



Investigating the neurocognitive background of speech perception with a fast multi-feature MMN paradigm

Ferenc Honbolygó^{1,2} · Borbála Zulauf² · Maria Ioanna Zavogianni^{1,3} · Valéria Csépe^{1,4}

Received: 13 July 2023 / Accepted: 11 April 2024 / Published online: 28 May 2024
© The Author(s) 2024

Abstract

The speech multi-feature MMN (Mismatch Negativity) offers a means to explore the neurocognitive background of the processing of multiple speech features in a short time, by capturing the time-locked electrophysiological activity of the brain known as event-related brain potentials (ERPs). Originating from Näätänen et al. (Clin Neurophysiol 115:140–144, 2004) pioneering work, this paradigm introduces several infrequent deviant stimuli alongside standard ones, each differing in various speech features. In this study, we aimed to refine the multi-feature MMN paradigm used previously to encompass both segmental and suprasegmental (prosodic) features of speech. In the experiment, a two-syllable long pseudoword was presented as a standard, and the deviant stimuli included alterations in consonants (deviation by place or place and mode of articulation), vowels (deviation by place or mode of articulation), and stress pattern in the first syllable of the pseudoword. Results indicated the emergence of MMN components across all segmental and prosodic contrasts, with the expected fronto-central amplitude distribution. Subsequent analyses revealed subtle differences in MMN responses to the deviants, suggesting varying sensitivity to phonetic contrasts. Furthermore, individual differences in MMN amplitudes were noted, partially attributable to participants' musical and language backgrounds. These findings underscore the utility of the multi-feature MMN paradigm for rapid and efficient investigation of the neurocognitive mechanisms underlying speech processing. Moreover, the paradigm demonstrated the potential to be used in further research to study the speech processing abilities in various populations.

Keywords Speech processing · Event-related brain potentials · Mismatch Negativity · Multi-feature paradigm · Segmental speech features · Prosodic speech features

Introduction

As György Ádám suggested in one of his articles (Ádám 2001, p. 1166), “input unconscious processes” are sensory processes that influence neural networks in a non-conscious way, and much of the information that enters the brain from the outside world remains outside of awareness. One of

Ádám's many interests included the study of unconscious or nonconscious processes in the human brain. As he wrote in his book *Visceral Perception: understanding internal cognition* (Ádám 1998, p. 10): “It's an old observation that when people are questioned about the detailed processes mediating the impact of a given set of stimuli in a series of responses, they are unable to report on the sequence of events. In other words, the *process* of coming to someone's knowledge remains unconscious, the *outcome* of the process, i.e., the *product* or the *consequence* may become reportable, i.e., conscious.” Language can be considered as one such process, as it involves several different mechanisms that are not available for consciousness: e.g., extracting phonetic information from speech sounds, establishing the transitional probabilities leading to the acquisition of word meaning, or implicitly acquiring the rules of language in the form of syntax.

✉ Ferenc Honbolygó
honbolygo.ferenc@ttk.hu

¹ HUN-REN Research Centre for Natural Sciences, Budapest, Hungary

² Institute of Psychology, Eötvös Loránd University, Budapest, Hungary

³ Faculty of Modern Philology and Social Sciences, Multilingualism Doctoral School, University of Pannonia, Veszprém, Hungary

⁴ University of Pannonia, Veszprém, Hungary

The input processes of language involve first and foremost the perception of speech. The goal of speech perception is to map sounds onto meaning. This is accomplished by storing speech related auditory information in auditory sensory memory and transforming it into representations that can be matched to the elements of the mental lexicon. These elements (i.e., words) are spelled from contrastive sound units (i.e., phonemes): abstract tokens that are essentially invariant across contexts and speakers (Port 2007). It can be argued that the processing of speech information is based on sensory processes that are largely unconscious, and one way to investigate them is by looking at the brain's electrical activity elicited by them.

The examination of the electrical activity of the brain has proven to be a productive method in understanding neural activity since it allows to follow the neural activity in a continuous online way. Despite their limited spatial resolution, event-related brain potentials (ERPs) offer insight into the real-time neural dynamics of linguistic processing (Steinhauer 2008). A particularly valuable ERP component in this area of research is the Mismatch Negativity (MMN) (Näätänen et al. 2007). The MMN is an ERP response elicited by any type of change in the stimulus sequence and can be observed peaking at 150–250 ms after the onset change. The MMN is characterized by a fronto-central voltage maximum with a negative polarity and is typically observed in passive oddball paradigms where frequently repeated standard stimuli are interspersed by infrequent deviant stimuli that differ from the standards in some discriminable feature. The MMN ERP component can be elicited without the participant's attention and is interpreted as an electrical correlate of the pre-attentive detection of regularity violations (Winkler et al. 2009).

Studies investigating the MMN elicited by speech sounds found that differences between speech sounds (i.e., phonetic contrasts) evoke MMN if the difference between the sounds is perceivable (Näätänen 2001). Processing is significantly modulated by the long-term representations (i.e., phoneme representations) of speech sounds. These long-term representations exist for native speech sounds, and foreign speech sounds that the listener is familiar with. In a study comparing Finnish and Hungarian participants (Winkler et al. 1999) speech sound pairs that were within-category contrasts in Hungarian but between-category contrasts in Finnish, and vice versa, were presented. According to the results, the MMN was present in all conditions, however it was larger for across- than the within-category contrasts for both groups. The larger MMN to the language-specific across-category difference suggests that the processing system considers not only the physical differences between the speech sounds, but also the native long-term linguistic representations. In their seminal paper, Näätänen et al. (1997) investigated the phoneme representations more closely in another cross-linguistic

MMN study. Using MEG, the authors studied the processing of prototypical speech sounds in Finnish and Estonian. The MMN component was observed in both conditions but was smaller in the non-prototypical vowel condition. A dipole analysis of the cortical sources of MMN showed enhanced activity in the left auditory cortex in the prototypical vowel, which was not observed in the non-prototypical vowel. This was interpreted as evidence of the existence of long-term phoneme representations. Similar results were found by Dehaene-Lambertz (1997) showing that native across-category deviants elicited a larger MMN than non-native or within-category deviants, suggesting that language-specific phonetic information has a separate neural representation, and it can serve as the basis for the mismatch detection. Ylinen et al. (2006) also found evidence for native-language phoneme representations influencing the MMN, by showing an enhanced MMN amplitude to across-category deviants in native as compared to non-native speakers.

Näätänen (2001) proposed a model to explain the MMN results in speech sound perception. According to the model, the recognition of speech sounds is facilitated by language-specific phoneme representations, called phoneme traces, which are abstract representations of the native speech sound categories. Research has shown that phoneme traces develop rapidly from birth on, undergo a delicate development in early childhood and are in effect from the age of 6 to 12 months (Cheour et al. 1998). The MMN response in speech perception experiments is therefore reflective not only of the acoustic differences between standard and deviant stimuli, but also of the differences between the incoming speech information and the long-term phoneme representations.

A more complex model for the MMN's emergence is proposed in the AERS (Auditory Event Representation System) predictive coding model suggested by Winkler and Schröger (2015). The predictive coding theory (Friston 2005, 2010) is probably the most influential current framework of cognitive neuroscience, positing that the human brain is an inference machine, actively predicting and explaining its sensations through internal representations that combine sensory input with top-down predictions. The AERS model suggests that the predictive model of the acoustic environment stores representations of the current regularities of the environment together with our long-term experience about it. In the case of speech, the long-term representations are available for different linguistic units: native speech sounds, prosodic patterns, words, etc. According to the model, the auditory system first analyzes incoming sounds, identifying basic speech-related features such as voicing, formant transitions, pitch, and intensity. These elements are then combined into sensory memory representations like phonemes and prosodic features. Finally, these representations are compared to the model's predictions of the surrounding

sound environment. The predictive model is the basis of the mechanisms comparing incoming information with the regularity representations. If predictions from the model fail, the model needs to be corrected via an updating process, which is reflected by the MMN.

As an alternative method for eliciting the MMN response beyond the traditional passive oddball paradigm, the multi-feature paradigm has been suggested by Näätänen et al. (2004). This paradigm involves the presentation of various types of acoustic changes where every other stimulus serves as the standard and every other stimulus is one of the different deviants. The multi-feature paradigm is based on the premise that each sound feature elicits a distinct MMN and simultaneously strengthens the memory trace of the standard stimulus with respect to the features shared between them (Pakarinen et al. 2013). The paradigm enables efficient recording of ERP responses to multiple deviants in a single stimulus sequence, and research suggests that there is no difference between the MMN elicited in the oddball and multi-feature paradigms (Näätänen et al. 2004; Pakarinen et al. 2009).

