



Predictions about prosody facilitate lexical access: Evidence from P50/N100 and MMN components

Hatice Zora^{a,*}, Janniek Wester^a, Valéria Csépe^b

^a Max Planck Institute for Psycholinguistics, P.O. Box 310 6500, AH, Nijmegen, the Netherlands

^b HUN-REN Research Centre of Natural Sciences, Brain Imaging Centre, P.O. Box 286 1519, Budapest, Hungary

ARTICLE INFO

Keywords:
Prosody
Predictive coding
Lexical access
P50/N100
MMN

ABSTRACT

Research into the neural foundation of perception asserts a model where top-down predictions modulate the bottom-up processing of sensory input. Despite becoming increasingly influential in cognitive neuroscience, the precise account of this predictive coding framework remains debated. In this study, we aim to contribute to this debate by investigating how predictions about prosody facilitate speech perception, and to shed light especially on lexical access influenced by simultaneous predictions in different domains, inter alia, prosodic and semantic. Using a passive auditory oddball paradigm, we examined neural responses to prosodic changes, leading to a semantic change as in Dutch nouns *canon* [ˈkaːnɔn] ‘canon’ vs *kanon* [kaːˈnɔn] ‘cannon’, and used acoustically identical pseudowords as controls. Results from twenty-eight native speakers of Dutch (age range 18–32 years) indicated an enhanced P50/N100 complex to prosodic change in pseudowords as well as an MMN response to both words and pseudowords. The enhanced P50/N100 response to pseudowords is claimed to indicate that all relevant auditory information is still processed by the brain, whereas the reduced response to words might reflect the suppression of information that has already been encoded. The MMN response to pseudowords and words, on the other hand, is best justified by the unification of previously established prosodic representations with sensory and semantic input respectively. This pattern of results is in line with the predictive coding framework acting on multiple levels and is of crucial importance to indicate that predictions about linguistic prosodic information are utilized by the brain as early as 50 ms.

1. Introduction

Dating from Helmholtz (1867), perception has been argued to operate in an inferential manner, rather than emerging from a purely sensory operation. Consolidated with Bayes theorem (Kersten et al., 2004; Knill and Pouget, 2004), the *predictive coding* (PC) theory (Friston, 2005, 2009, 2010) has been momentarily influential in cognitive neuroscience, integrating action, perception, attention, and learning (Winkler and Schröger, 2015), and hence seen as a unified theory of cortical function (Friston, 2010; Heilbron and Chait, 2018; Millidge et al., 2021). The PC theory postulates that the brain relies on a generative model, combining top-down predictions with bottom-up sensory input (de Lange et al., 2018; Heilbron and Chait, 2018; Mumford, 1992; Winkler and Schröger, 2015). At the neuronal level, the generative model pre-activates the cortical representation of a predicted stimulus, and this pre-activated representation is compared with the sensory input. A match between the sensory input and representation-based

prediction induces a suppression in the neural response, a phenomenon called *expectation suppression* (Bell, 2001; Garrido et al., 2018; Summerfield et al., 2008; Todorovic and de Lange, 2012; Wacongne et al., 2011); whereas a mismatch results in a *prediction error* signal (Friston, 2005; Summerfield and de Lange, 2014). Evidence for predictive mechanisms has been found at several levels of the linguistic hierarchy, including phonological (DeLong et al., 2005), morphosyntactic (van Berkum et al., 2005; Wicha et al., 2004; Wicha et al., 2003), lexical-semantic/discourse (Hasson et al., 2006; Lau et al., 2016; Orfanidou et al., 2006; Otten and Van Berkum, 2008; Poppenk et al., 2016), and syntactic contexts (Arai and Keller, 2013; Bornkessel-Schlesewsky and Schlewsky, 2013; Kuperberg and Jaeger, 2016; Kutas et al., 2011; Matchin et al., 2017; Rohde et al., 2011; Weber et al., 2016). Despite the growing number of studies, the question of how the PC principles apply to language in general and speech processing especially remains open. The present paper aims to contribute to this debate by investigating whether predictions about linguistic prosodic information facilitate

* Corresponding author.

E-mail address: Hatice.Zora@mpi.nl (H. Zora).

<https://doi.org/10.1016/j.ijpsycho.2023.112262>

Received 17 June 2023; Received in revised form 23 August 2023; Accepted 28 October 2023

Available online 2 November 2023

0167-8760/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

lexical access, involving simultaneous predictions at the sensory and cognitive domains.

Prosody is associated with physical properties of the auditory signal such as fundamental frequency (f_0), intensity and duration (Bolinger, 1961; Fry, 1958; Lehiste, 1970), and is a sine qua non for an efficient communication both at the lexical and the phrase/sentence levels. Variations in prosodic features support diverse communicative functions ranging from distinguishing lexical meaning and encoding syntactic structure to managing discourse (for reviews see, Arvaniti, 2020; Cutler, 2005; Cutler et al., 1997; Gussenhoven and Chen, 2020; Wagner and Watson, 2010). This paper concerns lexical level prosody, and specifically deals with *lexical stress*, a prosodic phenomenon that specifies which syllable in the word is more prominent than any of the others (for reviews see Cutler, 2005; Cutler, 2015; Cutler and Jesse, 2021; Zora, 2016). Stressed syllables typically have longer duration, greater intensity/spectral emphasis, and/or higher f_0 than unstressed syllables. Besides encoding perceptual salience at the sensory level, lexical stress contributes towards semantic content at the cognitive level. Different placement of stress for instance occasionally creates lexically distinct minimal pairs as in English stress-alternating homographs 'forebear – fore'bear (Cutler, 2015; Cutler and Jesse, 2021), and constrains lexical processing (Cutler and van Donselaar, 2001; van Donselaar et al., 2005). Languages however differ in the realization and function of lexical stress, and above all might favor one of the stress patterns over the other (for an overview, see Zora, 2016). Depending on prominence relations between syllables, two stress patterns emerge: trochaic (strong-weak) and iambic (weak-strong). Sensitivity to specific stress patterns has previously been documented in both infants (Friederici et al., 2007; Weber et al., 2004) and adults (Honbolygó et al., 2004, 2020; Ylinen et al., 2009) using the *mismatch negativity* (MMN) component of event-related potentials (ERPs). Besides, relative importance of cues to stress perception and their contribution towards lexical access were substantiated for various languages in our previous MMN research (Zora et al., 2015; Zora et al., 2016a).

The MMN component is an ideal tool to investigate the predictive processes involved in lexical stress processing as it not only signals the brain's automatic response to prosodic changes in the auditory sensory input (e.g., changes in f_0 , intensity and duration) (Näätänen et al., 1978; Näätänen et al., 1989) but also has often been referred to as a neurophysiological signature of prediction error (Friston, 2005; Näätänen et al., 2007; Winkler et al., 1996; Winkler, 2007). The MMN is typically elicited by infrequent deviant stimuli interspersed among frequent standard stimuli in an oddball paradigm. The brain establishes regularities based on standards and generates expectations of the upcoming stimuli; a deviant stimulus, mismatching these expectations, elicits an MMN response (Garrido et al., 2009). Since its discovery (Näätänen et al., 1978), several mechanisms have been postulated to explain the MMN component (for a review, see Garrido et al., 2009; Winkler and Schröger, 2015), centered around two major hypotheses: the *adaptation hypothesis* (Jääskeläinen et al., 2004; see also Larsson and Smith, 2012; Matsuzaki et al., 2012) and the *model adjustment hypothesis* (Winkler et al., 1996; Näätänen and Winkler, 1999). The adaptation hypothesis interprets MMN as a product of N100 adaptation in the auditory cortices, reflecting simple bottom-up processing. The N100 amplitude is reduced for the repetitive standard stimuli due to *repetition suppression* and *lateral inhibition*, while the deviant stimulus elicits enhanced N100, reflecting the activation of non-adapted N100 neurons (Butler, 1968; May and Tiitinen, 2004). By contrast, the model-adjustments hypothesis postulates that the MMN response reflects on-line updating of a perceptual model as a result of a discrepancy between the predicted and the actual sensory input (Winkler et al., 1996; Näätänen and Winkler, 1999), and long-term experience about the sensory input have modulating effects on the model (Winkler and Schröger, 2015). Given being larger for familiar items such as speech sounds (Dehaene-Lambertz, 1997; Näätänen et al., 1997) and words (Pulvermüller et al., 2001; Shtyrov and Pulvermüller, 2002), the MMN component proves to reflect

