

# Sensitivity to syllable stress regularities in externally but not self-triggered speech in Dutch

Alexandra K. Emmendorfer<sup>1,2,3</sup>  | Milene Bonte<sup>1,2</sup> | Bernadette M. Jansma<sup>1,2</sup> | Sonja A. Kotz<sup>3</sup>

<sup>1</sup>Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands

<sup>2</sup>Maastricht Brain Imaging Center, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands

<sup>3</sup>Department of Neuropsychology and Psychopharmacology, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands

## Correspondence

Alexandra K. Emmendorfer, Department of Neuropsychology and Psychopharmacology, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands. Email: [alexandra.emmendorfer@donders.ru.nl](mailto:alexandra.emmendorfer@donders.ru.nl)

## Present address

Alexandra K. Emmendorfer, Donders Institute for Brain, Cognition and Behaviour, Radboud University/Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

## Funding information

Universiteit Maastricht; Nederlandse Organisatie voor Wetenschappelijk Onderzoek, Grant/Award Number: 452-16-004

## Abstract

Several theories of predictive processing propose reduced sensory and neural responses to anticipated events. Support comes from magnetoencephalography/electroencephalography (M/EEG) studies, showing reduced auditory N1 and P2 responses to self-generated compared to externally generated events, or when the timing and form of stimuli are more predictable. The current study examined the sensitivity of N1 and P2 responses to statistical speech regularities. We employed a motor-to-auditory paradigm comparing event-related potential (ERP) responses to externally and self-triggered pseudowords. Participants were presented with a cue indicating which button to press (motor-auditory condition) or which pseudoword would be presented (auditory-only condition). Stimuli consisted of the participant's own voice uttering pseudowords that varied in phonotactic probability and syllable stress. We expected to see N1 and P2 suppression for self-triggered stimuli, with greater suppression effects for more predictable features such as high phonotactic probability and first-syllable stress in pseudowords. In a temporal principal component analysis (PCA), we observed an interaction between syllable stress and condition for the N1, where second-syllable stress items elicited a larger N1 than first-syllable stress items, but only for externally generated stimuli. We further observed an effect of syllable stress on the P2, where first-syllable stress items elicited a larger P2. Strikingly, we did not observe motor-induced suppression for self-triggered stimuli for either the N1 or P2 component, likely due to the temporal predictability of the stimulus onset in both conditions. Taking into account previous findings, the current results suggest that sensitivity to syllable stress regularities depends on task demands.

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

Milene Bonte, Bernadette M. Jansma and Sonja A. Kotz contributed equally to this study.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *European Journal of Neuroscience* published by Federation of European Neuroscience Societies and John Wiley & Sons Ltd.

**KEYWORDS**

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

**1 | INTRODUCTION**

The brain's capacity to formulate predictions of upcoming events in the 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-generated 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 (e.g., Christoffels et al., 2011). 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), and prototypicality of vowels (Niziolek et al., 2013). 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.

One aspect of MIS that is poorly understood is the extent to which it depends on the predictability of the timing of stimulus onset. Studies investigating MIS typically do so by comparing identical stimuli to each other that are either self-generated (e.g., via button press or speech production) or externally presented (e.g., Knolle et al., 2013a; Niziolek et al., 2013; Ott & Jäncke, 2013; Pinheiro et al., 2018). This will often be done in a blocked design, where the externally presented stimuli are presented at the same time intervals as the self-triggered stimuli. Although this approach preserves the temporal structure of the stimulus streams across conditions, it does not fully control for the temporal predictability (Hughes et al., 2013). Some studies have attempted to address this question, by either introducing temporal uncertainty in the self-generated stimuli (Bäss et al., 2008; Lange, 2011; Pinheiro et al., 2019) or enhancing the temporal predictability of the externally presented stimuli through external cues (Harrison et al., 2021; Ody et al., 2022; Sowman et al., 2012). Their results suggest that the effect of temporal predictability on MIS may depend on design parameters: Introducing temporal uncertainty in self-generated stimuli typically leads to preserved MIS, whereas temporal predictability of the externally presented stimuli seems to attenuate this suppression effect.

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 (faster and slower), 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 behavioural 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 and *present* = to give a presentation; Cutler, 2005). In production tasks, such as nonword repetition, both phonotactic probability (Edwards et al., 2004; Munson, Edwards, & Beckman, 2005; Munson, Kurtz, & Windsor, 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 behavioural observations of facilitated processing of more regular items in speech perception, with variations in phonotactic probability and stress patterns modulating 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 are sparse. Functional magnetic resonance imaging (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 blood oxygenation level-dependent (BOLD) signal for items with higher frequency of occurrence within the language (Papoutsis 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', whereas 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 are 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 metre during speech production, rather than a function of typical/atypical stress patterns, which is further supported by behavioural findings in trisyllabic stimuli (Schiller et al., 2006). Currently, there seem to be no studies investigating the effect of variations in phonotactic probabilities in 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 on MIS as this allows investigating how such (ir)regularities modulate the accuracy of the predictions. Prior studies have shown that the suppression effect is sensitive to subconscious variations in the predictability of the speech signal (Behroozmand & Larson, 2011; Niziolek et al., 2013). Although some studies have investigated this phenomenon in overt speech production (e.g., Christoffels et al., 2011; Niziolek et al., 2013), this comes with challenges due to artefacts 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 triggers the presentation of speech stimuli via button press (e.g., Knolle et al., 2019; Ott & Jäncke, 2013; Pinheiro et al., 2018). Although participants were able to consciously anticipate the upcoming stimulus in our paradigm, our aim was to investigate whether implicit knowledge of statistical regularities in speech (high vs. low phonotactic probability and first- vs. second-syllable stress) would influence the strength of these predictions, resulting in modulations of the N1 and P2 suppression effects similar to those seen for subconscious variation in the speech signal in paradigms using overt speech (Behroozmand & Larson, 2011; Niziolek et al., 2013).

The classical design in these experiments employs three conditions: an auditory-only (AO) condition, where participants are passively presented with auditory pseudowords; a motor-auditory (MA) condition, where participants trigger the presentation of self-produced pseudowords through a button press; and finally, a

motor-only (MO) control condition used to correct for the motor activity ( $MA - MO = MA$  corrected [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. An observed reduction of N1 amplitude in response to self-triggered stimuli is thought to reflect an unconscious, automatic prediction resulting from the efference copy/corollary discharge, whereas P2 suppression reflects a more conscious differentiation between self-generated and externally generated events (e.g., Bolt & Loehr, 2021; 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. If implicit knowledge of statistical regularities of speech influences the accuracy of the prediction generated by the corollary discharge, we would expect this to result in modulations of N1 and P2 suppression effects. 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 [AO vs. MAC], MIS); (2) this reduction in amplitude is modulated by phonotactic probability and syllable stress (i.e., interactions between phonotactic probability [high vs. low] and condition, and syllable stress [first vs. second] 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 MIS (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

Thirty-four 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 electroencephalography (EEG) session.

Five 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 =  $\pm 3.8$ ), who completed both sessions of the experiment. The stimulus recording procedures and variations of the EEG paradigm were piloted in an additional nine 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, adapted from Bonte et al., 2005; Emmendorfer et al., 2020), which differed from each other in phonotactic probability (notsal vs. notfal, quantified based on the log-frequency counts of the consonant clusters at the syllable boundary, calculated as 5.83 and 4.72 for /ts/ and /tf/, respectively; Bonte et al., 2005) and syllable stress (first vs. second syllable). 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

TABLE 1 Stimuli.

	SylStr	
	SylS1	SylS2
PhonProb		
High (HPP)	<b>Notsal</b>	<b>Notsal</b>
Low (LPP)	<b>Notfal</b>	<b>Notfal</b>

Note: Bold font indicates stressed syllable (SylS1, first syllable; SylS2, second syllable). The phoneme combination 'ts-' constitutes the HPP and 'tf-' the LPP.

Abbreviations: HPP, high phonotactic probability; LPP, low phonotactic probability; PhonProb, phonotactic probability; SylStr, syllable stress.



and stress pattern as the target pseudowords (e.g., /badz-out/ → /**not**zout/ and /badsal/ → **not**sal, /ontslag/ → /**not**slag/ and /ontsal/ → /**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.

