Phonological and temporal regularities lead to differential ERP effects in self- and externally generated speech

Alexandra K. Emmendorfer [1,2,3]*, Milene Bonte [1,2], Bernadette M. Jansma [1,2]°, Sonja A. Kotz [3]°

1   Dept of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands

2   Maastricht Brain Imaging Center, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands

3   Dept of Neuropsychology and Psychopharmacology, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands

° contributed equally

Running Title: ERP effects of sublexical regularities

* Corresponding Author:

Alexandra K. Emmendorfer, M.Sc.

Faculty of Psychology and Neuroscience, Maastricht University

Oxfordlaan 55, 6229 EV Maastricht, The Netherlands

E-mail: a.emmendorfer@maastrichtuniversity.nl

Tel: +31 433 88 21 59

ABSTRACT

Some theories of predictive processing propose reduced sensory and neural responses to anticipated events. Support comes from M/EEG studies, showing reduced auditory N1 and P2 responses to self- compared to externally generated events, or when stimulus properties are more predictable (e.g. prototypical). The current study examined the sensitivity of N1 and P2 responses to statistical regularities of speech. We employed a motor-to-auditory paradigm comparing ERP responses to externally and self-generated pseudowords, varying in phonotactic probability and syllable stress. We expected to see N1 and P2 suppression for self-generated stimuli, with greater suppression effect for more predictable features such as high phonotactic probability and first syllable stress in pseudowords. We observe an interaction between phonotactic probability and condition on the N1 amplitude, with an enhanced effect of phonotactic probability in processing self-generated stimuli. However, the directionality of this effect was reversed compared to what was expected, namely a larger N1 amplitude for high probability items, possibly indicating a perceptual bias toward the more predictable item. We further observed an effect of syllable stress on the P2 amplitude, with greater amplitudes in response to first syllable stress items. The current results suggest that phonotactic probability plays an important role in processing self-generated speech, supporting feedforward models involved in speech production.

Keywords: N1, P2, phonotactic probability, lexical stress, motor-induced suppression, forward model

Abbreviations: AO = auditory-only, BOLD = blood oxygenation level dependent, Cond = condition, ERP = event-related potential, fMRI = functional magnetic resonance imaging, HPP = high phonotactic probability, ICA = independent component analysis, IQR = inter-quartile range, LPP = low phonotactic probability, MA = motor-auditory, MAC = motor-auditory corrected, M/EEG = magneto-/electroencephalography, MIS = motor-induced suppression, MO = motor-only, PhonProb = phonotactic probability, ROI = region of interest, SylStr = syllable stress, SylS1 = first syllable stress, SylS2 = second syllable stress.

1   Introduction

The brain's capacity to formulate predictions of upcoming events in our environment is one of the most studied phenomena across sensory modalities (e.g., Baldeweg, 2006; Blakemore et al., 2000; Rao & Ballard, 1999). These predictions may relate to the timing ('when', temporal prediction) and content/quality ('what', formal prediction) of upcoming sensory events (Arnal & Giraud, 2012; Kotz & Schwartze, 2010), and are based on our acquired knowledge and experience of the world. A special form of prediction generated by the brain is related to the sensory consequences of our own actions. The underlying mechanism is described by the internal forward model of motor control (e.g., Wolpert & Miall, 1996). According to this model, when a motor plan is formulated, an internal copy of the command, termed "efference copy" is used to generate a prediction of the anticipated sensory feedback. This prediction, or "corollary discharge", is then compared to the actual sensory feedback (reafference signal), allowing the system to distinguish between self- and externally generated sensations, and to monitor and adapt our own motor output more readily. This model has also been applied to speech production, linking psycholinguistic models of feedback monitoring at the phoneme and syllable level, to general motor control mechanisms (e.g., Hickok, 2012; Kotz & Schwartze, 2016)

As a consequence of this mechanism, the sensory response to internally generated stimulation is suppressed, leading to well-known phenomena such as the inability to tickle oneself (Blakemore et al., 2000). This perceived sensory suppression, termed motor-induced suppression (MIS), goes hand in hand with the suppression of sensory-related neural activity, shown across multiple sensory domains, including somatosensory (Blakemore et al., 2000) and auditory (e.g., Christoffels et al., 2011; Knolle et al., 2012; Niziolek et al., 2013). The degree of MIS reflects the accuracy of the generated prediction: the better the match between predicted feedback and actual sensory feedback, the greater the suppression. This pattern is supported by observations of an inverse relationship between noise level and suppression in fMRI: with higher noise levels, less suppression is observed (Christoffels et al., 2011). The magnitude of the neural activity in response to self-generated sensations is therefore thought to reflect the prediction error, or the mismatch between predicted and actual feedback: Noisy situations result in less clear and less predictable feedback, leading to less suppression (i.e., more neural activity).

MIS is modulated by stimulus properties, including the predictability of the frequency and timing of tones (Bäss et al., 2008; Knolle et al., 2013a) or manipulations of voice identity

(Johnson et al., 2021), voice quality and timing in speech (Aliu et al., 2009; Behroozmand & Larson, 2011). In a study comparing different utterances of the same vowel, self- or externally generated, Niziolek and colleagues (2013) observed greater suppression when the utterance showed formant ratios more prototypical for the individual speaker. Crucially, when the utterance deviated from the speaker's prototype, the auditory cortical response predicted the correction of the articulation. This observation confirms that this mechanism may be involved in monitoring and correcting behavior. MIS is further modulated by experience, with musicians showing different suppression patterns than non-musicians (Ott & Jäncke, 2013). In summary, these findings suggest that greater suppression is indicative of more predictable sensory events, and that this suppression may be modulated by experience.

These observations suggest that MIS may be a suitable measure to investigate the brain's sensitivity to regularities in the formal and temporal structure of speech during production. Within speech and language, regularities exist at multiple timescales, allowing the formulation of formal (e.g. phonotactic probability) and temporal (e.g. syllable stress) predictions across different processing levels. These predictions are established through exposure to regularities in speech throughout development, and evidence of sensitivity to these regularities is found already in infancy (Nazzi et al., 1998; Saffran et al., 1996). This sensitivity may provide an important foundation in the early stages of language acquisition, by allowing infants to segment the continuous speech signal into words (Jusczyk et al., 1999; Mattys & Jusczyk, 2001; Thiessen & Saffran, 2003), and continues to facilitate speech processing throughout the lifespan, as indicated by both behavioral and neural evidence.

Phonotactic probability modulates primarily sublexical language processes (i.e. independent of lexical/conceptual processing) in speech perception, such as nonword recognition (Luce & Large, 2001; Vitevitch & Luce, 1999) and recall (Thorn & Frankish, 2005). In contrast, lexical stress can guide the resolution of lexical conflict in spoken word recognition (e.g. present = gift, present = to give a presentation; Cutler, 2005). In production tasks, such as nonword repetition, both phonotactic probability (Edwards et al., 2004; Munson, Edwards, et al., 2005; Munson, Kurtz, et al., 2005; Vitevitch & Luce, 1998, 2005) and syllable stress (Vitevitch et al., 1997) modulate performance, with high phonotactic probability items and more frequently occurring stress patterns (e.g. in Dutch, first syllable stress) being repeated more accurately. Furthermore, lexical stress can guide the learning of novel phonotactic constraints (Bian & Dell, 2020).

There is ample neural evidence supporting the aforementioned behavioral observations during speech perception, with variations in phonotactic probability and stress patterns modulations neural processing (Bonte et al., 2005; Di Liberto et al., 2019; Emmendorfer et al., 2020; Rothermich et al., 2012; Tremblay et al., 2016). However, data on the neural correlates of these features in speech production is sparse. fMRI investigations have shown sensitivity to distributional statistics such as phonotactic probability, syllable frequencies or mutual information in speech production tasks across the speech network, including auditory as well as motor regions, with reduced BOLD signal for items with higher frequency of occurrence within the language (Papoutsi et al., 2009; Tremblay et al., 2016). These findings are in line with psycholinguistic models proposing that motor plans of more frequently occurring structures are stored in a "mental syllabary", while less frequent articulatory representations need to be compiled from smaller units on the spot (Levelt, 1999; Levelt & Wheeldon, 1994; Schiller et al., 1996). Electrophysiological data on these features in speech production tasks is sparse. In a go/no-go task, where "go" decision was based on lexical stress position, N200 latency was earlier for words with first syllable stress (Schiller, 2006). However, this was proposed to be related to the incremental encoding (i.e. from word onset to end) of the meter during speech production, rather than a function of typical/atypical stress patterns, which is further supported by behavioral findings in trisyllabic stimuli (Schiller et al., 2006). Currently, we do not know of any studies investigating the effect of variations in phonotactic probabilities during speech production with electrophysiological methods.

