*Studies in Second Language Acquisition*, **44** (2022), 1260–1278
doi:10.1017/S0272263121000978

# *Research Article*

## DISCRIMINABILITY AND PROTOTYPICALITY OF NONNATIVE VOWELS

***Yasuaki Shinohara*** *

*Faculty of Commerce, Waseda University, Tokyo, Japan*

***Chao Han***

*Department of Linguistics and Cognitive Science, University of Delaware, DE, United States*

***Arild Hestvik***

*Department of Linguistics and Cognitive Science, University of Delaware, DE, United States*

**Abstract**

This study examined how discriminability and prototypicality of nonnative phones modulate the amplitude of the Mismatch Negativity (MMN) event-related brain potential. We hypothesized that if a frequently occurring (standard) stimulus is not prototypical to a listener, a weaker predictive memory trace will be formed and a smaller MMN will be generated for a phonetic deviant, regardless of the discriminability between the standard and deviant stimuli. The MMN amplitudes of Japanese speakers hearing the English vowels /æ/ and /ɑ/ as standard stimuli and /ʌ/ as a deviant stimulus in an oddball paradigm were measured. Although the English /æ/-/ʌ/ contrast was more discriminable than the English /ɑ/-/ʌ/ contrast for Japanese speakers, when Japanese speakers heard the /æ/ standard stimulus (i.e., less prototypical as Japanese /a/) and the /ʌ/ deviant stimulus, their MMN amplitude was smaller than the one elicited when they heard /ɑ/ as a standard stimulus

This work was supported by JSPS KAKENHI Grant Nos. 16K16884 and 19K13169 and Waseda University Grants for Special Research Projects Nos. 2019E-030, 2018K-159, and 2017K-221. We are grateful for the kind support of Prof. Hiromu Sakai, who generously allowed us to use his EEG equipment. We also thank our research assistants, Mr. Yu Nakajima, Ms. Manami Oya, Ms. Kana Seki (Waseda University), and Ms. Qing Xu (University of Delaware), who helped us collect data and recruit participants for our experiments.

* Corresponding author. Email: y.shinohara@waseda.jp

(i.e., more prototypical as Japanese /a/) and /ʌ/ as a deviant stimulus. The prototypicality of the standard stimuli in listeners' phonological representations modulates the MMN amplitude more robustly than does the discriminability between standard and deviant stimuli.

## INTRODUCTION

Mismatch Negativity (MMN), a component of auditory Event-Related Potentials (ERP), indicates sound-change detection. In a typical oddball paradigm, where listeners perceive a series of "standard" stimuli, the brain generates a short-term memory trace of the stimuli and uses it to predict upcoming ones. If listeners hear a sound that is different from their prediction (i.e., a sound change from a standard to a deviant stimulus), a prediction error occurs, as reflected by the MMN (Luck, 2005; Näätänen et al., 2019). The MMN amplitude is affected by the frequency of standard and deviant stimuli that listeners hear in the experiment (Garrido et al., 2009; Imada et al., 1993; Javitt et al., 1998; May et al., 1999; Näätänen, 1992; Sams et al., 1983), the acoustic difference between standards and deviants (Lang et al., 1990; Näätänen et al., 1993; Tervaniemi et al., 1997), prototypicality of standard and deviant stimuli (e.g., native vs. nonnative phones: Grimaldi et al., 2014; Näätänen et al., 1997; Peltola et al., 2003; allophonic variants: Bühler et al., 2017), and discriminability between standards and deviants (Dehaene-Lambertz & Baillet, 1998; Lovio et al., 2009; Näätänen, 2001; Näätänen et al., 1997; Pakarinen et al., 2009; Rinker et al., 2010; Yang Zhang et al., 2005). We address how the MMN is modulated according to the conscious discriminability between standard and deviant stimuli and the prototypicality of standard and deviant stimuli in listeners' native languages.

