Auditory predictions and prediction errors in response to self-initiated vowels

Running title: Prediction error indicates vowel type

Franziska Knolle[1], Michael Schwartze[2], Erich Schröger[3], and Sonja A. Kotz[2]

[1]Department of Psychiatry, Cambridge, UK; [2]Department of Neuropsychology and Psychopharmacology, Maastricht University; [3]Institute of Psychology, Leipzig University, Leipzig, Germany

Corresponding Author:

Franziska Knolle
Department of Psychiatry
University of Cambridge
Douglas House
18B Trumpington Road, Cambridge, CB2 8AH
fk291@cam.ac.uk
+44 1223 465 247

Acknowledgement: The research was funded by a DFG-Reinhart-Koselleck grant to ES and a DFG KO 2268/6-1 grant to SAK. We thank K. Ina Koch and Eleni Beyer for their support in data collection. Furthermore, we thank Kerstin Flake for graphics support and Helga Smallwood for proofreading.

Abstract

It has been suggested that speech production is accomplished by an internal forward model, reducing processing activity directed to self-produced speech in the auditory cortex. The current study uses an established N1-suppression paradigm comparing self- and externally-initiated natural speech sounds to answer two questions:

(1) Are forward predictions generated to process complex speech sounds, such as vowels, initiated via a button press?

(2) Are prediction errors regarding self-initiated deviant vowels reflected in the corresponding ERP components?

Results confirm an N1-suppression in response to self-initiated speech sounds. Furthermore, our results suggest that predictions leading to the N1-suppression effect are specific, as self-initiated deviant vowels do not elicit an N1-suppression effect. Rather, self-initiated deviant vowels elicit an enhanced N2b and P3a compared to externally-generated deviants, externally-generated standard, or self-initiated standards, again confirming prediction specificity.

Results show that prediction errors are salient in self-initiated auditory speech sounds, which may lead to more efficient error correction in speech production.

Keywords: N1 attenuation, self-generated speech, vowels, novelty, forward prediction, prediction error

Introduction

Speaking is a highly complex human capacity: It does not only involve a motor act, but also leads to the perception and the monitoring of one's own voice. The distinction between self-produced speech from speech of others is proposed to be accomplished by a "motor-to-sensory discharge" (Paus, Perry, Zatorre, Worselyy, & Evans, 1996) or an internal forward model (Hickok, 2012; Tian & Poeppel, 2010; Ventura, Nagarajan, & Houde, 2009). The idea of an internal forward model suggests that an efference copy (von Holst & Mittelstädt, 1950) of a motor act is generated that predicts its sensory consequences (Wolpert, Ghahramani, & Jordan, 1995). The prediction prepares a respective cortical area to perceive the predicted sensory input. Consequently, brain activity directed to incoming sensation is suppressed (Chen et al, 2011).

Interestingly, the suppression effect has been reported in many vocalization studies. It was shown that speech production elicits smaller event-related potentials (ERPs) or fields (ERFs) than passively perceived speech (Curio, Neuloh, Numminen, Jousmaki, & Hari, 2000; Ford, Gray, Faustman, Roach, & Mathalon, 2007; Gunji, Hoshiyama, & Kakigi, 2001; Numminen & Curio, 1999; Numminen, Salmelin, & Hari, 1999; Ott & Jäncke, 2013; Ventura et al., 2009). Based on a non-human primate study (Müller-Preuss & Ploog, 1981; replication and extension of non-human primate investigations by Eliades & Wang, 2003), Creutzfeldt and colleagues (1989) recorded intracranial neuronal activity from the right and left superior, middle and inferior temporal gyri in patients undergoing surgery for epilepsy. Results revealed suppressed activity in response to vocalization. Relatedly, Chen and colleagues (2011) conducted an electrocorticography (ECoG) study during human vocalization. They reported neural phase synchrony in the gamma band between Broca's area and the auditory cortex. This phase synchrony that preceded a speaker's speech onset was greater during vocalizing than when listening to their own speech passively (i.e., pre-recorded), indicating that phase synchrony in the gamma band between the two brain regions may describe the transmission of a motor efference copy.

In a PET-study, Hirano and colleagues (1996, 1997) found strong cerebellar activation when investigating self-produced speech that was either normal or distorted (i.e., delayed or changed in pitch), possibly indicating the role of the cerebellum in generating internal forward predictions

based on an efference copy. Along similar lines, we have shown that the cerebellum is involved in generating motor-to-auditory predictions when processing self-initiated sounds (Knolle, Schröger, Baess, & Kotz, 2012; Knolle, Schröger, & Kotz, 2013a). We utilized a N1-suppression paradigm (Schäfer & Marcus, 1973) to compare self-initiated (via finger tap) with externally-generated sinusoidal tones. We found that patients with focal cerebellar lesions did not show a significant N1-suppression effect in response to self-initiated sounds. This indicates that the cerebellum is involved in generating auditory forward predictions – or their (sensory) attenuation.

In a further study (Knolle, Schröger & Kotz, 2013b), we compared self- and externally-generated sounds including 30% unexpected deviant, or 'oddball', sounds (i.e., sounds altered in frequency). We investigated the violation of a prediction when processing a self-generated deviant sound; furthermore we investigated the reflection of such prediction errors in the ERP. The results revealed that precise predictions concerning the acoustic features of a self-generated sound were formed (i.e., modulated N1 suppression). The result was supported by an enhancement of "auditory responsiveness" towards a self-generated deviant shown in an increase in amplitude of the N2b and P3a components. The increased saliency of self-generated deviant sounds may lead to more efficient processing compared to externally-generated deviants. In other words, the increased salience of certain stimuli may reverse the sensory attenuation induced by agency. Thus, the study provides a more complete picture concerning auditory forward prediction and prediction errors with respect to self-generated sinusoidal sounds. Note that the detection of prediction errors is not in the service of detecting the content or cause of a stimulus – but in evaluating how predictable that stimulus was. This is a subtle but important point, which speaks to the predictive processing of the precision or predictability of different stimuli in different settings.

Based on this study the question arose, if precise predictions are also generated in response to complex auditory stimuli, such as speech sounds. Former studies (Behroozmand, Liu, & Larson, 2011; Behroozmend, Sangtina, Korzyukov, & Larson, 2016; Christoffels, van de Ven, Formisano, & Schiller, 2011; Fu et al., 2006; Heinks-Maldonado, Mathalon, Gray, & Ford, 2005; Heinks-

Maldonado, Nagarajan, & Houde, 2006) which compared natural and altered self-produced vocalizations suggest that this is indeed the case. They reported a reduced N1-suppression effect in response to altered compared to unchanged vocal feedback, indicating the generation of a precise prediction. Although altered auditory feedback creates prediction errors, these studies did not aim to investigate the detection of deviance (i.e., altered auditory feedback), nor have they discussed feedback alterations in terms of prediction errors. Although these studies used a carefully conducted design which aimed at reducing as many confounding factors as possible (i.e., participants were asked to hold head, jaw, and tongue in a stationary position to reduce muscle contraction to a minimum), an inherent limitation of these studies is that it is impossible to control for motor activity induced by self-produced vocalization. Hence, a motor command is conducted in order to produce the sound. Based on the motor command, a sensory input is predicted. Consequently, the motor command may influence the ERP in response to the auditory output. Thus, we believe that an appropriate motor control condition (e.g., internal sound production) is necessary in order to control for such effects in the auditory ERPs. Furthermore, Heinks-Maldonado et al. (2005) point out that the differences in sound quality between a speaking and a listening condition could substantially influence the results, and possibly create the suppression effect. Nonetheless, these findings provide support the notion that predictions are generated based on precise patterns, also capturing complex auditory stimuli. However, as the motor-to-auditory links are much tighter with self-vocalization than the motor-to-auditory links involved when generating a sound via a finger tap, it is of interest whether the results obtained with self-vocalization can be generalized to other forms of self-generation of speech sounds.