The current experiment aimed to investigate how predictability of phonotactic probability and syllable stress contribute to speech production, extending our knowledge from previous studies investigating speech perception (e.g., Bonte et al., 2005; Emmendorfer et al., 2020) and production (e.g., Schiller, 2006; Tremblay et al., 2016). To approach this question, we focused our attention on motor-induced suppression, as this allows investigating how such (ir)regularities modulate the accuracy of the prediction generated through the efference copy. While some studies have investigated this phenomenon in overt speech production (e.g., Aliu et al., 2009; Christoffels et al., 2011; Niziolek et al., 2013), this comes with challenges due to artifacts caused from engaging the facial muscles during articulation. Furthermore, overt production leads to variability in the pronunciation of the individual utterances, which can lead to changes in the degree of suppression (Niziolek et al., 2013). This is a particularly relevant constraint in the current design, as less familiar features may show more variability in articulation as well as more speech errors (Heisler & Goffman, 2016; Munson, 2001; Sasisekaran et al., 2010). To circumvent these challenges, we employed a button-press

paradigm, or motor-to-auditory paradigm, where the participant elicits the presentation of speech stimuli via button-press (e.g. Knolle et al., 2013a, 2019; Ott & Jäncke, 2013; Pinheiro et al., 2018).

The classical design in these experiments employs three conditions: an auditory-only condition (AO), where participants are passively presented with auditory stimuli, a motor-auditory (MA) condition, where participants trigger the generation of self-produced pseudowords through a button-press, and finally a motor-only (MO) control condition used to correct for the motor component (MA – MO = MAC). This design has been applied to investigate MIS in response to a range of stimulus types, including tones (Knolle et al., 2013a), voices (Pinheiro et al., 2018), vowels (Knolle et al., 2019), and single syllables (Ott & Jäncke, 2013). These designs typically elicit modulations of the auditory N1 and P2 components. Observed reduction of N1 amplitude in response to self-generated stimuli is thought to reflect an unconscious, automatic prediction resulting from the efference copy/corollary discharge, while P2 suppression reflects a more conscious differentiation between self- and externally generated events (e.g. Knolle et al., 2013a, 2019; Pinheiro et al., 2018). Here we investigated the effect of phonotactic and syllable stress regularities on MIS of the N1 and P2 components, using prerecorded utterances of bisyllabic Dutch pseudowords from each participant. Specifically, we aimed at testing the following hypotheses: (1) N1 and P2 amplitudes are reduced for self-generated stimuli compared to externally generated stimuli (i.e. main effect of condition, MIS), (2) this reduction in amplitude is modulated by phonotactic probability and syllable stress (i.e. interactions between phonotactic probability and condition, and syllable stress and condition), with high phonotactic probability and first syllable stress items leading to greater amplitude reduction due to greater predictability, and (3) phonotactic probability and syllable stress may interactively modulate motor-induced suppression (i.e. three-way interaction between phonotactic probability, syllable stress and condition), where we do not have precise predictions about the nature of this interaction.

2   Methods

2.1   Participants

34 right-handed native Dutch speakers participated in the study after giving their informed consent. The study was approved by the Ethical Committee of the Faculty of Psychology and Neuroscience at Maastricht University (ERCPN-OZL 205_17_03_2019) performed in accordance with the approved guidelines and the Declaration of Helsinki. Participants were

invited to complete two sessions: one for recording the stimulus materials, followed by the EEG session. 5 participants completed the stimulus recording but did not complete the EEG session due to the COVID-19 pandemic. One participant was excluded from the EEG session due to failure to accurately reproduce stimuli. One participant was excluded due to excessive noise in the EEG signal (< 100 trials remaining per stimulus per condition). This led to a final sample of 27 participants (9 male, mean age: 21.9; standard deviation +/- 3.8), who completed both sessions of the experiment. The stimulus recording procedures and variations of the EEG paradigm were piloted in an additional 9 participants. The stimuli generated from these pilot participants were used to determine the criteria for stimulus selection as described in the following section.

2.2   Stimulus generation

The stimuli for the EEG experiment were prepared on an individual basis. Participants were invited for an initial stimulus recording session scheduled several days prior to the EEG session. The stimuli consisted of four pseudowords (Table 1), which differed from each other in phonotactic probability (notsal vs. notfal) and syllable stress (first vs. second syllable; adapted from Bonte et al., 2005; Emmendorfer et al., 2020). During the EEG experiment, each participant was presented with stimuli in their own voice. As second syllable stress is rare in Dutch, "natural" pronunciation of bisyllabic pseudowords with this stress pattern is challenging. To circumvent this issue, participants were presented with the target words, which were generated using a splicing procedure. The target words were spoken by a female Dutch speaker, who produced the syllables of interest by replacing them individually with syllables in existing bisyllabic Dutch words containing the same (spoken) consonant cluster and stress pattern as the target pseudowords (e.g. /**bad**zout/ → /**not**zout/ and /**bad**sal/ → **not**sal; /ont**slag**/ → /**not**slag/, and /ont**sal**/ → /**not**sal/; for more details see Emmendorfer et al., 2020). These spliced target words were presented to the participants of the current experiment. After ensuring the participant could hear and reproduce the differences between the pseudowords, each target was presented 15 times in random order, and the participants were asked to repeat them as accurately as possible. Participants were not explicitly instructed to attend to the stress pattern as this could lead to exaggerated expression of syllable stress.

From the 15 repetitions of each pseudoword, one item was selected as the stimulus for the EEG experiment. To ensure comparability across participants, without having to manipulate the recording to deviate from the participants own naturally produced utterance, we selected items such that they were comparable in the timing of the perceptual centers (p-

centers) of the syllables. P-centers are thought to represent the perceived "beat" of the speech stimulus. The timing of the p-centers was estimated with a beat detection algorithm (custom Matlab script adapted from Cummins & Port, 1998). Here, the beat, or p-center, is defined as the midpoint of each local rise in the amplitude envelope of the recorded signal, representing the vocalic nucleus of a syllable. The duration of the interval between the p-centers of each syllable in the bisyllabic pseudowords was calculated, and from 10 participants (9 pilot participants and 1 from the final sample), the average interval was calculated for each pseudoword. These values were used to select the best fitting stimulus for the participants who completed the subsequent EEG session. For each pseudoword, the item with the closest matching interval was selected. If this item contained acoustic artifacts or a mispronunciation, it was discarded, and the next best item was selected. This procedure allowed the selection of temporally comparable stimuli, while preserving each participant's own pronunciation without editing or manipulating the timing. A representation of the stimuli included in the experiment can be found in Figure 1. Stimuli were filtered with a Hann bandpass filter (80 – 10500 Hz), and intensity scaled to 60 dB. Mean stimulus duration was 0.640 s (standard deviation: 0.056 s), and the mean interval between p-centers of the stimuli was 0.319 s (standard deviation: 0.042 s).

Insert Figure 1 here

2.3 EEG paradigm

The paradigm (adapted from Johnson et al., 2021; Ott & Jäncke, 2013) consisted of three conditions (Figure 2A). In all three conditions, the trial began with the presentation of a fixation cross, followed by a cue (< left, > right) at 0.4 – 1.0 s after trial onset. In the motor-auditory condition (MA), participants pressed a button (left or right), which triggered the presentation of a stimulus. In the auditory-only condition (AO), participants were presented with the same cue, but the stimulus presentation occurred without button press, 0.5 s after cue onset. In the motor-only condition (MO), the participants pressed the cued button, but no stimulus was presented. This condition was included to correct for the motor component in the MA condition. This corrected motor-auditory condition (MAC) was calculated as MA – MO, thus allowing the comparison of neural activity in response to self-generated (MAC) and externally generated auditory stimuli (AO). A reduction in N1 and P2 amplitudes for MAC relative to AO is then interpreted as motor-induced suppression.