One factor that affects the MMN amplitude is conscious discriminability between standard and deviant stimuli (Amenedo & Escera, 2000; Lang et al., 1990; Näätänen, 2001; Näätänen et al., 2007; Näätänen & Alho, 1997; Pakarinen et al., 2007). Conscious discriminability refers to whether listeners can behaviorally detect an auditory difference between stimuli; for example, people can often discriminate native vowels better than nonnative ones. For the perception of nonnative phones, listeners tend to assimilate unfamiliar nonnative phones into the most articulatorily similar native-language (L1) phonemes. According to the Perceptual Assimilation Model (PAM; Best, 1994a, 1994b, 1995), when two nonnative phones are assimilated into two different L1 phonemes (Two Category assimilation), discrimination of the two nonnative phones is excellent. When two nonnative phones are assimilated into a single L1 phoneme and are equally good exemplars of it (Single Category assimilation), discrimination is poor. Finally, when two nonnative phones are assimilated into a single L1 phoneme, but one is a better exemplar than the other (Category-Goodness difference assimilation), discrimination is moderate to good. Previous studies have shown that the behavioral discriminability predicted by those assimilation patterns is correlated with the MMN amplitude (Grimaldi et al., 2014; Näätänen et al., 1997; Peltola et al., 2003). Nonnative phonetic contrasts belonging to the Category-Goodness difference assimilation type are relatively easy to discriminate and elicit a larger MMN when compared to the less discriminable contrasts belonging to the Single Category assimilation type (Grimaldi et al., 2014).

Discriminability and prototypicality of speech sounds are often correlated. For example, two separate native phonemes are more prototypical (i.e., frequently used) and discriminable than are two nonnative phones (e.g., less prototypical allophonic variants

https://doi.org/10.1017/S0272263121000978 Published online by Cambridge University Press

in a native phoneme). However, under the standard-deviant oddball paradigm where MMN is elicited, the prototypicality of standard stimuli affects the MMN amplitude regardless of the discriminability between standard and deviant stimuli. When listeners hear a series of standard stimuli, they generate a short-term memory trace of them and predict upcoming ones. As this involves mapping the standards to a listener's phonological category (Näätänen et al., 1997; Phillips et al., 2000), their phonological categories and the phonetic representations in their long-term memory affect generating the short-term memory trace of the standard stimuli (Näätänen et al., 2007; Shafer et al., 2021). Shafer et al. (2004) found that a larger MMN is elicited when people hear a more prototypical consonant as standard and a less prototypical consonant as deviant, compared to the reversed order. Regarding vowels, Shafer et al. (2021) revealed that a larger MMN is elicited when the standard stimulus is a vowel sharing common phonetic features of a native vowel and the deviant stimulus is a vowel having less common phonetic features, compared to the reversed order. Shafer et al. (2004, 2021) described these results as a potential reflection of the strength of the short-term memory traces constructed for standard stimuli based on long-term phonological representations of listeners. That is, prototypical phones used as standard stimuli lead to more stable memory traces that generate more precise predictions.

Based on the theoretical explanation of MMN (Näätänen et al., 2007) and the prototypicality effects (Shafer et al., 2004, 2021), we claim that the prototypicality of a standard stimulus modulates the MMN amplitude more robustly than does the discriminability of standard and deviant stimuli. If standard stimuli are good exemplars of the listener's L1 phoneme, a more stable short-term memory trace will be formed, thus making more precise predictions about upcoming stimuli. As a result of this neural processing, when listeners perceive a deviant stimulus, a large prediction error occurs, leading to a large MMN. In contrast, if standard stimuli are not prototypical to listeners, the process of generating a short-term memory trace will be less certain, leading to less precise predictions about upcoming ones, resulting in a smaller MMN (Jacobsen, 2004; Peltola et al., 2003; see predictive coding: Baldeweg, 2006; Friston, 2002, 2005; Garrido et al., 2009; Wacongne et al., 2012; Winkler & Czigler, 2012). We hypothesized that this prototypicality effect happens even when both standard and deviant stimuli are categorized as a single native phoneme.

This study examined which of the two, namely the standard-stimulus prototypicality and discriminability between standards and deviants, modulates the MMN responses more. The Japanese speakers' English vowel perception was tested in a previous behavioral study (Shinohara et al., 2019), and their MMN responses were measured in this study. Table 1 presents the predictions from the two opposing accounts. Although the English /æ/, /ʌ/, and /ɑ/ are all categorized as a single Japanese /a/ phoneme, it is more difficult for Japanese speakers to discriminate between the English /ɑ/ and /ʌ/ stimuli than between the English /æ/ and /ʌ/, because the former are equally good exemplars of the Japanese /a/ (English /ɑ/-/ʌ/: Single Category assimilation), and the English /æ/ is a worse exemplar of the Japanese /a/ than is the English /ʌ/ (English /æ/-/ʌ/: Category-Goodness difference assimilation) (Lengeris, 2009; Shinohara et al., 2019; Strange et al., 1998). The discriminability account predicts that a larger MMN response would be observed when the English /æ/ is used as a standard and the English /ʌ/ is a deviant, compared to when the English /ɑ/ is used as a standard and the English /ʌ/ is a deviant. In contrast, according to

https://doi.org/10.1017/S0272263121000978 Published online by Cambridge University Press