In contrast to previous studies, the current study investigated whether predictions are formed on a precise patterns also in response to manually initiated complex, natural sounds (i.e., speech sounds), and whether the generation of a precise prediction leads to differential processing of self- compared to externally-produced complex deviant stimuli, seen in prediction errors specific to stimuli type. In a first experiment, a standard N1-suppression paradigm was used which has been studied extensively using click sounds, sinusoidal sounds or complex instrumental sounds (Baess, Jacobson, & Schröger, 2008; Baess, Horváth, Jacobson, & Schröger, 2011; Knolle et al.,

2012; Lange, 2011; Martikainen, Kaneko, & Hari, 2005 (MEG); McCarthy & Donchin, 1976; Schäfer & Marcus, 1973). Here, we compared complex, natural vowels that were self-initiated via a finger tap to the same vowels, externally-produced (i.e., pre-recorded vowels /a:/), similarly to a recent study by Pinheiro and colleagues (Pinheiro, Schwartze & Kotz, 2018). This first experiment examined whether the N1-suppression paradigm was applicable to complex stimuli, such as speech sounds in a highly controlled setup. Based on our preceding studies (Knolle et al., 2012; Knolle et al., 2013a) investigating sinusoidal tones in a N1-suppression paradigm, we expected to find a N1-suppression followed by a P2-reduction in response to self-initiated vowels. Whereas the N1-suppression may reflect the unconscious, automatic formation of a prediction, preparing the auditory cortex to receive sensory input, the P2-reduction may reveal a later, more conscious processing stage of the generation of a prediction (i.e., the conscious detection of a self-initiated vowel).

In a second experiment, we investigated the violation of a precise prediction regarding self-generated complex speech sounds by introducing surprising, unpredictable events, and furthermore we explored the reflection of such prediction errors in the corresponding ERP components. Thus, we modified the N1-suppression paradigm of experiment one corresponding to a former study on self- and externally-generated deviant sounds (Knolle et al., 2013b). We compared self-initiated and externally-produced natural vowels, of which 30% were altered in quality (either /a:/ which is an open front unrounded vowel, or /o:/ which is a mid-close back rounded vowel), creating self-initiated and externally-produced deviants (Knolle et al., 2013b). If precise predictions were generated to process self-initiated vowels, the N1-suppression effect should be modified when a self-initiated deviant vowel is elicited. Furthermore, if self-initiated deviant vowels which created a prediction error were more salient than externally-produced deviant vowels they should reveal an enhanced N2b and P3a (Knolle et al., 2013b). These results would provide additional support for the generation of a precise prediction, which impacts the neural suppression and the detection of prediction errors in self-generated complex speech sounds.

Methods

Participants: Sixteen volunteers (8 females) participated in the current study. All participants were right-handed according to the Edinburgh Handedness Inventory (Oldfield, 1971). The mean age was 24.9 years (SD: 1.8 years) and ranged from 23 to 27 years. Participants were students of the University of Leipzig and were recruited via the participants' database of the Max-Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany. None of the participants reported any neurological dysfunction, but normal or corrected-to-normal visual acuity, and normal hearing. All participants gave their written informed consent and were paid for their participation. The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the Leipzig University.

Speech Stimuli: In order to obtain individual vowels, we recorded vowel samples from all participants using the program AlgoRec™ TerraTec Edition. Thus, throughout the experiment each participant listened to the vowels that they had previously produced. We asked the participants to produce "ah" and "oh" as in the German words /a:bɐ/ (aber; engl. but) and /o:bɐ/ (Ober; engl. waiter). We recorded 10 trials for each vowel per participant. For each participant we picked the vowel that best matched the average characteristics within the 10 trials. Using Praat 5.2.03 (1992-2010 by Paul Boersma and David Weenink; University of Amsterdam, Amsterdam, NL) we applied minimal normalization procedures to duration, intensity and pitch to maintain natural sound quality. The vowel duration of /a:/ was approximately 360.69ms (SD: 65.66ms) and pitch around 77.44 Hz (SD: 3.80 Hz). The vowel /o:/ had an average duration of 366.31ms (SD: 77.75ms) and an average pitch of 80.76 Hz (SD: 3.84 Hz). For both vowels, the sound intensity was calibrated at about 80 dB SPL and an average loudness of 70 dB was maintained by all subjects.

Experimental conditions – Experiment 1: The first experiment contained two experimental conditions and one control condition (Figure 1, white background). In the vowel-motor condition (VMC-1) participants induced finger taps about every 2.4s (see Knolle et al., 2012 for a detailed description of the paradigm). Each tap elicited an immediate presentation of the vowel /a:/ (delay of 2-4ms due to the loading of the stimuli) via headphones. The acoustic stimulation, including

self-initiated vowels, was recorded online and used as an 'external vowel sequence' in the vowel-only condition (VOC-1). During VOC-1 participants did not produce finger taps, but were simply asked to listen and attend to the vowels. Lastly, participants carried out a motor-only condition (MOC-1), in which they also performed self-paced finger taps every 2.4s. However, in contrast to VMC-1, no sound was induced via the finger tap. This condition controlled for motor activity in VMC-1.

Experimental conditions – Experiment 2: In experiment 2, two experimental and one control condition were presented (Figure 1b). In the vowel-motor condition (VMC-2) participants induced finger taps about every 2.4s. Each tap elicited an immediate presentation of either the vowel /a:/ or /o:/ via headphones. In 30% of the taps a deviant was presented. If the standard stimulus was the vowel /a:/, the deviant was the vowel /o:/ and vice versa. The timing of the acoustic stimulation was recorded online. This information was used to produce an identical but externally generated vowel sequence in the vowel-only condition (VOC-2). Thus, participants received exactly the same set of stimuli in both experimental conditions. During VOC-2 participants did not produce finger taps, but were simply asked to listen attentively to the auditory stimuli. As in experiment 1 MOC-2 served as a control condition for motor activity in VMC-2.

Both experimental runs were preceded by two training blocks each. In the first block, participants practiced to tap every 2.4s. The second training block included visual feedback to indicate whether a trial was too slow (tapping interval longer than 3s) or too fast (tapping interval shorter than 1.8s). The feedback ensured that participants had learned to estimate the time between two successive finger taps without counting. Trials outside the range of 1.8s – 2.4s were treated as errors. During the experimental run, no feedback was given.

[Figure 1 about here]

Experimental procedure: Participants were comfortably seated in an electrically shielded and sound-attenuated experimental chamber. A fixation cross was displayed in the middle of a computer screen. To ensure that the motor activity was comparable across participants in the auditory-motor and the motor-only condition, they were instructed to change hands (index finger) whenever indicated on the screen. Hence all participants tapped in equal parts with left and right hand. The order of tapping hands was randomized across participants. Each tap triggered the instantaneous presentation of a vowel via headphones (Sennheiser HD 202) to both ears in VMC-1/2 and VOC-1/2. An in-house built, highly sensitive tapping device was used to record the finger taps. Participants wore headphones in order to cover them up from all sounds possibly emitted by the taps. In the second experiment, the participants performed a combination of VMC-2 and VOC-2: during one run the standard vowel that was triggered via a tap was the vowel /a:/ and the deviant was the vowel /o:/. In the other run, the allocation of standard and deviant vowel was a reversed. The first experiment consisted of 100 trials in each condition. Additionally, we collected 100 trials in MOC-1/2. In the second experiment, we recorded 200 trials in each the VMC-2 and VOC-2 with 70% /a:/ as standard and 30% /o:/ as deviant and vice versa. In total, 700 trials were recorded in both experiments. Experimental conditions were presented in blocks of 100 trials each. Block order was restricted: The VMC-1/2 always preceded VOC-1/2, but the MOC-1/2 was randomized across participants.

Electrophysiological recordings: The electroencephalogram (EEG) was recorded continuously from 59 Ag–AgCl electrodes according to the International 10–20 system. In addition, activity from the left and right mastoids and the sternum (ground electrode) was recorded. The EEG was sampled at a rate of 500Hz (Refa amplifiers system, TMS international, Enschede, NL) and an anti-aliasing filter of 135 Hz was applied. To control for eye movements, vertical and horizontal electrooculograms (EOG) were recorded bipolarly. The impedance of all electrodes was kept below 5 kΩ. The recordings were online referenced to the left mastoid. EEP 3.2.1 Max-Planck-Institute of Cognitive Neuroscience, Leipzig, Germany was used to process the data.

Data Analysis – Behavioral Data: Tapping intervals shorter than 1.8s or longer than 3.0s were treated as errors, and were excluded from further EEG analysis. We acquired tapping intervals

for VMC-1/2 and MOC-1/2 using the Presentation software (Neurobehavioral Systems, Inc., Albany, CA, USA). For each participant we generated the mean length of the tapped interval per condition and the overall performance accuracy (percent correct; ACCURACY) separately for VMC-1/2 and MOC-1/2.