higher cognitive processes such as the activation of long-term memory traces associated with familiar items, going beyond sensory processing. The *neurocomputational* model by Garagnani and Pulvermüller (2011) indeed asserts that the neuronal adaptation and local inhibition can explain the MMN response to simple acoustic changes in non-speech items but not in meaningful speech. The MMN response to meaningful familiar items can rather be accounted for because of distributed neuronal memory circuits. Following the Hebb's (1949) postulate, this model posits that co-activated neurons develop into memory circuits with feedback and feedforward connections, and it is the activation of these strongly connected cell assemblies that accounts for the MMN increase for learned cognitive representations (Näätänen, 2001; Pulvermüller and Shtyrov, 2006). All these models demonstrate different perspectives for the role of prediction in speech perception and give rise to a particular question about the MMN response in the PC context (Garrido et al., 2008; Garrido et al., 2009; Grisoni et al., 2019). Drawing inspiration from the PC framework, in this study we investigate how lexical stress information is analyzed by different predictive mechanisms – be it adaptation or long-term memory activation – of the brain before giving rise to lexical access as auditory processing unfolds over time.

Despite previous MMN research indicating the contribution of lexical stress towards lexical access (Zora et al., 2015; Zora et al., 2016a), further research is warranted to better understand how the MMN response translates into predictive processes involved not only in lexical stress processing but also in lexical access. Moreover, a direct investigation of the interplay between the MMN component and other ERP components – the P50 and the N100 – that are indicative of predictive processes (Schwartz et al., 2013) is missing. The P50 is considered as obligatory ERP component, being primarily influenced by the physical features of stimuli (Picton et al., 1974), and found to be larger to infrequent deviant stimuli in oddball paradigms (Boutros et al., 1995; Boutros and Belger, 1999) like the MMN component. It has also been claimed to index inhibition of irrelevant stimuli, a phenomenon called *repetition suppression* (Boutros et al., 2011) or *sensory gating* (Boutros and Belger, 1999), *habituation* (Rosburg, 2004), *neural adaptation* (Larsson and Smith, 2012; Matsuzaki et al., 2012). All these phenomena are characterized by an attenuation in the P50 response to repeated stimuli, typically in a paired-stimulus paradigm, and the underlying process is to filter out the redundant or irrelevant auditory information in order not to overload the higher-order stages of information processing (Korzyukov et al., 2007). As referred to above, the N100 component is another type of repetition-induced neural decrement (Fruhstorfer et al., 1970; Woods and Elmasian, 1986). Besides pure refractory account (Näätänen and Winkler, 1999), there are findings indicating some top-down modulation in the repetition-attenuation (Font-Alaminos et al., 2020; Herrmann et al., 2018; Öhman and Lader, 1972; Todorovic et al., 2011; Woods and Elmasian, 1986). For instance, the N100 amplitude was shown to be more attenuated to speech sounds than to non-speech sounds (Woods and Elmasian, 1986) and to expected tones than to unexpected tones (Todorovic et al., 2011). The N100 component has also been reported to reflect violation of rhythmical expectations (the alternation of stressed and unstressed syllables) (Cason and Schön, 2012; Zhang and Zhang, 2019). Cason and Schön (2012) observed a large N100 response in response to target words with mismatching stress patterns with primes, showing the detection of auditory events that do not match the expectations.

By examining predictive processes as reflected by the P50/N100 complex as well as the MMN component, the present study endeavors to provide a more comprehensive understanding of the cortical processing stages of lexical stress and lexical access. Given that segmental information might outweigh prosodic information in lexical processing (Bond and Small, 1983; Cooper et al., 2002; Cutler, 1986), the experimental material consist of words that are segmentally identical but differ in prosody, where the location of stress on the first or second syllable leads to different lexical semantics. To establish whether the presence of lexical semantics modulates the amplitude of the P50/N100 complex

and the MMN component, it is crucial to distinguish between lexical processing and general auditory processing. Words are therefore compared to physically matched pseudowords. Although language-specific prosodic rules may also be applied to pseudowords (Honbolygó et al., 2004; Honbolygó et al., 2020), semantic processing can best be investigated using real words (Zora et al., 2015; Zora et al., 2016a). We expect to see ERP modulations associated with the P50/N100 complex and the MMN component to prosodic changes in both words and pseudowords, however with different morphology and amplitude, indexing the role of prosodic and semantic familiarity in this process and different underlying mechanisms. Given the presence of an association between prosody and semantics, a more attenuated P50/N100 response is hypothesized to the prosodic changes in words than in pseudowords that lack such an association, indicating that changes in lexical stress pattern are more expected in words than in pseudowords at the early stages of auditory processing. Predictions might however also be modulated by abstract prosodic regularities, encoded not only at the word but also pseudoword levels, and accordingly, at a later stage, the MMN elicitation is expected to indicate the impact of long-term memory representations of prosodic information in the processing of meaningful and meaningless items.

2. Materials and methods

2.1. Participants

Thirty Dutch speakers (15 male, 15 female; age range 18–32 years, $M = 23.7$, $SD = 3.33$) participated in the experiment. The participants were recruited from the subject pool of the Max Planck Institute for Psycholinguistics in Nijmegen and remunerated for their participation. All participants were right-handed, and reported normal vision, hearing and language development. Datasets from two participants were excluded due to the noisy reference channel. Therefore, data from twenty-eight participants (age range 18–32 years, $M = 23.8$, $SD = 3.30$) were included in the analysis.

2.2. Ethics statement

The study conformed to the ethical guidelines on human subject research, and the ethical approval for the experimental protocol had been given by the Ethics Board of the Social Sciences Faculty of Radboud University (ECSW-2020-049). The participants were informed about the procedure both orally and in writing and gave written consent before the data collection.

2.3. Stimuli recording and manipulation

The experimental material was a monomorphemic disyllabic Dutch minimal pair, *canon* ['ka:nɔn] vs *kanon* [ka:'nɔn], where the lexical stress pattern (trochaic vs iambic respectively) encoded different linguistic meaning, 'canon' and 'cannon' respectively. To isolate the activity associated with the ERP components of interest, acoustically matched pseudowords ['ta:nɔn]* vs [ta:'nɔn]* were used as control. Recordings were conducted in a room equipped with a sound proof-booth using the Audacity software (version 2.4.2, 1999–2021 Audacity Team) at a rate of 44.1 kHz with 16 bits/per sample. Each stimulus was repeatedly produced in a semantically neutral frame sentence by a female native speaker of Dutch (from Overijssel province, 39 years old, with neuro- and psycho-linguistics background). The speaker was seated comfortably, with her mouth approximately 5 cm from the microphone (Sennheiser ME64) and was instructed to produce each sentence with a neutral prosody and at a natural pace. Out of 120 instances (60 for words and 60 for pseudowords) acoustically analyzed, the best exemplars of each stimulus type were selected, extracted from the frame sentence, and manipulated in Praat (version 6.2.04) (Boersma and Weenink, 2014).