TABLE 1. Predictions of the MMN indices for Japanese speakers perceiving English vowels, based on the two accounts (i.e., discriminability and prototypicality)

| Two factors affecting MMN responses | Standard | Deviant | MMN amplitude |
|---|---|---|---|
| Discriminability | /æ/ | /ʌ/ | Larger |
| | /ɑ/ | /ʌ/ | Smaller |
| Prototypicality | /æ/ | /ʌ/ | Smaller |
| | /ɑ/ | /ʌ/ | Larger |

the prototypicality effect, when Japanese speakers hear the English /ɑ/ as standard and the English /ʌ/ as deviant, a larger MMN is expected. Under the condition in which Japanese speakers hear a series of English /ɑ/ as standard stimuli, given its prototypicality as the Japanese /a/ (Lengeris, 2009; Shinohara et al., 2019; Strange et al., 1998), they easily form a short-term memory trace while hearing standards, which results in a stronger prediction. Thus, a deviant stimulus /ʌ/ would elicit a larger MMN. However, if they hear the English /æ/ as standard, owing to the poor fit to the Japanese /a/ (Lengeris, 2009; Shinohara et al., 2019; Strange et al., 1998), a less robust memory trace will be generated. A smaller MMN amplitude is predicted for the /æ/-/ʌ/ contrast by the prototypicality account.

When examining the discriminability of natural recordings with MMN responses, care must be taken to control for confounding acoustic difference. The discrimination accuracy of the /æ/-/ʌ/ contrast is higher than that of the /ɑ/-/ʌ/ contrast for Japanese speakers (Lengeris, 2009; Strange et al., 1998), but this result is often attributed to both perceptual assimilation and acoustic distance. The English /æ/ and /ʌ/ are acoustically more different from each other than the /ɑ/-/ʌ/ contrast (Hillenbrand et al., 1995). In this study, we controlled for this confound by using resynthesized stimuli of English /æ/, /ʌ/, and /ɑ/. This study used two English vowel contrasts (/æ/-/ʌ/, /ɑ/-/ʌ/) with the acoustic distances equalized (see the "Method" section for details) and measured the MMN amplitudes of the /æ/-/ʌ/ and /ɑ/-/ʌ/ contrasts in Japanese speakers. We claim that as the MMN is elicited by a regularity violation, the prototypicality of standard stimuli that generate the regularity (i.e., short-term memory trace) was predicted to be a factor modulating the MMN amplitude more than the discriminability between standard and deviant stimuli. Native English speakers were also recruited as a control group. The two MMN amplitudes should be about the same for English speakers, who have three separate phonological categories for /æ/, /ʌ/, and /ɑ/. The results describe how the MMN mechanism is phonetically driven and present the factors that need to be considered when measuring the MMN amplitude.

## METHOD

### *PARTICIPANTS*

This study was approved by the ethics review boards at Waseda University (Tokyo, Japan) and University of Delaware (Delaware, US); all participants signed informed consent forms. A total of 56 people participated in the electroencephalography (EEG) recording sessions at two different laboratories. All participants at the University of Delaware were native monolingual speakers of American English and all participants at Waseda University were native monolingual speakers of Japanese. Participants reported

https://doi.org/10.1017/S0272263121000978 Published online by Cambridge University Press

TABLE 2. Participants' information

| Native language | Number of subjects (female, male) | Age range in years (mean, median, *SD*) | Handedness (number) |
|---|---|---|---|
| English | 26 (25, 1) | female 18–27 (20.7, 21, 1.7) male 29–29 (29, 29, 29) | right (24), mixed (0), left (2) |
| Japanese | 30 (16, 14) | female 19–24 (20, 19, 1.6) male 19–23 (20.9, 20.5, 1.5) | right (29), mixed (0), left (1) |

no history of speech or hearing impairments, no experience living outside their home country for more than 4 months, and spoke only their native language in their daily lives. In addition, each participant's parents were native speakers of the participant's native language. Table 2 shows the age, gender, and handedness data of the participants. Although gender was not balanced between the language groups, age was nearly the same.