Data Analysis – EEG Data: The EEG data were filtered with a 0.3-15Hz bandpass filter (1601 Hamming windowed filter). The EEG data were re-referenced to linked mastoids. ERPs were time-locked to the stimulus onset of all critical trials. Each analyzed epoch lasted 600 ms including a 100 ms pre-stimulus baseline. The critical epochs were automatically scanned to reject horizontal and vertical eye-movements, muscle artifacts, and electrode drifts. Trials exceeding 30 µV at the eye channels and 40 µV at CZ were rejected. This automatic rejection was corrected manually by applying an eye-movement correction.

We controlled for motor activity by computing a difference-wave between VMC-1/2 and MOC-1/2 to compare the sensory activity elicited in the two experimental conditions. This corrected condition was labeled vowel-corrected condition (VCC-1, VCC-2). In the first experiment, we only compared fully predictable self-initiated and externally-produced vowels (i.e., /a:/). In the second experiment, we investigated two types of self-initiated vowels – standard (VCS-2) and deviant (VCD-2) vowels, as well as two types of externally-produced vowels – self-initiated standard and deviant vowels (VOS-2; VOD-2). As the statistical analysis of the ERP-results in response to standard vowel /a:/ and standard vowel /o:/ as well as to deviant vowel /a:/ and deviant vowel /o:/ did not differ significantly, we combined the standard vowels /a:/ and /o:/ as well as deviant vowel /a:/ and /o:/ for all further statistical analyses.

Group-average ERPs were generated for standards and deviants. In the first experiment, the difference waves (i.e., responses to externally generated minus self-initiated vowels) revealed two ERP responses, a negative one in the time window of the N1 peaking at approximately 90ms, and a positive response in the P2 time window peaking at approx. 190ms. Statistical analyses were calculated based on individual amplitudes, in the time windows of 70-110ms for the N1 and 170-210 ms for the P2. In experiment 2, difference waves revealed two ERP responses for standard vowels: a negative one in the N1 time window, peaking at approx. 90 ms followed by a

positive one in the P2 time window peaking at approx. 190ms; and four ERP responses to deviant vowels: a N1 and P2, as well as an N2b peaking at approx. 170 ms and a later positive response in the time window of the P3a peaking at approx. 310ms. Statistical analyses were calculated based on individual amplitudes in the following time windows: 70-110 ms for the N1, 150-190 ms for the N2b, 170-210 ms for the P2, and 260-360 ms for the P3a. Furthermore, we applied a regions of interest (ROI) analysis using five ROIs (central (ZZ): FZ, CZ, PZ, FCZ, CPZ; left lateral (LL): F7, T7, P7, FT7, TP7; left medial (LM): F3, C3, P3, FC3, CP3; right lateral (RL): F8, T8, P8, FT8, TP8; right medial (RM): F4, C4, P4, FC4, CP4).

Statistical analyses: For the statistical analysis, the SAS 8.20.20 (Statistical Analysis System, SAS Institute Inc., Cary, North Carolina, USA) software package was used. Only significant results are presented and where required, the Greenhouse–Geisser correction was applied. Furthermore, we conducted Bonferroni corrected post-hoc tests.

In the first experiment, comparing self-initiated /a:/-vowels to externally-produced /a:/-vowels, we ran two analyses of variance (ANOVA), one for each ERP component, including the within-subject factors CONDITION (self-initiated: VCC-1; externally-produced: VOC-1) and ROI (RL,RM, ZZ, LM, LL). For the statistical analyses of each individual ERP component in the second experiment, we ran a 2x2x5 ANOVA using the within-subject factors CONDITION (self-initiated: VCC-2; externally-produced: VOC-2), TYPE (standard vs. deviant) and ROI (LL, LM, RM, RL, ZZ).

Results

Behavioral data: In the first experiment, the average tapping interval duration was 2469.79 ms (SD: 360.77 ms) in VMC-1. Furthermore, participants tapped with an overall accuracy of 85.80% (SD: 18.73%). The Kolmogorov-Smirnov-Test revealed no deviance from the normal distribution (p = .18). In the second experiment, the results were similar. The average tapping interval was 2233.15 ms (SD: 257.43 ms) in VMC-2. Furthermore, participants tapped with an overall correctness of 92.92% (SD: 6.06%). The Kolmogorov-Smirnov-Test revealed a normal distribution (p = .82). In MOC-1/2 we found similar results, the average length of the tapping interval was 2388.94 ms (SD: 267.44 ms). Participants performed with an overall accuracy of 93.25% (SD:

8.14%; normal distribution: $p = .52$).

ERP data – experiment 1: The statistical results are presented in Table 1, and summarized below. The CONDITION (self-initiated, externally-initiated) x ROI (LL, LM, ZZ, RM, RR) ANOVA revealed a significant effects in the time window of N1 and P2. In the N1 time window, results confirmed significant differences between the two conditions of self-initiated and externally initiated vowels; as well as significant interaction between conditions and region. Self-initiated vowels elicited a N1-suppression compared to externally-produced vowels (N1: mean amplitude VCC-1: -2.20 µV, mean amplitude VOC-1: -3.09 µV) (Figure 2).

In the P2 time window, similarly to the N1, we found a significant difference between the two conditions and a significant interaction of condition and region, with self-initiated vowels being significantly suppressed compared to externally-produced vowels (P2: mean amplitude VCC-1: 1.73 µV, mean amplitude VOC-1: 3.44 µV) (Figure 2).

Experiment 1 revealed the expected N1-suppression effect as well as a reduced P2 in response to self-initiated vowels.

[Table 1 about here]

[Figure 2 about here]

ERP data – experiment 2: All statistical results are presented in Table 2. For all ERP components we conducted a CONDITION (self-initiated, externally-initiated) x ROI (LL, LM, ZZ, RM, RR) x TYPE (standard, deviant).

In the N1 time window, the ANOVA revealed significant differences between condition and region. According to our hypothesis, we conducted a planned analysis, investigating condition effects within the different vowel types: In the standard vowels (Figure3a) we found a significant

suppression effect in response to self-initiated vowels which was shown by an interaction of condition and region (mean amplitudes standard vowels: VCS-2 mean amplitude: -2.98 µV, VOS-2 mean amplitude: -3.77 µV). In the deviant vowels (Figure 3B), the suppression effect did not reach significance (mean amplitudes deviant vowels: VCD-2 mean amplitude: -3.81 µV, VOD-2 mean amplitude: -4.04 µV).

The results show that we only find a N1-suppression effect in response to self-initiated standard.

[Figure 3 about here]

In the P2 time window, the ANOVA analysis revealed a significant difference between conditions – self-initiated versus externally initiated – and types – standard versus deviants, as well as significant interactions between condition and region aa well as a marginally significant effect between condition and type. In response to standard vowels (Figure3a), we found a significant suppression effect in response to self-initiated vowels compared to externally initiated vowels (Standard: VCS-2 mean amplitude: 1.05 µV, VOS-2 mean amplitude: 2.62 µV) shown by a condition effect as well as a condition by region interaction. Deviant vowels showed a different pattern (Figure3b). In response to externally-generated deviant vowels we found a P2 (VOD-2 mean amplitude: 2.55 µV) that did not differ significantly from externally-produced standards (VOS-2 mean amplitude: 2.62 µV) (Figure 4A). The visual inspection of the ERPs in response to the self-initiated deviants showed a very small shoulder following the N1, which may reflect a P2. However, this component was overlaid by a strong N2b effect, which was elicited in the time window of the P2, and described below.

Taken together, the results show a significant P2-reduction in response to self-initiated standard sounds compared to externally-produced standards. Although externally-produced deviant and standard vowels elicited a similar P2 effect, a possible P2 in response to self-initiated deviant vowels cannot be statistically evaluated due to potential overlay effects.

[Figure 4 about here]

In the N2b time window we found a significant difference between conditions (self-initiated, externally-generated) and types (standard, deviant), as well as a significant interaction between condition and region, type and region, condition and type, and condition, type and region. The post-hoc analysis of condition, revealed a significant difference between vowel types, as well as a significant interaction between vowel type and region, showing that only self-initiated deviant vowels elicited an N2b effect (standard vowels: VCS-2 mean: -.29 µV; deviant vowels: VCD-2 mean: -1.48 µV).

In contrast (Figure 4a), we did not find a significant difference when comparing externally-produced standard and deviant vowels, indicating similar processing effort, also seen in the mean amplitude values (standard vowel: VOS-2 mean: .85 µV; deviant vowels: VOD-2 mean: .69 µV).

In conclusion, the N2b was only elicited in response to self-initiated deviant vowels.