Initial auditory and visual analysis of the stimuli entailed some adjustments to pitch and intensity parameters. To eliminate sharp jumps in the pitch contour, pitch range of the speaker, i.e., the minimum and maximum f_0 values, was adjusted to 100–600 Hz, recommended setting for a female voice. Using scale intensity command, the amplitude of each stimulus was multiplied in such a way that their average (i.e., root-mean-square) intensity became 70 dB SPL. To eliminate spurious clicks, 5 ms ramps were added to at the beginning and the end of the stimuli. Syllables were extracted from each stimulus type using the periodicity in the waveforms and wide-band spectrograms. Each syllable was cut at the nearest zero crossing (onset/offset) and saved separately.

To eliminate acoustic differences beyond interest and relevance, and to keep the deviant and standard identical to a great extent, the stimulus with iambic pattern served as a base for the rest of the experimental stimuli. The 1st syllable of the stimulus with iambic pattern (170 ms) was kept constant, and the 2nd syllable was manipulated to create the stimulus with trochaic pattern. To this end, the pitch and intensity parameters of the 2nd syllable was lowered in accordance with the mean pitch (~70 Hz) and mean energy difference (~5 dB) between the 1st and 2nd syllables of the trochaic pattern. Since the durational changes were in the same direction in both stimuli, no manipulation was carried out in terms of duration. The length of each stimulus was 565 ms; the divergence point was therefore at 170 ms between the iambic and trochaic patterns.

To keep the difference minimal across the blocks, pseudowords were created out of the real words by simply replacing the initial segment /k/ with /t/, which is identical to the word equivalent in manner of articulation. To preserve the natural flow of the waveform, the critical segments were extracted from the relevant context and spliced at zero-crossings. The words and pseudowords differed from each other at 55 ms. The acoustic quality of the stimuli was validated by independent judgment of three listeners. See Fig. 1 for the waveforms as well as pitch and intensity contours across all the stimuli.

2.4. Experimental design

A passive auditory oddball paradigm, where frequent ($p = 8/10$) standard stimuli were randomly replaced by rare ($p = 2/10$) deviant stimuli, served as experimental paradigm (see Fig. 2). The experiment was built on two blocks, words and pseudowords, with stimuli of iambic pattern used as standards, and stimuli of trochaic pattern as deviants. Each block consisted of 800 stimuli – 640 standards and 160 deviants. The deviants were semi-randomly placed among the standards in a way that at least two intervening standards were presented between two consecutive deviants. To avoid the rhythmicity effect, five different interstimulus intervals (ISI) – 400 ms, 450 ms, 500 ms, 550 ms, 600 ms – were used. With these ISIs, the shortest stimulus onset asynchrony (SOA) was 965 ms and the longest 1165 ms, given that the stimulus length was 565 ms.

The experiment took place in an electrically shielded and sound-attenuated recording booth. The experiment was programmed in Presentation software system (version 22.1), and the auditory stimuli were delivered via loudspeakers at a comfortable listening level (set at -17 dB; Yamaha, HTR-4072). The task of the participants was to focus on a silent ocean documentary (without subtitles), and disregard the auditory stimuli. The documentary covered only a quarter of the computer screen to minimize the saccadic eye movements. To ensure that the participants' attention was exclusively on the documentary, a short questionnaire was administered at the end of the session. Each block took ~14 min, and if requested, a short break was given to the participants between the blocks. The order of the blocks was counterbalanced, and the experiment, including breaks and electrode application, lasted ~1.5 h.

2.5. Electroencephalography recordings and data analysis

The electroencephalography (EEG) recordings were performed by

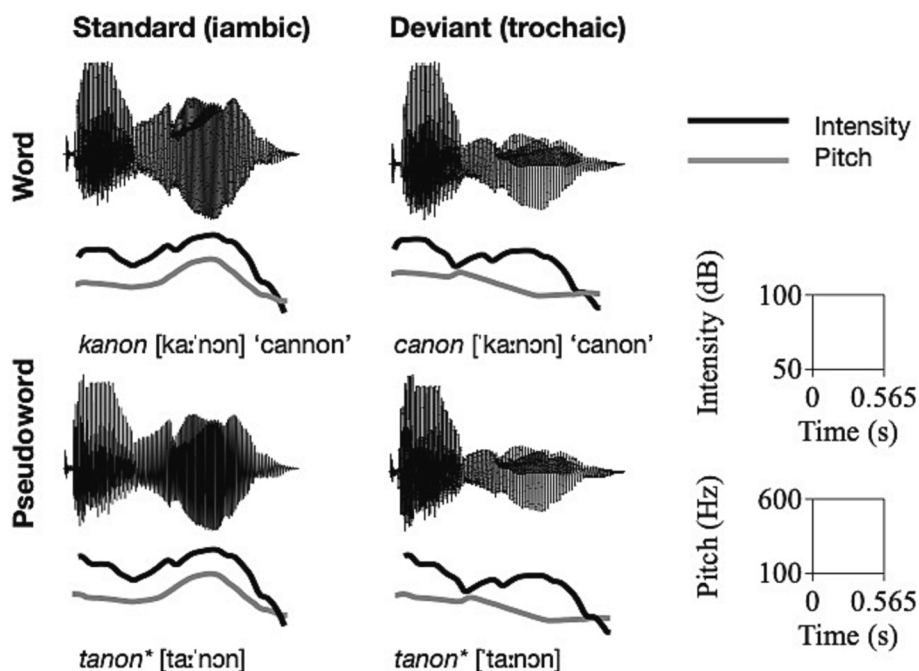


Fig. 1. Waveforms, and pitch and intensity contours of all stimuli in the word and the pseudoword blocks. Phonetic transcriptions as well as English translations are given below. Grey solid line, pitch; Black solid line, intensity. Duration of stimuli is given in seconds (s) and frequency in hertz (Hz) and intensity in decibels (dB).

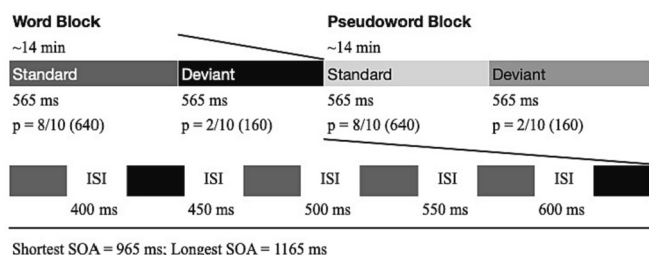


Fig. 2. Illustration of the oddball paradigm across the word and the pseudoword blocks. ISI: Interstimulus interval; SOA: Stimulus onset asynchrony; Total number of standards per block: 640; Total number of deviants per block: 160. Time is given in milliseconds (ms).

using the standard 10–20 ActiCAP montage for 32 channels and the BrainVision Recorder software with a BrainAmps DC amplifier (Brain Products, Munich, Germany). Eye movements were measured both horizontally and vertically using the electrooculography electrodes. The EEG signal was referenced online to the left mastoid and recorded at a sampling rate of 500 Hz and band-pass filtered at 0.016–150 Hz (10 s time constant). The data were preprocessed using the BrainVision Analyzer (version 2.2.0; Brain Products, Munich, Germany). The continuous EEG data were filtered with a zero-phase Butterworth filter (bandpass 0.5–30 Hz) as well as with a notch filter (50 Hz) to remove line noise. The channels were then re-referenced to the average of the left and right mastoid, and the EEG data were epoched from –100 to 900 ms, relative to the word onset. An independent component analysis (Jung et al., 2000) was performed for artifact identification and rejection (Infomax Restricted ICA, 512 steps). When applicable, noisy EEG channels were interpolated through spherical splines to avoid unnecessary data loss. A 100 ms pre-onset interval was used for baseline correction and activation exceeding $\pm 100 \mu\text{V}$ at any epochs was automatically removed. ERP quantification was computed as a mean voltage within a 50-ms-window centered at peaks in the grand-average waveforms. Time windows were defined to optimally capture ERP modulations related to prosodic changes, and accordingly three consecutive time windows

were chosen: 185–235 ms (15–65 ms after deviance onset at 170 ms), 300–350 ms (130–180 ms after deviance onset), and 375–425 ms (205–255 ms after deviance onset). To plot the ERP waveforms, grand averages were obtained for standards and deviants in each block.