Several studies have applied the multi-feature paradigm to explore the MMN elicited by speech sounds (Fisher et al. 2011; Kuuluvainen et al. 2014; Lovio et al. 2009; Pakarinen et al. 2009, 2013; Sorokin et al. 2010). Pakarinen et al. (2009) and Sorokin et al. (2010) examined the processing of feature changes in consonant–vowel (CV) syllables in both passive oddball and multi-feature paradigms and found that changes in all five aspects (f0, intensity, vowel duration, vowel change, and consonant changes in the first study and vowel, vowel duration, consonant, syllable intensity, and frequency changes in the second study) elicited similar MMN ERP responses in both paradigms. Pakarinen et al. (2013) reported that the amplitude and latency of the MMN followed the magnitude of deviation in several acoustic and phonetic features in vowel stimuli: the larger the deviation was, the larger and earlier the MMN peaked. Kuuluvainen et al. (2014) demonstrated that the MMN/MMNm (i.e., the magnetic counterpart of MMN) was enhanced for the same features in speech CV syllables compared to their non-speech counterparts, providing further evidence for the activation of phoneme traces during speech perception. Partanent et al. (2011) found that the MMN was elicited by acoustic (f0, intensity) and phonetic (vowel duration, vowel identity) changes in all syllables of a three-syllable pseudoword. The results of these studies support that the multi-feature paradigm is an effective tool for investigating the neural mechanisms underlying speech sound processing.

In the present study, we aimed to create a slightly modified version of multi-feature MMN paradigms used in previous studies. Our goal was to devise a paradigm that reliably elicits the MMN component to multiple speech features,

including both segmental and suprasegmental (prosodic), allowing to profile the speech perception abilities of participants. In subsequent studies, we plan to use this paradigm to study the speech perception abilities of children with speech and language difficulties; therefore, it was crucial to establish if the paradigm 1) can reliably elicit the MMN and 2) can show individual differences.

One novelty of the paradigm was the inclusion of a prosodic feature, word stress. Previous studies already established that changes in word stress elicit the MMN component in classic oddball paradigms (Honbolygó et al. 2004, 2020; Honbolygó and Csépe 2013; Weber et al. 2004; Ylinen et al. 2009; Zora et al. 2015, 2016, 2023). Multi-feature studies also presented linguistically relevant prosody changes, i.e., changes in f0 and intensity in speech sounds (Pakarinen et al. 2009, 2013; Partanent et al. 2011; Sorokin et al. 2010), and in affective prosody (Kostilainen et al. 2020; Pakarinen et al. 2014), but there is only one study that examined linguistic prosodic changes in a multi-feature paradigm (Zavogianni et al. 2024). In this study, the authors included segmental, prosodic, and acoustic changes to investigate the processing of a wide range of speech related features, while in the present study, we wanted to focus on a more in-depth examination of segmental speech features. Therefore, we included two types of consonant changes: place, and place and mode of articulation, and two types of vowel changes: mode and place of articulation. We selected these phonetic contrasts because we wanted to introduce systematic changes in the stimuli: instead of randomly selecting deviant consonants and vowels, we selected particular phonetic features in which the deviants differed from the standards (see Table 1). We also included a contrast in the word stress pattern (iambic as compared to trochaic), and to allow the investigation of word stress change, we created two-syllable long pseudowords.

To follow the main goals of the study first we investigated if the paradigm can reliably elicit the MMN component. Based on previous studies, we expected that all phonetic contrasts will elicit the MMN component, including the word stress change (Honbolygó et al. 2017, 2020; Honbolygó and Csépe 2013; Zavogianni et al. 2024). However, we didn't have any specific hypothesis concerning the phonetic features. Moreover, we wanted to investigate whether the paradigm is sensitive enough to reveal individual differences. We collected information about participants' music and foreign language learning background, and we expected to find that participants with better proficiency in music and/or language learning will show enhanced MMNs (see Kühnis et al. 2013; Marie et al. 2012; Shahin 2011).

Table 1 Characteristics of experimental stimuli

Stimulus (IPA transcription)	Phonetic contrast	Phonetic features	Total duration (ms)	Duration of 1st and 2nd syllable (ms)
keke (/ˈkɛkɛ/)	Standard	/k/: voiceless, plosive, velar; /e/: front, unrounded, mid; trochaic	600	280 320
peke (/ˈpɛkɛ/)	Consonant deviant—place of articulation (PoA)	/p/: voiceless, plosive, labial	600	270 330
feke (/ˈfɛkɛ/)	Consonant deviant—place and mode of articulation (PMoA)	/f/: voiceless, fricative, labial	680	380 300
kike (/ˈkikɛ/)	Vowel deviant—mode of articulation (MoA)	/i/: front, unrounded, close	590	255 335
koke (/ˈkokɛ/)	Vowel deviant—place of articulation (PoA)	/o/: back, mid	680	290 390
ke'ke (/kɛˈkɛ/)	Stress deviant	iambic	650	290 360

Methods

Participants

A total of 29 native Hungarian-speakers (25 females) participated in the research; all of them were right-handed. The mean age of participants was 22.4 years (SD: ± 3.75). The oldest participant was 39 years old, and the youngest was 19. Participants had normal hearing and did not report any neurological, auditory, or speech-related issues. Participants were all students of Eötvös Loránd University, and they participated in the experiment for course credits. The study was approved by the United Ethical Review Committee for Research in Psychology (EPKEB) in Hungary and was conducted in accordance with the Declaration of Helsinki.

Stimuli and procedure

The stimuli were audio recordings of two syllable long pseudowords, voiced by a female actor. There were altogether six stimuli: one standard and five deviants. The standard stimulus was the pseudoword 'keke'. The deviant stimuli were modifications of the standard stimulus in the following characteristics: changes in one consonant or vowel, and changes in stress pattern (see Table 1). We chose the following stimuli: 'kike' and 'koke' for the vowel deviants, 'peke' and 'feke' for the consonant deviants, and the iambic version (i.e., stress on the second syllable) of 'keke' for the stress deviant.

The stimuli were presented in a multi-feature paradigm (following the paradigm of Partanent et al. 2011). The stimulus onset asynchrony (SOA) was randomly varied between 750 and 950 ms. We presented 630 standard and 120 deviant stimuli per condition (altogether 1230 stimuli) in two blocks. Each block started with the presentation of 15 standard stimuli, and every other stimulus was a deviant

presented in a random order. During the experiment, participants were comfortably seated in a chair inside a soundproof and electrically shielded experiment booth. They were asked to watch a movie of their preference without any sound or subtitles while the stimulus sequence was presented to them through headphones (Sennheiser PX200). The duration of the recording was about 17 min. The presentation of the experiment was performed by using Presentation software (v. 21.1, Neurobehavioral Systems).

Participants completed a questionnaire in which we inquired about their formal music education: whether they received it or not (*Music education*) and if yes for how long they received it (*Music education years*); if they were actively engaged in music activities at the time of the data collection (*Current music activity*). We also inquired about their foreign language proficiency, i.e., highest proficiency level according to the CEFR (Common European Framework of Reference for Languages): beginner: B1-B2; advanced: C1-C2 (*Foreign language proficiency*).

EEG recording and analysis

We recorded the EEG activity with a 64-channel recording system (BrainAmp amplifier and BrainVision Recorder software, BrainProducts GmbH) with a sampling rate of 1000 Hz. The Ag/AgCl sintered active electrodes were mounted in an electrode cap (actiCAP) on the scalp of each participant according to the 10% equidistant system at the following positions Fp1, Fp2, AF7, AF3, AFz, AF4, AF8, F7, F3, F1, F2, Fz, F4, F6, F8, FC5, FC1, FC3, FCz, FC2, FC4, FC6, FT7, FT9, FT8, FT10, T7, C5, C3, C1, Cz, C2, C4, C6, T8, TP9, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, TP10, P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO3, POz, PO4, PO8, O1, Oz, O2, Iz. The reference and ground electrodes were at the FCz and AFz locations, respectively.