[Table 2 about here]

In the P3a time window (Figure 3b; 4 a&b), we found a significant difference between the vowel types (TYPE $F(1,15) = 22.68$; $p = .0005$). In a post-hoc analysis we resolved the type effect, analyzing standard and deviant sounds separately. We found a significant difference between the two conditions in the deviant vowels (CONDITION $F(1,15) = 5.24$; $p = .037$), revealing a significantly enhanced P3a in response to self-initiated deviant vowels (VCC-2 mean amplitude: 1.60 µV) compared to externally-produced deviant vowels (VOC-2 mean amplitude: .64 µV) (Figure 3b). However, we did not find a significant difference between self-initiated standard

vowels and externally-produced standard vowels (VCC-2 mean amplitude: -.44 µV, VOC-2 mean amplitude: -.48 µV) (Figure 3a). The result suggested that only deviant vowel elicited a significant P3a effect which was significantly enhanced in response to self-initiated compared to externally-produced deviant vowels.

Discussion

The current study investigated the questions whether precise predictions are generated to process self-initiated complex speech sounds (i.e., vowels) by applying an internal forward model and whether prediction errors regarding self-generated deviant vowels are reflected in corresponding ERP components. In order to address these questions, two experiments were conducted. In the first experiment, we used a standard N1-suppression paradigm comparing self- and externally-produced individually pre-recorded vowels in order to test whether the N1-suppression paradigm suffices to investigate complex auditory stimuli such as speech sounds. The results revealed a strong N1-suppression effect in response to self-initiated vowels. This finding confirms the successful generation of a forward prediction independent of the complexity of an anticipated stimulus.

Moreover, we found a reduced P2 response elicited by self-initiated vowels. Although the literature is very diverse regarding the P2, with some studies showing a suppression effect (Houde et al., 2002; Knolle et al., 2012, 2013a and b; Wang et al., 2014), whereas others do not (Baess et al., 2008; Behroozmand, Liu, & Larson, 2011; Chen et al., 2013; Martikainen et al., 2005), the pattern we find in the current data is comparable to our previous results utilizing sounds (Knolle et al., 2012; Knolle et al., 2013a, 2013b). Thus, we suggest that the present data may indicate two processing stages of forming a prediction: Whereas the N1 reflects a fast and automatic forward prediction that prepares the auditory cortex to receive predicted sensory input, the P2 effect represents a more cognitive response (De Chicchis et al., 2002; Crowley & Colrain, 2004), as in distinguishing self- from externally-produced vowels by consciously detecting a self-initiated sensation. This notion is supported by patient studies (Knolle et al.,

2012, 2013a). The results of these studies revealed that patients with cerebellar lesions did not show an N1-suppression effect in response to self-initiated sinusoidal sounds, indicating an inability to generate a fast and automatic forward prediction. However, the patients showed a reduced P2 indicating that they had consciously recognized a respective sound as a self-generated one.

Interestingly, a recent study (Pinheiro et al., 2018) testing non-clinical voice hearers using a similar setup as the first experiment in the current study, also using natural vowel sounds, showed that the N1-supression effect is reversed in non-clinical voice hearers with high symptom scores compared to those with low symptom scores whereas the P2-supression effect is maintained in both groups. This is in accordance with findings in schizophrenia patients (Ford, et al., 2014) using a similar paradigm, which indicates that alterations in generating motor-to-auditory predictions might be linked to developing auditory hallucinations (Brebion et al., 2016; Pinheiro et al., 2017).

In the second experiment, we adapted the N1-suppression paradigm by comparing self- and externally-produced standard and deviant vowels. The deviant vowels (30%) were either an /a:/ or an /o:/ dependent on which of these two vowels represented the standard vowel. Comparable to the results of the first experiment, standard vowels elicited a suppressed N1 and P2 component in response the self-initiated vowels indicating that motor-to-auditory predictions are generated in order to process self-initiated vowels. Self-initiated deviant vowels, on the other hand, did not elicit a clear N1-suppression effect, indicating the violation of a precise prediction. In contrast, externally-produced deviants did not differ from externally-produced standard vowels showing that the difference in the N1-suppression effect was not elicited due to deviant detection. Additionally, our results show that the P2 in response to deviant vowels reveals a more complicated pattern: The P2 elicited by externally-produced deviants is well pronounced and very similar to the P2 in response to externally-produced standard vowels. In contrast, the potential P2 component in response to self-initiated deviant vowels is overlaid by an N2b response (Näatänen et al., 1982; for a review, see Näatänen & Gaillard, 1983), and cannot easily be interpreted.

Furthermore, deviant vowels elicited an N2b effect, which suggests the conscious detection of an unexpected, infrequent stimulus (Horvath et al., 2008; Näätänen et al., 1982). The effect was enhanced in response to self- compared to externally-generated deviants. A further study (Kotz, Stockert, & Schwarzte, 2014) revealed an increased N2b response to deviant sounds. Although Kotz and colleagues (2014) did not use a self-generation paradigm they investigated predictability by changing timing and context information to create different degrees of predictability. Their results show that irregular deviants elicit the biggest N2b response showing the detection of a prediction error. This is in accordance to our finding which provides further support for the notion that precise predictions in response to complex stimuli are generated, because only if a prediction concerning a specific feature of the auditory input (i.e., vowel quality) or temporal structure exists, its violation can be detected faster compared to unpredictable auditory input. As these deviants create a prediction error, the result suggests that prediction errors with respect to self-initiated stimuli are more salient, as a self-initiated stimulus is still temporally predictable. This suggestion is supported by our findings of a P3a response to infrequent, unexpected stimuli, to which attention is drawn (Linden, 2005; Polich, 2007; Snyder & Hillyard, 1976; Squires et al., 1975). Here, we report an enhanced P3a effect in response to self-initiated deviants, providing further evidence for the saliency of self-initiated prediction errors (Ford et al., 2010; Nittono & Ullsperger, 2000).

The results of the current study replicated the pattern of components found in a previous study on deviancy processing in sinusoidal sounds (Knolle et al., 2013b). This strongly suggests that predictive processing and the detection of prediction errors occurs independently of stimulus complexity. Additionally, the results reveal that the suppression paradigm is applicable to complex, speech-like sounds, suggesting that the internal forward model provides a theoretical explanation for processing self-produced speech. It can be postulated that by applying an internal forward model, the amplitude of the N1 is modulated when a speech sound is self-initiated compared to when this same speech sound is externally-triggered (Paus et al., 1996; Ventura et al., 2009). In the same line of thought, many studies, using different methods and paradigms, compared spontaneous, self-produced speech to externally-produced speech. They consistently find that spontaneously self-produced speech elicits suppressed cortical responses compared to

recorded speech (fMRI: Christoffels, Formisano, & Schiller, 2007; Christoffels et al. 2011; Hashmoto & Sakai, 2003; MEG: Aliu, Houde, & Nagarajan, 2009; Curio, et al. 2000; Heinks-Maldonado, Nagarajan, & Houde, 2006; Houde, Nagarajan, Sekihara, & Merzenich, 2002; Kauramäki, Jääskeläinen, Hari, Möttönen, Rauschecker, & Sams, 2010; Ventura et al., 2009; EEG: Ford, Mathalon, Heinks, Kalba, Faustman, & Roth, 2001; Heinks-Maldonalo, Mathalon, Gray, & Ford, 2005).

However as we hypothesized that the N1-suppression effect reflects the precision of the prediction, we modulated the standard paradigm introducing deviant vowels (second experiment) to reduce the prior knowledge concerning an anticipated stimulus and engenders a prediction error. The results show that precise predictions are generated (Bendixen, SanMiguel & Schröger, 2012; Ford & Mathalon, 2012; Knolle et al., 2013b) as the N1-suppression effect is modified in response to self-initiated deviant vowels. We propose that a prediction generated to process a self-initiated vowel holds a concrete representation of the vowel including for example, its frequency, onset, and intensity. This prediction is violated when a deviant is elicited. It is less precise, as it contains incorrect information on a vowel's acoustic quality. However, as the concept is still correct with regard to vowel intensity and temporal occurrence, the suppression effect is maintained but modulated. The vocalization literature investigating altered auditory feedback consistently reports a similar effect. For example, Behroozmand and colleagues (2011) reported a reduced N1-suppression effect in response to self-produced but pitch- or onset-altered vocalizations. Thus, participants formed a precise prediction concerning the temporal and acoustic appearance of their voice. When the auditory feedback was altered in frequency or in onset, the prediction was violated. Consequently, the N1-suppression effect was reduced in response to altered auditory feedback, compared to unchanged feedback. Interestingly, two very similar MEG studies show a reduction of the suppression effect of the N1m with regard to altered self-produced speech sounds (Niziolek, Nagaraja, & Houde, 2013; Ylinen et al., 2014),. In accordance to our interpretation Ylinen and colleagues (2014) argue that the reduced suppression reflects a precise motor-to-auditory forward prediction used for concrete speech monitoring. Based on very similar results, Niziolek and colleagues (2013) argue that specific speech monitoring allows error detection as well as concrete error correction mechanisms, which

is in accordance to our interpretation.

