnitude of this N100 amplitude modulation did not differ for responses evoked by the speech vs. nonspeech stimuli used here. Third, we also found a statistically significant modulation of the auditory P200 amplitude, but this P200 modulation was exclusive to responses evoked by speech stimuli and did not occur for nonspeech stimuli. Thus, the overall results from this study indicate that immediately prior to speech onset, modulation of the auditory system has: (a) a general effect on early auditory processing stages (as evident by similar magnitudes of N100 modulation for responses elicited by speech and nonspeech stimuli); but (b) a speech-

specific effect on later processing stages (as evident by significant modulation of P200 amplitude for response elicited by speech stimuli but not nonspeech stimuli). This pattern of results is consistent with the proposed hypothesis that pre-speech auditory modulation may play a role in priming the auditory system for its role in monitoring auditory feedback during speech production.

Although this study cannot answer the question why the modulating influence of speech planning is general during the early stages of auditory processing (100 ms after stimulus onset) but speech-specific during later stages of auditory processing (200 ms after stimulus onset), we offer two possible, although not necessarily mutually exclusive, explanations. The first explanation is based on differences in the role of such separate stages of auditory processing. It has been suggested previously that, for speech stimuli, early stages of auditory processing encode basic physical aspects of the auditory input whereas later stages are involved in speech-specific, higher-level processing (Näätänen and Picton, 1987; Crowley and Colrain, 2004; Tremblay et al., 2009; Pratt and Lightfoot, 2012). If, during speech planning, the CNS already prepares the auditory system for its role in processing upcoming auditory feedback, it may modulate neuronal populations involved in processing the basic physical properties of any acoustic input as well as neuronal populations only involved in higher levels of processing speech-specific information. Thus, probing the auditory system during the speech planning stage with stimuli that do not have speech-like characteristics (such as our pure tone stimulus) may reveal only the modulation of the former population of neurons (resulting in decreased N100 amplitude prior to speaking) whereas probing with speech stimuli (such as our syllable stimulus) may reveal the modulation of both populations of neurons (resulting in decreases in both N100 and P200 amplitude prior to speaking).

The second proposed explanation relates to the time course of auditory processing. Given that the LLAEP P200 component occurs approximately 100 ms after the N100 component, our P200 measurements are extracted at a time point closer to movement onset. If the CNS incrementally refines its motor commands during the movement planning stage, its modulating signals to auditory cortex may—in parallel—also become more specific over time. Thus, in comparison with the neural generators of the N100 component, the neural generators of the P200 component may receive modulating signals that carry more refined information about the expected input. As we have suggested previously (Daliri and Max, 2015a,b), this hypothesis can be tested empirically by examining pre-speech auditory modulation at different time points relative to movement onset.

Some authors have suggested that the phenomenon of auditory modulation may reflect general attentional processes rather than motor-to-auditory processes (see Jones et al., 2013; Horvath et al., 2015; Schröger et al., 2015a). If so, it could be argued that our present paradigm's use of a delayed-response task (i.e., participants actively withhold a planned utterance until the go signal is presented) reduces auditory attention in the

speaking task. As a result of such reduced attention allocation, the LLAEP amplitudes could also be reduced. However, the findings from other studies as well as aspects of our own overall methodology make an attention-based interpretation of the present results highly unlikely. First, several studies have shown that experimentally manipulating the allocation of attention (i.e., attending to the auditory stimuli vs. attending to the motor task vs. attending to unrelated visual stimuli) does not influence the magnitude of auditory modulation, at least not during movement execution (SanMiguel et al., 2013; Saupe et al., 2013; Timm et al., 2013). Second, our most important result relates to a difference in pre-speech auditory modulation for pure tone vs. speech stimuli. These two types of stimuli were presented in randomized (i.e., non-predictable) order within the trial blocks for both the speaking condition and the silent reading condition. When a participant was planning and withholding a speech response, no cues were available to indicate whether a pure tone or a speech syllable (or no auditory stimulus at all) would be heard during that trial. Thus, auditory attention would have affected the responses to the different stimuli in similar ways.

Lastly, one potential caveat related to our auditory stimuli should be acknowledged. Although the two stimuli had the same duration, intensity, and rate of presentation, there were differences in several physical characteristics, including the rise/fall time, maximum amplitude, temporal envelope and spectral complexity. It could be argued that the differential modulation of auditory processing that is reflected in the P200 component may be a result of, or influenced by, such basic stimulus characteristics. The methodology used here cannot rule out this possibility. Future research along these lines should include studies examining the effects on auditory modulation of stimuli that differ in only one of these characteristics.