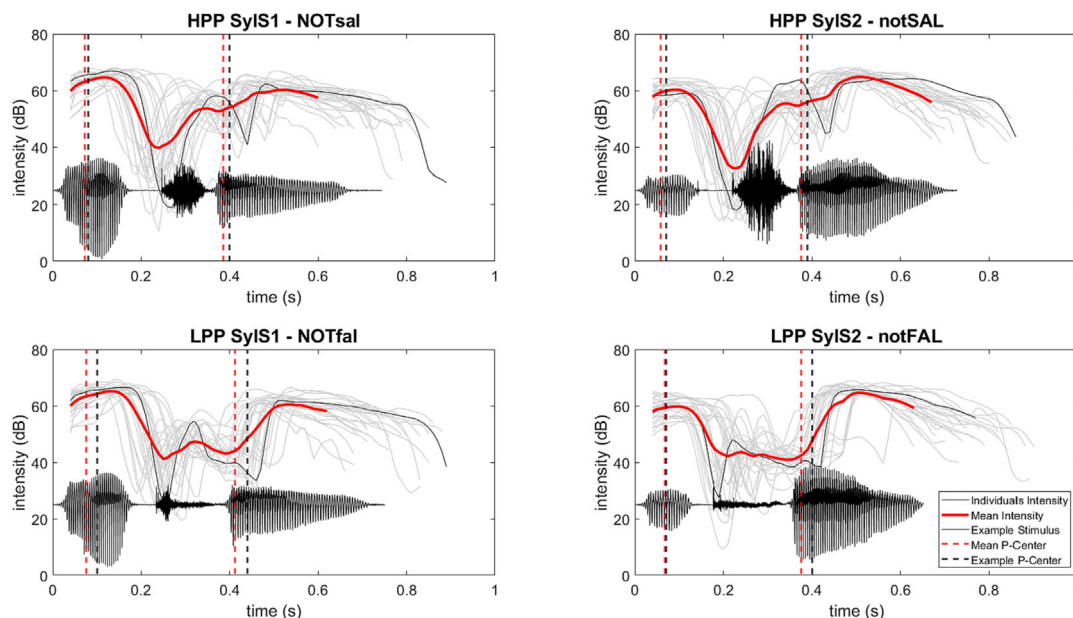
From the 15 repetitions of each pseudoword, one item was selected as the stimulus for the EEG experiment. To ensure comparability across participants, without having to manipulate the recording to deviate from the participants own naturally produced utterance, we selected items such that they were comparable in the timing of the perceptual centres (p-centres) of the syllables. P-centres are thought to represent the perceived 'beat' of the speech stimulus. The timing of the p-centres was estimated with a beat detection algorithm (custom MATLAB script adapted from Cummins & Port, 1998). Here, the beat or p-centre 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-centres 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 artefacts 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 (80–10,500 Hz) and intensity scaled to 60 dB. Mean stimulus duration was .640 s (standard deviation = .056 s), and the mean interval between p-centres of the stimuli was .319 s (standard deviation = .042 s).

## 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 and > right) at .4–1.0 s after trial onset. In the MA condition, participants pressed a button (left or right), which triggered the presentation of a stimulus (due to technical limitations, the stimulus presentation was delayed by 12.36 ms on average; however, such a delay is far below the detection threshold for trained musicians, which lies around 100 ms [van Vugt & Tillman, 2014] and would thus be perceived as simultaneous). In the AO condition, participants were presented with the same cue, but the stimulus presentation occurred without button press, .5 s after cue onset. In the MO condition, 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 MAC condition 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 MIS.

The EEG recording occurred over the course of six experimental runs, each consisting of 18 blocks (8 MA, 8 AO and 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, allowing the participant to anticipate the upcoming token in a mini-block. Each pair was presented twice per run and condition, with the cue/button assignment counterbalanced across blocks. Within each block, the first four trials (always including two left and two right) were excluded from analysis to allow the participant to form an association between cue and word. In four blocks per run (two MA and two AO), four catch trials were included at the end of the block, where the cue–stimulus pairing was switched; that is, 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. Note that this task required participants to memorize the button–stimulus association and thus consciously anticipate the specific token (notsal vs. notfal and first- vs. second-syllable stress). In the example presented in Figure 2a, the stimulus pair is **not**sal–**not**fal, so the participant would form the association of right button press corresponding to **not**sal and left button press corresponding to **not**fal. If this block contained a switch, the last four trials would have the reversed association (left button **not**sal and right button **not**fal). We



**FIGURE 1** Stimuli. Stimuli selected for the electroencephalography (EEG) experiment. Individual intensity contours of the stimuli are represented in grey and mean intensity contours across participants in red. Stimuli from an exemplary participant are represented in black. Timing of the p-centres, representing the onset of the vocalic nucleus, is represented by dashed lines (red: averaged across participants; black: exemplary participant). Individual tokens to be used for each participant were selected based on the interval between the p-centres for first and second syllables. Stimulus onset  $t = 0$  is equivalent to  $t = 0$  in the subsequent event-related potential (ERP) plots. Subplot titles indicate which stimulus is represented: HPP, high phonotactic probability; LPP, low phonotactic probability; SylS1, first-syllable stress; SylS2, second-syllable stress. Notsal and notfal represent stimulus tokens, where capitalization indicates the stressed syllable.

hypothesized that implicit knowledge of the statistical regularities of the more abstract stimulus features (high vs. low phonotactic probability and typical vs. atypical syllable stress) would modulate the strength of these predictions as reflected in N1 and P2 suppression effects. 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).

## 2.4 | EEG recording

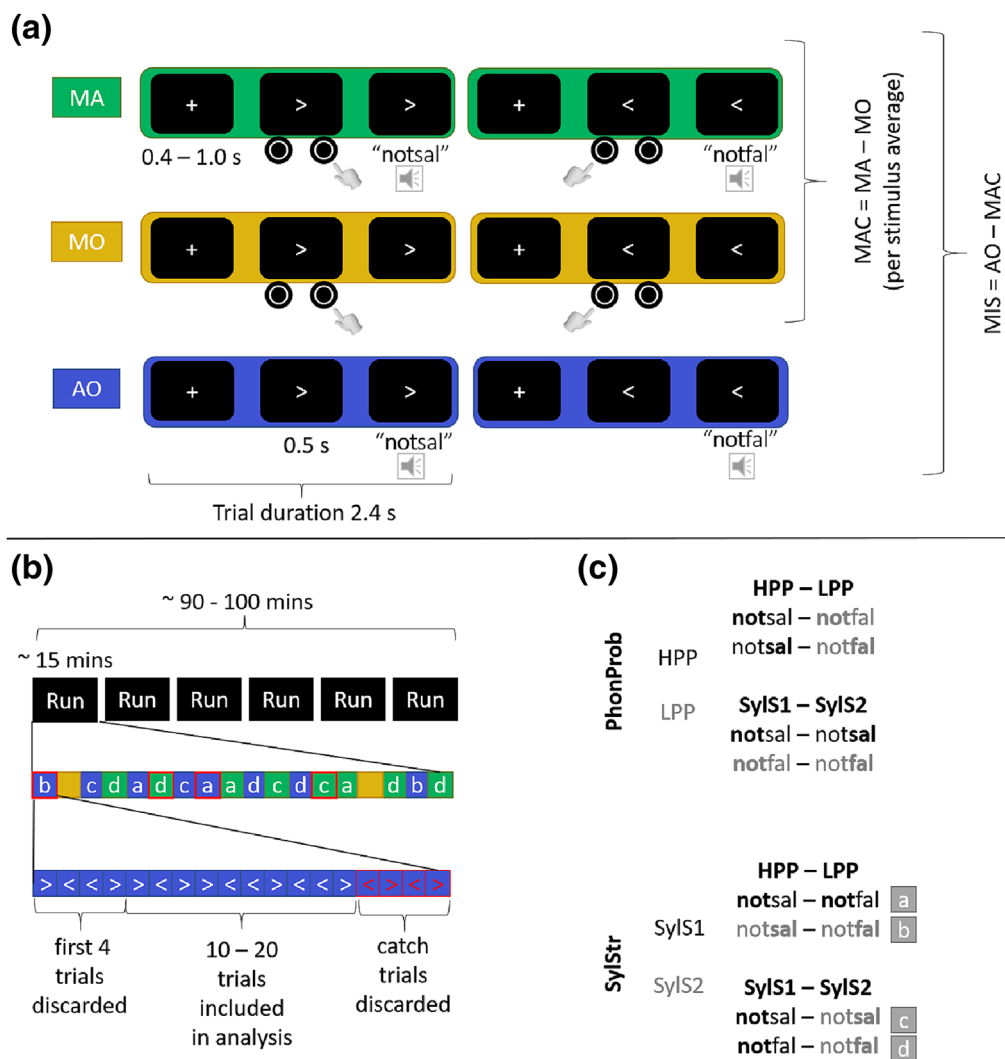
EEG was recorded with BrainVision Recorder (Brain Products, Munich, Germany) using a 63-channel recording setup. Ag/AgCl sintered electrodes were mounted according to the 10% equidistant system, including 57 scalp electrodes, left and right mastoids for offline re-referencing and four electrooculogram (EOG) electrodes to facilitate removal of artefacts caused by eye movements (two placed on the outer canthi and two above and below the right eye). The scalp was cleaned at electrode sites, and electrodes were filled with electrolyte gel to keep impedances below 10 k $\Omega$ . Data were acquired