### *STIMULI*

Using linear predictive coding (LPC) analysis and resynthesis in Praat (Boersma & Weenink, 2017), three stimuli were generated. LPC is often used in signal processing to control acoustic cues, such as formant frequencies (i.e., one of the acoustic cues that people use for identifying vowels). In this study, a neutral LPC residual (i.e., a female voice with formant frequencies cancelled out) was filtered using a spectral envelope with F1 to F4 information. Shinohara et al. (2019) examined the categorization with goodness-rating test for 28 English and 30 Japanese speakers, using a stimulus continuum varying in only F2. The F1, F3, and F4 were set at 979 Hz, 2,886 Hz, and 4,151 Hz, respectively. The stimuli with F2 at 2,017 Hz, at 1,755 Hz, and at 1,493 Hz were most frequently identified as /æ/, /ʌ/, and /ɑ/, respectively, by English speakers, whereas the same stimuli were all categorized as Japanese /a/ by Japanese speakers. The goodness-rating test results for the 30 Japanese speakers demonstrated that the /æ/ stimulus was a significantly worse exemplar of the Japanese /a/ than the /ʌ/ stimulus, while there was no significant difference in the goodness rating between the /ɑ/ and /ʌ/ stimuli. Thus, it was confirmed that the English /æ/-/ʌ/ contrast belongs to the Category-Goodness difference assimilation type, whereas the English /ɑ/-/ʌ/ contrast belongs to the Single Category assimilation type for Japanese speakers. Shinohara et al. (2019) examined the discriminability of the English /æ/-/ʌ/ and /ɑ/-/ʌ/ contrasts for Japanese speakers and found that it was significantly more difficult to discriminate the /ɑ/-/ʌ/ contrast than the /æ/-/ʌ/ contrast, whereas the acoustic distance between the sounds in each contrast was the same in Hertz.

Table 3 shows the acoustic information of the stimuli resynthesized for the present auditory ERP experiment. Figure 1 displays the F1 and F2 frequencies of those stimuli in the Bark scale (i.e., a frequency scale corresponding with human perception). To minimize the discrepancy between the acoustic and perceptual distance, the F2 frequency of the three stimuli of the English /æ/, /ʌ/, and /ɑ/ used in Shinohara et al. (2019) were

https://doi.org/10.1017/S0272263121000978 Published online by Cambridge University Press

TABLE 3. Resynthesized stimuli information for the auditory ERP experiment

| Stimuli | Duration (ms) | F0 (Hz) | F1 (Hz) | B1 | F2 (Hz) | B2 | F3 (Hz) | B3 | F4 (Hz) | B4 |
|---|---|---|---|---|---|---|---|---|---|---|
| æ | 138 | 220 | 979 | 70 | 2,027 | 90 | 2,886 | 170 | 4,151 | 200 |
| ʌ | | | | | 1,746 | | | | | |
| ɑ | | | | | 1,502 | | | | | |
| Random 1 | | | | | 2,463 | | | | | |
| Random 2 | | | | | 1,229 | | | | | |
| Random 3 | | | 644 | | 2,463 | | | | | |
| Random 4 | | | | | 1,746 | | | | | |
| Random 5 | | | | | 1,229 | | | | | |
| Random 6 | | | 378 | | 2,463 | | | | | |
| Random 7 | | | | | 1,746 | | | | | |
| Random 8 | | | | | 1,229 | | | | | |

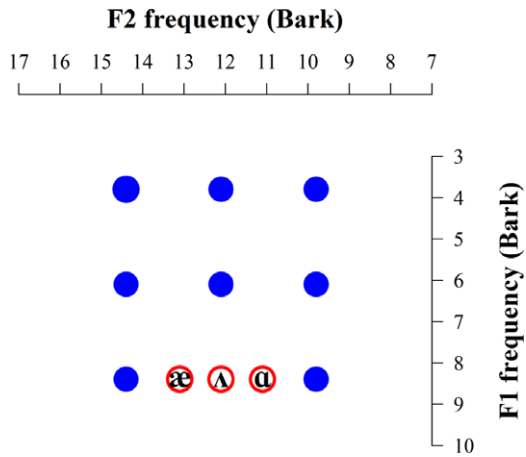


FIGURE 1. Stimuli used for the auditory ERP experiment.
*Note*: The three red circles represent the standard stimuli of the English /æ/ and /ɑ/, and the deviant stimulus of the English /ʌ/, used in the MMN-testing conditions. The blue dots represent the random-standard stimuli used in the control condition (see "Procedure" section).