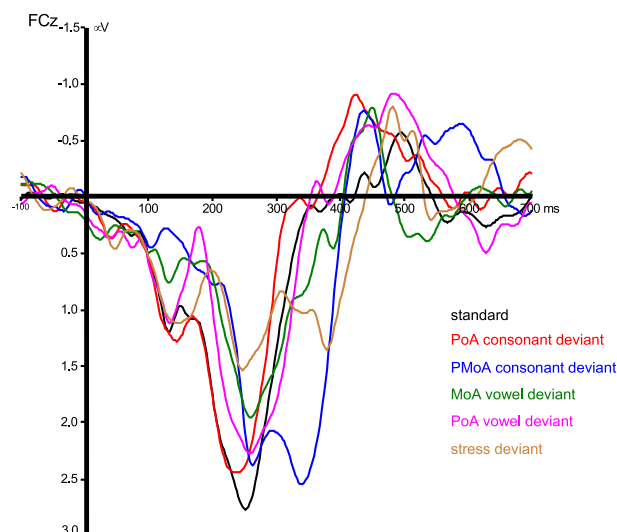


Fig. 1 Grand average ERPs of the standard and five deviants on the FCz electrode. Negativity is plotted upwards here and on the following figures

During the recording, electrode contact impedances were kept below 10 k Ω .

We analyzed the EEG data offline with the BrainVision Analyzer 2.2 software (Brain Products GmbH, Munich, Germany). The analysis pipeline consisted of the following steps. Continuous EEG was filtered with a band-pass filter between 1 and 30 Hz (48 db/oct), and with a notch filter at 50 Hz to remove additional electrical noise. This was followed by the removal of horizontal and vertical eye-movement artifacts as well as heartbeats by applying an Independent Component Analysis (ICA) (Delorme et al. 2007). During ICA, the EEG signals were first decomposed into 32 independent components, and components matching the waveform and amplitude distribution of blinks, horizontal eye-movements and heartbeats were removed manually (1–3 per participant). Finally, the EEG data were recomposed based on the remaining components. After that, EEG data were re-referenced to the average activity of all electrodes. This was followed by the segmentation of continuous EEG time locked to the onset of the stimuli. Segments were created separately for the standard and five deviants from 100 ms before stimulus onset to 700 ms after. Next, to remove potentially remaining artifacts, we applied an automatic artifact rejection algorithm on the segmented data, which rejected segments where the activity exceeded $\pm 80 \mu\text{V}$ at any of the electrode sites (minimum number of retained segments per condition per participant: 104 out of 120, 86.7%). Finally, remaining epochs were baseline corrected (-100 – 0 ms) and averaged (Fig. 1).

Statistical analysis

To quantify the MMN components, difference curves were created individually by subtracting the ERP waveform of the standard from that of the deviants. We performed a point-by-point paired samples *t*-test (see Fig. 2) to determine those time windows where the deviants elicited a significantly different ERP activity compared to the standard. This method ensures an unbiased selection on temporal regions to quantify ERP components and has been previously used by for example Kostilainen et al. (2020). Of the several significant time windows demonstrating a significant difference between the standard and deviant stimuli, we selected the earliest ones, corresponding to the expected time window of the MMN (150–250 ms after the onset of change, see Näätänen et al. 2007).

These time windows were the following: PoA consonant: 245–330 ms; PMoA consonant: 115–250 ms; MoA vowel: 125–280 ms; PoA vowel: 160–255 ms; stress: 185–300 ms. We used these time windows to direct the automatic detection of MMN component peaks and exported the peak amplitude and latency values. One participant's amplitude value in the PMoA consonant deviant condition were deemed as outlier; therefore, we removed the participant from further statistical analysis, and only the data of 28 participants were considered for further analysis.

To investigate whether MMN peaks were elicited in a statistically significant manner, we performed one-sample *t*-tests against 0 on the peak amplitude values. We also compared the latency and amplitude values of deviants using repeated measures ANOVA, using a factor of Stimulus with five levels.

To study individual differences, we calculated correlations (in the case of continuous variables) and point biserial correlations (in the case of binary variables) between the MMN peak amplitude and latency values and the participant's musical background and foreign language proficiency (see above). As some of the variables were not normally distributed, according to the Shapiro–Wilk test, we performed Spearman correlations. Because multiple comparisons were conducted, we corrected significance levels according to the method described by Benjamini and Hochberg (1995), which controls for the False Discovery Rate (FDR) using sequential Bonferroni-type corrections. The statistical analyses were performed using JASP (Version 0.16.4).

Results

The grand average ERPs of all stimuli (see Fig. 1) depict a series of positive going ERPs on the frontal electrodes in the first 300 ms after stimulus onset, followed by a series

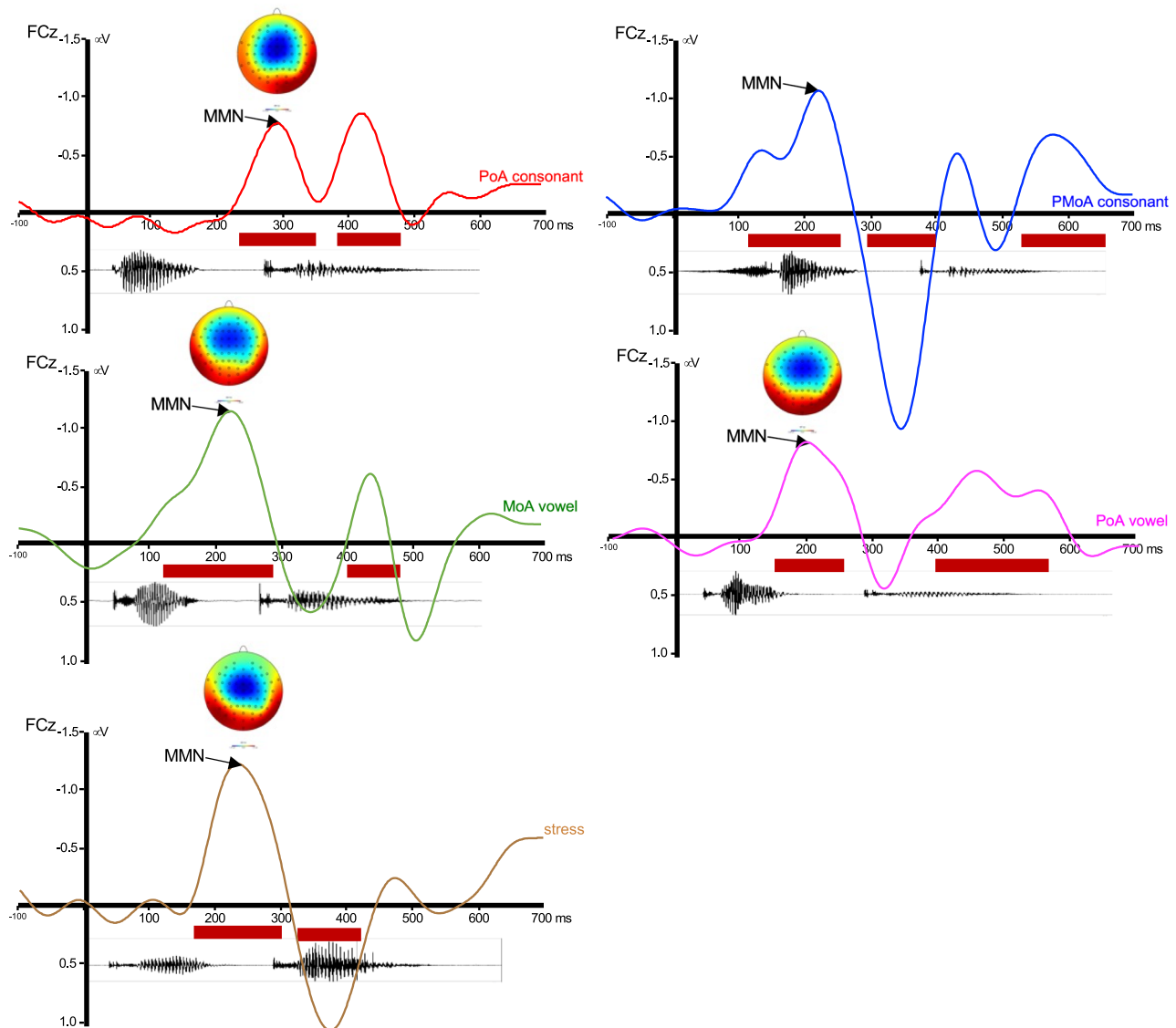


Fig. 2 Difference curves of the five deviants on the FCz electrode. Red horizontal lines depict the time periods where the point-by-point paired *t*-test comparing the amplitudes of the deviant and standard

showed a significant difference. Topoplots present the amplitude distribution of the MMN peaks at the time point indicated by the arrows. Below the time axis, the audio waveform of the stimuli is shown

of negative deflections. ERPs elicited by the stimuli start to deviate from each other after around 100 ms.