2.6. Statistical analysis

The statistical analysis was performed in SPSS (version 20.0.0.1) (International Business Machines Corp., Armonk, New York, United States). The region of interest (ROI) covered frontal (F3, Fz, F4), frontocentral (FC1, FCz, FC2) and central (C3, Cz, C4) electrodes, conforming to the typical frontocentral MMN topography. The average number of included trials (epochs) was 634.8 (out of 640; min = 620, max = 640) for Word Standard stimuli; 158.6 (out of 160; min = 152, max = 160) for Word Deviant stimuli; 634.4 (out of 640; min = 583, max = 640) for Pseudoword Standard stimuli; and 158.7 (out of 160; min = 141, max = 160) for Pseudoword Deviant Stimuli. To examine whether the ERP responses significantly differed from zero, deviant-minus-standard difference amplitudes were tested against zero with one-sample *t*-tests. Deviant-minus-standard difference amplitudes were also used to carry out post hoc paired samples *t*-tests when needed. A two-way repeated-measures ANOVA with factors of *Block* (Word and Pseudoword) and *Stimuli* (Standard and Deviant) was then performed in the three time windows. The significance level was set at $p < 0.05$. For significant interactions, follow-up ANOVAs were performed and effect sizes are reported with η^2 (partial η^2).

3. Results

Fig. 3 illustrates the grand average ERP waveforms elicited by the standards and deviants at the Fz, FCz and Cz electrode sites across the Word and Pseudoword blocks. Topographic maps are displayed for deviants in the Word block to provide a rough estimate of spatial distribution in each time window. Visual analysis of the grand averages indicates that the earliest ERP response is a positive-going wave peaking almost immediately after the deviance onset (at 170 ms), and seems to be larger to pseudowords than words. Given the latency and

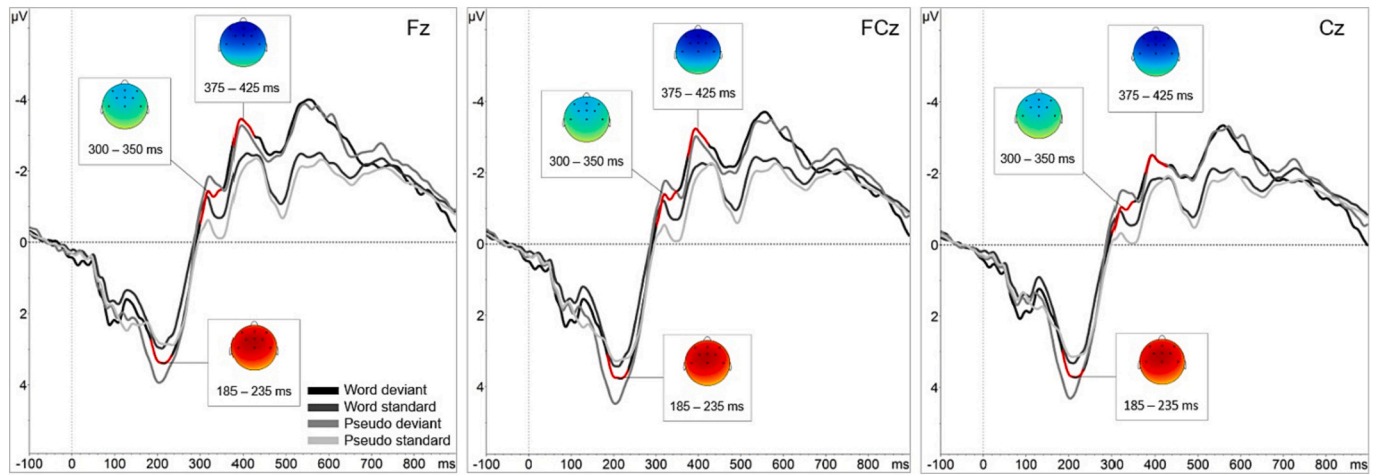


Fig. 3. Grand average ERP waveforms elicited by standards and deviants in Fz, FCz and Cz across the word and the pseudoword blocks. Amplitude is given in microvolts [μV , (-5, 5)] and time in milliseconds [ms, (-100, 900)]. Black line: Word deviant; Light black line: Word standard; Grey line: Pseudoword deviant; Light grey line: Pseudoword standard; Red highlight: Time windows selected for the statistical analysis. [Color in print]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

morphology, this response is considered to a P50 response, and followed by two negative waveforms. These negative waves can be classified as N100 and MMN responses respectively – N100 with somewhat larger amplitude to pseudowords, and MMN being slightly enhanced for words.

An overview of the statistical results is presented in Table 1. Results of the one-sample *t*-tests indicated significant difference from zero for the ERPs in both the first time window [words ($t(27) = 2.998, p = 0.006$) and pseudowords ($t(27) = 6.634, p = 0.000$)], and the third time window [words ($t(27) = -3.980, p = 0.000$) and pseudowords ($t(27) = -4.247, p = 0.000$)]. In the second time window, the ERP responses differed from zero for pseudowords ($t(27) = -7.045, p = 0.000$) but not

Table 1

The results of one-sample *t*-test: deviant-minus-standard difference amplitudes were tested against zero in three time windows; The results of two-way repeated-measures ANOVA: Two way interaction between *Block* (Word and Pseudoword) and *Stimuli* (Standard and Deviant) was analyzed in three time windows; The results of follow-up ANOVAs: The significant interactions between *Block* and *Stimuli* in the first and second time windows were further analyzed.

| | First time window | Second time window | Third time window |
|--------------------------|--|--|--|
| One sample t-test | | | |
| Words | $t(27) = 2.998, p = 0.006$ | $t(27) = -1.878, p = 0.071$ | $t(27) = -3.980, p = 0.000$ |
| Pseudowords | $t(27) = 6.634, p = 0.000$ | $t(27) = -7.045, p = 0.000$ | $t(27) = -4.247, p = 0.000$ |
| Two-way ANOVA | | | |
| Block | $F(1, 27) = 1.281, p = 0.268, \eta^2 = 0.045$ | $F(1, 27) = 0.262, p = 0.613, \eta^2 = 0.010$ | $F(1, 27) = 0.687, p = 0.414, \eta^2 = 0.025$ |
| Stimuli | $F(1, 27) = 51.935, p = 0.000, \eta^2 = 0.658$ | $F(1, 27) = 34.785, p = 0.000, \eta^2 = 0.563$ | $F(1, 27) = 26.507, p = 0.000, \eta^2 = 0.495$ |
| Block X Stimuli | $F(1, 27) = 7.176, p = 0.012, \eta^2 = 0.210$ | $F(1, 27) = 24.864, p = 0.000, \eta^2 = 0.479$ | $F(1, 27) = 0.398, p = 0.533, \eta^2 = 0.015$ |
| Follow-up ANOVA | | | |
| Words | $F(1, 27) = 8.987, p = 0.006, \eta^2 = 0.250$ | $F(1, 27) = 3.525, p = 0.071, \eta^2 = 0.115$ | |
| Pseudowords | $F(1, 27) = 44.015, p = 0.000, \eta^2 = 0.620$ | $F(1, 27) = 49.629, p = 0.000, \eta^2 = 0.648$ | |