modified to suit the Bark scale for the auditory ERP experiment in this study. As depicted by the red circles in Figure 1, the F2 of the English /æ/ was set at 2,027 Hz (13.1 in Bark), that of the English /ʌ/ was at 1,746 Hz (12.1 in Bark), and that of the English /ɑ/ was at 1,502 Hz (11.1 in Bark). Another eight stimuli (Random 1–8 in Table 3) were created and are represented by the blue dots in Figure 1. F1 was set at 378 Hz (3.8 in Bark) for the three close vowels (Random 6, 7, and 8 in Table 3), 644 Hz (6.1 in Bark) for the three mid

https://doi.org/10.1017/S0272263121000978 Published online by Cambridge University Press

vowels (Random 3, 4, and 5), and 979 Hz (8.4 in Bark) for the two open vowels (Random 1 and 2). F2 was set at 2,463 Hz (14.4 in Bark) for the three front vowels (Random 1, 3, and 6), 1,746 Hz (12.1 in Bark) for the two central vowels (Random 4 and 7), and 1,229 Hz (9.8 in Bark) for the three back vowels (Random 2, 5, and 8). Other acoustic cues (F3, F4, bandwidth for each formant, duration) were the same as those in the behavioral auditory discrimination test of Shinohara et al. (2019). The sound intensity was normalized among the 11 stimuli by the root mean square method in Praat (Boersma & Weenink, 2019).

## *PROCEDURE*

### *EEG Recording*

A total of 56 people (26 American English and 30 Japanese speakers) participated in the EEG recording sessions, conducted at two locations (United States, Japan). One English-speaking participant's data were not saved because of a technical problem, which resulted in the data of 55 participants being analyzed. The participants were tested in a soundproof booth, where they were seated in a comfortable reclining chair. For English speakers, continuous EEG was recorded from 128 carbon fiber core/silver-coated electrodes in an elastic electrode net (HydroCel Geodesic Sensor Net) at the University of Delaware. The continuous EEG was digitized with the EGI Net Station software v. 4.5 with a sampling rate of 250 Hz. Before data acquisition, electrode impedances were lowered to below 50 kΩ. Participants' electroocular activity was recorded from four bipolar channels. The vertical electrooculogram (EOG) was recorded with the supraorbital and infraorbital electrodes of both eyes; the horizontal EOG was recorded with the electrodes located at the outer canthi of both eyes. Channel E129 (corresponding to Cz electrode in the 10-10 system) placed at the center point of the scalp was used as the online reference site. The data of English speakers were passed through a 0.3 Hz FIR high-pass filter after recording, and the channel set was later remapped to match the 32 channels of the BrainAmp (Brain Products GmbH) used in the other laboratory in Japan, so that both datasets were analyzed together.

For Japanese speakers, the continuous EEG was recorded from 32 sintered Ag/AgCl passive electrodes of BrainAmp which adhered to the subject's scalp using an EEG recording cap (Easy Cap 40 Asian cut, Montage No. 24) at Waseda University. Of the 32 channels, one was used for recording horizontal eye movement (HEOG) and placed next to the outer canthus of the right eye. One channel (otherwise used as AFz) was used for grounding, and one was used as an online reference electrode attached to the FCz (i.e., fronto-central point of the head). The remaining 29 channels were mounted onto the cap, according to the 10/20-system, with the electrode adaptor A06 using high-chloride, abrasive electrolyte gel. The impedance level for all electrodes was below 5 kΩ. The analog signal was digitized at 250 Hz. As the data had passed through an online 0.016 Hz high-pass filter in the Brain Vision Recorder software as a default setting, it was not necessary to use an offline filter for the Japanese speakers' data.

There were three blocks in the EEG recording session. The first was the random-standard control condition (Horváth et al., 2008), which comprised a randomized sequence of eight resynthesized vowel sounds (as illustrated in Figure 1) and the /ʌ/ stimulus that was used as a deviant in the following two blocks. As the eight random standards stimuli did not form a single category, participants could not predict the upcoming stimuli. Therefore, the /ʌ/

https://doi.org/10.1017/S0272263121000978 Published online by Cambridge University Press

stimulus should not elicit MMN in this condition, and the Auditory Evoked Potential (AEP) it generates should just be the ERP response to that particular sound, not affected by MMN modulation. It therefore serves as a control condition to show that the attenuation of the AEP in the following two MMN-testing conditions is because of the MMN mechanism. Thus, the /ʌ/ stimulus in the random-standard control condition was not a deviant because it just appeared among the eight random standards. However, it is described as "deviant" for the statistical analysis to compare the AEP change from standards to deviant with that in the following two MMN-testing conditions.