with a sampling rate of 1000 Hz, 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 were processed using the EEGLAB toolbox (Delorme & Makeig, 2004) and custom MATLAB scripts. 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 independent component analysis (ICA). Two to four independent components, reflecting blinks and horizontal eye movements, were removed for each participant.

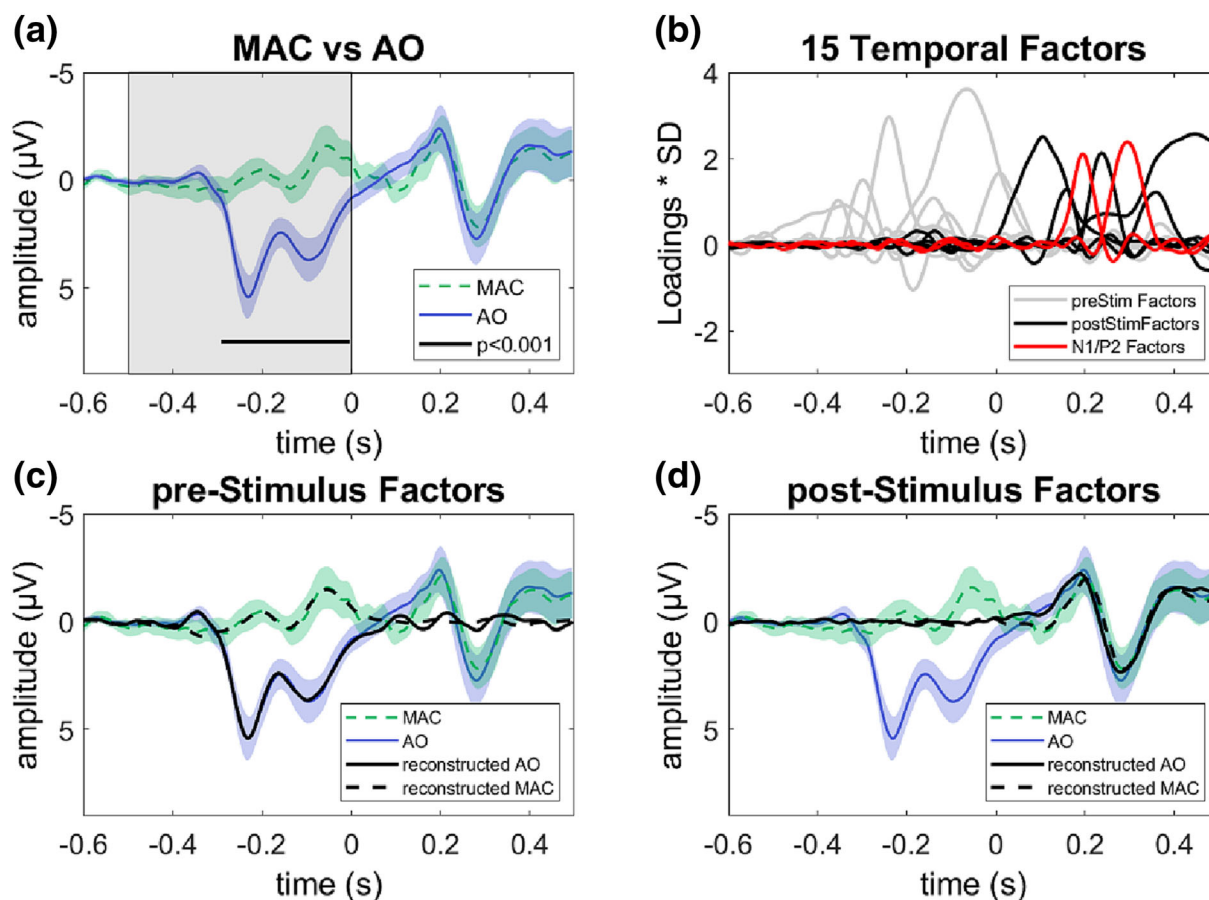
Initial inspection of the reconstructed data, using a typical pre-stimulus baseline correction, revealed an unexpected negative shift in the AO condition relative to



**FIGURE 2** Experimental design. (a) Three experimental conditions. In all three conditions, the trial started with the presentation of a fixation cross for .4–1.0 s, followed by a visual cue < or >. In the motor–auditory (MA) condition, participants pressed the button corresponding to the cue (< left or > right), triggering the presentation of the pseudoword stimulus via speakers. In the motor-only (MO) condition, the participants pressed the button, but no stimulus was presented. In the auditory-only (AO) condition, the stimulus was presented .5 s after cue onset without button press. MAC, motor–auditory corrected; MIS, motor-induced suppression. (b) Overview of the electroencephalography (EEG) paradigm timeline. The total EEG measurement lasted 90–100 min, consisting of six runs of approximately 15 min each. Within each run, 18 mini-blocks were presented (8 AO, 8 MA and 2 MO). In each mini-block, one stimulus pair was presented (letters a–d correspond to the stimulus pair presented as denoted in [c]), where one stimulus was associated with the < (left) cue/button press and one with the > (right) cue/button press. Within the eight mini-blocks of AO and MA, each pair was presented twice, with the cue/hand assignment counterbalanced across mini-blocks. AO and MA conditions consisted of 14–28 trials per mini-block. The first four trials were discarded from analysis. In four mini-blocks per run (two AO and two MA), four catch trials where cue/hand assignment was switched were included at the end of a mini-block. These trials were also discarded from analysis, leading to a final 10–20 trials per mini-block. (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.

the MAC condition. To explore whether this may be due to systematic differences between conditions in the baseline window, we expanded the pre-stimulus window. This revealed a positive deflection preceding the stimulus/button press in all three conditions, which appeared time locked to the visual cue (Figure S1). This deflection could not be removed through high-pass filtering

(Figure S2) or ICA (Figure S3). Although the deflection was present in all three conditions, it was effectively removed during the MA – MO subtraction (Figure 3a). The remaining deflection in the AO condition likely reflects a combination of visual processing of the cue, as well as anticipatory processes and temporal orientation to the upcoming stimulus due to the fixed temporal



**FIGURE 3** Overview of pre-stimulus deflection and temporal principal component analysis (PCA) results. (a) Motor–auditory-corrected (MAC, green) and auditory-only (AO, blue) conditions, time locked to stimulus onset and averaged across stimuli ( $\pm 95\%$  confidence interval [CI] of the mean) in a frontocentral region of interest (ROI) (FCz, FC1, FC2, FC3 and 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) Unstandardized factor loadings for 15 temporal factors obtained from temporal PCA with Promax rotation. Factors related to the pre-stimulus deflection in grey. Factors related to post-stimulus activity in black and red, where red indicates the two factors identified as N1 and P2 components based on their timing and polarity. (c) Reconstructed event-related potential (ERP) (loadings  $\times$  SD  $\times$  factor scores) based on all pre-stimulus factors (black) overlaid on original ERP data (blue and green). (d) Reconstructed ERP based on all post-stimulus factors.