The assumption that precise predictions are generated also in response to complex self-initiated stimuli receives further support from the enhanced N2b effect in response to self-initiated deviant vowels. Similar to our previous study (Knolle et al., 2013b), we consider that the N2b indicates the conscious detection of a prediction error, which represents an infrequent stimulus (Horvath et al., 2008; Näätänen et al., 1982). As the N2b is very much enhanced in response to self-initiated deviant vowels, it can be suggested that a prediction error increased the saliency of a deviant (Ford et al., 2010). More generally speaking, when a prediction error is generated, the detection of such violation is processed more efficiently in self-produced compared to externally-produced deviant vowels. This is supported by our finding that externally-produced deviant vowels also show a reduced P3a compared to self-produced deviants, revealing a less salient response (Ford et al., 2010).

The results concerning the detection of deviants nicely complements the results presented in the literature on speech monitoring and processing of speech errors (Christoffels et al., 2011; Postma, 2000; Tourville, Reilly, & Guenther, 2008; Zheng, Munhall, & Johnsrude, 2010) which most reliably show an increased blood oxygenation level dependent (BOLD) response during altered auditory feedback compared to normal feedback, indicating that increased activity is necessary to accomplish the monitoring effort or error coding. In the current study we show that violated predictions – prediction errors – are reflected in specific ERP components (i.e., enhanced N2b and P3a), which can be compared to error coding in fMRI studies.

In the current study we have investigated motor-to-auditory predictions, framed within a forward model account, which is specific to motor control and uses the concept of a efference copy to generate a prediction. Other current accounts of surprising (e.g., oddball or deviant) responses in a more general, modality-independent framework are usually cast in terms of predictive coding (e.g. Brown et al. 2013; Friston, 2018; Shipp, 2016). In these models, the brain uses a hierarchical forward or generative model inferring the causes of sensations from its sensory consequences. Violations or mismatches from these predicted sensations result in a prediction error that can be differently weighted based on their precision (Haarsma et al., 2018)

and is used for belief updating in cortical hierarchies. How does this concept link to agency? To explain self-produced action, such as speech, current accounts treat motor commands as predictions of proprioceptive and somatosensory consequences of an intended act, while the efference copy or corollary discharge corresponds to the predictions in the exteroceptive (e.g. auditory or visual) domain (Sterzer et al., 2018). Crucially, to act, it is necessary to attenuate the gain or precision of ascending prediction errors. Psychologically, this manifests as sensory attenuation as measured psychophysically. Physiologically, this is usually manifest as an attenuation or suppression of evoked responses that are generated by self, relative to another, as shown in this study.

In conclusion, the present study presents two experiments: The first investigates whether forward predictions are generated to process self-initiated complex speech sounds (i.e., vowel)? And the second addresses the question whether prediction error regarding self-initiated deviant, are these predictions based on precise patterns, as in holding a concrete representation of the anticipated vowel? Addressing the first question, the results revealed N1 and P2 suppression elicited by self-initiated vowels, indicating a successful generation of predictions in response to complex self-initiated speech sounds. This finding supports the notion that processing of self-produced speech mirrors components of a forward model, preparing respective cortical areas for incoming sensory consequences of self-produced speech. Investigating the second question, we found N1-suppression in response to self-initiated vowels compared to externally-generated vowels. In addition, we report an enhanced N2b and P3 effect in response to self-initiated compared to externally-produced deviant vowels. These findings imply that specific predictions are generated. Furthermore, we show that prediction errors are more salient in self-initiated speech sounds compared to externally-produced sounds. More generally, our results speak to a key role of agency in predictive processing formulations; namely, a key role in mediating sensory attention and its reversal when attending to the consequences of self-generated acts.

Figure Caption

Figure 1. Schematic illustration of the three different conditions used in the studies. A) represents the vowel-motor condition (VMC-1/2): A vowel is self-initiated via a finger tap. In experiment 2, 30% of the finger taps elicit a vowel deviant, indicated in the illustration via the diagonal shading. B) represents the vowel-only condition (VOC-1/2): The vowel sequence is presented externally, containing all vowels from the VMC-1/2 accordingly. The diagonal shading presents the inclusion of deviant vowels used in experiment 2. C) illustrates the motor-only control condition (MOC-1/2): taps are required, but no sound is elicited. The motor-only condition is identical in experiment 1 and 2.

Figure 2. Results of the first experiment: ERP responses: Brain responses elicited by self-initiated and externally-produced vowels in the central region. The blue solid line represents externally-produced vowels (VOC-1), whereas the red solid line shows responses elicited by self-initiated vowels (VCC-1). Brain maps: Grand average scalp maps showing the spatial distribution of the difference waves (VOC-1 – VCC-1) in the analyzed N1 and P2 time window.

Figure 3. Results of the second experiment: A) Brain responses elicited by self-initiated and externally-produced standard vowels in the central region. The blue solid line represents externally-produced standard vowels (VOS-2), whereas the red solid line shows responses elicited by self-initiated standard vowels (VCS-2). B) Brain responses elicited by self-initiated and externally-produced deviant vowels also in the central region. The blue solid line shows externally-produced standard vowels (VOD-2), and the red solid line shows responses to self-initiated deviant vowels (VCD-2). The brain maps show the distribution of the effect (for standard vowels: VOS-2 minus VCS-2; for deviant vowels: VOD-2 minus VCD-2). Grey bar reflect analyzed in the time window of the N1, N2b and P3a.

Figure 4. A) Brain responses elicited by externally-produced standard and deviant vowels in the central region. The blue solid line represents externally-produced standard vowels (VOS-2), whereas the red solid line shows responses elicited by externally-produced deviant vowels (VOD-2). B) Brain responses elicited by self-initiated standard and deviant vowels also in the central

region. The blue solid line shows self-produced standard vowels (VCS-2), and the red solid line shows responses to self-initiated deviant vowels (VCD-2). The brain maps show the distribution of the effect (for externally-produced vowels: VOS-2 minus VOD-2; for self-initiated vowels: VCS-2 minus VCD-2). Grey bar reflect analyzed in the time window of the N1, N2b and P3a

References

Aliu, S.O., Houde, J.F., & Nagarajan, S.S. (2009). Motor-induced suppression of the auditory cortex. *Journal of Cognitive Neuroscience, 21,* 791-802.

Baess, P., Jacobsen, T., & Schröger, E. (2008). Suppression of the auditory N100 event-related potential component with unpredictable self-initiated tones: evidence for internal forward models with dynamic stimulation. *International Journal of Psychophysiology, 70,* 137-143.

Baess, P., Horváth, J., Jacobsen, T., & Schröger, E. (2011). Selective suppression of self-initiated sounds in an auditory stream: An ERP study. *Psychophysiology, 48,* 1276-1283.

Behroozmand, R., Liu, H., & Larson, C.R. (2011). Time-dependent neural processing of auditory feedback during voice pitch error detection. *Journal of Cognitive Neuroscience, 23,* 1205–1217.

Behroozmand, R., Sangtian, S., Korzyukov, O., & Larson, C.R. (2016). A temporal predictive code for voice motor control: Evidence from ERP and behavioral responses to pitch-shifted auditory feedback. *Brain Research, 1636,* 1-12.

Bendixen, A., SanMiguel, I., & Schröger, E. (2012). Early electrophysiological indicators for predictive processing in audition: A review. *International Journal of Psychophysiology, 83,* 120-131.

Brébion, G., Stephan-Otto, C., Ochoa, S., Roca, M., Nieto, L., & Usall, J. (2016). Impaired self-monitoring of inner speech in schizophrenia patients with verbal hallucinations and in non-clinical individuals prone to hallucinations. *Frontiers in psychology*, *7*, 1381.

Brown, H., Adams, R.A., Parees, I., Edwards, M., and Friston, K. (2013). Active inference, sensory attenuation and illusions. Cognitive processing 14, 411-427.

Chen, C.M., Mathalon, D.H., Roach, B.J., Cavus, I., Spencer, D.D., & Ford, J.M. (2011). The corollary discharge in humans is related to synchronous neural oscillations. *Journal of Cognitive Neuroscience, 23,* 2892-904.