The EEG recording occurred over the course of 6 experimental runs, each consisting of 18 blocks (8 MA, 8 AO, 2 MO) (Figure 2B). In each MA and AO block, one stimulus pair was presented. The stimuli within the pair differed from each other in either phonotactic probability or syllable stress (Figure 2C), and each cue/button press corresponded to one stimulus. Each pair was presented twice per run and condition, with the cue/button assignment counterbalanced across blocks. Within each block, the first 4 trials (always including 2 left, 2 right) were excluded from analysis to allow the participant to form an association between cue and word. In four blocks per run (2 MA, 2 AO), four catch trials were included at the end of the block, where the cue-stimulus pairing was switched, i.e., the left cue was followed by the stimulus previously associated with the right cue. Participants were instructed to attend to the cue-stimulus pairing and were asked to report at the end of each block whether they noticed a switch. This task was included to ensure the participants were correctly associating the presented stimulus with the cue/button-press, and these trials were excluded from analysis. The total number of trials per block varied between 14 and 28 trials such that the participant could not anticipate when the catch trials would occur by counting. This resulted in 10 – 20 trials per block, and a total of 90 trials per condition/stimulus/cue assignment included in the analysis (Figure 2B).

Insert Figure 2 here

2.4   EEG recording

EEG was recorded with BrainVision Recorder (Brain Products, Munich, Germany) using a 63-channel recording setup. Ag/AgCl sintered electrodes were mounted according to the 10% equidistant system, including 57 scalp electrodes, left and right mastoids for offline re-referencing, and four EOG electrodes to facilitate removal of artefacts caused by eye movements (2 placed on the outer canthi, 2 above and below the right eye). The scalp was cleaned at electrode sites and electrodes were filled with electrolyte gel to keep impedances below 10kΩ. Data was acquired with a sampling rate of 1000Hz, using Fpz as an online reference and AFz as ground. During recording, participants were seated on a comfortable chair in an acoustically and electrically shielded room.

2.5   EEG processing

EEG data was processed using the EEGLAB toolbox (Delorme & Makeig, 2004) and custom MATLAB scripts (MATLAB, 2018) The continuous EEG data were filtered using a bandpass

filter of 1 – 30 Hz, and then downsampled to a sampling rate of 250 Hz. Noisy channels were identified, removed and interpolated using the EEGLAB plugin clean_rawdata, and the data were re-referenced to the average signal of the two mastoid electrodes. The data were then epoched 0 – 2.4 s relative to the onset of the trial to remove noisy break intervals, while still including the entire duration of the experimental blocks. The data were then decomposed using ICA. 2 – 4 independent components, reflecting blinks and horizontal eye-movements were removed for each participant. The reconstructed data were then baseline corrected to a window 0.2 s prior to the onset of the *cue*, and epoched -0.6 – 0.5 s relative the onset of the *stimulus* or *button-press* (this includes 0.1 s of the baseline window in the AO condition). This baseline window prior to the *cue* rather than the *stimulus* onset was selected due to an observed positivity in the AO response (Figure 3A), which was removed during the MA – MO subtraction. This deflection in the AO condition could not be removed from the data through highpass filtering or ICA, and likely reflects processes related to the visual cue (see Supplementary Figures S1, S2 and S3). This observation violates the assumption of baseline correction that there are no systematic differences across conditions in the selected window. Therefore, the pre-cue baseline window was deemed more appropriate. This notion is also supported by previous findings showing differences between self- and externally generated auditory stimuli already in the pre-stimulus window (Reznik et al., 2018). The pre-cue baseline correction is used throughout the analysis steps.

Individual N1 and P2 peaks were manually determined from single subject average waveforms at electrode FCz for each stimulus and condition separately. While the auditory N1 and P2 are classically measured over the vertex electrode (Cz), we opted for a slightly more frontal site as visual inspection of the ERPs suggested the amplitudes to be less influenced by the pre-stimulus deflection at this channel (see Supplementary Figure S2). The N1 peak was determined as a negativity in the time window 100 – 300 ms following stimulus onset, P2 as a positivity following the N1 peak up until 400 ms. These time windows are later than the classically observed N1 and P2 windows, however, a relative delay is consistent with the nature of the stimuli due to their complexity (Conde et al., 2018) and slow onset rise time (Onishi & Davis, 1968). Furthermore, broad time windows were selected to determine the individual peak as we anticipated variability in their timing due to the variability of the individual stimuli (i.e., variations in rise time of first syllable between participants and between first and second syllable stress). If there was ambiguity in the selection of the peak within a waveform (e.g., 2 peaks within the given window), the peak with the more appropriate topography and timing relative to the participant's average as well as the grand

average across participants was selected, to ensure the inclusion of comparable neural events across conditions and participants. The amplitude in a window +/-24 ms surrounding this latency was extracted for all scalp electrodes.

Insert Figure 3 here

2.6 Statistical analyses

Due to the pre-stimulus deflection observed in the AO condition (Figure 3A), we first investigated whether this indeed reflected a systematic difference between AO and MAC. Such a systematic difference between conditions would render a direct comparison of the N1 and P2 amplitudes of these two conditions invalid, as we cannot exclude that any observed modulations of these components might be driven by this deflection rather than true motor-induced suppression as hypothesized. We tested this via a cluster-based permutation analysis (Maris & Oostenveld, 2007). One-sided paired-samples t-tests between AO and MAC were performed at each time-point in the time-window -0.5 - 0 s relative to stimulus onset for 1000 random partitions using the ft_timelockstatistics function of the Fieldtrip toolbox (Oostenveld et al., 2011). This analysis revealed a significant difference between AO and MAC. The observed cluster started at approximately 0.3 s prior to stimulus onset, with a broad topographic distribution. Based on this unexpected observation, a direct comparison of AO vs MAC at N1 and P2 components could not be interpreted as motor-induced suppression. Due to this finding, we were unable to test our specific hypotheses regarding the modulation of motor-induced suppression by phonotactic probability and syllable stress. However, as the broader aim of this research was to investigate the role of these regularities in speech production, we pursue analyses to answer the question of how phonotactic probability and syllable stress modulate speech processing, and whether this differed across self- and externally generated speech.

Statistical analyses on N1 and P2 amplitudes were performed in R version 3.6.3 (R Core Team, 2013) using the rstatix package (Kassambara, 2019). Normal distribution of the N1 and P2 mean amplitude values was confirmed for all conditions via Shapiro-Wilk test (Supplementary Tables S1, S4), and outlier identification via boxplot methods did not reveal any extreme outliers (points beyond Q1 – 3*IQR, Q3 + 3*IQR). In a 2x2x2 (high vs. low PhonProb x first vs. second SylStr x AO vs. MAC Cond) repeated-measures ANOVA, we

tested the following hypotheses for both N1 and P2 mean amplitudes averaged across electrodes in a frontocentral ROI (FCz, FC1, FC2, FC3, FC4): The N1/P2 amplitudes are modulated by the predictability of stimulus features (1) PhonProb and (2) SylStr, where more predictable utterances (i.e., high phonotactic probability and first syllable stress) lead to smaller amplitudes. Furthermore, (3) these features may interactively modulate N1/P2 amplitudes (PhonProb x SylStr interaction), and (4) may differ across conditions (PhonProb x Cond or SylStr x Cond interaction), where we would expect the MAC condition to show greater effects of these features due to error-monitoring. ANOVA results were corrected for multiple comparisons with Bonferroni-Holm correction using the adjust_pvalue() function (Cramer et al., 2016), and follow-up t-tests of simple effects were Bonferroni corrected.

3   Results

Visual inspection of the ERP grand averages (Figure 4A, 5A) reveals an N1/P2 morphology, with the N1 peaking around 200 ms and the P2 around 300 ms. When adjusted for the timing of the p-center of the first syllable of participants' pseudoword pronunciations, the N1 and P2 latencies are shorter, at approximately 125 and 212 ms, respectively. For our analyses we kept the time-locking to stimulus onset as it resulted in delayed but better aligned N1 and P2 responses across participants. In the following sections, we present the results of the statistical analyses. Here, we report only significant or otherwise noteworthy main effects and interactions, as well as post-hoc simple effects. The full results of the statistical analyses can be found in the supplementary materials (Supplementary Tables S1, S2 and S3 for N1, Tables S4 and S5 for P2 results).