interval between cue and stimulus in this condition (Figure 2a). This observation violates the assumption of baseline correction that there are no systematic differences between conditions in the selected window. Therefore, the data were instead first epoched  $-0.2$  to  $1.3$  s relative to the onset of the *cue*, and baseline correction was applied relative to  $0.2$  s prior to the onset of the *cue*. The data were then subsequently epoched  $-0.6$  to  $0.5$  s relative to the onset of the *stimulus* or *button press*. Although this window does not include the full stimulus duration (on average  $0.640$  s), it is sufficient to present the N1 and P2 components while ensuring that the window presented only contains data from the current trial and is not contaminated by the presentation of the fixation cross from the subsequent trial. Given the  $0.5$ -s interval between

cue and stimulus onset in the AO condition, this means that the first  $0.1$  s of the epoch ( $-0.6$  to  $-0.5$  s) includes a portion of the baseline window. Due to the variability of the interval duration in the MA and MO conditions, the timing of the baseline window relative to *button press* is variable; however, as in the AO condition, the baseline window is prior to the *cue* onset, during the presentation of the fixation cross. Previous findings have shown differences between self-generated and externally generated auditory stimuli already in the pre-stimulus window (Reznik et al., 2018); thus, a pre-stimulus baseline window may not be appropriate for this type of paradigm, even without the positivity we observe. The pre-cue baseline correction was used throughout the analysis steps. Epochs with reaction times shorter than  $400$  ms or longer



than 800 ms after cue onset were excluded from analysis. As this resulted in a smaller number of trials in the MA condition compared to the AO condition, trial numbers were equalized across stimuli and conditions for each participant, to avoid unequal numbers of trials biasing the results. A minimum of 100 trials per stimulus per condition (900 trials total) were retained per participant.

Before moving to the planned analysis of the N1 and P2 components, we first explored the pre-stimulus positivity through cluster-based permutation analysis. 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 MIS 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  $-0.5$  to  $0$  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  $.3$  s prior to stimulus onset, with a broad topographic distribution.

To isolate the components of interest from this pre-stimulus deflection, we followed a temporal principal component analysis (PCA) approach (Korka et al., 2019) using the event-related potential (ERP) PCA toolkit (Dien, 2010). Average waveforms per participant, condition and stimulus were entered into a temporal PCA. In total, this resulted in nine average waveforms per participant: four AO (one per stimulus), four MA and one MO. The input to the PCA was thus a matrix of 275 variables (time points) and 243 observations (9 waveforms \* 27 participants). Based on the results of Horn's parallel test, 15 temporal components (Table S1) explaining 94.99% of the variance in the data were retained for Promax rotation ( $k = 3$ ) with a covariance matrix structure and Kaiser weighting. PCA decomposition yields two sets of coefficients to describe the EEG signal: Factor loadings correspond to the time course of a factor, which is constant across all conditions, participants and electrodes; factor scores correspond to the contribution of each factor to the EEG signal of each observation (participant, condition and electrode) and can be directly used for statistical analyses to quantify differences across observations. The ERP can thus be described as the sum of unstandardized factor loadings multiplied by the corresponding factor scores (Scharf et al., 2022). The unstandardized factor loadings (loadings \* standard deviation) of these 15 components are presented in Figure 3b. Components relating to pre- and post-stimulus/button-press activity

(represented in grey and black, respectively) were identified based on their peak latency. Figure 3c,d illustrates that the reconstructed ERPs (factor loadings \* standard deviation \* factor scores) from these component groups accurately align with the pre- and post-stimulus/button-press activity in the original ERPs, while minimizing modulations in the other time window. Two temporal components reflecting the N1 and P2 observed in the dataset were identified based on their timing and polarity: temporal component 7 peaking at 196-ms post-stimulus/button-press onset as the N1 and temporal component 5 peaking at 296-ms post-stimulus/button-press onset as the P2 (represented in red in Figure 3b). These timings are later than the classically observed N1 and P2 latencies. 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). When adjusted for the timing of the p-centre 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.

## 2.6 | Statistical analyses

PCA factor scores corresponding to the N1 and P2 components were entered into statistical analysis in R Version 4.1.0 (R Core Team), using functions from the `rstatix` (Kassambara, 2021), `ggpubr` (Kassambara, 2020), `gridExtra` (Auguie, 2017) and `ggplot2` (Wickham, 2016) packages. Normal distribution of the N1 and P2 factor scores was confirmed for all conditions via Shapiro-Wilk test (Tables S1 and S4), and outlier identification via boxplot methods did not reveal any extreme outliers (points beyond  $Q1 - 3 * \text{inter-quartile range [IQR]}$  and  $Q3 + 3 * \text{IQR}$ ). In two separate  $2 \times 2 \times 2$  repeated-measures analyses of variance (ANOVAs) (high vs. low PhonProb  $\times$  first vs. second SylStr  $\times$  AO vs. MAC Cond), we tested the following hypotheses for both N1 and P2 factor scores averaged across electrodes within a region of interest (ROI) determined by the PCA component's peak channel (FCz, FC1, FC2, FC3 and FC4 for N1 and Cz, C1, C2, C3 and C4 for P2): (1) N1 and P2 amplitudes are reduced for self-triggered stimuli compared to externally generated stimuli (i.e., main effect of Cond, MIS); (2) this amplitude reduction is modulated by phonotactic probability and syllable stress (i.e., interactions between PhonProb and Cond, and SylStr and Cond), with high phonotactic probability and first-syllable stress items leading to a greater amplitude reduction due to greater

predictability; and (3) phonotactic probability and syllable stress may interactively modulate MIS (i.e., three-way interaction between PhonProb, SylStr and Cond), where we did not have precise predictions about the nature of this interaction. ANOVA results were corrected for multiple comparisons with Bonferroni–Holm correction (Cramer et al., 2016), and follow-up *t* tests of simple effects were Bonferroni corrected.

## 3 | RESULTS

### 3.1 | Behavioural results

Due to non-normal distribution of the response time and sensitivity scores ( $d'$ ), behavioural analyses were performed using non-parametric tests. For two participants, behavioural data were not recorded for one out of six experimental runs due to a technical error. Excluding these participants from the analysis did not change the outcome of behavioural or ERP results. The average time of button presses was 508.45 ms after the cue presentation (standard deviation = 44.77 ms). Non-parametric paired-samples Wilcoxon tests revealed a significant difference in response time across conditions ( $W = 299$ ,  $p = .007$ ), with longer response times in the MA condition (mean = 518.62 ms, standard deviation = 43.54 ms) than the MO condition (mean = 499.29 ms, standard deviation = 44.90 ms). This difference likely reflects increased attentional demands in the MA condition, where participants had to direct their attention to the upcoming stimulus.

Sensitivity to the catch trials was assessed with  $d'$  calculated with the psycho R package. Mean  $d'$  across conditions revealed that participants were able to perform the task well (mean = 3.21, standard deviation = .77). Non-parametric paired-samples Wilcoxon tests revealed a significant difference in sensitivity to catches across conditions ( $W = 49.5$ ,  $p = .0129$ ), with improved sensitivity in the MA condition (mean = 3.36, standard deviation = .77) compared to the AO condition (mean = 3.08, standard deviation = .76). The active button-press task in the MA condition likely facilitated directing attention to the upcoming stimulus, resulting in this improved behavioural performance.