In the second and third blocks, the participants heard either /æ/ or /ɑ/ as a standard (or frequent) stimulus, whereas /ʌ/ was always the deviant. The block order was counterbalanced between the participants in each language group. If the second block presented /æ/ as the standard and /ʌ/ as the deviant stimuli, the third block presented /ɑ/ as standard and /ʌ/ as deviant, or vice versa. In both blocks, as the standard stimulus is of one kind, participants can predict upcoming stimuli while hearing a repetition of the standard stimulus. When they hear a deviant, an MMN is expected to be elicited (i.e., the MMN-testing conditions).

In summary, two native-language groups (English and Japanese) heard two stimulus types (standard and deviant) in three vowel contrast conditions (control, front, and back). Each of the three blocks had 900 tokens (800 standards and 100 deviants), resulting in a total of 2,400 standards and 300 deviants. A continuous sequence of standards and deviants were presented through ER1 insert earphones (Etymotic Research) in Japan, but from two analog speakers in the United States. At both sites, the stimuli were played at 70 dB on average, and the interstimulus interval was varied randomly, around 717 ms (median = 716 ms, $SD = 50$ ms). Each block lasted for about 11 minutes, and the entire EEG recording took about 45 minutes, including breaks. During the EEG recording, the subjects were instructed to ignore the auditory input and watch a movie, *Wall-E* (Stanton, 2008) with no sound, because the MMN is elicited even in the absence of attention (Atienza & Cantero, 2001; Atienza et al., 1997, 2000, 2001, 2004, 2005; Nashida et al., 2000; Sallinen et al., 1994, 1996; Sculthorpe et al., 2009).

Table 4 presents the six cells that resulted from the three testing conditions with two types of stimuli: (1) random standards in the control condition, (2) /ʌ/ in the random standards control condition, (3) standard /æ/ in the front /æ/-/ʌ/ vowel condition, (4) deviant /ʌ/ in the front /æ/-/ʌ/ vowel condition, (5) standard /ɑ/ in the back /ɑ/-/ʌ/ vowel condition, and (6) deviant /ʌ/ in the back /ɑ/-/ʌ/ vowel condition. We hypothesized that the MMN elicited when the sound changes from (5) standard /ɑ/ to (6) deviant /ʌ/ is larger

TABLE 4. Six cells separated for the ERP analysis (3 vowel contrast conditions × 2 stimulus types)

| Vowel contrast condition | Standards | Deviants |
|---|---|---|
| Control | (1) Random standards (8 vowels) | (2) /ʌ/ |
| Front | (3) /æ/ | (4) /ʌ/ |
| Back | (5) /ɑ/ | (6) /ʌ/ |

https://doi.org/10.1017/S0272263121000978 Published online by Cambridge University Press

than the one that is elicited when the sound changes from (3) standard /æ/ to (4) deviant /ʌ/ for Japanese speakers, but there is no such difference in the MMN effects between the two conditions for English speakers.

### *EEG Signal Processing*

*Segmentation and Artifact Correction*. After recording, the raw continuous EEG data of the electrodes were imported into the ERP PCA toolkit v. 2.77 (Dien, 2010) run on MATLAB R2019b (Delorme & Makeig, 2004). The continuous EEG was first segmented into epochs from –200 ms to 800 ms relative to the stimulus onset. The segmented data were baseline-corrected by subtracting the mean of the 200 ms baseline period (i.e., –200 ms to 0 ms from the onset of stimuli) from the whole segment. The data were then submitted to an automatic process of eyeblink subtraction using Independent Component Analysis (ICA). An eyeblink template was automatically generated for each subject. An eyeblink component was marked and subtracted from the data if it was correlated at $r = 0.9$ or greater with the eyeblink template. Next, the bad channels were marked if their best absolute correlation with their neighboring channels fell below 0.4 across all time points. Those bad channels were replaced using spline interpolation from neighboring good channels. A channel was also declared globally bad if it was bad for more than 20% of the trials. A trial was marked bad and zeroed out if it contained more than 10% bad channels. All channels were rereferenced to the average of two mastoid electrodes. Finally, the remaining amplitude data from all tokens were categorized into the six cells in Table 4 and were averaged for each participant. If there were more than 10% global bad channels or fewer than 15 good trials in any of the six categories, those participants' data were not included in the statistical analysis. None of the American English speakers' data were excluded from the analysis based on this, whereas four Japanese speakers' data were excluded.