3.1   N1

A 2x2x2 repeated measures ANOVA (high vs. low PhonProb x first vs. second SylStr x AO vs. MAC Cond) on N1 mean amplitudes (+/- 24 ms surrounding peak) averaged within the frontocentral ROI revealed a significant interaction between phonotactic probability and condition ($F(1,26) = 8.463$, p.adj = 0.049, $\eta^2_P = 0.246$). In the AO condition, LPP stimuli had a slightly larger N1 mean amplitude compared to HPP stimuli, while the reverse directionality was observed in the MAC condition (Figure 4A, B). This interaction was resolved by means of post-hoc paired samples t-tests testing the effect of phonotactic probability at each level of condition (Bonferroni corrected), which showed no significant effects for either AO ($t(26) = 0.840$, p.adj = 0.818, d = 0.162) or MAC ($t(26) = -2.04$, p.adj = 0.104, d = -0.392). Thus, the observed effect seems to reflect a crossover interaction, where the difference between HPP

and LPP is different across conditions, but in neither AO or MAC do HPP and LPP differ from each other significantly. However, the effect size and mean amplitude difference is larger for MAC (d = -0.392, N1 mean amplitude HPP = -3.17 µV vs LPP = -2.67 µV) than AO (d = 0.162, N1 mean amplitude HPP = -2.87 µV vs LPP = -3.05 µV). No other main effects or interactions on N1 mean amplitude were significant (Figure 4B). We note that there was a small but significant difference in N1 latency across conditions (t(26) = -2.43, p = 0.022, d = -0.468), with the N1 peaking slightly earlier in the AO condition (mean latency = 190 ms) compared to MAC (mean latency = 197 ms).

Insert Figure 4 here

3.2   P2

A 2x2x2 repeated measures ANOVA (high vs. low PhonProb x first vs. second SylStr x AO vs. MAC Cond) on P2 mean amplitude (+/-24 ms surrounding peak) averaged within the frontocentral ROI revealed a significant main effect of syllable stress (F(1,26) = 22.993, p.adj < 0.001, η^2_P = 0.469). Stimuli with first syllable stress elicit a larger P2 (mean amplitude = 3.30 µV) compared to those with second syllable stress (mean amplitude = 2.39 µV; Figure 5). No other main effects or interactions were significant.

Insert Figure 5 here

4   Discussion

The current study aimed to investigate whether motor-induced suppression of the N1 and P2 amplitudes is modulated by formal (phonotactic probability) and temporal (syllable stress) predictability in the speech signal. We used a motor-to-auditory paradigm, where participants triggered the generation of self-produced pseudowords through a button-press. This approach was intended as a step towards investigating speech production, while limiting the interference of motor artifacts and speech errors present during overt production of pseudowords. We expected to observe a motor-induced suppression effect, with larger N1 and P2 amplitudes in the auditory-only condition, compared to the motor-auditory condition. Furthermore, we expected this suppression effect to be modulated by phonotactic probability and/or syllable stress, where high probability items (high phonotactic probability and first syllable stress) would elicit greater suppression, as they might be more "prototypical" items in the language (Niziolek et al., 2013). Due to an observed pre-stimulus deflection in the auditory-only condition, not present in the motor-auditory condition after correcting for motor

output (Figure 3), we were not able to test our specific a priori hypotheses regarding motor-induced suppression in the current data. However, the design still allowed investigating the broader question as to how phonotactic probability and syllable stress contribute to the processing of self- and externally generated speech. We observed a modulation of N1 amplitude by phonotactic probability, which was enhanced in response to self-generated stimuli, and a modulation of P2 amplitude by syllable stress, with first syllable stress eliciting a larger P2 compared to second syllable stress. While these analyses are post-hoc in nature, they provide insights into the differences in processing phonotactic and temporal regularities in self- and externally produced speech that can be followed up upon in future experimental designs.

N1

We investigated the effect of variations of phonotactic probability and syllable stress across self- and externally generated conditions on N1 mean amplitude. Here, we observed a significant interaction between phonotactic probability and condition, but no significant main effects or simple effects. Inspecting the data revealed that this interaction was a crossover effect, indicating that the direction of the effect of phonotactic probability differs across conditions (i.e. HPP > LPP in MAC, LPP > HPP in AO). However, the simple effects did not reach significance in either the auditory-only or motor-auditory condition. It is noteworthy though that visual inspection of amplitudes as well as comparison of effect sizes revealed the difference between high and low phonotactic probability to be larger for self-generated words compared to externally generated ones. Thus, the pattern we observed in this interaction is in line with the notion that such regularities have greater weight in speech production due to feedforward processes. Variations in phonotactic probability of planned utterances may require different degrees of monitoring, as they may be more or less likely to result in mispronunciation. Indeed, it has been shown that phonotactic probability modulates accuracy and speed in speech production (Edwards et al., 2004; Munson, Edwards, et al., 2005; Munson, Kurtz, et al., 2005; Vitevitch & Luce, 2005). However, we hypothesized a larger amplitude for the less probable item as this would generate greater surprise, while our data suggest the opposite pattern. One could assume that these effects might be driven by differences in the motor-only (MO) condition, however we can exclude this possibility, as this condition does not include variations in stimulus type (i.e. same trials of MO are subtracted from all MA averages).

A closer look at theories of predictive processing may explain this discrepancy. While cancellation theories predict suppression of predicted sensory events to highlight novel or unexpected events (e.g., Blakemore et al., 2000), Bayesian theories suggest a perceptual bias or gain for predictable events (e.g., de Lange et al., 2018). Recent developments propose a two-process model to resolve the conflict between these contradictory theories (Press et al., 2020). Here it is proposed that the perceptual system is tuned toward expected events when there is a large overlap between prior and posterior probabilities (i.e. low surprise), resulting in perceptual gain for expected events, as suggested by Bayesian accounts (e.g., Thomas et al., 2020). When there is little overlap between these prior and posterior distributions (i.e. high surprise), this suggests that the model of the environment must be updated, resulting in higher activation to signal the unexpected event, in line with cancellation theories.

The pseudowords used in our design all consist of legal phonotactic structures and lexical stress patterns. Furthermore, the trials included in the analysis did not include any violations of predicted stimuli. Thus, the surprise generated by any given stimulus was low, and would not require the system to update their model of the world. Instead, given the acoustic similarity of high and low phonotactic stimuli, it is more likely that perception was biased toward the high probability item *notsal*. The trend toward a larger N1 amplitude for high probability items in the motor-auditory condition supports this notion. Interestingly, the timing of this modulation around 200 ms suggests that it may occur prior to, or concurrently with, the actual manipulation of phonotactic probability, which occurs at the syllable boundary (occurring around 200 – 250 ms, see Figure 1). This is in line with the opposing process theory proposed by Press and colleagues, as effects of perceptual sharpening are often observed prior to or within 50 ms of the expected stimulus, preceding cancellation effects (e.g., Press & Yon, 2019; Yon & Press, 2017). The perceptual sharpening may also render the system more sensitive to coarticulatory cues already present within the first syllable.

We did not observe a comparable modulation of the N1 by syllable stress, despite this feature also varying in probability in the Dutch language, with first syllable stress being the more common pattern in bisyllabic words. This observation is in line with a previous study in speech perception (Emmendorfer et al., 2020), where variations in syllable stress did not modulate MMN amplitudes. A range of other studies conducted in languages with a fixed stress pattern, such as Hungarian (Honbolygó & Csépe, 2013) or Finnish (Ylinen et al., 2009) does however show a modulation that is in line with a violation response (i.e. larger MMN to the illegal stress pattern). The divergent results here indicate that Dutch speakers process

variations in lexical stress patterns differently than their Hungarian or Finnish speaking counterparts. While syllable stress may be exploited during development (Weber et al., 2004), and continues to play a role in resolving lexical conflict in spoken word recognition (Cutler & Van Donselaar, 2001), predictions generated relating to stress patterns may be weaker compared to those relating to phonotactic probability, particularly in the case of pseudoword stimuli in languages without a fixed lexical stress pattern.

P2

We further investigated the effect of variations of phonotactic probability and syllable stress across self- and externally generated conditions on P2 mean amplitude. Here, we observe a main effect of syllable stress, where first syllable stress stimuli elicit a larger P2 amplitude compared to second syllable stress stimuli. This observation may be explained by acoustic differences in first and second syllable stress items (see Figure 1). The main acoustic markers of lexical stress are intensity, pitch, and duration of the syllable. Thus, while the stimuli were equalized in intensity across the whole word, they differed in the first syllable, with first syllable stress items having a greater intensity compared to second syllable stress items. The timing of our P2 at around 280 ms suggests that this component reflects information from the first syllable. We do not observe an interaction with condition, which would suggest a conscious differentiation between self- and externally generated events (Knolle et al., 2013a, 2019; Pinheiro et al., 2018). Therefore, this pattern is more in line with previous observations of P2 amplitude being modulated by stimulus intensity (for review, see Crowley & Colrain, 2004). However, if this is indeed a purely acoustic effect, one would expect to find a similar modulation in the N1 component, which we do not observe. It is possible that the amplitude modulation from the syllable stress pattern is masked due to distortion of the overall N1 amplitude from overlap of the pre-stimulus deflection. However, assuming that this distortion is equal across stimuli we would still expect to observe an effect in the N1 amplitude across first and second syllable stress stimuli.