### 3.2 | ERP results

Visual inspection of the ERP grand averages (Figures 4a and 5a) revealed an N1/P2 morphology, with the N1 peaking around 200 ms and the P2 around 300 ms. When

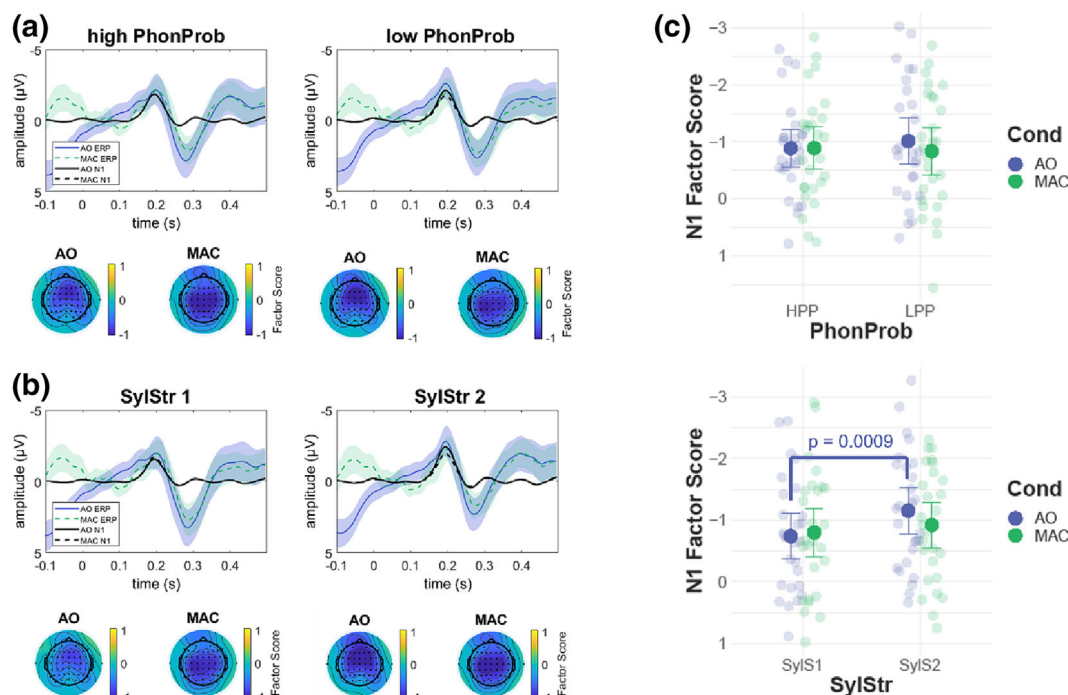
adjusted for the timing of the p-centre of the first syllable of participants' pseudoword pronunciations, the N1 and P2 latencies were shorter, at approximately 125 and 212 ms, respectively. For our analyses, we kept the time locking to stimulus onset as it resulted in delayed but better aligned N1 and 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 as well as descriptive statistics can be found in Tables S2–S4 for N1 and Tables S5 and S6 for P2 results.

#### 3.2.1 | N1

A  $2 \times 2 \times 2$  repeated-measures ANOVA (high vs. low PhonProb  $\times$  first vs. second SylStr  $\times$  AO vs. MAC Cond) on N1 factor scores averaged within the frontocentral ROI revealed a significant interaction between syllable stress and condition ( $F(1, 26) = 12.468$ ,  $p_{\text{adj}} = .014$ ,  $\eta^2_p = .324$ ). There was no main effect of condition ( $F(1, 26) = .330$ ,  $p_{\text{adj}} = 1.000$ ,  $\eta^2_p = .013$ ). However, a main effect of syllable stress was approaching significance ( $F(1, 26) = 7.402$ ,  $p_{\text{adj}} = .066$ ,  $\eta^2_p = .222$ ). Follow-up tests of simple effects revealed no effect of condition (AO vs. MAC) for either first ( $t(26) = .395$ ,  $p_{\text{adj}} = 1.000$ ,  $d = .0760$ ) or second-syllable stress items ( $t(26) = -1.42$ ,  $p_{\text{adj}} = .332$ ,  $d = .274$ ), indicating no MIS. Instead, this interaction was driven by an effect of syllable stress in the AO condition ( $t(26) = 4.02$ ,  $p_{\text{adj}} = .0009$ ,  $d = .773$ ), where second-syllable stress items showed a larger amplitude (factor score =  $-1.16$ , standard deviation = .957) compared to first-syllable stress items (factor score =  $-.741$ , standard deviation = .927). This effect was not present in the MA condition ( $t(26) = 1.09$ ,  $p_{\text{adj}} = .572$ ,  $d = .209$ ). No other main effects or interactions on N1 mean amplitude were significant (Figure 4).

#### 3.2.2 | P2

A  $2 \times 2 \times 2$  repeated-measures ANOVA (high vs. low PhonProb  $\times$  first vs. second SylStr  $\times$  AO vs. MAC Cond) on P2 factor scores averaged within the frontocentral ROI revealed a significant main effect of syllable stress ( $F(1, 26) = 24.498$ ,  $p_{\text{adj}} = .0003$ ,  $\eta^2_p = .485$ ). Stimuli with first-syllable stress elicit a larger P2 (factor score = 1.17, standard deviation = .995) compared to those with second-syllable stress (factor score = .786, standard deviation = .906; Figure 5). No other main effects or interactions were significant.

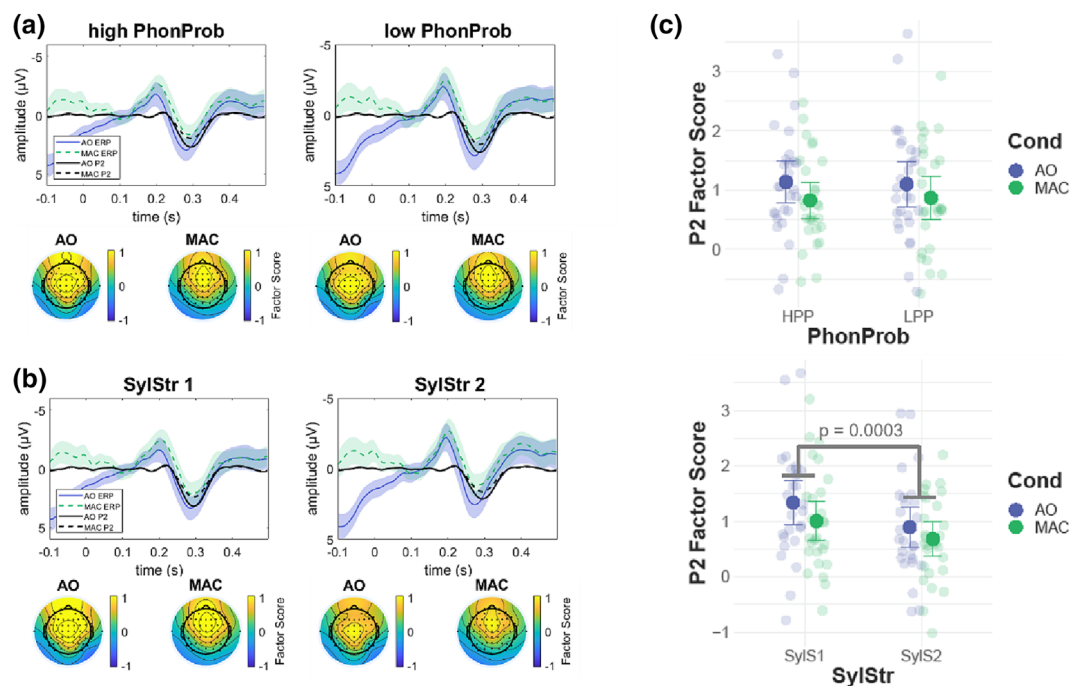


**FIGURE 4** Effects of phonotactic probability and syllable stress on N1 factor scores. Plots represent the event-related potential (ERP) waveforms (blue and green) and reconstructed N1 (black; factor loadings \* SD \* factor score) in microvolt scale and time locked to stimulus onset, topographic distribution of N1 factor scores and mean N1 factor scores averaged within a frontocentral region of interest (ROI) (FCz, FC1, FC2, FC3 and FC4). Effects of phonotactic probability are presented in (a) and for syllable stress in (b). A significant interaction between syllable stress and condition was observed, which was driven by an effect of syllable stress in the auditory-only (AO) condition, where second-syllable stress items elicited a larger N1 compared to first-syllable stress items (c, bottom,  $p = .0009$ ). Note that baseline correction was performed in 200-ms window prior to cue onset (approximately  $-0.5$  s). Due to the pre-stimulus positivity (see Figure 3), AO and motor-auditory corrected (MAC) are different in the pre-stimulus window depicted here. Cond, condition; HPP, high phonotactic probability; LPP, low phonotactic probability; PhonProb, phonotactic probability; SylStr, syllable stress; SylS1, first-syllable stress; SylS2, second-syllable stress.