The MMN component appeared on the difference curves (see Fig. 2). Figure 2 shows the difference curves on the FCz electrode, together with the results of the point-by-point paired *t*-test, the amplitude distribution of the MMN peaks and the audio waveform of the stimuli. As visible in Fig. 2., the deviants elicited slightly different ERP waveforms, with different morphology and temporal characteristics, which necessitated quantifying the MMN component in time windows with slightly different latencies (see above). As it visible on the topoplots, each deviant elicited the MMN component, with a characteristic fronto-centrally negative

amplitude distribution, accompanied by a polarity reversal on the occipital electrode sites.

Additionally, according to the point-by-point *t*-tests, deviants elicited significantly different ERPs compared to the standard in later time windows. These could be identified either as a P300 (when positive around 300 ms) or Late Discriminative Negativity (LDN, when negative around 400 ms). As the focus of the current study was the MMN, we did not analyze or interpret these additional components.

To demonstrate if the MMN components were significantly elicited by the deviant stimuli, we performed a series of one-sample *t*-tests against 0 on the peak amplitude values of MMN. Figure 3 demonstrates the

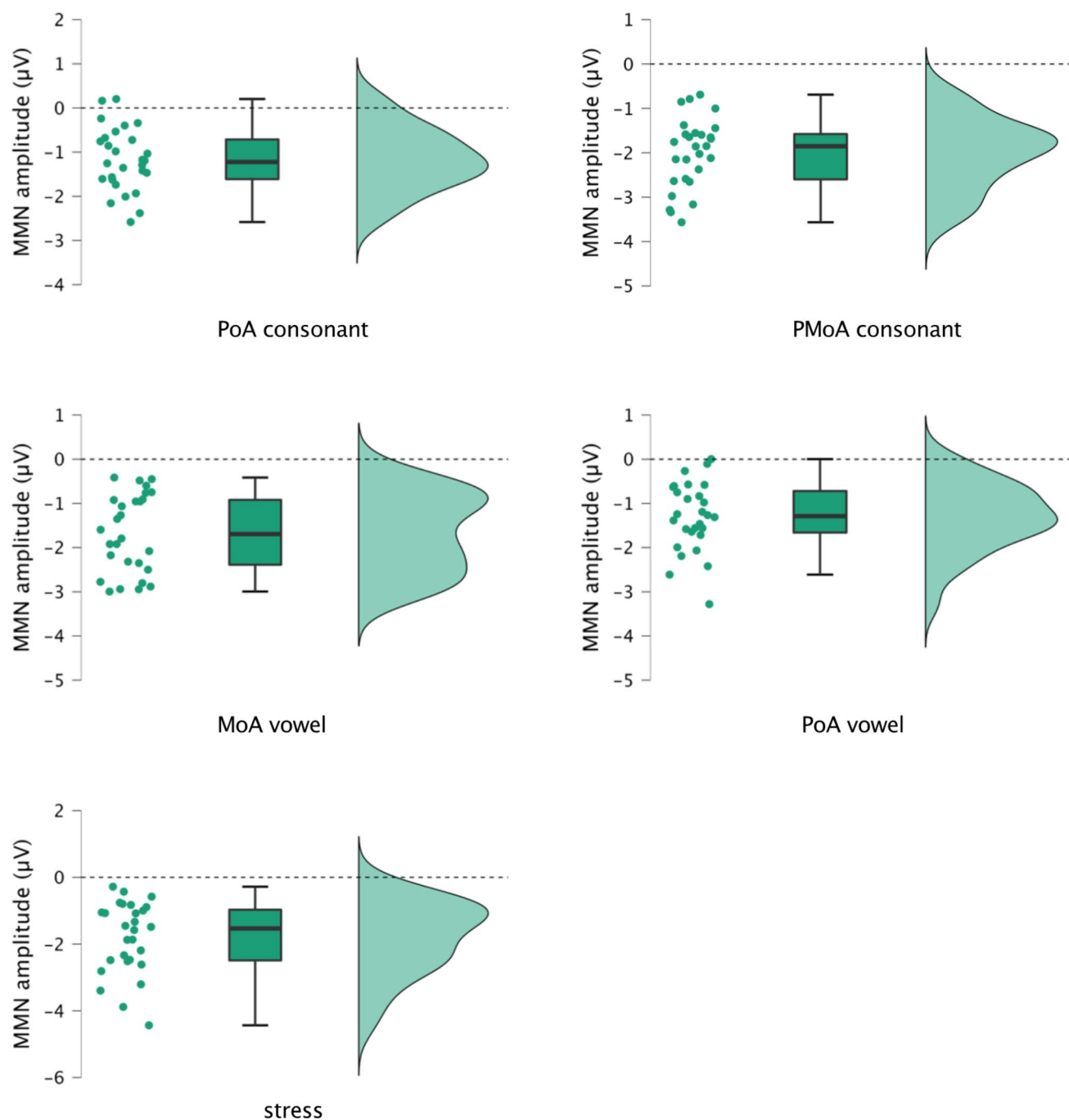


Fig. 3 Raincloud plots showing a combination of jittered individual values, the boxplot and the probability density plot of the MMN amplitudes in the five conditions. *NB* negativity is plotted downwards

Table 2 Results of the one-sample *t*-test

	<i>t</i>	df	<i>p</i>
PoA consonant	−8.71	27	<.001
PMoA consonant	−13.53	27	<.001
MoA vowel	−10.01	27	<.001
PoA vowel	−8.94	27	<.001
Stress	−8.85	27	<.001

distribution of the amplitude values in each condition, and Table 2 summarizes the results of the *t*-tests. Figure 3 depicts a so-called “raincloud plot”, which is the combination of several forms of data visualization: raw jittered individual data points, a boxplot, i.e. the visualization of central tendency, and a “split-half violin”, which is un-mirrored probability density plot (i.e., a plot of the likelihood of occurrence of a data point using the probability

Fig. 4 Raincloud plots showing the MMN latencies in the five conditions. Connected data points belong to the same participant

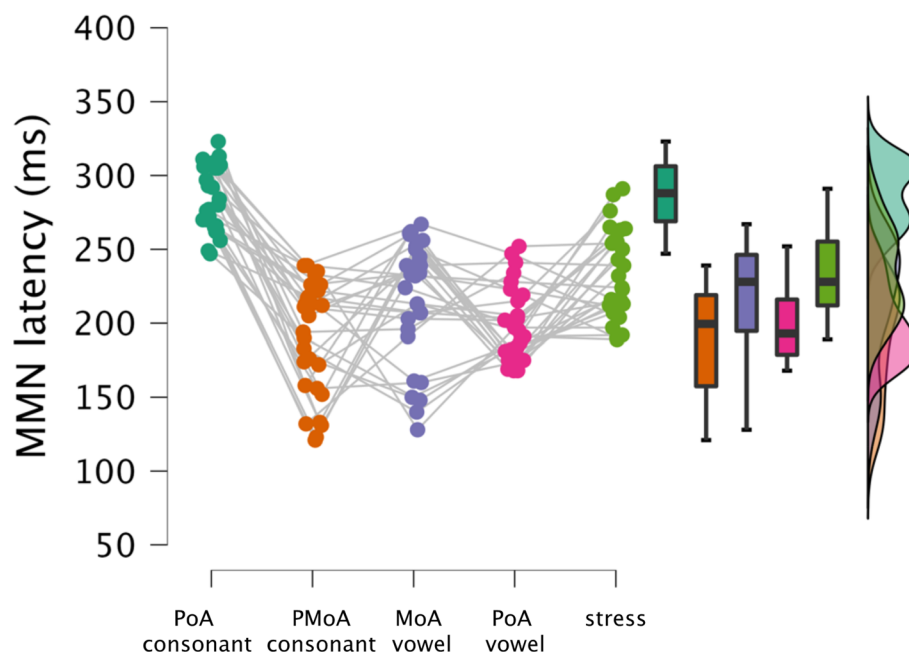


Table 3 Results of the Holm post-hoc test for the MMN latencies

Stimuli		Mean difference	SE	t	P_{holm}
PoA cons	PMoA cons	96.929	8.415	11.518	<.001***
	MoA vowel	71.607	8.415	8.509	<.001***
	PoA vowel	87.929	8.415	10.449	<.001***
	Stress	52.857	8.415	6.281	<.001***
PMoA cons	MoA vowel	-25.321	8.415	-3.009	0.013*
	PoA vowel	-9	8.415	-1.07	0.287
	Stress	-44.071	8.415	-5.237	<.001***
MoA vowel	PoA vowel	16.321	8.415	1.94	0.11
	Stress	-18.75	8.415	-2.228	0.084
PoA vowel	Stress	-35.071	8.415	-4.168	<.001***

P -value adjusted for comparing a family of 10.* $p < .05$, *** $p < .001$

density function) plotted against the redundant data axis (Allen et al. 2023).