for words ($t(27) = -1.878, p = 0.071$). Results of ANOVA showed significant two-way interactions of Block with Stimuli in the first time window ($F(1, 27) = 7.176, p = 0.012, \eta^2 = 0.210$) and in the second time window ($F(1, 27) = 24.864, p = 0.000, \eta^2 = 0.476$). In the third time window only Stimuli main effect was found ($F(1, 27) = 26.507, p = 0.000, \eta^2 = 0.495$). Follow-up ANOVA in the first time window indicated significant differences between standards and deviants in words ($F(1, 27) = 8.987, p = 0.006, \eta^2 = 0.250$) and pseudowords ($F(1, 27) = 44.015, p = 0.000, \eta^2 = 0.620$). Paired samples *t*-test, using the deviant-minus-standard data, revealed a significant difference ($t(27) = -2.679, p = 0.012$) between words ($M = 0.400, SD = 0.706$) and pseudowords ($M = 0.945, SD = 0.753$). This enhanced positivity for pseudowords corresponds to a P50 response, indicating that the prosodic change was more unexpected in pseudowords than in words. Follow-up ANOVA in the second time window indicated significant differences between standard ($M = -0.105, SD = 1.611$) and deviants ($M = -1.262, SD = 1.887$) in pseudowords ($F(1, 27) = 49.629, p = 0.000, \eta^2 = 0.648$). There was no statistical significance between standard ($M = -0.682, SD = 1.903$) and deviant ($M = -0.934, SD = 2.06$) in words ($F(1, 27) = 3.525, p = 0.071, \eta^2 = 0.115$) in this time window. Additionally, paired samples *t*-test, using the deviant-minus-standard data, indicated a significant difference ($t(27) = 4.986, p = 0.000$) between words ($M = -0.251, SD = 0.710$) and pseudowords ($M = -1.156, SD = 0.868$) in the second time window. This negative response is consistent with an N100 response, and since being reduced to words, it indexes the suppression of prosodic information that has already been encoded in words. Main effect in the third time window indicated that deviant ($M = -2.649$) elicited a larger negative response than standards ($M = -1.751$) regardless of block. This negative response present in both words and pseudowords is best justified by the MMN response, and demonstrates the activation of long-term memory traces associated with abstract prosodic regularities. Despite the fact that the word block seems to elicit a larger negativity ($M = -2.316$) irrespective of the stimuli in comparison to the pseudoword block ($M = -2.084$), this effect is not significant.

4. Discussion

Spoken words are neither simple perceptual entities, nor rigidly defined by the auditory sensory input. They rather reflect complex forms substantiated with the cognitive and neural representations of a particular language. This statement is compatible with the principal insight of predictive coding (PC): to derive an optimal interpretation of

incoming data, the brain unifies auditory sensory input with prior knowledge and experiences, representations in cognitive terms. In the present study, we investigated the inferential processes of lexical access and the impact of stimulus features on it. Moreover, we aimed at identifying those sensory and cognitive processes that persist over different time scales. As a working model we choose prosody, coding semantic information through acoustic modulations (Zora, 2016), and as such requiring inferences that are both signal and semantic-driven. Using an auditory passive oddball paradigm, we compared brain responses to lexical stress pattern changes and their consequences in the processing of words and pseudowords. The results indicated enhanced P50/N100 response to pseudowords and MMN responses to both words and pseudowords. In the following, we discuss these results within the PC framework.

As reviewed in the introduction, the P50/N100 complex has mostly been investigated using repetition suppressions paradigms, and argued to reflect adaptive neural response pattern in the auditory modality (Boutros et al., 2011). However, it has also been identified as a neural marker of deviance detection elicited in auditory oddball paradigms (Boutros et al., 1995; Boutros and Belger, 1999) in a similar fashion to the MMN component. Given recent ERP studies showing larger positive (Slabu et al., 2010) and negative (Grimm et al., 2011) responses to deviants than to standards within the first 50 ms after sound onset, preceding the well-studied MMN component, Grimm and Escera (2012) suggested the existence of a multi-level system for deviance detection. Similarly, Boutros and Belger (1999) argued that the repetition suppression acts on multiple levels, and the P50, N100 and MMN components are part of a complex multistage and multicomponent sensory gating system. As such, it is not trivial to postulate that the P50/N100 complex, going beyond a simple form of adaptation, not only detects stimulus change but also reflects the selection of relevant information for further processing, while simultaneously inhibiting irrelevant information (see Boutros and Belger, 1999). Furthermore, in contrast to a bottom-up flow of perceptual information, repetition suppression measures are found to be modulated by top-down perceptual expectations (Font-Alaminos et al., 2020; Herrmann et al., 2018; Ohman and Lader, 1972; Todorovic et al., 2011; Woods and Elmasian, 1986), and are claimed to index a reduction in prediction error that occurs when sensory input matches with a more probable, previously experienced, percept in comparison to a less probable one (Summerfield et al., 2008). The attenuated suppression to the stress pattern changes in the pseudowords in the present study indicates that they might be unexpected compared to the prosodic changes in the words. Given the previously established link between prosody and semantics, the reduced P50/N100 response reported to the words might reflect the suppression of information that has already been encoded and as such more probable, whereas the enhanced response to the pseudowords might indicate that all relevant auditory information still stay in competition since the brain is still trying to interpret the input.

A second larger negative deflection is present to both words and pseudowords. One might argue that the enhanced neural activation to meaningless pseudowords might be a product of N100 enhancement, reflecting the activation of non-adapted neurons and thus simple bottom-up processing. However, we argue that the negative response to pseudowords is best justified by the MMN response and by the PC framework (Friston, 2010; Millidge et al., 2021) arguing for modulating effects of our experience on sensory processing. In line with the previous accounts demonstrating the existence of an abstract representation of stress assignment (Honbolygó et al., 2004; Honbolygó and Csépe, 2013; Honbolygó et al., 2020), we believe that the MMN response to the pseudowords reflects predictions of prosodic regularities for speech-like stimuli. Dutch prefers trochaic stress pattern (Cutler, 2012; Rietveld et al., 2004; van Oostendorp, 2012), and a change from iambic to trochaic stress pattern might activate long-term memory traces associated with the typical stress pattern. Predictive processes for prosodic information in pseudowords have also been demonstrated through fMRI

research. Honbolygó et al. (2020), for instance, indicated that expectations about word stress in pseudowords modulate neural activity in the posterior and middle parts of the superior temporal gyrus (STG). The authors argued that this finding is in accordance with the assumption of a *dual auditory stream model* (Bornkessel-Schlesewsky and Schlesewsky, 2013), asserting a postero-dorsal stream for the predictive sequential processing of linguistic information and for the prosodic segmentation. The authors linked the lack of activation in areas involved in cognitive processes such as the inferior frontal gyri (IFG) to the use of meaningless pseudowords.

The present paper sheds light on the cognitive processes by looking into the use of real words in addition to the use of meaningless pseudowords. The negative response to words is claimed to be an MMN response indexing the activation of not only the typical stress pattern but also another word associated with the prosodic change (for the role of lexical stress in Dutch word recognition, see Cutler and van Donselaar, 2001; van Donselaar et al., 2005). In accordance with our previous research (Zora et al., 2015, 2016a, 2020), we claim that the MMN to words indicates the integration of their semantic representation with embodied prosodic information, and their predictions based on long-term memory processes. It should be noted that previous studies have documented differences in MMN modulations related to wordedness (Garagnani et al., 2009; Jacobsen et al., 2004; Korpilahti et al., 2001; Pulvermüller et al., 2001) and language specific prosodic representations (Zora et al., 2015, 2016a, 2016b, 2019, 2020). The fact that no MMN amplitude differences were found between the words and pseudowords in the present study does not eliminate the presence of differences as distinct cortical sources, temporal and frontal, have been suggested to underlie MMN elicitation to unfamiliar and familiar items respectively (Garagnani and Pulvermüller, 2011). We however refrain from drawing conclusions on the cortical sources that might contribute to the present ERP patterns as no source estimation was carried out in our study. Given that prediction might influence areas that play a role in integration of upcoming/perceived information (Brennan and Pylkkänen, 2012; Hagoort, 2013; Willems et al., 2016), we encourage further research to design proper paradigms for investigating the role of the IFG, claimed to have a special role in unification (Hagoort, 2013) and cognitive operations at the word level (Bornkessel-Schlesewsky and Schlesewsky, 2013; Honbolygó et al., 2020).