## 4 | DISCUSSION

The current study aimed to investigate whether MIS 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 artefacts, variability in speech production and speech errors present during overt production of pseudowords. We expected to observe an MIS effect, with larger N1 and P2 amplitudes in the AO condition, compared to the MA 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 cue-locked pre-stimulus deflection in the AO condition, not present in the MA condition after correcting for motor output (Figure 3a), we applied a temporal PCA approach to isolate the N1 and P2 components of interest. Our hypotheses were tested on the factor scores from two temporal factors that aligned with the N1 and P2 components in the current dataset. We observed an interaction between syllable stress and condition on N1 factor scores, where second-syllable stress items elicited a larger N1 compared to first-syllable stress items, but only in the AO condition. Syllable stress further modulated P2 factor scores, where now first-syllable stress items elicited greater activation compared to second-syllable stress. Strikingly, we did not observe any MIS effect for self-triggered (MAC) compared to externally triggered (AO) stimuli, unlike a large body of past literature comparing self-triggered and externally triggered auditory stimuli (e.g., Bäss et al., 2008; Knolle et al., 2012, 2013a, 2013b, 2019; Niziolek et al., 2013; Pinheiro et al., 2018). In the following, we offer



**FIGURE 5** Effects of phonotactic probability and syllable stress on P2 factor scores. Plots represent the event-related potential (ERP) waveforms (blue and green) and reconstructed P2 (black; factor loadings \* SD \* factor score) in microvolt scale and time locked to stimulus onset, topographic distribution of P2 factor scores and mean P2 factor scores averaged within a frontocentral region of interest (ROI) (FCz, FC1, FC2, FC3 and FC4). Effects of phonotactic probability are presented in (a) and for syllable stress in (b). A main effect of syllable stress on P2 factor scores was observed, where first-syllable stress items elicited a larger P2 compared to second-syllable stress items (c, bottom,  $p = .0003$ ). Note that baseline correction was performed in 200-ms window prior to cue onset (approximately  $-0.5$  s). Due to the pre-stimulus positivity (see Figure 3), auditory only (AO) and motor-auditory corrected (MAC) are different in the pre-stimulus window depicted here. Cond, condition; HPP, high phonotactic probability; LPP, low phonotactic probability; PhonProb, phonotactic probability; SylStr, syllable stress; SylStr1, first-syllable stress; SylStr2, second-syllable stress.

interpretations for the effects of syllable stress on the N1 and P2 components and provide possible explanations for the lack of MIS effect.

#### 4.1 | Syllable stress variations modulate N1 and P2 components

First- and second-syllable stress items differ from each other in terms of both acoustic properties of the stimuli as well as their likelihood of occurrence in the language at hand. Pitch, intensity and duration serve as perceptual markers of stress in Dutch, with stressed syllables exhibiting increased pitch, intensity and duration relative to their unstressed counterparts. As obligatory auditory evoked potentials, both N1 and P2 have been shown to be modulated by the acoustic features related to perceptual markers of syllable stress. Increased stimulus amplitude and duration have both been associated with increased N1 and P2 amplitudes (Alain et al., 1997; Ostroff et al., 2003; Paiva et al., 2016). Changes in stimulus frequency show the reversed effect: Increases in tonal

frequency are associated with decreased amplitudes (Antinoro et al., 1969; Jacobson et al., 1992; Pantev et al., 1995; Wunderlich & Cone-Wesson, 2001). However, in these studies, frequency manipulations occurred predominantly in ranges above 400 Hz, which is outside of the typical  $f_0$ s of human speech, typically between 100 and 250 Hz (e.g., Pépiot, 2014; Pisanski et al., 2020). If the observed N1 and P2 modulations are related to acoustic differences, this would most likely be reflected in increases in N1 and P2 factor scores for first-syllable stress items relative to second-syllable stress items, which should be consistent across conditions (AO vs. MAC). The main effect of syllable stress on the P2 component, which does not differ across conditions, is therefore consistent with an effect of acoustic stimulus properties.

The pattern of modulations observed on the N1 component tells a different story. Here, we observe an *interactive* effect of syllable stress and condition, with the effect of syllable stress being present only in the AO condition. The directionality of this effect, with a larger N1 for the less probable, second-syllable stress items, further supports the notion that statistical regularities in the



temporal structure of speech may contribute to the observed effect. Theories of predictive processing propose that greater surprisal, or unexpected input, lead to increased neural responses (Aitchison & Lengyel, 2017; Friston, 2005; Rao & Ballard, 1999; Spratling, 2017). Additionally, the forward model proposes that self-triggered stimuli are associated with reduced neural responses as the sensory outcome can be anticipated based on the motor plan (Blakemore et al., 2000; Miall & Wolpert, 1996). The lack of effect of syllable stress in the MA condition suggests that first- and second-syllable stress items could be anticipated equally well when they are self-triggered. However, when externally presented in the AO condition, the less regular second-syllable stress item may have elicited greater surprisal, resulting in a larger neural response.

An alternative explanation for the differences across conditions may lie in attentional demands differing across conditions. The attention hypothesis proposes that the classically observed suppression effect is a result of attentional resources being divided between the motor output and sensory input (Horváth, 2015). Although we did not observe a suppression effect, this hypothesis may still explain the lack of sensitivity to syllable stress observed in the MA condition. However, the behavioural results suggest that participants were in fact able to attend *better* to the upcoming stimuli when they followed a button press rather than when they were externally triggered, which is not consistent with the attention hypothesis. Thus, if the current results are related to differences in attentional demands, it is more likely that the increased attention to the stimuli in the MA condition resulted in more accurate predictions of the upcoming stimulus, such that both first- and second-syllable stress items could be anticipated equally well.

## 4.2 | No effect of phonotactic probability

The results of the current study stand in contrast to findings from research investigating the effect of sublexical regularities on the mismatch negativity (MMN) using similar stimuli in a passive oddball paradigm (Emmendorfer, 2022; Emmendorfer et al., 2020). These studies observed modulations of the MMN by variations in phonotactic probability, but not syllable stress, in contrast to the present results that indicate an effect of syllable stress, but not phonotactic probability, on N1 and P2 amplitudes. We first focus on the effects of syllable stress and examine the role that lexical stress plays in Dutch. Although they differ in their probability of occurrence, both first- and second-syllable stress patterns are legal in Dutch, and variations in stress can be used to resolve

lexical conflict in the cases where the meaning of the word differs depending on which syllable is stressed (e.g., **present** vs. **present**; Cutler, 2005; Cutler & Van Donselaar, 2001). The prior studies interpret the lack of an effect of syllable stress as an indicator that, for Dutch speakers, variations in stress patterns are not an important factor in passive listening to individual pseudowords in the absence a sentence context. Although the stimuli in the current paradigm also consisted of individual pseudowords, the paradigm differed from prior passive oddball studies in that participants had to distinguish between first- and second-syllable stress in order to correctly perform the task. These increased attentional demands relative to passive listening in an oddball task likely contributed to the difference in effects of syllable stress across studies.

Moving to the effects of phonotactic probability, prior studies employing similar stimuli showed a facilitative effect of high phonotactic probability relative to low phonotactic probability items on deviance detection (Bonte et al., 2005; Emmendorfer, 2022; Emmendorfer et al., 2020). Here, we did not observe any N1 or P2 modulation by variations in phonotactic probability. The reason for this likely lies in the stimulus selection: Phonotactic probability is modulated at the syllable boundary (~200–250 ms after stimulus onset), whereas the components of interest here peaked around 200 and 300 ms. Thus, the manipulation of phonotactic probability likely occurred too late to contribute to N1 and P2 modulations (though co-articulatory cues may already be present in the first syllable). Additionally, the consonant clusters /ts/ and /tf/ differed from each other less in both phonotactic probability and phonetic properties compared to the clusters /ts/ and /tk/ used in Emmendorfer et al. (2020). Although Bonte et al. (2005) reported MMN modulations for both notsel vs. notkel and notsel vs. notfel contrasts, these differences were larger for the /ts/ vs. /tk/ contrast. Thus, the combination of the relative timing of the consonant clusters within the pseudowords and that of the ERP components of interest, together with the smaller phonotactic difference between the stimuli, likely contributed to the absence of an effect of this feature on the N1 and P2 components.