An alternative explanation for the P2 modulation may lie in categorical perception of speech. The neural correlates of categorical perception around the typical P2 time-window, indicated by investigations of phoneme processing comparing tokens that vary along a continuum across phoneme boundaries (e.g. Bidelman et al., 2013, 2020). Here, ambiguous speech sounds show a smaller P2 amplitude compared to speech sounds that clearly fall within a phoneme category. As second syllable stress is atypical for Dutch bisyllabic words, this may come with variability in the pronunciation, including variability in vowel quality.

However, as the current experiment did not specifically modulate categorization, and we also do not have data on the perceptual categorization of the vowels in the current design, the current data cannot address this question completely, thus this interpretation remains speculative.

General discussion

The temporal dissociation of the observed effects, with phonotactic probability modulating N1 and syllable stress P2 amplitude, may suggest differences in the time course of processing of these features. In the current design, the features phonotactic probability and syllable stress are manipulated at different points in time of the stimuli: first and second syllable stress items differ from each other in principle from stimulus onset on, while phonotactic probability is varied at the syllable boundary. Furthermore, as the stimuli are naturally produced, we have little control over the precise timing of acoustic markers of phonotactic probability and syllable stress relative to the timing of the ERP components of interest. This makes it difficult to disentangle differences in the time course of the neural processing of these linguistic features from differences in when the information relevant to these features becomes available in the specific pseudoword stimuli. Thus, it is not possible to draw general conclusions about the relative time course of processing of phonotactic probability and syllable stress beyond the current design.

Although we originally set out to test hypotheses relating to motor-induced suppression, limitations to the current design hinder us from following the original analysis plan. The pre-stimulus deflection observed in the auditory-only condition (Figure 3) draws attention to the cue as a confound. This deflection appears to be time-locked to the cue onset, covers a broad time window and is present across the scalp, though larger in amplitude at more parietal regions. While the cue is identical in all conditions, it is effectively subtracted out from the motor-auditory condition along with the motor component (MAC = MA - MO) but remains present in the auditory-only condition. Including a visual control to subtract from the auditory-only condition may ameliorate this issue, however this would only account for purely visual processes. The deflection likely also represents attentional and anticipatory processes, as the participant was instructed to explicitly attend to the stimulus and could anticipate not only *which item* would be presented, but also *when* it would be presented, due to the constant timing between cue and stimulus. Thus, an additional adjustment to the current paradigm could include jittering the timing of these events to dissociate the processes associated with the cue and the stimulus. Varying the time between cue and stimulus could

also address the question of whether the suppression effect is driven by the temporal predictability of the stimulus (Hughes et al., 2013; Sowman et al., 2012).

We do not observe any suppression in either the N1 or P2 components for the motor-auditory vs auditory-only condition. This observation differs from the bulk of similar studies applying this type of paradigm comparing the processing of self- and externally-generated stimuli (e.g., Bäss et al., 2008; Knolle et al., 2012, 2013a; Pinheiro et al., 2018). The lack of suppression may be explained by variations in the design between previous studies and the current experiment. The typical approach in this paradigm does not include a cue. Instead, the paradigm is typically applied as a blocked design (but see Knolle et al., 2013b for an event-related variation), where the button-presses generating the stimulus presentation are self-initiated in the motor-auditory condition. The auditory stimuli are then presented at the same temporal intervals in the auditory-only condition, again without a cue. Thus, a crucial difference between the auditory-only and motor-auditory conditions in these approaches, in addition to whether the stimulus is self-generated or externally presented, is the predictability of the stimulus timing: in the motor-auditory condition, the participant can accurately predict the timing, while some temporal uncertainty remains in the auditory-only condition. A considerable portion of the suppression effect observed in previous research may therefore be driven by the temporal predictability of the events. In the current study, the timing of the stimulus in the auditory-only condition is predictable due to the cue, thus this difference between the auditory-only and motor-auditory conditions does not exist. If temporal predictability indeed drives the suppression effect, it is therefore not surprising that we do not observe this effect in the current paradigm. Future studies investigating the suppression effect should therefore consider not only varying the formal predictability of the stimulus, but also its temporal predictability.

In conclusion, the present experiment provides preliminary insights into differences in processing phonotactic and temporal regularities in speech perception and production, by comparing self-generated (via button press) to externally generated (own) speech. Our findings suggest that phonotactic regularities play an important role in processing self-generated speech, with a perceptual bias toward more probable phonotactic structures, in line with Bayesian accounts of predictive processing, or a combined model incorporating both Bayesian and cancellation theories. We further observe an effect of syllable stress, which is likely explained by the acoustic differences between the first syllable in pseudowords with first and second syllable stress. To summarize, the current research suggests that a sensitivity

to regularities in phonotactic and temporal structure of speech may be differently exploited in speech perception and production processes. Further investigations controlling for some of the limitations observed in the current paradigm are needed to confirm the results of the current post-hoc analyses.

Data availability: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Author Contributions: AKE, MB, BMJ and SAK designed the experiment, AKE prepared materials, collected and analyzed the data, and wrote the first draft of the manuscript, AKE, MB, BMJ and SAK refined the manuscript.

Competing Interests: The authors declare that they have no competing interests.

References

Aliu, S. O., Houde, J. F., & Nagarajan, S. S. (2009). Motor-induced suppression of the auditory cortex. *Journal of Cognitive Neuroscience*, *21*(4), 791–802. https://doi.org/10.1162/jocn.2009.21055

Arnal, L. H., & Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences*, *16*(7), 390–398. https://doi.org/10.1016/j.tics.2012.05.003

Baldeweg, T. (2006). Repetition effects to sounds: Evidence for predictive coding in the auditory system. *Trends in Cognitive Sciences*, *10*(3), 93–94. https://doi.org/10.1016/j.tics.2006.01.010

Bäss, P., Jacobsen, T., & 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. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *70*(2), 137–143. https://doi.org/10.1016/j.ijpsycho.2008.06.005

Behroozmand, R., & Larson, C. R. (2011). Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. *BMC Neuroscience*, *12*, 54. https://doi.org/10.1186/1471-2202-12-54

Bian, Y., & Dell, G. S. (2020). Novel stress phonotactics are learnable by English speakers: Novel tone phonotactics are not. *Memory & Cognition*, *48*(2), 176–187. https://doi.org/10.3758/s13421-019-01000-9

Bidelman, G. M., Bush, L. C., & Boudreaux, A. M. (2020). Effects of Noise on the Behavioral and Neural Categorization of Speech. *Frontiers in Neuroscience*, *14*, 153. https://doi.org/10.3389/fnins.2020.00153

Bidelman, G. M., Moreno, S., & Alain, C. (2013). Tracing the emergence of categorical speech perception in the human auditory system. *NeuroImage*, *79*, 201–212. https://doi.org/10.1016/j.neuroimage.2013.04.093

Blakemore, S. J., Wolpert, D., & Frith, C. (2000). Why can't you tickle yourself? *Neuroreport*, *11*(11), R11-16. https://doi.org/10.1097/00001756-200008030-00002

Bonte, M. L., Mitterer, H., Zellagui, N., Poelmans, H., & Blomert, L. (2005). Auditory cortical tuning to statistical regularities in phonology. *Clinical Neurophysiology*, *116*(12), 2765–2774. https://doi.org/10.1016/j.clinph.2005.08.012

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*(5), e18307. https://doi.org/10.1371/journal.pone.0018307

Conde, T., Gonçalves, Ó. F., & Pinheiro, A. P. (2018). Stimulus complexity matters when you hear your own voice: Attention effects on self-generated voice processing. *International Journal of Psychophysiology*, *133*, 66–78. https://doi.org/10.1016/j.ijpsycho.2018.08.007

Cramer, A. O. J., van Ravenzwaaij, D., Matzke, D., Steingroever, H., Wetzels, R., Grasman, R. P. P. P., Waldorp, L. J., & Wagenmakers, E.-J. (2016). Hidden multiplicity in exploratory multiway ANOVA: Prevalence and remedies. *Psychonomic Bulletin & Review*, *23*(2), 640–647. https://doi.org/10.3758/s13423-015-0913-5

Crowley, K. E., & Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: Age, sleep and modality. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *115*(4), 732–744. https://doi.org/10.1016/j.clinph.2003.11.021

Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, *26*(2), 145–171. https://doi.org/10.1006/jpho.1998.0070

Cutler, A. (2005). *Lexical stress*.