The present study provides sound evidence to claim that the ERP patterns to pseudowords may reflect both lower-level auditory processing and higher-level though limited linguistic processing in contrast to a more complex processing of words reflecting higher level lexical processing and predictions complex in nature. The documented neural responses fit with the theories arguing for modulating effects of our long-term experience on sensory and linguistic processing, be it the PC framework (Friston, 2010; Millidge et al., 2021), the model-adjustment hypothesis (Näätänen and Winkler, 1999; Winkler and Schröger, 2015), the neurocomputational model (Garagnani and Pulvermüller, 2011), or the dual auditory stream model (Bornkessel-Schlesewsky and Schlesewsky, 2013). The findings are thus of crucial importance not only for providing further support to studies suggesting that prediction can occur already at the word level (Dikker and Pylkkänen, 2013), but also for substantiating the impact of long-term memory representations of sensory input in the processing of meaningful and meaningless items.

5. Conclusions

To conclude, the present study indicates that depending on whether the auditory input maps onto existing lexical items, either linguistic features (and cognitive mechanisms) or acoustic parameters (and sensory mechanisms), are of critical relevance (see also Bornkessel-Schlesewsky and Schlesewsky, 2019). The brain does not simply respond to changes in the acoustic environment, such as fundamental frequency and intensity deviations, but also evaluates them based on previously established representations to a large extent. This transitional process

can successfully be verified by the P50/N100 and MMN components as shown and documented by the present study. It seems that a common view of cognitive neuroscience on processing words postulating that hundreds of milliseconds after the sensory input are needed to interpret meaningful linguistic information should be revisited. Our results show that linguistic prosodic information can be weighted by importance as early as 50 ms (see also Shtyrov and Lenzen, 2017). In sum, in agreement with previous indications, the present study demonstrates that the neural circuits rely on present and past predictions during lexical access and a combination of predictive and integrative functions is required for a meaningful representation of speech.

Declaration of competing interest

None.

Acknowledgments

Part of this work was presented at 14th Annual Meeting of the Society for the Neurobiology of Language, 6–8 October 2022 Philadelphia, USA.