## 4.3 | Lack of MIS due to temporal predictability

The current results also differ crucially from a large body of research investigating MIS (e.g., Bäss et al., 2008; Knolle et al., 2012, 2013a; Pinheiro et al., 2018) in that we do not observe a significant suppression effect. The lack of suppression may be explained by variations in the

design between previous studies and the current experiment. The typical approach involves 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 MA condition. The auditory stimuli are then presented at the same temporal intervals in the AO condition. In the current study, participants were instead instructed to press a button following the presentation of a visual cue. This same visual cue was presented in the AO condition, with the stimulus following at the fixed interval of 500 ms after cue presentation, introducing temporal predictability of the stimulus in the AO condition. This is a crucial difference in design compared to prior studies: In the classical approach, the AO and MA conditions differ not only in whether the stimulus is self-generated or externally presented but also in the predictability of the stimulus timing: In the MA condition, the participant can accurately predict the timing, whereas some temporal uncertainty remains in the AO condition.

A growing body of literature suggests that a considerable portion of the suppression effect observed in previous research may be driven by the temporal predictability of the events (Harrison et al., 2021; Hughes et al., 2013; Ody et al., 2022; Sowman et al., 2012; Storch & Zimmermann, 2022). However, several studies have also reported preserved suppression effects when manipulating temporal predictability (Bäss et al., 2008; Lange, 2011; Pinheiro et al., 2019). These studies differ in how they manipulated temporal predictability: While some did so by introducing variable delays between button press and sound presentation to make the timing of the stimuli in the MA condition more unpredictable (Bäss et al., 2008; Lange, 2011; Pinheiro et al., 2019), others did so by also introducing temporal predictability in the AO condition with cues from different modalities (Harrison et al., 2021; Ody et al., 2022; Sowman et al., 2012). Thus, precisely how temporal predictability influences the suppression effect may depend heavily on design parameters. Although the current study did not explicitly manipulate the temporal predictability of the stimulus presentation, the visual cue included in the AO condition led to a close match in temporal predictability across conditions. Thus, the current findings are in line with recent studies reporting a role of temporal prediction in MIS and highlight the need for this effect to be more carefully characterized in future studies.

#### 4.4 | Limitations

Although the lack of MIS is consistent with the temporal predictability of the stimulus presentation in the AO

condition, this was not the intended outcome of the current experiment. The introduction of the visual cue that was presented in all three conditions (adapted from Johnson et al., 2021) led to an undesired positive deflection in the AO condition, which was effectively removed from the MA condition when correcting for motor output (MA – MO). PCA can deal with the separation of overlapping components but cannot erase the possibility that the visual cue affected the processing of the stimuli. Including a visual control condition to subtract from the AO condition may have ameliorated this issue. However, this would only account for purely visual processes. The deflection likely also represents attentional and anticipatory processes, as the participants were 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 (Bäss et al., 2008; Harrison et al., 2021; Hughes et al., 2013; Lange, 2011; Ody et al., 2022; Pinheiro et al., 2019; Sowman et al., 2012).

A further limitation of the current design is in the manipulation of phonotactic probability. We selected stimuli similar to those used in our prior oddball paradigms (Bonte et al., 2005; Emmendorfer et al., 2020) to allow comparing results across studies. However, the timing of the manipulation of phonotactic probability was likely too late in the stimulus (~200–250 ms after stimulus onset) to influence the processes underlying N1 and P2 generation. Although subtle co-articulatory cues may already be present within the first syllable, these may not be salient enough to elicit differences in surprisal between stimuli. Therefore, the current results do not allow us to draw final conclusions about whether auditory evoked potentials including the N1 and P2 are sensitive to variations in phonotactic probability. Future studies may investigate this using different stimuli that are more suited to answer this question, for example, by manipulating phonotactic probability at stimulus onset. Alternatively, manipulating local stimulus regularities similarly to an oddball paradigm would allow us to test for modulations of the MMN elicited by self-triggered deviants (e.g., Korka et al., 2019), facilitating closer comparisons between experimental approaches.

Finally, we note that the current paradigm, using button press-triggered presentation of self-produced pseudowords, is only an indirect measure of speech production. We selected this approach to avoid interference from

motor artefacts associated with overt speech, as well as to control for variability in utterances and errors that may differ across stimulus features (high vs. low phonotactic probability and first- vs. second-syllable stress; Heisler & Goffman, 2016; Munson, 2001; Sasisekaran et al., 2010). Thus, although this approach results in strong experimental control, it does not account for the full complexity of overt speech production. Although other studies have followed a similar approach using button press-triggered speech (Conde et al., 2018; Knolle et al., 2019; Ott & Jäncke, 2013; Pinheiro et al., 2018), future research should consider adapting this paradigm to overt speech production to directly investigate the neural correlates of phonotactic probability and syllable stress during speech production.

## 5 | CONCLUSION

In conclusion, the present experiment provides insights into processing differences for phonotactic and syllable stress regularities in speech perception and production, by comparing self-triggered (via button press) to externally triggered (own) speech. We report novel observations suggesting that syllable stress regularities influence speech perception in Dutch speakers, where processing of more regular syllable stress patterns is facilitated. Considering previous findings that showed no such effect in a passive oddball paradigm (Emmendorfer et al., 2020), the role of syllable stress regularities appears to depend on whether task demands require attention to the stimuli. To summarize, the current results suggest 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 analyses.

## AUTHOR CONTRIBUTIONS

All authors designed the experiment and refined the manuscript. AKE prepared the materials, collected and analysed the data and wrote the first draft of the manuscript.

## ACKNOWLEDGEMENTS

The authors would like to thank Joseph Johnson and Stevan Nikolin for sharing the code that was adapted for the current experiment; Dana Hermsen and Alexa Holfelder for their assistance with data collection; and Pia Brinkmann, Suvarnalata Xanthate Duggirala, Francesco Gentile, Lisa Goller, Lars Hausfeld and Michael Schwartze for discussions on data analysis. This research was supported by Maastricht University

(Universiteit Maastricht) (personal grants to BMJ and SAK) and the Netherlands Organization for Scientific Research (Nederlandse Organisatie voor Wetenschappelijk Onderzoek) (Vidi Grant 452-16-004 to MB).

## CONFLICT OF INTEREST STATEMENT

The authors declare that they have no competing interests.

## DATA AVAILABILITY STATEMENT

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

## ORCID

Alexandra K. Emmendorfer  <https://orcid.org/0000-0002-0651-7620>

## PEER REVIEW

The peer review history for this article is available at <https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/ejn.16003>.