Cutler, A., & Van Donselaar, W. (2001). Voornaam is not (really) a Homophone: Lexical Prosody and Lexical Access in Dutch. *Language and Speech*, *44*(2), 171–195. https://doi.org/10.1177/00238309010440020301

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How Do Expectations Shape Perception? *Trends in Cognitive Sciences*, *22*(9), 764–779. https://doi.org/10.1016/j.tics.2018.06.002

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Di Liberto, G. M., Wong, D., Melnik, G. A., & de Cheveigné, A. (2019). Low-frequency cortical responses to natural speech reflect probabilistic phonotactics. *NeuroImage*, *196*, 237–247. https://doi.org/10.1016/j.neuroimage.2019.04.037

Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research: JSLHR*, *47*(2), 421–436. https://doi.org/10.1044/1092-4388(2004/034)

Emmendorfer, A. K., Correia, J. M., Jansma, B. M., Kotz, S. A., & Bonte, M. (2020). ERP mismatch response to phonological and temporal regularities in speech. *Scientific Reports*, *10*(1), 9917. https://doi.org/10.1038/s41598-020-66824-x

Heisler, L., & Goffman, L. (2016). The Influence of Phonotactic Probability and Neighborhood Density on Children's Production of Newly Learned Words. *Language Learning and Development*, *12*(3), 338–356. https://doi.org/10.1080/15475441.2015.1117977

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, *13*(2), 135–145. https://doi.org/10.1038/nrn3158

Honbolygó, F., & Csépe, V. (2013). Saliency or template? ERP evidence for long-term representation of word stress. *International Journal of Psychophysiology*, *87*(2), 165–172.

Hughes, G., Desantis, A., & Waszak, F. (2013). Mechanisms of intentional binding and sensory attenuation: The role of temporal prediction, temporal control, identity prediction, and motor prediction. *Psychological Bulletin*, *139*(1), 133–151. https://doi.org/10.1037/a0028566

Johnson, J. F., Belyk, M., Schwartze, M., Pinheiro, A. P., & Kotz, S. A. (2021). Expectancy changes the self-monitoring of voice identity. *European Journal of Neuroscience*, *53*(8), 2681–2695. https://doi.org/10.1111/ejn.15162

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in english-learning infants. *Cognitive Psychology*, *39*(3–4), 159–207. https://doi.org/10.1006/cogp.1999.0716

Kassambara, A. (2019). *rstatix: Pipe-Friendly Framework for Basic Statistical Tests*.

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*(3), 698–706. https://doi.org/10.1162/jocn_a_00167

Knolle, F., Schröger, E., & Kotz, S. A. (2013a). Prediction errors in self- and externally-generated deviants. *Biological Psychology*, *92*(2), 410–416. https://doi.org/10.1016/j.biopsycho.2012.11.017

Knolle, F., Schröger, E., & Kotz, S. A. (2013b). Cerebellar contribution to the prediction of self-initiated sounds. *Cortex*, *49*(9), 2449–2461. https://doi.org/10.1016/j.cortex.2012.12.012

Knolle, F., Schwartze, M., Schröger, E., & Kotz, S. A. (2019). Auditory Predictions and Prediction Errors in Response to Self-Initiated Vowels. *Frontiers in Neuroscience*, *13*, 1146. https://doi.org/10.3389/fnins.2019.01146

Kotz, S. A., & Schwartze, M. (2010). Cortical speech processing unplugged: A timely subcortico-cortical framework. *Trends in Cognitive Sciences*, *14*(9), 392–399. https://doi.org/10.1016/j.tics.2010.06.005

Kotz, S. A., & Schwartze, M. (2016). Chapter 57—Motor-Timing and Sequencing in Speech Production: A General-Purpose Framework. In G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 717–724). Academic Press. https://doi.org/10.1016/B978-0-12-407794-2.00057-2

Levelt, null. (1999). Models of word production. *Trends in Cognitive Sciences*, *3*(6), 223–232. https://doi.org/10.1016/s1364-6613(99)01319-4

Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, *50*(1–3), 239–269. https://doi.org/10.1016/0010-0277(94)90030-2

Luce, P. A., & Large, N. R. (2001). Phonotactics, density, and entropy in spoken word recognition. *Language and Cognitive Processes*, *16*(5–6), 565–581. https://doi.org/10.1080/01690960143000137

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, *78*(2), 91–121. https://doi.org/10.1016/s0010-0277(00)00109-8

Munson, B. (2001). Phonological pattern frequency and speech production in adults and children. *Journal of Speech, Language, and Hearing Research: JSLHR*, *44*(4), 778–792. https://doi.org/10.1044/1092-4388(2001/061)

Munson, Benjamin, Edwards, J., & Beckman, M. E. (2005). Relationships between nonword repetition accuracy and other measures of linguistic development in children with phonological disorders. *Journal of Speech, Language, and Hearing Research: JSLHR*, *48*(1), 61–78. https://doi.org/10.1044/1092-4388(2005/006)

Munson, Benjamin, Kurtz, B. A., & Windsor, J. (2005). The influence of vocabulary size, phonotactic probability, and wordlikeness on nonword repetitions of children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research: JSLHR*, *48*(5), 1033–1047. https://doi.org/10.1044/1092-4388(2005/072)

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology. Human Perception and Performance*, *24*(3), 756–766. https://doi.org/10.1037//0096-1523.24.3.756

Niziolek, C. A., Nagarajan, S. S., & Houde, J. F. (2013). What does motor efference copy represent? Evidence from speech production. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(41), 16110–16116. https://doi.org/10.1523/JNEUROSCI.2137-13.2013

Onishi, S., & Davis, H. (1968). Effects of duration and rise time of tone bursts on evoked V potentials. *The Journal of the Acoustical Society of America*, *44*(2), 582–591. https://doi.org/10.1121/1.1911124

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, 1–9. https://doi.org/10.1155/2011/156869

Ott, C. G. M., & Jäncke, L. (2013). Processing of self-initiated speech-sounds is different in musicians. *Frontiers in Human Neuroscience*, *7*, 41. https://doi.org/10.3389/fnhum.2013.00041

Papoutsi, M., de Zwart, J. A., Jansma, J. M., Pickering, M. J., Bednar, J. A., & Horwitz, B. (2009). From Phonemes to Articulatory Codes: An fMRI Study of the Role of Broca's Area in Speech Production. *Cerebral Cortex*, *19*(9), 2156–2165. https://doi.org/10.1093/cercor/bhn239

Pinheiro, A. P., Schwartze, M., & Kotz, S. A. (2018). Voice-selective prediction alterations in nonclinical voice hearers. *Scientific Reports*, *8*(1), 14717. https://doi.org/10.1038/s41598-018-32614-9

Press, C., Kok, P., & Yon, D. (2020). The Perceptual Prediction Paradox. *Trends in Cognitive Sciences*, *24*(1), 13–24. https://doi.org/10.1016/j.tics.2019.11.003

Press, C., & Yon, D. (2019). Perceptual Prediction: Rapidly Making Sense of a Noisy World. *Current Biology*, *29*(15), R751–R753. https://doi.org/10.1016/j.cub.2019.06.054

R Core Team. (2013). *R: A language and environment for statistical computing*.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Reznik, D., Simon, S., & Mukamel, R. (2018). Predicted sensory consequences of voluntary actions modulate amplitude of preceding readiness potentials. *Neuropsychologia*, *119*, 302–307. https://doi.org/10.1016/j.neuropsychologia.2018.08.028

Rothermich, K., Schmidt-Kassow, M., & Kotz, S. A. (2012). Rhythm's gonna get you: Regular meter facilitates semantic sentence processing. *Neuropsychologia*, *50*(2), 232–244.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science (New York, N.Y.)*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Sasisekaran, J., Smith, A., Sadagopan, N., & Weber-Fox, C. (2010). Nonword repetition in children and adults: Effects on movement coordination. *Developmental Science*, *13*(3), 521–532. https://doi.org/10.1111/j.1467-7687.2009.00911.x