### *PCA Preprocessing and Selection of the Time Windows and Electrode Regions*

We took two steps to measure the MMN amplitudes. First, the time windows and electrode regions were objectively selected using a sequential temporospatial Principal Component Analysis (PCA), with a Promax rotation for the temporal PCA and an Infomax rotation for the spatial PCA. PCA decomposes the temporal and spatial dimensions into a linear combination of a smaller set of abstract ERP factors based on covariance patterns among time points and electrode sets. Before conducting the PCA, the electrode montage used to collect the English speakers' data was remapped to the same montage used for the Japanese speakers in the ERP PCA toolkit 2.93 (Dien, 2010), so that the data could be combined into a single dataset with the speaker's language as a between-subject variable. Using the combined data of Japanese and English speakers, three difference waves were calculated by subtracting the absolute waveforms of standard stimuli (i.e., random standards, /æ/, /ɑ/) from those of the deviant ones (i.e., /ʌ/) in three vowel conditions (control, front, and back). Then, the PCA was conducted with the three difference waves to identify the temporal and spatial distribution of the MMN. The temporal PCA generated 22 temporal factors, accounting for 83% of the total variance, and the spatial PCA identified four spatial factors for each of the temporal factors,

https://doi.org/10.1017/S0272263121000978 Published online by Cambridge University Press

accounting for 69% of the total variance. The temporal factors that accounted for less than 5% of the total variance were excluded, leaving only four temporal factors (TF1 = 16.7%, TF2 = 12.8%, TF3 = 7.3%, TF4 = 6.8%).

Figure 2 displays the difference waves reconstructed as voltage based on a temporospatial factor loading. The PCA identified the temporospatial factors that showed attenuation of AEP between standard and deviant stimuli in the MMN-testing condition (front and back) compared to that in the control condition. The four temporospatial factors (TF1SF1, TF2SF1, TF3SF1, and TF4SF1) showed negative responses in the front and back conditions, but not in the control condition. Those factors' electrode regions also corresponded to the MMN responses. However, one factor (TF2SF1), which peaked at 744 ms during the 632–796 ms time window was excluded because of the late response (Garrido et al., 2009; Luck, 2005; Yun Zhang et al., 2018). Thus, the three temporospatial