REFERENCES

- Bäess, P., Horváth, J., Jacobsen, T., and Schröger, E. (2011). Selective suppression of self-initiated sounds in an auditory stream: an ERP study. *Psychophysiology* 48, 1276–1283. doi: 10.1111/j.1469-8986.2011.01196.x
- Bäess, P., Jacobsen, T., and Schröger, E. (2008). Suppression of the auditory N1 event-related potential component with unpredictable self-initiated tones: evidence for internal forward models with dynamic stimulation. *Int. J. Psychophysiol.* 70, 137–143. doi: 10.1016/j.ijpsycho.2008.06.005
- Baumann, S. B., Rogers, R. L., Papanicolaou, A. C., and Saydjari, C. L. (1990). Intersession replicability of dipole parameters from three components of the auditory evoked magnetic field. *Brain Topogr.* 3, 311–319. doi: 10.1007/bf01135440
- Beal, D. S., Cheyne, D. O., Gracco, V. L., Quraan, M. A., Taylor, M. J., and De Nil, L. F. (2010). Auditory evoked fields to vocalization during passive listening and active generation in adults who stutter. *Neuroimage* 52, 1645–1653. doi: 10.1016/j.neuroimage.2010.04.277
- Behroozmand, R., and Larson, C. R. (2011). Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. *BMC Neurosci.* 12:54. doi: 10.1186/1471-2202-12-54
- Boersma, P., and Weenink, D. (2016). *Praat: Doing Phonetics by Computer [Computer Program]*. Available online at: <http://www.praat.org/>.
- Chang, E. F., Niziolek, C. A., Knight, R. T., Nagarajan, S. S., and Houde, J. F. (2013). Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proc. Natl. Acad. Sci. U S A* 110, 2653–2658. doi: 10.1073/pnas.1216827110
- Creutzfeldt, O., Ojemann, G., and Lettich, E. (1989). Neuronal-activity in the human lateral temporal-lobe. II. Responses to the subjects own voice. *Exp. Brain Res.* 77, 476–489. doi: 10.1007/bf00249601
- Crowley, K. E., and Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: age, sleep and modality. *Clin. Neurophysiol.* 115, 732–744. doi: 10.1016/j.clinph.2003.11.021
- Curio, G., Neuloh, G., Numminen, J., Jousmaki, V., and Hari, R. (2000). Speaking modifies voice-evoked activity in the human auditory cortex. *Hum. Brain Mapp.* 9, 183–191. doi: 10.1002/(sici)1097-0193(200004)9:43.0.co;2-z
- Daliri, A., and Max, L. (2015a). Electrophysiological evidence for a general auditory prediction deficit in adults who stutter. *Brain Lang.* 150, 37–44. doi: 10.1016/j.bandl.2015.08.008
- Daliri, A., and Max, L. (2015b). Modulation of auditory processing during speech movement planning is limited in adults who stutter. *Brain Lang.* 143, 59–68. doi: 10.1016/j.bandl.2015.03.002
- Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009
- Desantis, A., Weiss, C., Schütz-Bosbach, S., and Waszak, F. (2012). Believing and perceiving: authorship belief modulates sensory attenuation. *PLoS One* 7:e37959. doi: 10.1371/journal.pone.0037959
- Eliades, S. J., and Wang, X. (2003). Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. *J. Neurophysiol.* 89, 2194–2207. doi: 10.1152/jn.00627.2002

AUTHOR CONTRIBUTIONS

AD and LM designed and conducted the experiment. AD analyzed the data. AD and LM interpreted the results and wrote the manuscript.

ACKNOWLEDGMENTS

This research was supported, in part, by grants R01DC007603 and P30DC004661 from the National Institute on Deafness and Other Communication Disorders and grant MOP-137001 from the Canadian Institutes of Health Research. The first author was also supported by a Graduate Student Scholarship from American Speech-Language-Hearing Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies. The authors thank Tamar Mentzer for assistance with subject recruitment and data collection.