## REFERENCES

- Aitchison, L., & Lengyel, M. (2017). With or without you: Predictive coding and Bayesian inference in the brain. *Current Opinion in Neurobiology*, 46, 219–227. <https://doi.org/10.1016/j.conb.2017.08.010>
- Alain, C., L. Woods, D., & Covarrubias, D. (1997). Activation of duration-sensitive auditory cortical fields in humans. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 104(6), 531–539. [https://doi.org/10.1016/S0168-5597\(97\)00057-9](https://doi.org/10.1016/S0168-5597(97)00057-9)
- 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>
- Antinoro, F., Skinner, P. H., & Jones, J. J. (1969). Relation between sound intensity and amplitude of the AER at different stimulus frequencies. *The Journal of the Acoustical Society of America*, 46(6B), 1433–1436. <https://doi.org/10.1121/1.1911881>
- 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>
- Auguie, B. (2017). gridExtra: Miscellaneous functions for “grid” graphics (2.3) [computer software]. <https://CRAN.R-project.org/package=gridExtra>
- 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>
- Blakemore, S. J., Wolpert, D., & Frith, C. (2000). Why can't you tickle yourself? *Neuroreport*, 11(11), R11–R16. <https://doi.org/10.1097/00001756-200008030-00002>
- Bolt, N. K., & Loehr, J. D. (2021). Sensory attenuation of the auditory P2 differentiates self- from partner-produced sounds during joint action. *Journal of Cognitive Neuroscience*, 33(11), 2297–2310. [https://doi.org/10.1162/jocn\\_a\\_01760](https://doi.org/10.1162/jocn_a_01760)
- 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., Steingrover, 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>
- 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*, The Handbook of Speech Perception. Wiley. <https://doi.org/10.1002/9780470757024.ch11>
- 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.
- 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>
- Dien, J. (2010). The ERP PCA Toolkit: An open source program for advanced statistical analysis of event-related potential data. *Journal of Neuroscience Methods*, 187(1), 138–145. <https://doi.org/10.1016/j.jneumeth.2009.12.009>
- 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](https://doi.org/10.1044/1092-4388(2004)034)
- Emmendorfer, A. K. (2022). *Electrophysiological correlates of phonological and temporal regularities in speech processing* (Doctoral Thesis. Ipskamp Printing BV. <https://doi.org/10.26481/dis.20220222ae>
- 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>
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456), 815–836. <https://doi.org/10.1098/rstb.2005.1622>
- Harrison, A. W., Mannion, D. J., Jack, B. N., Griffiths, O., Hughes, G., & Whitford, T. J. (2021). Sensory attenuation is modulated by the contrasting effects of predictability and control. *NeuroImage*, 237, 118103. <https://doi.org/10.1016/j.neuroimage.2021.118103>
- 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>
- Horváth, J. (2015). Action-related auditory ERP attenuation: Paradigms and hypotheses. *Brain Research*, 1626, 54–65.
- 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>
- Jacobson, G. P., Lombardi, D. M., Gibbens, N. D., Ahmad, B. K., & Newman, C. W. (1992). The effects of stimulus frequency and recording site on the amplitude and latency of multichannel cortical auditory evoked potential (CAEP) component N1. *Ear and Hearing*, 13(5), 300–306. <https://doi.org/10.1097/00003446-199210000-00007>
- 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. (2020). ggpubr: “ggplot2” based publication ready plots (0.4.0) [computer software].
- Kassambara, A. (2021). rstatix: Pipe-friendly framework for basic statistical tests. <https://CRAN.R-project.org/package=rstatix>
- 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](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., Schwartz, 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>
- Korka, B., Schröger, E., & Widmann, A. (2019). Action intention-based and stimulus regularity-based predictions: Same or different? *Journal of Cognitive Neuroscience*, 31(12), 1917–1932. [https://doi.org/10.1162/jocn\\_a\\_01456](https://doi.org/10.1162/jocn_a_01456)
- Kotz, S. A., & Schwartz, 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., & Schwartz, 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>
- Lange, K. (2011). The reduced N1 to self-generated tones: An effect of temporal predictability? *Psychophysiology*, 48(8), 1088–1095. <https://doi.org/10.1111/j.1469-8986.2010.01174.x>
- Levelt, W. J. M. (1999). Models of word production. *Trends in Cognitive Sciences*, 3(6), 223–232. [https://doi.org/10.1016/s1364-6613\(99\)01319-4](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](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](https://doi.org/10.1016/s0010-0277(00)00109-8)
- Miall, R. C., & Wolpert, D. M. (1996). Forward models for physiological motor control. *Neural Networks*, 9(8), 1265–1279. [https://doi.org/10.1016/S0893-6080\(96\)00035-4](https://doi.org/10.1016/S0893-6080(96)00035-4)
- 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\(2001/061\)](https://doi.org/10.1044/1092-4388(2001/061))
- Munson, B., 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\(2005/006\)](https://doi.org/10.1044/1092-4388(2005/006))
- Munson, B., 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\(2005/072\)](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>
- Ody, E., Straube, B., He, Y., & Kircher, T. (2022). Perception of self- and externally-generated visual stimuli: Evidence from EEG and behaviour. *bioRxiv*, 2022.05.30.492531. <https://doi.org/10.1101/2022.05.30.492531>
- 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, 156869. <https://doi.org/10.1155/2011/156869>
- Ostroff, J. M., McDonald, K. L., Schneider, B. A., & Alain, C. (2003). Aging and the processing of sound duration in human auditory cortex. *Hearing Research*, 181(1), 1–7. [https://doi.org/10.1016/S0378-5955\(03\)00113-8](https://doi.org/10.1016/S0378-5955(03)00113-8)
- 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>
- Paiva, T. O., Almeida, P. R., Ferreira-Santos, F., Vieira, J. B., Silveira, C., Chaves, P. L., Barbosa, F., & Marques-Teixeira, J. (2016). Similar sound intensity dependence of the N1 and P2 components of the auditory ERP: Averaged and single trial evidence. *Clinical Neurophysiology*, 127(1), 499–508. <https://doi.org/10.1016/j.clinph.2015.06.016>
- Pantev, C., Bertrand, O., Eulitz, C., Verkindt, C., Hampson, S., Schuierer, G., & Elbert, T. (1995). Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings. *Electroencephalography and Clinical Neurophysiology*, 94(1), 26–40. [https://doi.org/10.1016/0013-4694\(94\)00209-4](https://doi.org/10.1016/0013-4694(94)00209-4)
- Papoutsis, 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>
- Pépiot, E. (2014). Male and female speech: A study of mean f0, f0 range, phonation type and speech rate in Parisian French and American English speakers. 305. <https://halshs.archives-ouvertes.fr/halshs-00999332>
- Pinheiro, A. P., Schwartz, M., Gutierrez, F., & Kotz, S. A. (2019). When temporal prediction errs: ERP responses to delayed action-feedback onset. *Neuropsychologia*, 134, 107200. <https://doi.org/10.1016/j.neuropsychologia.2019.107200>
- Pinheiro, A. P., Schwartz, 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>

- Pisanski, K., Raine, J., & Reby, D. (2020). Individual differences in human voice pitch are preserved from speech to screams, roars and pain cries. *Royal Society Open Science*, 7(2), 191642. <https://doi.org/10.1098/rsos.191642>
- 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. <https://doi.org/10.1016/j.neuropsychologia.2011.10.025>
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 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>
- Scharf, F., Widmann, A., Bonmassar, C., & Wetzels, N. (2022). A tutorial on the use of temporal principal component analysis in developmental ERP research—Opportunities and challenges. *Developmental Cognitive Neuroscience*, 54, 101072. <https://doi.org/10.1016/j.dcn.2022.101072>
- 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>
- Spratling, M. W. (2017). A review of predictive coding algorithms. *Brain and Cognition*, 112, 92–97. <https://doi.org/10.1016/j.bandc.2015.11.003>
- Storch, D., & Zimmermann, E. (2022). Temporal context modulates sensory attenuation magnitude. *Acta Psychologica*, 229, 103703. <https://doi.org/10.1016/j.actpsy.2022.103703>
- 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>
- 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>
- van Vugt, F. T., & Tillmann, B. (2014). Thresholds of auditory-motor coupling measured with a simple task in musicians and non-musicians: Was the sound simultaneous to the key press? *PLoS One*, 9(2), e87176.
- 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. <https://doi.org/10.1177/002383099704000103>
- Wickham, H. (2016). ggplot2: Elegant graphics for data analysis.
- 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](https://doi.org/10.1016/s0893-6080(96)00035-4)
- Wunderlich, J. L., & Cone-Wesson, B. K. (2001). Effects of stimulus frequency and complexity on the mismatch negativity and other components of the cortical auditory-evoked potential. *The Journal of the Acoustical Society of America*, 109(4), 1526–1537. <https://doi.org/10.1121/1.1349184>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Emmendorfer, A. K., Bonte, M., Jansma, B. M., & Kotz, S. A. (2023). Sensitivity to syllable stress regularities in externally but not self-triggered speech in Dutch. *European Journal of Neuroscience*, 58(1), 2297–2314. <https://doi.org/10.1111/ejn.16003>