Chen, Z., Jones, J.A., Liu, P., Li, W., Huang, D., et al. (2013) Dynamics of Vocalization-Induced Modulation of Auditory Cortical Activity at Mid-utterance. *PLoS ONE* 8(3): e60039. doi:10.1371/journal.pone.0060039

Christoffels, I.K., Formisano, E., & Schiller, N.O. (2007). Neural correlates of verbal feedback processing: an fMRI study employing overt speech. *Human Brain Mapping, 28,* 868-879.

Christoffels, I.K., van de Ven, V., Waldorp, L.J., Formisano, E., & Schiller, N.O. (2011). The sensory consequences of speaking: parametric neural cancellation during speech in auditory cortex. *PLoS One, 6,* e18307.

Creutzfeldt, O., Ojemann, G., & Lettich, E. (1989). Neuronal activity in the human lateral temporal lobe. II. Responses to the subjects own voice. *Experimental Brain Research, 77,* 476-89.

Crowley, K.E. & Colrain, I.M. (2004). A review of the evidence for P200 being an independent component process: age, sleep and modality. *Clinical Neurophysiology, 115,* 732-744.

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

De Chicchis, A.R., Carpenter, M., Cranford, J.L., & Hymel, M.R. (2002). Electrophysiologic correlates of attention versus distraction in young and elderly listeners. *Journal of the American Acadamy of Audiology, 13,* 383-91; quiz 400-1.

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

Ford, J.M., Mathalon, D.H., Heinks, T., Kalba, S., Faustman, W.O., & Roth, W.T. (2001). Neurophysiological evidence of corollary discharge dysfunction in schizophrenia. *American Journal of Psychiatry,* 158, 2069-2071.

Ford, J.M., Gray, M., Faustman, W.O., Roach, B.J., & Mathalon, D.H. (2007). Dissecting corollary discharge dysfunction in schizophrenia. *Psychophysiology, 44,* 522-9.

Ford, J.M., Roach, B.J., Miller, R.M., Duncan, C.C., Hoffman, R.E., & Mathalon, D.H. (2010). When it's time for a change: failures to track context in schizophrenia. *International Journal of Psychophysiology, 78,* 3-13.

Ford, J.M. & Mathalon, D.H. (2012). Anticipating the future: automatic prediction failures in schizophrenia. *International Journal of Psychophysiology, 83*, 232-239.

Friston, K. (2018). Does predictive coding have a future?. *Nature neuroscience*, *21*(8), 1019.

Fu, C.H.Y., Vythelingum, G.N., Brammer, M.J., Williams, S.C.R., Amaro, E., Andrew, C.M., Yágüez, L., van Haren, N.E., Matsumoto, K., McGuire, P.K. (2006). An fMRI study of verbal self-monitoring: neural correlates of auditory verbal feedback. *Cerebral Cortex 16,* 969–977.

Grimm, S., Recasens, M., Althen, H., & Escera, C. (2012). Ultrafast tracking of sound location changes as revealed by human auditory evoked potentials. *Biological Psychology, 89,* 232-239.

Gunji, A., Hoshiyama, M., & Kakigi, R. (2001). Auditory response following vocalization: a magnetoencephalographic study. *Clincial Neurophysiology, 112,* 514–520.

Haarsma, J., Fletcher, P.C., Ziauddeen, H., Spencer, T.J., Diederen, K.M.J., and Murray, G.K. (2018). Precision weighting of cortical unsigned prediction errors is mediated by dopamine and benefits learning. bioRxiv.

Hashimoto, Y. & Sakai, K.L. (2003). Brain activations during conscious self-monitoring of speech production with delayed auditory feedback: an fMRI study. *Human Brain Mapping, 20,* 22–28.

Hazemann, P., Audin, G., & Lille, F. (1975). Effect of voluntary self-paced movements upon auditory and somatosensory evoked potentials in man. *Electroencephalography and Clinical Neurophysiology, 39,* 247-54.

Heinks-Maldonado, T.H., Mathalon, D.H., Gray, M., & Ford, J.M. (2005). Fine-tuning of auditory cortex during speech production. *Psychophysiology, 42,* 180-90.

Heinks-Maldonado, T.H., Nagarajan, S.S., & Houde, J.F. (2006). Magnetoencephalographic evidence for a precise forward model in speech production. Neuroreport, 17, 1375-1379.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, *13*(2), 135.

Hirano, S., Kojima, H., Naito, Y., Honjo, I., Kamoto, Y., Okazawa, H., Ishizu, K., Yonekura, Y., Nagahama, Y., Fukuyama, H., & Konishi, J. (1996). Cortical speech processing mechanisms while vocalizing visually presented languages. *Neuroreport, 8,* 363-367.

Hirano, S., Kojima, H., Naito, Y., Honjo, I., Kamoto, Y., Okazawa, H., Ishizu, K., Yonekura, Y., Nagahama, Y., Fukuyama, H., & Konishi, J. (1997). Cortical processing mechanism for vocalization with auditory verbal feedback. *Neuroreport, 8,* 2379–2382.

Horváth, J., Roeber, U., Bendixen, A., & Schröger, E. (2008). Specific or general? The nature of attention set changes triggered by distracting auditory events. *Brain Research, 1229,* 193-203.

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,* 1125-1138.

Kauramäki, J., Jääskeläinen, I.P., Hari, R., Möttönen, R., Rauschecker, J.P., & Sams, M. (2010). Lipreading and covert speech production similarly modulate human auditory-cortex responses to pure tones. *Journal of Neuroscience, 30,* 1314-1321.

Katayama, J. & Polich, J. (1998). Stimulus context determines P3a and P3b. *Psychophysiology, 35,* 23–33.

Knolle, F., Schröger, E., Baess, P., & Kotz, S.A. (2012). The cerebellum generates motor-to-auditory predictions: ERP lesion evidence. *Journal of Cognitive Neuroscience*, *24,* 698–706.

Knolle, F., Schröger, E., & Kotz, S.A. (2013a). Cerebellar contribution to the prediction of self-initiated sounds. *Cortex*. 2013 Oct;49(9):2449-61. doi: 10.1016/j.cortex.2012.12.012. PubMed PMID: 23318086.

Knolle, F., Schröger, E., & Kotz, S.A. (2013b). Prediction errors in self- and externally-generated deviants. *Biological Psychology*, 92(2):410-6. doi: 10.1016/j.biopsycho.2012.11.017. PubMed PMID: 23246535.

Kotz S.A, Stockert, A, & Schwartze, M. (2014). Cerebellum, temporal predictability and the updating of a mental model. *Phil. Trans. R. Soc.* 369(1658). pii: 20130403. doi: 10.1098/rstb.2013.0403.

Lange, K. (2011). The reduced N1 to self-generated tones: An effect of temporal predictability? *Psychophysiology, 48(8)*, 1088–1095. doi:10.1111/j.1469-8986.2010.01174.x)

Linden, D.E.J. (2005). The P300: where in the brain is it produced and what does it tell us? *The Neuroscientist 11,* 563–576.

Martikainen, M.H., Kaneko, K., & Hari, R. (2005). Suppressed responses to self-triggered sounds in the human auditory cortex. *Cerebral Cortex 15,* 299-302.

Müller-Preuss, P., & Ploog, D. (1981). Inhibition of auditory cortical neurons during phonation. *Brain Research, 215,* 61-76.

Näätänen, R., Simpson, M., & Loveless, N.E. (1982). Stimulus deviance and evoked potentials. *Biological Psychology, 14,* 53–98.

Näätänen, R. & Gaillard, A.W.K. (1983). The N2 deflection of the ERP and the orienting reflex. In A.W.K. Giallard & W. Ritter (eds.) *EEG correlates of information processing: Theoretical issues*, pp. 119-141. Amsterdam: Elsevier.

Nittono, H., & Ullsperger, P. (2000). Event-related potentials in a self-paced novelty oddball task. *Neuroreport*, 11, 1861–1864.

Nittono, H. (2006). Voluntary stimulus production enhances deviance processing in the brain. International Journal of Psychophysiology, 59, 15–21.

Niziolek, C.A., Nagarajan, S.S.,& Houde, J.F.(2013). What does motor efference copy represent? Evidence from speech production. *Journal of Neuroscience*. 33(41):16110-6. doi: 10.1523/JNEUROSCI.2137-13.2013. PubMed PMID: 24107944; PubMed Central PMCID: PMC3792453.

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,* 29–32.

Numminen, J., Salmelin, & R., Hari, R. (1999). Subject's own speech reduces reactivity of the human auditory cortex. *Neuroscience Letters, 265,* 119–122.