- Ford, J. M., Mathalon, D. H., Kalba, S., Whitfield, S., Faustman, W. O., and Roth, W. T. (2001). Cortical responsiveness during talking and listening in schizophrenia: an event-related brain potential study. *Biol. Psychiatry* 50, 540–549. doi: 10.1016/s0006-3223(01)01166-0
- Gilmore, R. L. (1994). American-electroencephalographic-society guidelines in electroencephalography, evoked-potentials and polysomnography. *J. Clin. Neurophysiol.* 11, 1–142.
- Godey, B., Schwartz, D., de Graaf, J., Chauvel, P., and Liégeois-Chauvel, C. (2001). Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: a comparison of data in the same patients. *Clin. Neurophysiol.* 112, 1850–1859. doi: 10.1016/s1388-2457(01)00636-8
- Greenlee, J. D., Jackson, A. W., Chen, F., Larson, C. R., Oya, H., Kawasaki, H., et al. (2011). Human auditory cortical activation during self-vocalization. *PLoS One* 6:e14744. doi: 10.1371/journal.pone.0014744
- Hari, R., Pelizzone, M., Mäkelä, J. P., Hällström, J., Leinonen, L., and Lounasmaa, O. V. (1987). Neuromagnetic responses of the human auditory cortex to on-and offsets of noisebursts. *Audiology* 26, 31–43. doi: 10.3109/00206098709078405
- Heinks-Maldonado, T. H., Nagarajan, S. S., and Houde, J. F. (2006). Magnetoencephalographic evidence for a precise forward model in speech production. *Neuroreport* 17, 1375–1379. doi: 10.1097/01.wnr.0000233102.43526.e9
- Horvath, P., Oliver, S. R., Zaldivar, F. P., Radom-Aizik, S., and Galassetti, P. R. (2015). Effects of intravenous glucose and lipids on innate immune cell activation in healthy, obese and type 2 diabetic subjects. *Physiol. Rep.* 3:e12249. doi: 10.14814/phy2.12249
- Houde, J. F., Nagarajan, S. S., Sekihara, K., and Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: an MEG study. *J. Cogn. Neurosci.* 14, 1125–1138. doi: 10.1162/089892902760807140
- Hyde, M. (1997). The N1 response and its applications. *Audiol. Neurootol.* 2, 281–307. doi: 10.1159/000259253
- Jones, A., Hughes, G., and Waszak, F. (2013). The interaction between attention and motor prediction. An ERP study. *Neuroimage* 83, 533–541. doi: 10.1016/j.neuroimage.2013.07.004
- Liotti, M., Ingham, J. C., Takai, O., Paskos, D. K., Perez, R., and Ingham, R. J. (2010). Spatiotemporal dynamics of speech sound perception in chronic developmental stuttering. *Brain Lang.* 115, 141–147. doi: 10.1016/j.bandl.2010.07.007
- Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique*. 2nd edn. Cambridge, MA: MIT press.
- Mäkelä, J. P., and Hari, R. (1990). Long-latency auditory evoked magnetic fields. *Adv. Neurol.* 54, 177–191.
- Martikainen, M. H., Kaneko, K., and Hari, R. (2005). Suppressed responses to self-triggered sounds in the human auditory cortex. *Cereb. Cortex* 15, 299–302. doi: 10.1093/cercor/bhh131
- Max, L., Daniels, J., Curet, K., and Cronin, K. (2008). “Modulation of auditory and somatosensory processing during the planning of speech movements,” in *Paper presented at the Proceedings of the 8th International Seminar on Speech Production* (Strasbourg, France), 41–44.
- Max, L., and Onghena, P. (1999). Some issues in the statistical analysis of completely randomized and repeated measures designs for speech, language and hearing research. *J. Speech Lang. Hear. Res.* 42, 261–270. doi: 10.1044/jslhr.4202.261
- Mock, J. R., Foundas, A. L., and Golob, E. J. (2011). Modulation of sensory and motor cortex activity during speech preparation. *Eur. J. Neurosci.* 33, 1001–1011. doi: 10.1111/j.1460-9568.2010.07585.x
- Mock, J. R., Foundas, A. L., and Golob, E. J. (2015). Speech preparation in adults with persistent developmental stuttering. *Brain Lang.* 149, 97–105. doi: 10.1016/j.bandl.2015.05.009
- Muller-Preuss, P., and Ploog, D. (1981). Inhibition of auditory cortical neurons during phonation. *Brain Res.* 215, 61–76. doi: 10.1016/0006-8993(81)90491-1
- Näätänen, R., and Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology* 24, 375–425. doi: 10.1111/j.1469-8986.1987.tb00311.x
- Nelson, A., Schneider, D. M., Takatoh, J., Sakurai, K., Wang, F., and Mooney, R. (2013). A circuit for motor cortical modulation of auditory cortical activity. *J. Neurosci.* 33, 14342–14353. doi: 10.1523/jneurosci.2275-13.2013
- Nizielek, C. A., Nagarajan, S. S., and Houde, J. F. (2013). What does motor efference copy represent? evidence from speech production. *J. Neurosci.* 33, 16110–16116. doi: 10.1523/jneurosci.2137-13.2013