To control for multiple comparisons, we used Bonferroni correction which set the critical p -value to 0.01. According to the results, the MMN amplitude was significantly different in all deviant conditions ($p < 0.001$).

We compared the latency and amplitude values of the MMN components elicited in the five conditions using repeated measures ANOVAs. In the case of latency, the main effect of Stimulus was significant, $F(4, 108) = 41.75$, $p < 0.001$, $\eta_p^2 = 0.607$ (see Fig. 4). The Holm post-hoc test (see Table 3) demonstrated that the PoA consonant deviant had the largest latency which significantly differed from the latency of all other deviants ($p < 0.001$). Additionally, the PoA consonant deviant differed from the MoA vowel and

stress deviants ($p < 0.05$), and the PoA vowel deviant differed from the stress deviant ($p < 0.001$).

In the case of amplitude, the main effect of Stimulus was significant, $F(4, 108) = 7.09$, $p < 0.001$, $\eta_p^2 = 0.208$ (see Fig. 5). The Holm post-hoc test (see Table 4) demonstrated that the amplitude of the PoA consonant deviant significantly differed from the PMoA consonant and the stress deviants ($p < 0.01$). Additionally, the PMoA consonant deviant significantly differed from the PoA vowel deviant ($p < 0.01$).

To investigate individual differences, we calculated correlations between the participants' musical background and foreign language proficiency, and the latency and amplitude values of MMN components in the five conditions. Correlations were calculated separately for the MMN latency (see Table 5) and amplitude (see Table 6) data. According to the results, in the case of MMN latencies, the latency of the stress deviant correlated negatively with *Foreign language proficiency*, $r = -0.46$, $p < 0.05$, indicating that participants who were more advanced speakers of a foreign language had an earlier MMN peak.

In the case of MMN amplitudes, the amplitude of the PMoA consonant deviant correlated positively with *Music education years*, $r = 0.58$, $p < 0.05$, indicating that participants who had more music education, showed a more positive, i.e., smaller MMN to this deviant. Moreover, the amplitude of the PoA vowel deviant correlated positively with *Music education*, $r = 0.41$, $p < 0.05$, meaning that participants who had a formal music education, showed also a more positive, MMN. Similarly, the amplitude of the PoA vowel deviant correlated positively with *Current music activity*, $r = 0.40$, $p < 0.05$, indicating that participants who

Fig. 5 Raincloud plots showing the MMN amplitudes in the five conditions. Connected data points belong to the same participant. *NB* negativity is plotted downwards

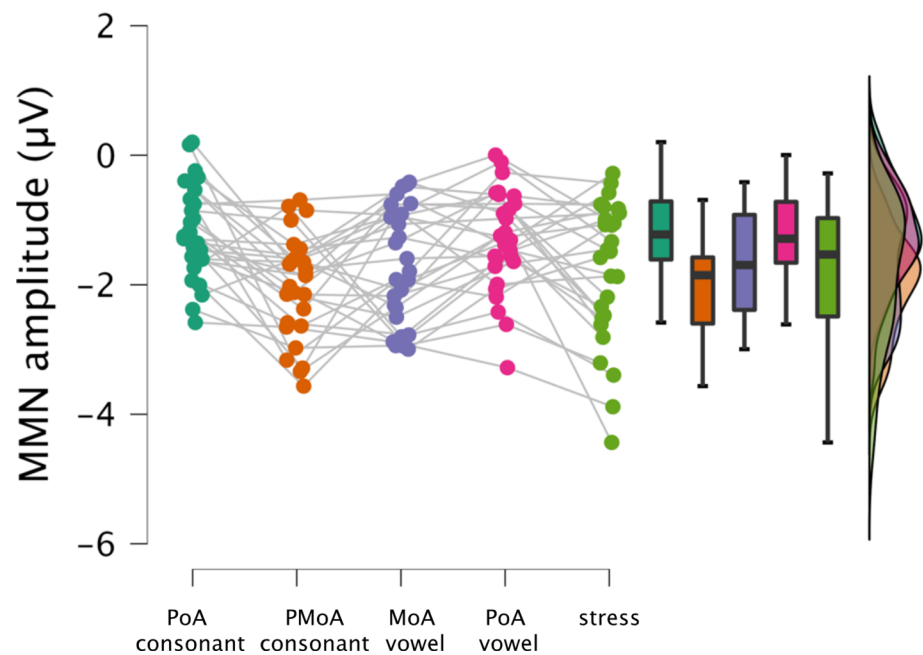


Table 4 Results of the Holm post-hoc test for the MMN amplitudes

Stimuli		Mean difference	SE	t	P _{holm}
PoA cons	PMoA cons	0.837	0.185	4.532	<.001***
	MoA vowel	0.499	0.185	2.701	0.055
	PoA vowel	0.135	0.185	0.73	0.923
	Stress	0.635	0.185	3.44	0.007**
PMoA cons	MoA vowel	−0.338	0.185	−1.832	0.279
	PoA vowel	−0.702	0.185	−3.802	0.002**
	Stress	−0.202	0.185	−1.093	0.831
MoA vowel	PoA vowel	−0.364	0.185	−1.97	0.257
	Stress	0.136	0.185	0.739	0.923
PoA vowel	Stress	0.5	0.185	2.709	0.055

P-value adjusted for comparing a family of 10. ** $p < .01$, *** $p < .001$

were engaged in musical activities, showed a more positive MMN amplitude.

Discussion

In the present study, we investigated the pre-attentive processes related to perceiving speech information, i.e., using Ádám's terms, we studied the input unconscious processes of speech perception (Ádám 2001). We created a speech multi-feature MMN paradigm that allowed the fast investigation of the neural mechanism underlying multiple segmental and prosodic linguistic features. We also studied whether individual differences of participants related to their musical and foreign language proficiency modulates the MMN response. In the experiment, participants were presented with five different kinds of deviant stimuli: two consonant deviants with place of articulation (peke, PoA), or place and mode of articulation (feke, PMoA) contrasts; two vowel deviants with mode of articulation (kike, MoA) or place of

Table 5 Correlation between MMN latencies and participant's musical background and foreign language proficiency

MMN latency	Music education years	Music education	Current musical activity	Foreign language proficiency
PoA consonant	0.19	−0.02	0.17	0.13
PMoA consonant	−0.38	−0.23	−0.23	−0.04
MoA vowel	−0.15	−0.04	−0.14	−0.20
PoA vowel	0.03	−0.16	−0.24	0.26
Stress	−0.06	−0.15	−0.03	−0.46*

All data present Spearman's rho values. Values in bold indicate correlations that survived FDR correction. * $p < .05$

Table 6 Correlation between MMN amplitudes and participant's musical background and foreign language proficiency

MMN latency	Music education years	Music education	Current musical activity	Foreign language proficiency
PoA consonant	0.19	−0.19	0.07	−0.02
PMoA consonant	0.58*	0.13	0.32	0.25
MoA vowel	0.18	0.14	0.33	0.05
PoA vowel	−0.18	0.41*	0.40*	0.12
Stress	0.39	0.18	0.33	0.12

All data present Spearman's rho values. Values in bold indicate correlations that survived FRD correction.