Schiller, N. O. (2006). Lexical stress encoding in single word production estimated by event-related brain potentials. *Brain Research*, *1112*(1), 201–212. https://doi.org/10.1016/j.brainres.2006.07.027

Schiller, N. O., Jansma, B. M., Peters, J., & Levelt, W. J. M. (2006). Monitoring metrical stress in polysyllabic words. *Language and Cognitive Processes*, *21*(1–3), 112–140. https://doi.org/10.1080/01690960400001861

Schiller, N. O., Meyer, A. S., Baayen, R. H., & Levelt, W. J. M. (1996). A comparison of lexeme and speech syllables in Dutch. *Journal of Quantitative Linguistics*, *3*(1), 8–28. https://doi.org/10.1080/09296179608590060

Sowman, P. F., Kuusik, A., & Johnson, B. W. (2012). Self-initiation and temporal cueing of monaural tones reduce the auditory N1 and P2. *Experimental Brain Research*, *222*(1–2), 149–157. https://doi.org/10.1007/s00221-012-3204-7

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, *39*(4), 706–716. https://doi.org/10.1037/0012-1649.39.4.706

Thomas, E. R., Yon, D., de Lange, F. P., & Press, C. (2020). *Action enhances predicted touch* [Preprint]. Neuroscience. https://doi.org/10.1101/2020.03.26.007559

Thorn, A. S. C., & Frankish, C. R. (2005). Long-term knowledge effects on serial recall of nonwords are not exclusively lexical. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*(4), 729–735. https://doi.org/10.1037/0278-7393.31.4.729

Tremblay, P., Deschamps, I., Baroni, M., & Hasson, U. (2016). Neural sensitivity to syllable frequency and mutual information in speech perception and production. *NeuroImage*, *136*, 106–121. https://doi.org/10.1016/j.neuroimage.2016.05.018

Vitevitch, M. S., & Luce, P. A. (1998). When Words Compete: Levels of Processing in Perception of Spoken Words. *Psychological Science*, *9*(4), 325–329. https://doi.org/10.1111/1467-9280.00064

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic Phonotactics and Neighborhood Activation in Spoken Word Recognition. *Journal of Memory and Language*, *40*(3), 374–408. https://doi.org/10.1006/jmla.1998.2618

Vitevitch, M. S., & Luce, P. A. (2005). Increases in phonotactic probability facilitate spoken nonword repetition. *Journal of Memory and Language*, *52*(2), 193–204. https://doi.org/10.1016/j.jml.2004.10.003

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech*, *40*(1), 47–62.

Weber, C., Hahne, A., Friedrich, M., & Friederici, A. D. (2004). Discrimination of word stress in early infant perception: Electrophysiological evidence. *Brain Research. Cognitive Brain Research*, *18*(2), 149–161. https://doi.org/10.1016/j.cogbrainres.2003.10.001

Wolpert, D. M., & Miall, R. C. (1996). Forward Models for Physiological Motor Control. *Neural Networks: The Official Journal of the International Neural Network Society*, *9*(8), 1265–1279. https://doi.org/10.1016/s0893-6080(96)00035-4

Ylinen, S., Strelnikov, K., Huotilainen, M., & Näätänen, R. (2009). Effects of prosodic familiarity on the automatic processing of words in the human brain. *International Journal of Psychophysiology*, *73*(3), 362–368.

Yon, D., & Press, C. (2017). Predicted action consequences are perceptually facilitated before cancellation. *Journal of Experimental Psychology. Human Perception and Performance*, *43*(6), 1073–1083. https://doi.org/10.1037/xhp0000385
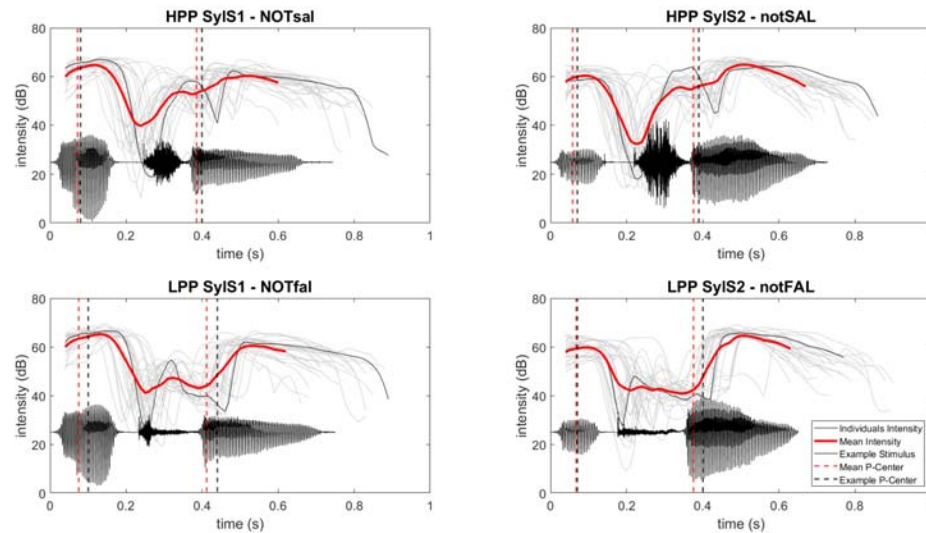
Table 1

Stimuli

	SylStr	
	SylS1	SylS2
PhonProb		
High (HPP)	**not**sal	not**sal**
Low (LPP)	**not**fal	not**fal**

*Note.* Bold font indicates stressed syllable (SylS1 = 1st syllable, SylS2 = 2nd syllable). PhonProb = phonotactic probability. SylStr = syllable stress. The phoneme combination '-ts-' constitutes the high phonotactic probability (HPP), and '-tf-' the low phonotactic probability (LPP).