Oldfield, R. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia 9,* 97-113.

Ott, C.G. & Jäncke, L. (2013) Processing of self-initiated speech-sounds is different in musicians. *Frontiers in Human Neuroscience, 7*, 41.

Paus, T., Perry, D.W., Zatorre, R.J., Worsley, K.J., Evans, A.C. (1996). Modulation of cerebral blood flow in the human auditory cortex during speech: role of motor-to-sensory discharges. *European Journal of Neuroscience*, 8, 2236-46.

Pinheiro, A. P., Rezaii, N., Rauber, A., Nestor, P. G., Spencer, K. M., & Niznikiewicz, M. (2017). Emotional self–other voice processing in schizophrenia and its relationship with hallucinations: ERP evidence. *Psychophysiology*, *54*(9), 1252-1265.

Pinheiro, A.P., Schwartze, M. & Kotz, S.A. (2018). Voice-selective prediction alterations in nonclinical voice hearers. *Scientific Reports,* 8, 14717.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology, 118,* 2128-2148.

Postma, A. (2000). Detection of errors during speech production: a review of speech monitoring models. *Cognition*, *77,* 97–132.

Ritter, W., Paavilainen, P., Lavikainen, J., Reinikainen, K., Alho, K., Sams, K., & Näätänen, R. (1992). Event-related potentials to repetition and change of auditory stimuli. *Electroencephalography and Clinical Neurophysiology, 83,* 306–321.

Schäfer, E.W., & Marcus, M.M. (1973). Self-stimulation alters human sensory brain responses. *Science, 181,* 175-177.

Shipp, S. (2016). Neural Elements for Predictive Coding. Front Psychol 7, 1792.

Snyder, E. & Hillyard, S. (1976). Long-latency evoked potentials to irrelevant, deviant stimuli. *Behavioral Biology, 16,* 319–331.

Sterzer, P., Adams, R. A., Fletcher, P., Frith, C., Lawrie, S. M., Muckli, L., ... & Corlett, P. R. (2018). The predictive coding account of psychosis. *Biological psychiatry*, *84*(9), 634-643.

Tian, X., & Poeppel, D. (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. *Frontiers in psychology*, *1*, 166.

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

Ventura, M.I., Nagarajan, S.S., & Houde, J.F. (2009). Speech target modulates speaking induced suppression in auditory cortex. *BMC Neuroscience 10,* 58.

von Holst, E. & Mittelstädt, H. (1950). Das Reafferenzprinzip. *Naturwissenschaften, 37,* 464–476.

Wang, J., Mathalon, D.H., Roach, B.J., Reilly, J., Keedy, S.K., Sweeney, J.A., Ford, J.M. (2014). Action planning and predictive coding when speaking. Neuroimage. 2014 May 1;91:91-8. doi: 10.1016/j.neuroimage.2014.01.003. Epub 2014 Jan 11. PubMed PMID: 24423729; PubMed Central PMCID: PMC3965634.

Wolpert, D.M., Ghahramani, Z. & Jordan, M.I. (1995). An internal model for sensorimotor integration. *Science, 269,* 1880-1882.

Ylinen, S., Nora, A., Leminen, A., Hakala, T., Huotilainen, M., Shtyrov, Y., Mäkelä, J.P., & Service, E. (2014). Two Distinct Auditory-Motor Circuits for Monitoring Speech Production as Revealed by Content-Specific Suppression of Auditory Cortex. Cerebral Cortex. [Epub ahead of print] PubMed PMID: 24414279.

Zheng, Z.Z., Munhall, K.G., & Johnsrude, I.S. (2010). Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production. *Journal of Cognitive Neuroscience, 22,* 1770–1781.

Table 1 - Results of Omnibus ANOVA in Experiment 1

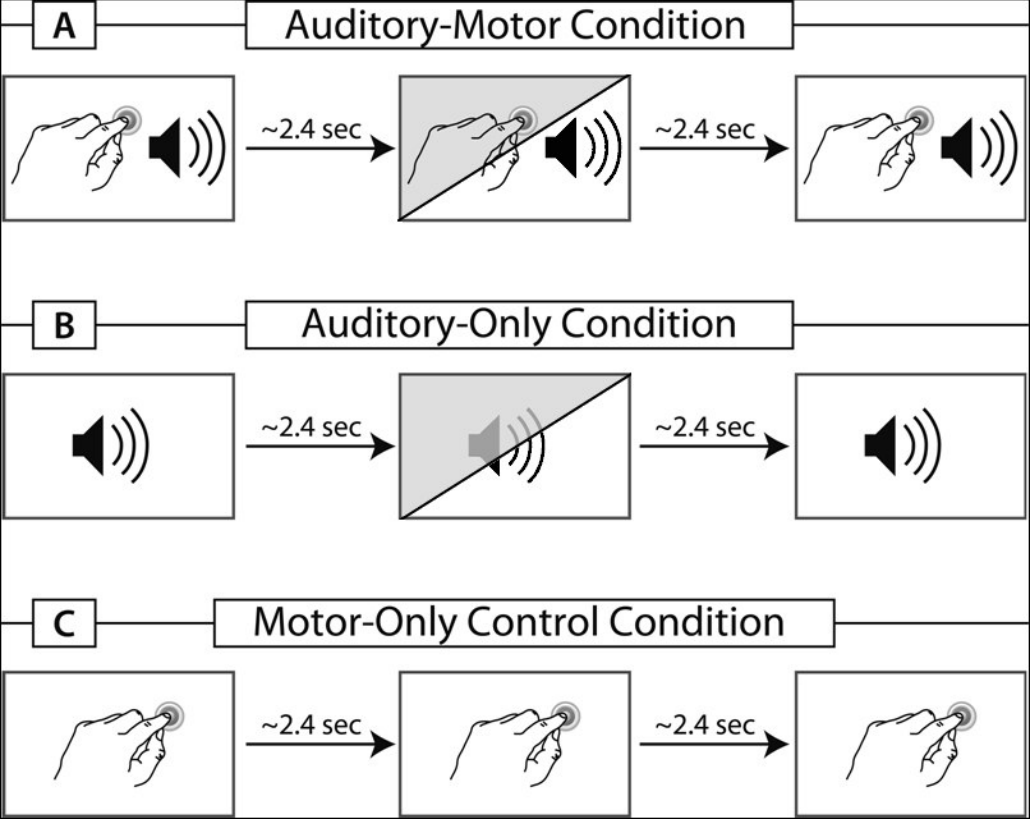
N1			**P2**		
	F	**p**		**F**	**p**
Condition[a]	4.66	**.048**	Condition[a]	10.16	**.006**
Condition x ROI[b]	6.70	**<.001**	Condition x ROI[b]	5.25	**.001**

[a]F(1,15); [b]F(4,60)