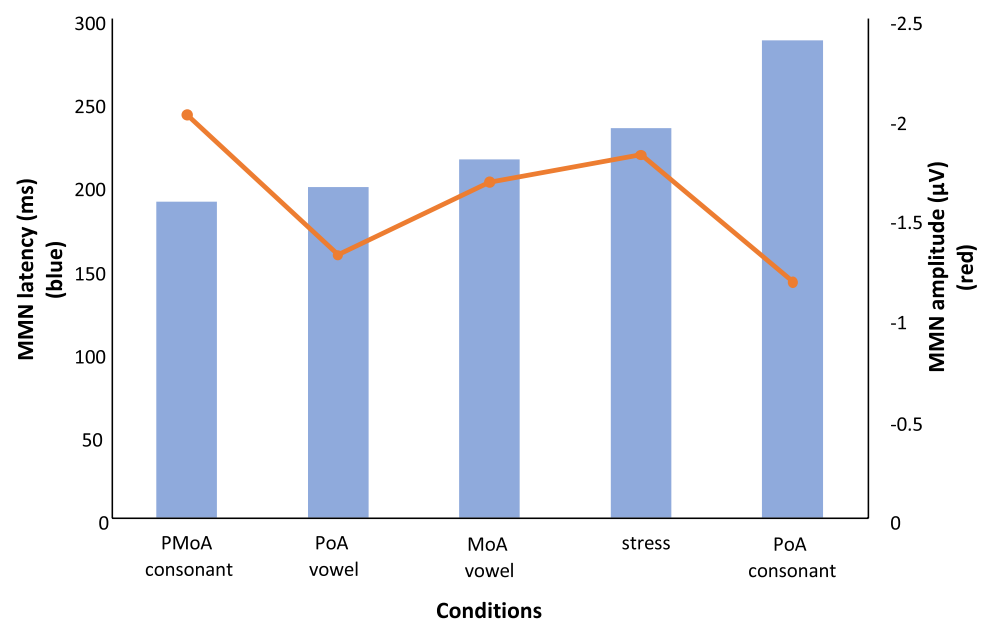
* $p < .05$

articulation (koke, PoA) contrasts, and a stress deviant. The results showed that all deviants elicited the MMN ERP component. This corroborates the findings of previous speech multi-feature studies (Fisher et al. 2011; Kuuluvainen et al. 2014; Lovio et al. 2009; Pakarinen et al. 2009, 2013; Sorokin et al. 2010) that also showed reliable MMN for different acoustic and phonetic contrasts. We also found an MMN to the word stress contrast, which confirms the results of Zavogianni et al. (2024). As the ERP correlates of word stress processing were mainly investigated in oddball paradigms (Honbolygó et al. 2004, 2020; Honbolygó and Csépe 2013; Weber et al. 2004; Ylinen et al. 2009; Zora et al. 2015, 2016, 2023), this finding supports that the multi-feature paradigm allows to investigate the processing of prosodic contrasts. The investigation of prosodic contrast is important, because in some developmental speech and language disorders (e.g., developmental dyslexia), apart from other speech features, the processing of word stress seems to be impaired (Schaadt and Männel 2019).

We also found that the segmental and prosodic deviants induced different MMN peak latencies and amplitudes. The

analysis of MMN peak latencies showed that the PoA consonant deviant elicited the MMN with the longest latency that significantly differed from all other latencies (see Fig. 6.). The PMoA consonant deviant elicited the shortest MMN latency, differing significantly from all, but the PoA vowel deviant. The analysis of MMN peak amplitudes revealed that the PMoA consonant deviant elicited the largest MMN that differed significantly from the PoA consonant and PoA vowel deviants, and the PoA consonant deviant elicited the smallest MMN that differed significantly from the PoA consonant and the stress deviants. These differences between the amplitude and latency values of the MMN component elicited by the different deviants show that the multi-feature paradigm is adequately sensitive to reveal stimulus related processing differences.

Concerning the interpretation of MMN differences, Näätänen et al. (2007) suggested that the MMN amplitude gets larger and the peak latency shorter with the increasing *magnitude of stimulus deviation*. This effect has been shown for several different stimulus features: frequency, intensity, duration, spectral complexity, spatial locus (for review, see

Fig. 6 Summary of latency and amplitude values of MMNs in the five conditions. Latency: left axis, blue bars; amplitude: right axis, red curve

Näätänen et al. 2007). Moreover, the MMN shows an *additivity effect*, being larger for a conjunction of two features, as compared to one feature (Takegata et al. 1999).

Studies using the multi-feature MMN paradigm also found stimulus related differences in the MMN peak latency and amplitude. Pakarinen et al. (2007) investigated the processing of four different types of deviants (frequency, intensity, duration, spatial location) in non-speech tones, and changed the magnitude of the deviation in six steps. The authors found that the MMN peak amplitude increased with increasing stimulus deviance for all deviants, but the MMN latency followed the magnitude of change only in the case of frequency and location deviations. The MMN amplitude change corresponded to the improvement of behavioral discrimination performance. Pakarinen et al., (2013) investigated the processing of small, medium and large deviations in the case of vowels by changing the vowel identity, intensity, duration and pitch, and found that the increase in the magnitude of deviation induced an increase in the amplitude, and a decrease in the latency of the MMN peak. Furthermore, Pakarinen et al. (2009) recorded speech and non-speech deviances in a multi-feature MMN paradigm presenting syllables, and found notable MMN peak amplitude and latency differences in stimuli: the MMN was the smallest and latest for the consonant change, and the largest for the vowel change. This pattern of MMN amplitude difference followed the behavioral discrimination performance, as participants had a low performance in the consonant change condition and a high performance in the vowel change condition.

In our study, the POA consonant deviant elicited the smallest and latest MMN peak, and the PMOA consonant elicited the largest and earliest MMN peak, while the vowels and stress deviants elicited MMNs in between (see Fig. 6). Pakarinen et al. (2009) used a place of articulation difference for the consonant deviant (/p/ vs. /t/) which is similar to our PoA consonant deviant (/p/ vs. /k/), suggesting that this contrast is difficult to discriminate in both languages. In our case, the other consonant deviant (PMoA) involving both a mode and place of articulation change, elicited the largest and earliest MMN peak, suggesting an easier discrimination when two consonant features are changed compared to the standard. These results support the assumption that the MMN follows the magnitude of stimulus deviance.

Alternatively, the additivity effect (Takegata et al. 1999) could also explain these findings: the MMN was larger for the PMoA consonant deviant involving the conjunction of two phonetic features, as compared to the PoA consonant deviant, involving a single phonetic feature. In any case, the results suggest that the MMN amplitude and latency might reflect how much the deviant differs from the standard in phonetic terms. However, we did not find a similar gradient in the case of vowel contrasts: this might be explained by

the fact that in the case of vowel contrast, there was a single feature difference, and no conjunction of several features. The reason for this is that there are no feature conjunctions in Hungarian vowel contrast: vowels can differ from each other either in the place of articulation, or in the mode of articulation, but not both.

In the Introduction, we referred to the AERS as a more complex model explaining the emergence of the MMN component (Winkler and Schröger 2015). Concerning the processing of speech information, the AERS model assumes that the auditory system extracts basic speech-related features from the incoming sounds, such as voicing, formant transitions, etc., and later combines these into sensory memory representations, such as phonemes and prosodic features. These representations are compared to the model's predictions that are based on the long-term representations of different linguistic units: native speech sounds, prosodic patterns, words, etc. We think that the assumptions of AERS are in accordance with the above interpretation of MMN differences related to the magnitude of stimulus deviance, or the additivity effect: the deviant speech features are compared against the standard speech features (relying on both the actual sensory information and the top-down representation of the features) and if the deviant and the standard differ from each other in more features (or if the magnitude of the deviance is larger) this leads to an enhanced MMN. The present study did not allow the systematic investigation of a wide range of different phonetic contrasts, which necessitates further studies to establish how the MMN reflects different kinds of phonetic contrasts.

We also found individual differences in the MMN peak latencies and amplitudes of the different stimuli: participants who were more advanced speakers of a foreign language had an earlier MMN peak to the stress deviant; moreover, participants who had more years of music education, showed a smaller MMN amplitude to the PMOA consonant deviant; furthermore, participants who had a formal music education, and participants who were currently participating in musical activities showed also a smaller MMN to the POA vowel deviant. These results contradicted somewhat our expectation that participants with more music training would show enhanced MMN. Based on previous studies, we assumed that apart from its sensitivity to stimulus features, the MMN also reflects discrimination accuracy (Näätänen et al. 2007), and we expected that participants with a musical training have a better acoustic discrimination. Lang et al. (1990) and Tervaniemi et al. (1997) found that the MMN amplitude in a pitch and temporal information processing task was larger for those participants who demonstrated a better behavioral discrimination accuracy. Amenedo and Escara (2000) showed that the MMN amplitude elicited by the duration changes of spectrally rich tones predicted the behavioural detection of these changes. Kujala et al. (2001)

investigated the discrimination of sound pairs with different silent inter-pair intervals and found that the appearance of the MMN was related to whether participants could behaviorally discriminate the sound pairs. Aaltonen et al. (1994) found a correlation between the MMN amplitude and the participant's behavioral performance of discriminating small, within-category changes in vowels. At the same time, Bazana and Stelmack (2002) could not demonstrate a correlation between performance accuracy and MMN measures in a tonal frequency discrimination task.

In our study, we found that participants with a more advanced knowledge of any foreign language (we did not take into account what exactly the language was) had an earlier MMN peak to the stress deviant. We might assume that more experience with a second language increased the sensitivity of participants to prosodic (stress) cues. However, this contradicts previous results on the stress processing deficit of speakers of fixed-stressed languages (Domahs et al. 2012; Honbolygó et al. 2019; Peperkamp and Dupoux 2002).