Figure 1: Stimuli. Stimuli selected for the EEG experiment. Individual intensity contours of the stimuli are represented in grey, mean intensity contours across participants in red. Stimuli from an exemplary participant are represented in black. Timing of the p-centers, representing the onset of the vocalic nucleus, are represented by dashed lines (red: averaged across participants, black: exemplary participant)
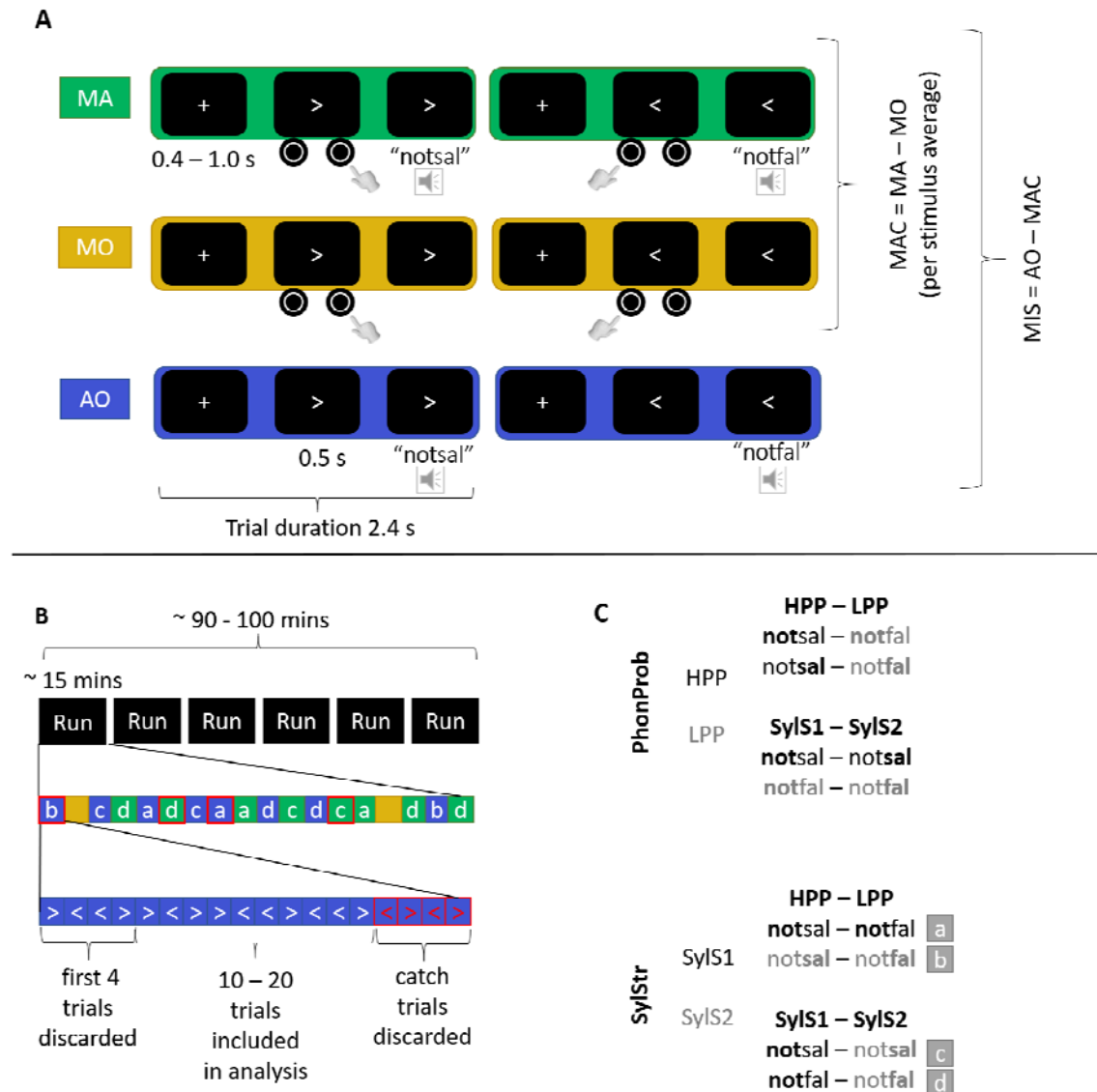
Figure 2: Experimental design. (A) Three experimental conditions: MA = motor-auditory, MO = motor only, AO = auditory only. (B) Overview of the EEG paradigm timeline. Letters a-d correspond to the stimulus pair presented as denoted in panel C. (C) Overview of stimuli and contrasted features: PhonProb = phonotactic probability, HPP = high phonotactic probability, LPP = low phonotactic probability, SylStr = syllable stress, SylS1 = first syllable stress, SylS2 = second syllable stress.
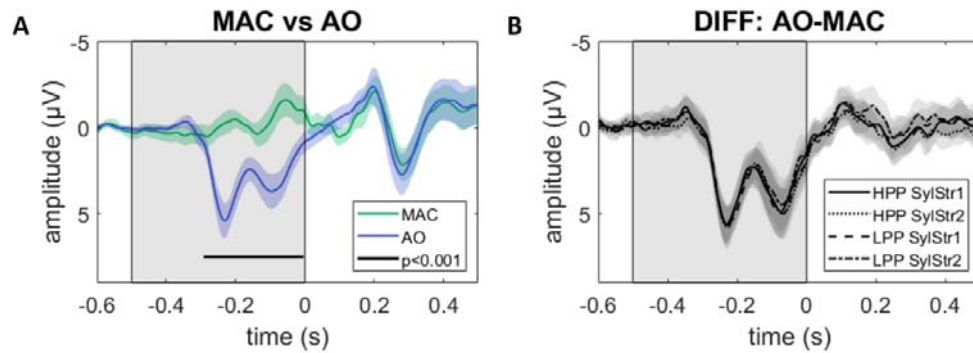
Figure 3. Overview of pre-stimulus deflection. (A) MAC (green) and AO (blue) conditions, time-locked to stimulus onset and averaged across stimuli (+/-95% CI of the mean) in a frontocentral ROI (FCz, FC1, FC2, FC3, FC4). The shaded window indicates the window where cluster-based permutation tests were performed. The black line indicates the timing of the observed cluster with a significant difference between MAC and AO. (B) AO – MAC difference wave for individual stimuli at the same frontocentral ROI. The shaded area indicates the window where cluster-based permutation tests were performed to test for systematic effects of PhonProb, SylStr or an interaction thereof. No significant differences were found.
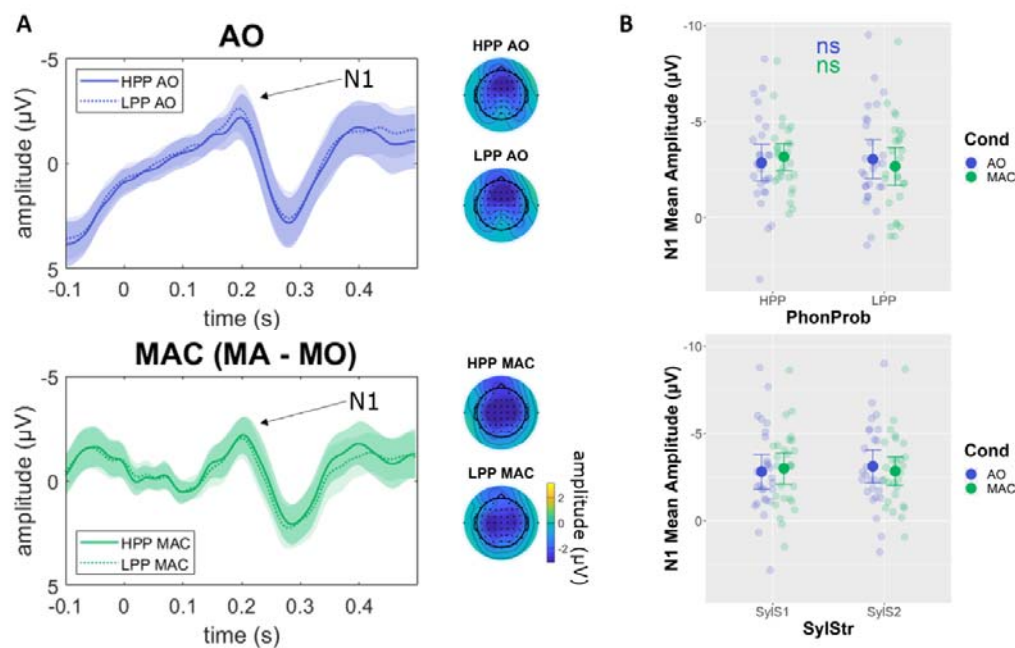
Figure 4: Interaction between phonotactic probability and condition on N1 amplitude.
(A) ERP waveforms at frontocentral ROI (FCz, FC1, FC2, FC3, FC4) time-locked to stimulus

onset, and corresponding topographies of N1 mean amplitudes at frontocentral ROI. (B) N1 mean amplitudes (+/- 24 ms surrounding individual peaks). HPP = high phonotactic probability, LPP = low phonotactic probability, AO = auditory only, MA = motor-auditory, MO = motor only, MAC = motor-auditory corrected, PhonProb = phonotactic probability, Cond = condition, ns = non-significant. Note: simple effects were not tested for SylStr x Cond as this interaction was not significant.
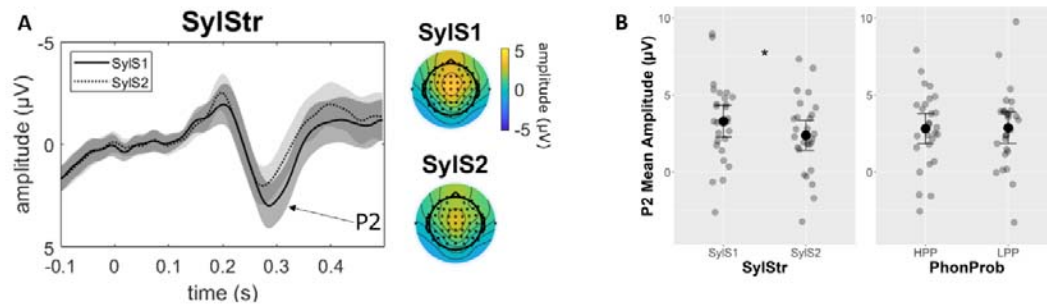
Figure 5: Main effect of syllable stress on P2 amplitude. (A) ERP waveforms at frontocentral ROI (FCz, FC1, FC2, FC3, FC4) time-locked to stimulus onset, averaged across phonotactic probability and condition, and corresponding topographies of individual P2 mean amplitudes (+/- 24 ms surrounding individual peaks). (B) P2 mean amplitudes (+/- 24 ms surrounding individual peaks). * $p < 0.001$. SylStr = syllable stress, SylS1 = first syllable stress, SylS2 = second syllable stress, PhonProb = phonotactic probability, HPP = high phonotactic probability, LPP = low phonotactic probability.