References

- Arai, M., Keller, F., 2013. The use of verb-specific information for prediction in sentence processing. *Lang. Cogn. Process.* 28 (4), 525–560.
- Arvaniti, A., 2020. The phonetics of prosody. In: Aronoff, M., et al. (Eds.), *Oxford Research Encyclopedia of Linguistics*. Oxford University Press, Oxford.
- Bell, C.C., 2001. Memory-based expectations in electrosensory systems. *Curr. Opin. Neurobiol.* 11, 481–487.
- Boersma, P., Weenink, D., 2014. *Doing Phonetics by Computer*. Retrieved from. <http://www.praat.org/>. Version 5.3.24.
- Bolinger, D., 1961. Contrastive accent and contrastive stress. *Language* 37, 83–96.
- Bond, Z.S., Small, L.H., 1983. Voicing, vowel, and stress mispronunciations in continuous speech. *Percept. Psychophys.* 34, 470–474.
- Bornkessel-Schlesewsky, I., Schlesewsky, M., 2013. Reconciling time, space and function: a new dorsal–ventral stream model of sentence comprehension. *Brain Lang.* 125 (1), 60–76.
- Bornkessel-Schlesewsky, I., Schlesewsky, M., 2019. Toward a neurobiologically plausible model of language-related, negative event-related potentials. *Front. Psychol.* 10, 298.
- Boutros, N.N., Belger, A., 1999. Midlatency evoked potentials attenuation and augmentation reflect different aspects of sensory gating. *Biol. Psychiatry* 45, 917–922.
- Boutros, N.N., Torello, M.W., Barker, B.A., Tueting, P.A., Wu, S.-C., Nasrallah, H.A., 1995. The P50 evoked potential component and mismatch detection in normal volunteers: implications for the study of sensory gating. *Psychiatry Res.* 57, 83–88.
- Boutros, N.N., Gjini, K., Urbach, H., Pflieger, M.E., 2011. Mapping repetition suppression of the N100 evoked response to the human cerebral cortex. *Biol. Psychiatry* 69, 883–889.
- Brennan, J., Pyllkänen, L., 2012. The time-course and spatial distribution of brain activity associated with sentence processing. *NeuroImage* 60, 1139–1148.
- Butler, R.A., 1968. Effect of changes in stimulus frequency and intensity on habituation of the human vertex potential. *J. Acoust. Soc. Am.* 44, 945–950.
- Cason, N., Schön, D., 2012. Rhythmic priming enhances the phonological processing of speech. *Neuropsychologia* 50, 2652–2658.
- Cooper, N., Cutler, A., Wales, R., 2002. Constraints of lexical stress on lexical access in English: evidence from native and non-native listeners. *Lang. Speech* 45, 207–228.
- Cutler, A., 1986. Forbear is a homophone: lexical prosody does not constrain lexical access. *Lang. Speech* 29 (3), 201–220.
- Cutler, A., 2005. *The Handbook of Speech Perception*. Blackwell, Malden, MA.
- Cutler, A., 2012. *Native Listening: Language Experience and the Recognition of Spoken Words*. MIT Press, Cambridge, MA.
- Cutler, A., 2015. Lexical stress in English pronunciation. In: Reed, M., Levis, J.M. (Eds.), *The Handbook of English Pronunciation*. John Wiley & Sons, New York, NY, pp. 106–124.
- Cutler, A., Jesse, A., 2021. Word stress in speech perception. In: Pardo, J.S., Nygaard, L. C., Remez, R.E., Pisoni, D.B. (Eds.), *The Handbook of Speech Perception*. John Wiley & Sons, New York, NY.
- Cutler, A., Dahan, D., van Donselaar, W., 1997. Prosody in the comprehension of spoken language: a literature review. *Lang. Speech* 40, 141–201.
- Cutler, A., van Donselaar, W., 2001. *Voornaam* is not a homophone: lexical prosody and lexical access in Dutch. *Lang. Speech* 44, 171–195.
- de Lange, F.P., Heilbron, M., Kok, P., 2018. How do expectations shape perception? *Trends Cogn. Sci.* 22, 764–779.
- Dehaene-Lambertz, G., 1997. Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport* 8, 919–924.
- DeLong, K., Urbach, T., Kutas, M., 2005. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat. Neurosci.* 8, 1117–1121.
- Dikker, S., Pyllkänen, L., 2013. Predicting language: MEG evidence for lexical preactivation. *Brain Lang.* 127, 55–64.
- Font-Alaminos, M., Cornella, M., Costa-Faidella, J., Hervás, A., Leung, S., Rueda, I., Escera, C., 2020. Increased subcortical neural responses to repeating auditory stimulation in children with autism spectrum disorder. *Biol. Psychol.* 149, 107807.
- Friederici, A.D., Friedrich, M., Christophe, A., 2007. Brain responses in 4-month-old infants are already language specific. *Curr. Biol.* 17, 1208–1211.
- Friston, K.J., 2005. A theory of cortical responses. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 360 (1456), 815–836.
- Friston, K., 2009. The free-energy principle: a rough guide to the brain? *Trends Cogn. Sci.* 13, 293–301.
- Friston, K.J., 2010. The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11 (2), 127–138.
- Fruhstorfer, H., Soveri, P., Jarvilehto, T., 1970. Short-term habituation of the auditory evoked response in man. *Electroencephalogr. Clin. Neurophysiol.* 28, 153–161.
- Fry, D.B., 1958. Experiments in the perception of stress. *Lang. Speech* 1, 126–152.
- Garagnani, M., Pulvermüller, F., 2011. From sounds to words: A neurocomputational model of adaptation, inhibition and memory processes in auditory change detection. *Neuroimage* 54, 170–181.
- Garagnani, M., Shtyrov, Y., Pulvermüller, F., 2009. Effects of attention on what is known and what is not: MEG evidence for functionally discrete memory circuits. *Front. Hum. Neurosci.* 3, 10.
- Garrido, M.I., Friston, K.J., Kiebel, S.J., Stephan, K.E., Baldeweg, T., Kilner, J.M., 2008. The functional anatomy of the MMN: a DCM study of the roving paradigm. *Neuroimage* 42, 936–944.
- Garrido, M.I., Kilner, J.M., Stephan, K.E., Friston, K.J., 2009. The mismatch negativity: a review of underlying mechanisms. *Clin. Neurophysiol.* 120 (3), 453–463.
- Garrido, M.I., Rowe, E.G., Halász, V., Mattingley, J.B., 2018. Bayesian mapping reveals that attention boosts neural responses to predicted and unpredicted stimuli. *Cereb. Cortex* 28 (5), 1771–1782.
- Grimm, S., Escera, C., 2012. Auditory deviance detection revisited: evidence for a hierarchical novelty system. *Int. J. Psychophysiol.* 85, 88–92.
- Grimm, S., Escera, C., Slabu, L., Costa-Faidella, J., 2011. Electrophysiological evidence for the hierarchical organization of auditory change detection in the human brain. *Psychophysiology* 48, 377–384.
- Grisoni, L., Mohr, B., Pulvermüller, F., 2019. Prediction mechanisms in motor and auditory areas and their role in sound perception and language understanding. *NeuroImage* 199, 206–216.
- Gussenhoven, C., Chen, A., 2020. *The Oxford Handbook of Language Prosody*. Oxford University Press, Oxford.
- Hagoort, P., 2013. MUC (memory, unification, control) and beyond. *Front. Psychol.* 4, 416.
- Hasson, U., Nusbaum, H.C., Small, S.L., 2006. Repetition suppression for spoken sentences and the effect of task demands. *J. Cogn. Neurosci.* 18, 2013–2029.
- Hebb, D.O., 1949. *The Organization of Behavior*. John Wiley, New York.
- Heilbron, M., Chait, M., 2018. Great expectations: is there evidence for predictive coding in auditory cortex? *Neuroscience* 389, 54–73.
- Helmholtz, H., 1867. *Handbuch der physiologischen Optik*. Voss, Leipzig.
- Herrmann, B., Maess, B., Johnsrude, I.S., 2018. Aging affects adaptation to sound-level statistics in human auditory cortex. *J. Neurosci.* 38 (8), 1989–1999.
- Honbolygó, F., Csépe, V., 2013. Saliency or template? ERP evidence for long-term representation of word stress. *Int. J. Psychophysiol.* 87, 165–172.
- Honbolygó, F., Csépe, V., Ragó, A., 2004. Suprasegmental speech cues are automatically processed by the human brain: a mismatch negativity study. *Neurosci. Lett.* 363, 84–88.
- Honbolygó, F., Kőbor, A., Hermann, P., Kettinger, Á.O., Vidnyánszky, Z., Kovács, G., Csépe, V., 2020. Expectations about word stress modulate neural activity in speech-sensitive cortical areas. *Neuropsychologia* 143, 107467.
- Jääskeläinen, I.P., Ahveninen, J., Bonmassar, G., Dale, A.M., Ilmoniemi, R.J., Levänen, S., Lin, F.H., May, P., Melcher, J., Stufflebeam, S., Tiitinen, H., Belliveau, J. W., 2004. Human posterior auditory cortex gates novel sounds to consciousness. *Proc. Natl. Acad. Sci. U. S. A.* 101 (17), 6809–6814.
- Jacobsen, T., Horvath, J., Schröger, E., Lattner, S., Widmann, A., Winkler, I., 2004. Pre-attentive auditory processing of lexicality. *Brain Lang.* 88, 54–67.
- Jung, T.P., Makeig, S., Humphries, C., Lee, T.W., McKeown, M.J., Iragui, V., et al., 2000. Removing electroencephalographic artifacts by blind source separation. *Psychophysiology* 37 (2), 163–178.
- Kersten, D., Mamassian, P., Yuille, A., 2004. Object perception as Bayesian inference. *Annu. Rev. Psychol.* 55, 271–304.
- Knill, D.C., Pouget, A., 2004. The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci.* 27 (12), 712–719.
- Korpiälhti, P., Krause, C.M., Holopainen, I., Lang, A.H., 2001. Early and late mismatch negativity elicited by words and speech-like stimuli in children. *Brain Lang.* 76, 332–339.
- Korzyukov, O., Pflieger, M.E., Wagner, M., Bowyer, S.M., Rosburg, T., Sundaresan, K., Elger, C.E., Boutros, N.N., 2007. Generators of the intracranial P50 response in auditory sensory gating. *Neuroimage* 35, 814–826.
- Kuperberg, G.R., Jaeger, T.F., 2016. What do we mean by prediction in language comprehension? *Lang. Cogn. Neurosci.* 31 (1), 32–59.
- Kutas, M., DeLong, K.A., Smith, N.J., 2011. A look around at what lies ahead: prediction and predictability in language processing. In: Bar, M. (Ed.), *Predictions in the Brain: Using Our Past to Generate a Future*. Oxford University Press, New York, NY, pp. 190–207.