Furthermore, studies investigating auditory processing in musicians consistently found more enhanced neural responses to different kinds of auditory deviances. Tervaniemi et al. (1997) found a larger MMN amplitude in musicians than non-musicians in a paradigm involving temporal changes. Nikjeh et al. (2009) investigated musicians and non-musicians in a paradigm involving pure tones, harmonic tone complexes, and speech sounds, and found shorter MMN latency in the case of musicians. Marie et al. (2012) showed that musicians showed a larger MMN to duration and frequency changes than non-musicians. Using the multi-feature paradigm, Tervaniemi et al. (2016) not only demonstrated that musicians had an enhanced processing of tuning, timbre, rhythm, melody transpositions, and melody contour, but also that musicians with different training history (i.e., classical and jazz musicians) show a specific auditory processing profile reflecting the importance of the sound features in the different music genres. Kühnis et al. (2013) applied a multi-feature paradigm to investigate the speech processing in trained musicians. Participants were presented with a CV syllable as standard and four deviants including a pitch, duration, vowel, and consonant change. The authors found increased MMN amplitude for musicians in the processing of vowel and duration changes and concluded that professional musicians demonstrated a general improvement in auditory processing as compared to non-musicians.

In our study, we found that participants with more musical training showed a smaller MMN in one of the consonant (PMoA) and one of the vowel (PoA) deviants. This clearly contradicts the findings of previous studies. An important caveat is that our study was not designed to show music training related differences, and we didn't assess the musical background of the participants carefully. Moreover, we used a continuous measure of years of training, instead of

a dichotomous measure to create groups of musicians and non-musicians as in previous studies. Still, it might be interesting to consider that participants with more music training showed a smaller MMN amplitude in the case of the PMoA consonant deviant, which was otherwise presumably the easiest to discriminate. We can speculate that participants with more music training were better at predicting the deviant because of their enhanced auditory perception skills, therefore they showed a smaller MMN to the deviant (as it was not as surprising). However, this does not explain why they also showed a smaller MMN in the case of the PoA vowel deviant, which was not specifically easy or difficult to discriminate. Irrespective of the precise explanation, our findings show that the MMN to multiple segmental and prosodic contrasts in the current paradigm can reflect individual differences. Further studies are needed to determine the effect of musical expertise and second language proficiency on the processing of speech contrasts.

In conclusion, the speech multi-feature paradigm we created to profile the neural background of processing different speech features was adequately sensitive to reveal differences in the MMN component's amplitude and latency in the different deviant conditions. The MMN component reflected the magnitude of stimulus deviance and was elicited by both segmental (vowel and consonant change) and prosodic (word stress) features. We also found individual differences in MMN amplitudes and latencies according to the second language and musical background of participants, although these differences did not necessarily coincide with the findings of the literature.

Although several studies have shown that the MMN component's magnitude correlates with the behavioral performance of the participants, it is important to emphasize that the neural mechanism involved in the MMN's generation are largely pre-attentive. Thus, as Näätänen et al. (2012, p. 443) suggests “the MMN provides a unique window to the neuropsychology of central auditory processing, and hence a possibility for the objective assessment of auditory discrimination”. We think that this resonates well with György Ádám's suggestion about the importance and prevalence of “input unconscious processes” in the human brain.

Acknowledgements This research was supported by the Research Programme for Public Education Development, Hungarian Academy of Sciences (SZKF-20/2021; PI: FH), the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant agreement No.813546, and the János Bolyai Research Scholarship of the Hungarian Academy of Sciences [BO/00771/21] awarded to FH. The authors are thankful for Dóra Bárány, Minka Petánszki, Marcell Püski, and Khakim Zulfikri for their help in data collection.

Funding Open access funding provided by HUN-REN Research Centre for Natural Sciences.

Declarations

Conflict of interest The authors declare no conflict of interest with the organizations founding the research.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aaltonen O, Eerola O, Lang AH, Uusipaikka E, Tuomainen J (1994) Automatic discrimination of phonetically relevant and irrelevant vowel parameters as reflected by mismatch negativity. *J Acoust Soc Am* 96(3):1489–1493
- Ádám G (1998) Visceral perception: understanding internal cognition. Springer Science & Business Media, Berlin
- Ádám, G. (2001). A tudattalan reneszánsza. *Magyar Tudomány*, 10.
- Allen M, Poggiali D, Whitaker K, Langen Van J, Kievit RA, Marshall TR (2023) Raincloud plots : a multi-platform tool for robust data visualization [version 2 ; peer review : 2 approved]. 1–52
- Amenedo E, Escera C (2000) The accuracy of sound duration representation in the human brain determines the accuracy of behavioural perception. *Eur J Neurosci* 12(7):2570–2574. <https://doi.org/10.1046/j.1460-9568.2000.00114.x>
- Bazana PG, Stelmack RM (2002) Intelligence and information processing during an auditory discrimination task with backward masking: an event-related potential analysis. *J Pers Soc Psychol* 83(4):998–1008. <https://doi.org/10.1037/0022-3514.83.4.998>
- Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Cheour M, Ceponiene R, Lehtokoski A, Luuk A, Allik J, Alho K, Näätänen R (1998) Development of language-specific phoneme representations in the infant brain. *Nat Neurosci* 1(5):351–353. <https://doi.org/10.1038/1561>
- Dehaene-Lambertz G (1997) Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport* 8(4):919–924. <https://doi.org/10.1097/00001756-199703030-00021>
- Delorme A, Sejnowski T, Makeig S (2007) Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *NeuroImage*, 34(4):1443–1449. <https://doi.org/10.1016/j.neuroimage.2006.11.004>
- Domahs U, Knaus J, Orzechowska P, Wiese R (2012) Stress “deafness” in a language with fixed word stress: an ERP study on polish. *Front Psychol* 3:1–15. <https://doi.org/10.3389/fpsyg.2012.00439>
- Fisher DJ, Grant B, Smith DM, Knott VJ (2011) Effects of deviant probability on the “optimal” multi-feature mismatch negativity (MMN) paradigm. *Int J Psychophysiol* 79(2):311–315. <https://doi.org/10.1016/j.ijpsycho.2010.11.006>
- Friston K (2010) The free-energy principle: a unified brain theory? *Nat Rev Neurosci* 11(2):127–138. <https://doi.org/10.1038/nrn2787>
- Friston K (2005) A theory of cortical responses. *Philos Trans Royal Soc Lond Series B Biol Sci* 360(1456):815–836. <https://doi.org/10.1098/rstb.2005.1622>
- Honbolygó F, Csépe V (2013) Saliency or template? ERP evidence for long-term representation of word stress. *Int J Psychophysiol* 87(2):165–172. <https://doi.org/10.1016/j.ijpsycho.2012.12.005>
- Honbolygó F, Csépe V, Ragó A (2004) Suprasegmental speech cues are automatically processed by the human brain: a mismatch negativity study. *Neurosci Lett* 363(1):84–88. <https://doi.org/10.1016/j.neulet.2004.03.057>
- Honbolygó F, Kóbor A, Csépe V (2019) Cognitive components of foreign word stress processing difficulty in speakers of a native language with non-contrastive stress. *Int J Biling* 23:366–380. <https://doi.org/10.1177/1367006917728393>
- Honbolygó F, Kolozsvári O, Csépe V (2017) Processing of word stress related acoustic information: a multi-feature MMN study. *Int J Psychophysiol* 118:9–17. <https://doi.org/10.1016/j.ijpsycho.2017.05.009>
- Honbolygó F, Kóbor A, German B, Csépe V (2020) Word stress representations are language-specific: evidence from event-related brain potentials. *Psychophysiology* 57(5):e13541. <https://doi.org/10.1111/psyp.13541>
- Kostilainen K, Partanen E, Mikkola K, Wikström V, Pakarinen S, Fellman V, Huotilainen M (2020) Neural processing of changes in phonetic and emotional speech sounds and tones in preterm infants at term age. *Int J Psychophysiol* 148:111–118. <https://doi.org/10.1016/j.ijpsycho.2019.10.009>
- Kühnis J, Elmer S, Meyer M, Jäncke L (2013) The encoding of vowels and temporal speech cues in the auditory cortex of professional musicians: an EEG study. *Neuropsychologia* 51(8):1608–1618. <https://doi.org/10.1016/j.neuropsychologia.2013.04.007>
- Kujala T, Kallio J, Tervaniemi M, Näätänen R (2001) The mismatch negativity as an index of temporal processing in audition. *Clin Neurophysiol* 112(9):1712–1719. [https://doi.org/10.1016/S1388-2457\(01\)00625-3](https://doi.org/10.1016/S1388-2457(01)00625-3)
- Kuuluvainen S, Nevalainen P, Sorokin A, Mittag M, Partanen E, Putkinen V, Seppänen M, Kähkönen S, Kujala T (2014) The neural basis of sublexical speech and corresponding nonspeech processing: a combined EEG–MEG study. *Brain Lang* 130:19–32. <https://doi.org/10.1016/j.bandl.2014.01.008>
- Lang H, Nyrke T, Ek M, Aaltonen O, Raimo I, Näätänen R (1990) Pitch discrimination performance and auditory event-related potentials. *Psychophysiol Brain Res* 1:294–298
- Lovio R, Pakarinen S, Huotilainen M, Alku P, Silvennoinen S, Näätänen R, Kujala T (2009) Auditory discrimination profiles of speech sound changes in 6-year-old children as determined with the multi-feature MMN paradigm. *Clin Neurophysiol* 120(5):916–921. <https://doi.org/10.1016/j.clinph.2009.03.010>
- Marie C, Kujala T, Besson M (2012) Musical and linguistic expertise influence pre-attentive and attentive processing of non-speech sounds. *Cortex* 48(4):447–457. <https://doi.org/10.1016/j.cortex.2010.11.006>
- 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):1–21. <https://doi.org/10.1111/1469-8986.3810001>
- Näätänen R, Lehtokoski A, Lennes M, Cheour M, Huotilainen M, Iivonen A, Vainio M, Alku P, Ilmoniemi RJ, Luuk A, Allik J, Sinkkonen J, Alho K (1997) Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385(6615):432–434. <https://doi.org/10.1038/385432a0>
- Näätänen R, Pakarinen S, Rinne T, Takegata R (2004) The mismatch negativity (MMN): towards the optimal paradigm. *Clin*