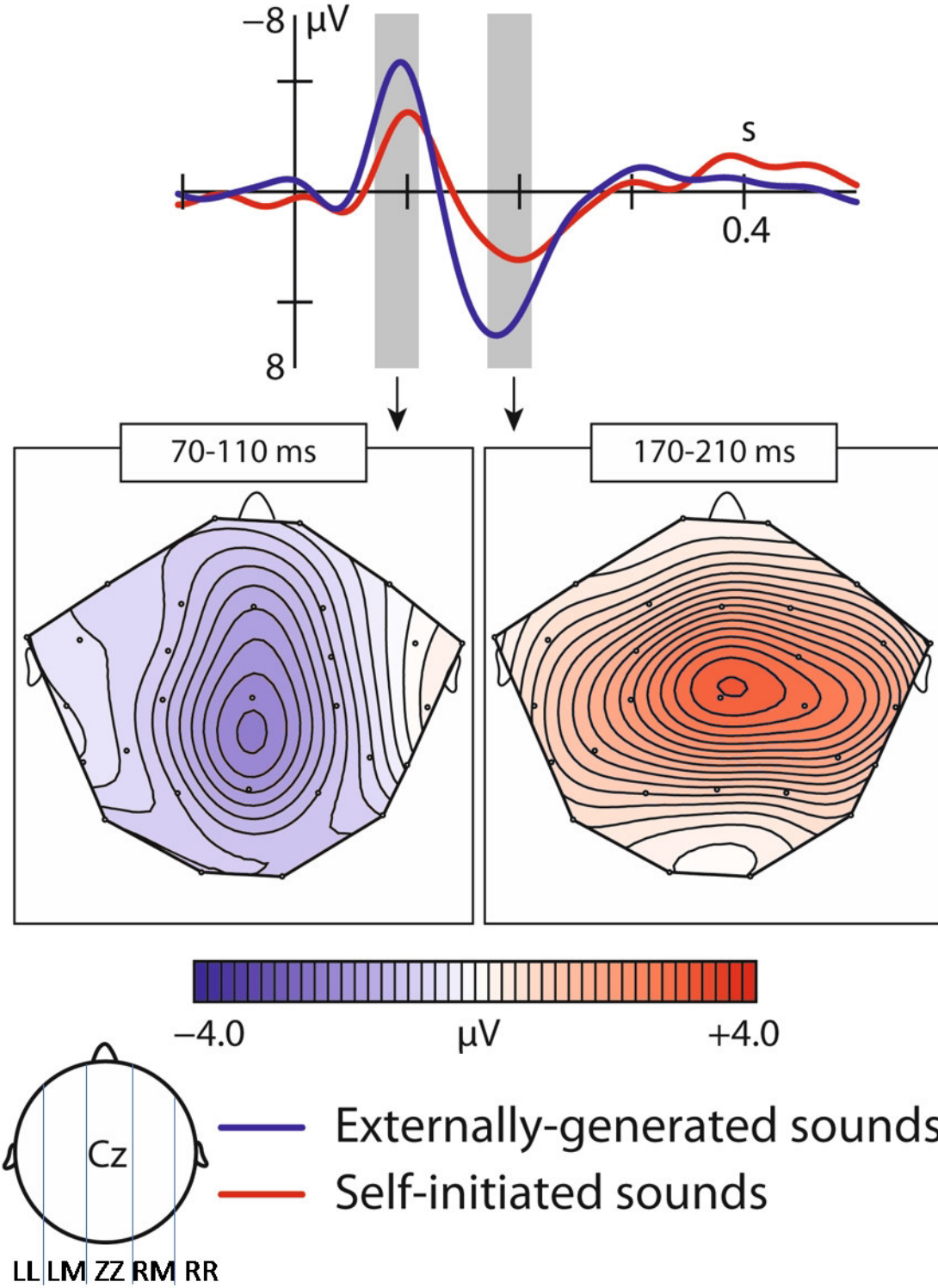
Table 2 - Results of Omnibus ANOVA in Experiment 2

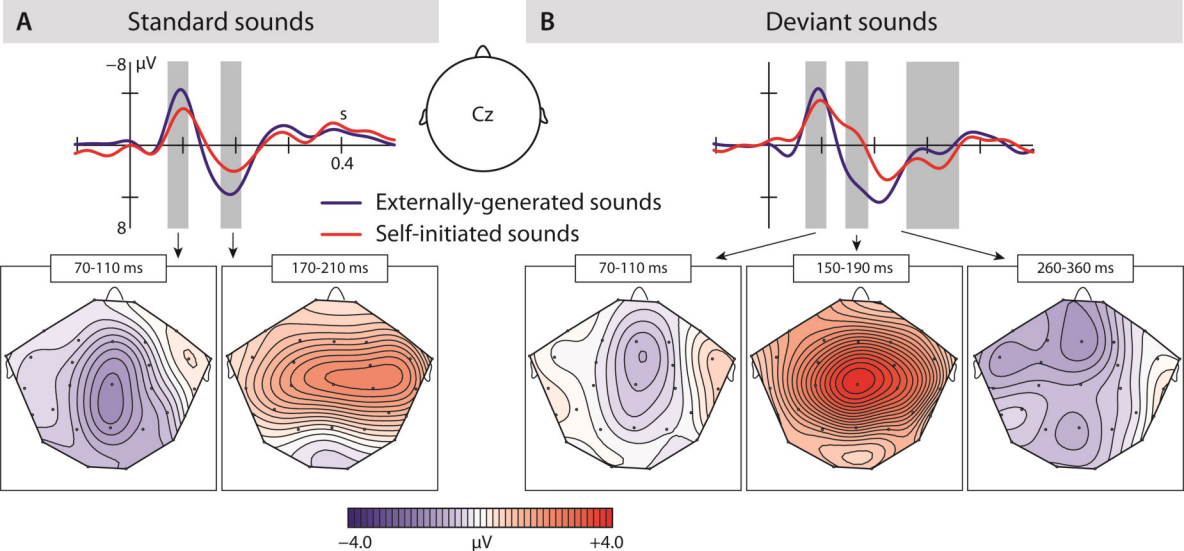
N1		
	F	**p**
Condition x ROI[b]	4.36	**.037**
Standard vowels		
Condition x ROI[b]	5.88	**<.001**
Deviant vowels		
Condition x ROI[b]	2.39	.061
P2		
	F	**P**
Condition[a]	22.42	**<.001**
Condition x ROI[b]	6.68	**<.001**
Conditon x Type[a]	4.34	.055
Standard vowels		
Condition[a]	19.36	**<.001**
Condition x ROI[b]	6.83	**.004**
N2b		
	F	**P**
Condition[a]	9.92	**.007**
Condition x ROI[b]	11.60	**<.001**
Type[a]	12.24	**.003**
Type x ROI[b]	4.60	**.026**
Condition x Type[a]	18.56	**<.001**
Condition x Type x ROI[b]	6.45	**<.001**
P3a		
Type[a]	22.68	**<.001**

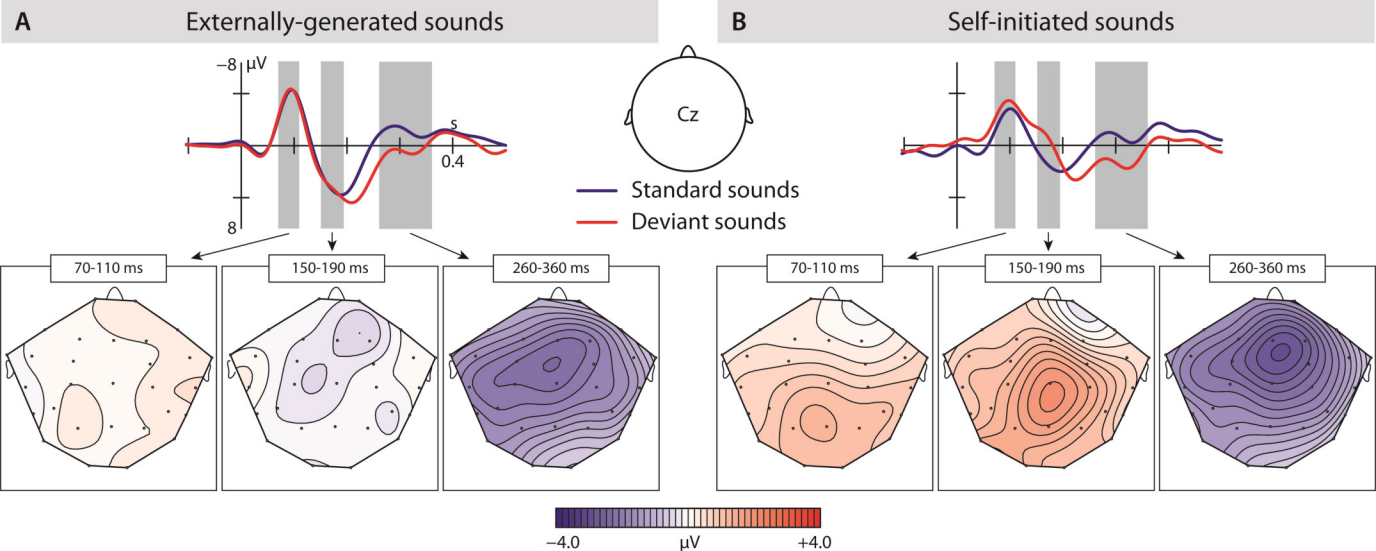
[a]F(1,15); [b]F(4,60)


A
Auditory-Motor Condition
~2.4 sec
~2.4 sec
B
Auditory-Only Condition
~2.4 sec
~2.4 sec
C
Motor-Only Control Condition
~2.4 sec
~2.4 sec

Standard Paradigm




A
Standard sounds
B
Deviant sounds
−8 μV
8
s
0.4
Cz
Externally-generated sounds
Self-initiated sounds
70-110 ms
170-210 ms
70-110 ms
150-190 ms
260-360 ms
−4.0
μV
+4.0


A
Externally-generated sounds
B
Self-initiated sounds
−8 μV
8
s
0.4
Cz
Standard sounds
Deviant sounds
70-110 ms
150-190 ms
260-360 ms
70-110 ms
150-190 ms
260-360 ms
−4.0
μV
+4.0