- Larsson, J., Smith, A.T., 2012. fMRI repetition suppression: neuronal adaptation or stimulus expectation? *Cereb. Cortex* 22, 567–576.
- Lau, E.F., Weber, K., Gramfort, A., Hämäläinen, M.S., Kuperberg, G.R., 2016. Spatiotemporal signatures of lexical-semantic prediction. *Cereb. Cortex* 26, 1377–1387.
- Lehiste, I., 1970. *Suprasegmentals*. The MIT Press, Cambridge.
- Matchin, W., Hammerly, C., Lau, E., 2017. The role of the IFG and pSTS in syntactic prediction: evidence from a parametric study of hierarchical structure in fMRI. *Cortex* 88, 106–123.
- Matsuzaki, N., Nagasawa, T., Juhász, C., Sood, S., Asano, E., 2012. Independent predictors of neuronal adaptation in human primary visual cortex measured with high-gamma activity. *Neuroimage* 59, 1639–1646.
- May, P.J.C., Tiitinen, H., 2004. The MMN is a derivative of the auditory N100 response. *Neurol. Clin. Neurophysiol.* 20, 1–5.
- Millidge, B., Seth, A., Buckley, C.L., 2021. Predictive Coding: A Theoretical and Experimental Review arXiv preprint arXiv:210712979.
- Mumford, D., 1992. On the computational architecture of the neocortex II. The role of cortico-cortical loops. *Biol. Cybern.* 66 (3), 241–251.
- Näätänen, R., 2001. The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology* 38, 1–21.
- Näätänen, R., Winkler, I., 1999. The concept of auditory stimulus representation in cognitive neuroscience. *Psychol. Bull.* 125 (6), 826–859.
- Näätänen, R., Gaillard, A.W., Mäntysalo, S., 1978. Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol.* 42, 313–329.
- Näätänen, R., Paavilainen, P., Alho, K., Reinikainen, K., Sams, M., 1989. Do event-related potentials reveal the mechanism of the auditory sensory memory in the human brain? *Neurosci. Lett.* 98, 217–221.
- Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., et al., 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385, 432–434.
- Näätänen, R., Paavilainen, P., Rinne, T., Alho, K., 2007. The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin. Neurophysiol.* 118, 2544–2590.
- Öhman, A., Lader, M., 1972. Selective attention and “habituation” of the auditory averaged evoked response in humans. *Physiol. Behav.* 8 (1), 79–85.
- Orfanidou, E., Marslen-Wilson, W.D., Davis, M.H., 2006. Neural response suppression predicts repetition priming of spoken words and pseudowords. *J. Cogn. Neurosci.* 18, 1237–1252.
- Otten, M., Van Berkum, J.J.A., 2008. Discourse-based word anticipation during language processing: prediction or priming? *Discourse Process.* 45 (6), 464–496.
- Picton, T.W., Hillyard, S.A., Krausz, H.I., Galambos, R., 1974. Human auditory evoked potentials. I. Evaluation of components. *Electroencephalogr. Clin. Neurophysiol.* 36 (2), 179–190.
- Poppenk, J., McIntosh, A.R., Moscovitch, M., 2016. fMRI evidence of equivalent neural suppression by repetition and prior knowledge. *Neuropsychologia* 90, 159–169.
- Pulvermüller, F., Shtyrov, Y., 2006. Language outside the focus of attention: the mismatch negativity as a tool for studying higher cognitive processes. *Prog. Neurobiol.* 79, 49–71.
- Pulvermüller, F., Kujala, T., Shtyrov, Y., Simola, J., Tiitinen, H., Alku, P., Alho, K., Martinkauppi, S., Ilmoniemi, R.J., Näätänen, R., 2001. Memory traces for words as revealed by the mismatch negativity. *Neuroimage* 14, 607–616.
- Rietveld, T.C., Kerkhoff, J., Gussenhoven, C., 2004. Word prosodic structure and vowel duration in Dutch. *J. Phon.* 32, 349–371.
- Rohde, H., Levy, R., Kehler, A., 2011. Anticipating explanations in relative clause processing. *Cognition* 118 (3), 339–358.
- Rosburg, T., 2004. Effects of tone repetition on auditory evoked neuromagnetic fields. *Clin. Neurophysiol.* 115, 898–905.
- Schwartz, M., Farrugia, N., Kotz, S.A., 2013. Dissociation of formal and temporal predictability in early auditory evoked potentials. *Neuropsychologia* 51, 320–325.
- Shtyrov, Y., Lenzen, M., 2017. First-pass neocortical processing of spoken language takes only 30 msec: electrophysiological evidence. *Cogn. Neurosci.* 8 (1), 24–38.
- Shtyrov, Y., Pulvermüller, F., 2002. Neurophysiological evidence of memory traces for words in the human brain. *Neuroreport* 13, 521–525.
- Slabu, L., Escera, C., Grimm, S., Costa-Faidella, J., 2010. Early change detection in humans as revealed by auditory brainstem and middle-latency evoked potentials. *Eur. J. Neurosci.* 32, 859–865.
- Summerfield, C., de Lange, F.P., 2014. Expectation in perceptual decision making: neural and computational mechanisms. *Nat. Rev. Neurosci.* 15, 745–756.
- Summerfield, C., Trittschuh, E.H., Monti, J.M., Mesulam, M.M., Egner, T., 2008. Neural repetition suppression reflects fulfilled perceptual expectations. *Nat. Neurosci.* 11 (9), 1004–1006.
- Todorovic, A., de Lange, F.P., 2012. Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. *J. Neurosci.* 32, 13389–13395.
- Todorovic, A., van Ede, F., Maris, E., de Lange, F.P., 2011. Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: a MEG study. *J. Neurosci.* 31, 9118–9123.
- van Berkum, J.J., Brown, C.M., Zwitserlood, P., Kooijman, V., Hagoort, P., 2005. Anticipating upcoming words in discourse: evidence from ERPs and reading times. *J. Exp. Psychol. Learn. Mem. Cogn.* 31 (3), 443–467.
- van Donselaar, W., Koster, M., Cutler, A., 2005. Exploring the role of lexical stress in lexical recognition. *Q. J. Exp. Psychol.* 58, 251–273.
- van Oostendorp, M., 2012. Quantity and the three-syllable window in Dutch word wtrress. *Lang Ling Compass* 6, 343–358.
- Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., Dehaene, S., 2011. Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proc. Natl. Acad. Sci. U. S. A.* 108 (51), 20754–20759.
- Wagner, M., Watson, D.G., 2010. Experimental and theoretical advances in prosody: A review. *Lang. Cogn. Process.* 25, 905–945.
- Weber, C., Hahne, A., Friedrich, M., Friederici, A.D., 2004. Discrimination of word stress in early infant perception: electrophysiological evidence. *Cogn. Brain Res.* 18, 149–161.
- Weber, K., Christiansen, M., Petersson, K.M., Indefrey, P., Hagoort, P., 2016. fMRI syntactic and lexical repetition effects reveal the initial stages of learning a new language. *J. Neurosci.* 36, 6872–6880.
- Wicha, N.Y., Bates, E.A., Moreno, E.M., Kutas, M., 2003. Potato not pope: human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neurosci. Lett.* 346 (3), 165–168.
- Wicha, N.Y., Moreno, E.M., Kutas, M., 2004. Anticipating words and their gender: an event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *J. Cogn. Neurosci.* 16 (7), 1272–1288.
- Willems, R.M., Frank, S.L., Nijhof, A.D., Hagoort, P., Van Den Bosch, A., 2016. Prediction during natural language comprehension. *Cereb. Cortex* 26, 2506–2516.
- Winkler, I., 2007. Interpreting the mismatch negativity. *J. Psychophysiol.* 21, 147–163.
- Winkler, I., Karmos, G., Näätänen, R., 1996. Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. *Brain Res.* 742, 239–252.
- Winkler, I., Schröger, E., 2015. Auditory perceptual objects as generative models: setting the stage for communication by sound. *Brain Lang.* 148, 1–22.
- Woods, D.L., Elmasian, R., 1986. The habituation of event-related potentials to speech sounds and tones. *Electroencephalogr. Clin. Neurophysiol.* 65 (6), 447–459.
- Ylinen, S., Strelnikov, K., Huotilainen, M., Näätänen, R., 2009. Effects of prosodic familiarity on the automatic processing of words in the human brain. *Int. J. Psychophysiol.* 73 (3), 362–368.
- Zhang, N., Zhang, Q., 2019. Rhythmic pattern facilitates speech production: an ERP study. *Sci. Rep.* 9, 12974.
- Zora, H., 2016. Mapping Prosody onto the Lexicon: Memory Traces for Lexically Specified Prosodic Information in the Brain (PhD Dissertation, Department of Linguistics, Stockholm University).
- Zora, H., Schwarz, I.-C., Heldner, M., 2015. Neural correlates of lexical stress: mismatch negativity reflects fundamental frequency and intensity. *Neuroreport* 26, 791–796.
- Zora, H., Heldner, M., Schwarz, I.-C., 2016a. Perceptual correlates of Turkish word stress and their contribution to automatic lexical access: evidence from early ERP components. *Front. Neurosci.* 10, 7.
- Zora, H., Riad, T., Schwarz, I.-C., Heldner, M., 2016b. Lexical specification of prosodic information in Swedish: evidence from event-related potentials. *Front. Neurosci.* 10, 533.
- Zora, H., Riad, T., Ylinen, S., 2019. Prosodically controlled derivations in the mental lexicon. *J. Neurolinguistics* 52.
- Zora, H., Rudner, M., Magnusson, A., 2020. Concurrent affective and linguistic prosody with the same emotional valence elicits a late positive ERP response. *Eur. J. Neurosci.* 51, 2236–2249.