- Neurophysiol 115(1):140–144. <https://doi.org/10.1016/j.clinph.2003.04.001>
- 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(12):2544–2590. <https://doi.org/10.1016/j.clinph.2007.04.026>
- Näätänen R, Kujala T, Escera C, Baldeweg T, Kreegipuu K, Carlson S, Ponton C (2012) The mismatch negativity (MMN)-a unique window to disturbed central auditory processing in ageing and different clinical conditions. *Clin Neurophysiol* 123(3):424–458. <https://doi.org/10.1016/j.clinph.2011.09.020>
- Nikjeh DA, Lister JJ, Frisch SA (2009) Preattentive cortical-evoked responses to pure tones, harmonic tones, and speech: influence of music training. *Ear Hear* 30(4):432–446. <https://doi.org/10.1097/AUD.0b013e3181a61bf2>
- Pakarinen S, Takegata R, Rinne T, Huottilainen M, Näätänen R (2007) Measurement of extensive auditory discrimination profiles using the mismatch negativity (MMN) of the auditory event-related potential (ERP). *Clin Neurophysiol* 118(1):177–185
- Pakarinen S, Lovio R, Huottilainen M, Alku P, Näätänen R, Kujala T (2009) Fast multi-feature paradigm for recording several mismatch negativities (MMNs) to phonetic and acoustic changes in speech sounds. *Biol Psychol* 82(3):219–226. <https://doi.org/10.1016/j.biopsycho.2009.07.008>
- Pakarinen S, Sokka L, Leinikka M, Henelius A, Korpela J, Huottilainen M (2014) Fast determination of MMN and P3a responses to linguistically and emotionally relevant changes in pseudoword stimuli. *Neurosci Lett* 577:28–33. <https://doi.org/10.1016/j.neulet.2014.06.004>
- Pakarinen S, Teinonen T, Shestakova A, Kwon MS, Kujala T, Hämäläinen H, Näätänen R, Huottilainen M (2013) Fast parametric evaluation of central speech-sound processing with mismatch negativity (MMN). *Int J Psychophysiol* 87(1):103–110. <https://doi.org/10.1016/j.ijpsycho.2012.11.010>
- Partanent E, Vainio M, Kujala T, Huottilainen M (2011) Linguistic multifeature MMN paradigm for extensive recording of auditory discrimination profiles. *Psychophysiology* 48(10):1372–1380. <https://doi.org/10.1111/j.1469-8986.2011.01214.x>
- Peperkamp S, Dupoux E (2002) A typological study of stress deafness. *Lab Phonol* 7:203–240
- Port R (2007) How are words stored in memory? Beyond phones and phonemes. *New Ideas Psychol* 25(2):145–172. <https://doi.org/10.1016/j.newideapsych.2007.02.001>
- Schaadt G, Männel C (2019) Phonemes, words, and phrases: tracking phonological processing in pre-schoolers developing dyslexia. *Clin Neurophysiol* 130(8):1329–1341. <https://doi.org/10.1016/j.clinph.2019.05.018>
- Shahin AJ (2011) Neurophysiological influence of musical training on speech perception. *Front Psychol* 2:1–10. <https://doi.org/10.3389/fpsyg.2011.00126>
- Sorokin A, Alku P, Kujala T (2010) Change and novelty detection in speech and non-speech sound streams. *Brain Res* 1327:77–90. <https://doi.org/10.1016/j.brainres.2010.02.052>
- Steinhauer K (2008) Event-related potentials in the study of language. In: Stemmer B, Whitaker HA (eds) *Handbook of the neuroscience of language*. Elsevier, New York, pp 91–104
- Takegata R, Paavilainen P, Näätänen R, Winkler I (1999) Independent processing of changes in auditory single features and feature conjunctions in humans as indexed by the mismatch negativity. *Neurosci Lett* 266(2):109–112. [https://doi.org/10.1016/S0304-3940\(99\)00267-0](https://doi.org/10.1016/S0304-3940(99)00267-0)
- Tervaniemi M, Ilvonen T, Karma K, Alho K, Näätänen R (1997) The musical brain: brain waves reveal the neurophysiological basis of musicality in human subjects. *Neurosci Lett* 226(1):1–4
- Tervaniemi M, Janhunen L, Kruck S, Putkinen V, Huottilainen M (2016) Auditory profiles of classical, jazz, and rock musicians: genre-specific sensitivity to musical sound features. *Front Psychol* 6:1–11. <https://doi.org/10.3389/fpsyg.2015.01900>
- Weber C, Hahne A, Friedrich M, Friederici AD (2004) Discrimination of word stress in early infant perception: electrophysiological evidence. *Cogn Brain Res* 18(2):149–161
- Winkler I, Schröger E (2015) Auditory perceptual objects as generative models: setting the stage for communication by sound. *Brain Lang* 148:1–22. <https://doi.org/10.1016/j.bandl.2015.05.003>
- Winkler I, Lehtokoski A, Alku P, Vainio M, Czigler I, Csépe V, Aaltonen O, Raimo I, Alho K, Lang H, Ilvonen A, Näätänen R (1999) Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cogn Brain Res* 7(3):357–369
- Winkler I, Denham SL, Nelken I (2009) Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends Cogn Sci* 13(12):532–540. <https://doi.org/10.1016/j.tics.2009.09.003>
- Ylinen S, Shestakova A, Huottilainen M, Alku P, Näätänen R (2006) Mismatch negativity (MMN) elicited by changes in phoneme length: a cross-linguistic study. *Brain Res* 1072(1):175–185. <https://doi.org/10.1016/j.brainres.2005.12.004>
- Ylinen S, Strelnikov K, Huottilainen 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. <https://doi.org/10.1016/j.ijpsycho.2009.05.013>
- Zavogianni MI, Honbolygó F, Kelic M (2024) Profiling the central auditory processing of speech sounds: a MMN multifeature study. In: Prep
- Zora H, Schwarz I-C, Heldner M (2015) Neural correlates of lexical stress. *NeuroReport* 26:791–796. <https://doi.org/10.1097/WNR.0000000000000426>
- Zora H, Heldner M, Schwarz IC (2016) Perceptual correlates of turkish word stress and their contribution to automatic lexical access: evidence from early ERP components. *Front Neurosci* 10:1–11. <https://doi.org/10.3389/fnins.2016.00007>
- Zora H, Wester J, Csépe V (2023) Predictions about prosody facilitate lexical access: evidence from P50/N100 and MMN components. *Int J Psychophysiol* 194:112262. <https://doi.org/10.1016/j.ijpsycho.2023.112262>