- Oostenveld, R., and Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clin. Neurophysiol.* 112, 713–719. doi: 10.1016/s1388-2457(00)00527-7
- Poulet, J. F., and Hedwig, B. (2002). A corollary discharge maintains auditory sensitivity during sound production. *Nature* 418, 872–876. doi: 10.1038/nature00919
- Poulet, J. F., and Hedwig, B. (2007). New insights into corollary discharges mediated by identified neural pathways. *Trends Neurosci.* 30, 14–21. doi: 10.1016/j.tins.2006.11.005
- Pratt, H., and Lightfoot, G. (2012). “Physiologic mechanisms underlying MLRs and cortical EPs,” in *Translational Perspectives in Auditory Neuroscience: Hearing Across the Life Span Assessment and Disorders*, eds K. Tremblay and R. Burkard (San Diego, CA: Plural Publishing), 243–282.
- SanMiguel, I., Todd, J., and Schröger, E. (2013). Sensory suppression effects to self-initiated sounds reflect the attenuation of the unspecific N1 component of the auditory ERP. *Psychophysiology* 50, 334–343. doi: 10.1111/psyp.12024
- Sato, A. (2008). Action observation modulates auditory perception of the consequence of others’ actions. *Conscious. Cogn.* 17, 1219–1227. doi: 10.1016/j.concog.2008.01.003
- Sato, A. (2009). Both motor prediction and conceptual congruency between preview and action-effect contribute to explicit judgment of agency. *Cognition* 110, 74–83. doi: 10.1016/j.cognition.2008.10.011
- Saupe, K., Widmann, A., Trujillo-Barreto, N. J., and Schröger, E. (2013). Sensorial suppression of self-generated sounds and its dependence on attention. *Int. J. Psychophysiol.* 90, 300–310. doi: 10.1016/j.ijpsycho.2013.09.006
- Scherg, M., Vajsaar, J., and Picton, T. W. (1989). A source analysis of the late human auditory evoked potentials. *J. Cogn. Neurosci.* 1, 336–355. doi: 10.1162/jocn.1989.1.4.336
- Schneider, D. M., Nelson, A., and Mooney, R. (2014). A synaptic and circuit basis for corollary discharge in the auditory cortex. *Nature* 513, 189–194. doi: 10.1038/nature13724
- Schröger, E., Kotz, S. A., and SanMiguel, I. (2015a). Bridging prediction and attention in current research on perception and action. *Brain Res.* 1626, 1–13. doi: 10.1016/j.brainres.2015.08.037
- Schröger, E., Marzecová, A., and SanMiguel, I. (2015b). Attention and prediction in human audition: a lesson from cognitive psychophysiology. *Eur. J. Neurosci.* 41, 641–664. doi: 10.1111/ejn.12816
- Steinschneider, M., and Dunn, M. (2002). “Electrophysiology in developmental neuropsychology,” in *Handbook of Neuropsychology*, eds S. Segalowitz and I. Rapin (Amsterdam, Netherlands: Elsevier), 91–146.
- Suga, N., and Schlegel, P. (1972). Neural attenuation of responses to emitted sounds in echolocating bats. *Science* 177, 82–84. doi: 10.1126/science.177.4043.82
- Suga, N., and Shimozaawa, T. (1974). Site of neural attenuation of responses to self-vocalized sounds in echolocating bats. *Science* 183, 1211–1213. doi: 10.1126/science.183.4130.1211
- Timm, J., SanMiguel, I., Saupe, K., and Schröger, E. (2013). The N1-suppression effect for self-initiated sounds is independent of attention. *BMC Neurosci.* 14:1. doi: 10.1186/1471-2202-14-2
- Towle, V. L., Yoon, H., Castelle, M., Edgar, J. C., Biassou, N. M., Frim, D. M., et al. (2008). ECoG gamma activity during a language task: differentiating expressive and receptive speech areas. *Brain* 131, 2013–2027. doi: 10.1093/brain/awn147
- Tremblay, K. L., Shahin, A. J., Picton, T., and Ross, B. (2009). Auditory training alters the physiological detection of stimulus-specific cues in humans. *Clin. Neurophysiol.* 120, 128–135. doi: 10.1016/j.clinph.2008.10.005

- Waszak, F., Cardoso-Leite, P., and Hughes, G. (2012). Action effect anticipation: neurophysiological basis and functional consequences. *Neurosci. Biobehav. Rev.* 36, 943–959. doi: 10.1016/j.neubiorev.2011.11.004
- Weiss, C., Herwig, A., and Schütz-Bosbach, S. (2011). The self in action effects: selective attenuation of self-generated sounds. *Cognition* 121, 207–218. doi: 10.1016/j.cognition.2011.06.011
- Zouridakis, G., Simos, P. G., and Papanicolaou, A. C. (1998). Multiple bilaterally asymmetric cortical sources account for the auditory N1m component. *Brain Topogr.* 10, 183–189. doi: 10.1023/A:1022246825461

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Daliri and Max. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.