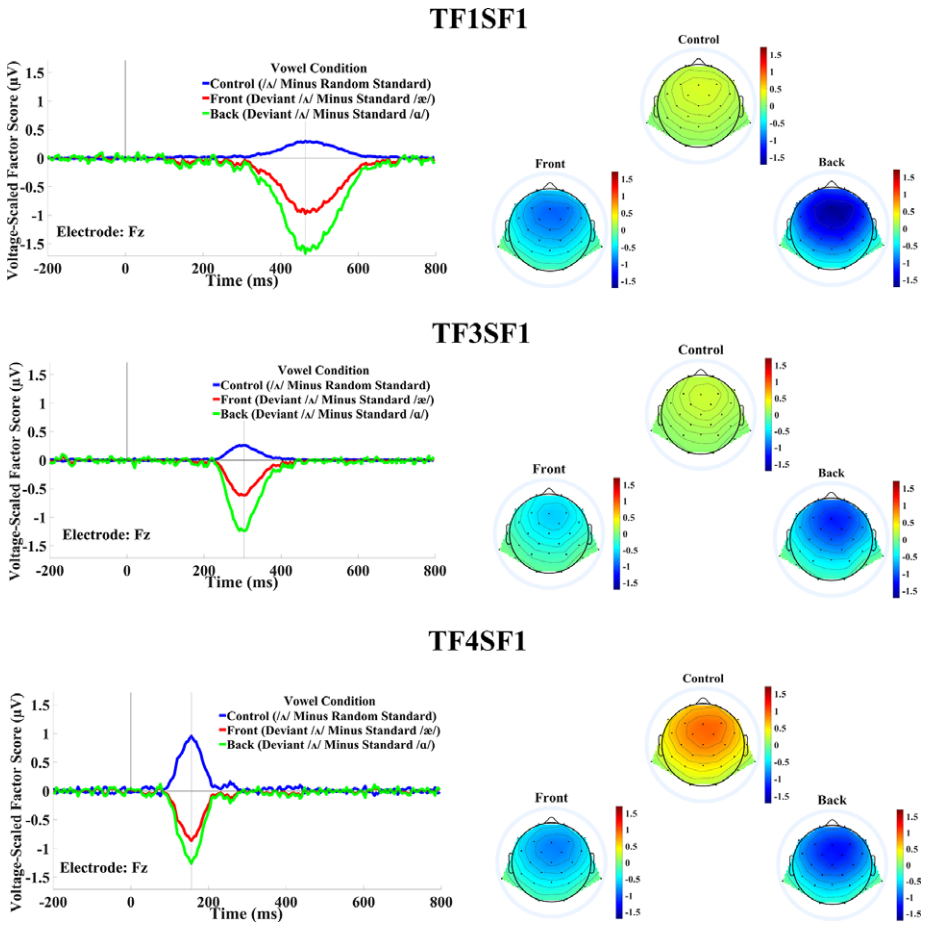


FIGURE 2. Temporospatial factor decompositions of the mean difference wave (deviants minus standards) in each vowel contrast condition (control, front, and back).
*Note*: English (N = 25) and Japanese speakers' (N = 26) waveforms were combined.

https://doi.org/10.1017/S0272263121000978 Published online by Cambridge University Press

factors displayed in Figure 2 (TF1SF1, TF3SF1, and TF4SF1) were selected for further analysis.

Table 5 describes the selected temporospatial factors. Their time windows carry a factor loading score of more than 0.6, and their electrodes carry a factor loading score of more than 0.9. In the final step, using the time windows and electrode regions identified by PCA, the amplitude of the absolute waveform to each token (2,400 standards and 300 deviants) was computed and categorized as one of the six cells (random standards in the control condition, /ʌ/ in the control condition, /æ/ standards in the front /æ/-/ʌ/ condition, /ʌ/ deviants in the /æ/-/ʌ/ condition, /ɑ/ standards in the back /ɑ/-/ʌ/ condition, and /ʌ/ deviants in the /ɑ/-/ʌ/ condition). Finally, we averaged the tokens in each cell for each participant for statistical analyses.

## RESULTS

Figure 3 displays the voltage amplitude of absolute waveforms of standards and deviants and their difference waves (deviant minus standard) in the control (random standards vs. /ʌ/), the front (standard /æ/ vs. deviant /ʌ/) and the back (standard /ɑ/ vs. deviant /ʌ/) vowel conditions. The boxplots show the voltage amplitude of difference waves of the standard and deviant stimuli collected from the time windows and electrode regions of the temporospatial factors reflecting MMN. The difference wave was calculated by subtracting the voltage amplitude of standard stimuli from that of deviant stimuli in each vowel condition (control, front, and back). A linear mixed-effects model was used for statistical analysis, with the difference wave as the dependent variable. The best-fitting model was selected through a top-down approach (i.e., excluding unnecessary fixed and random factors from the model that included all potential factors), using the R package *lme4* (Bates et al., 2015). The linear mixed-effects model included the fixed factors of language group (English, Japanese), vowel contrast condition (control: random vowels-/ʌ/, front: /æ/-/ʌ/, back: /ɑ/-/ʌ/), and their interactions. Orthogonal contrast was set for each fixed factor. The random factors were the crossed intercepts of participant and temporospatial factor.

Table 6 presents the results of the planned contrast analyses of the linear mixed-effects model. The significant effect of vowel condition (control vs. the front and back MMN-testing conditions), $\beta = -0.48$, $SE = 0.03$, $t = -15.56$, $p < .001$, suggests that the attenuation of the absolute waveforms from standard to deviant stimuli in the MMN-testing (i.e., front and back) conditions was significantly larger than that in the control condition. Another significant vowel condition contrast (front vs. back), $\beta = -0.24$, $SE = 0.05$, $t = -4.49$, $p < .001$, suggests that the MMN effect was larger in the back condition

TABLE 5. Time windows and electrodes of the temporospatial factors selected for analysis

| Temporospatial Factors | Time Window | Electrodes |
|---|---|---|
| TF1SF1 | 392–564 ms | F3, F4, C4, Fz, Cz, FC1, FC2 |
| TF3SF1 | 260–348 ms | Fp1, Fp2, F3, F4, Fz, Cz, FC1, FC2 |
| TF4SF1 | 120–188 ms | F3, F4, C3, C4, Fz, Cz, FC1, FC2, CP1, CP2, FC6 |

https://doi.org/10.1017/S0272263121000978 Published online by Cambridge University Press

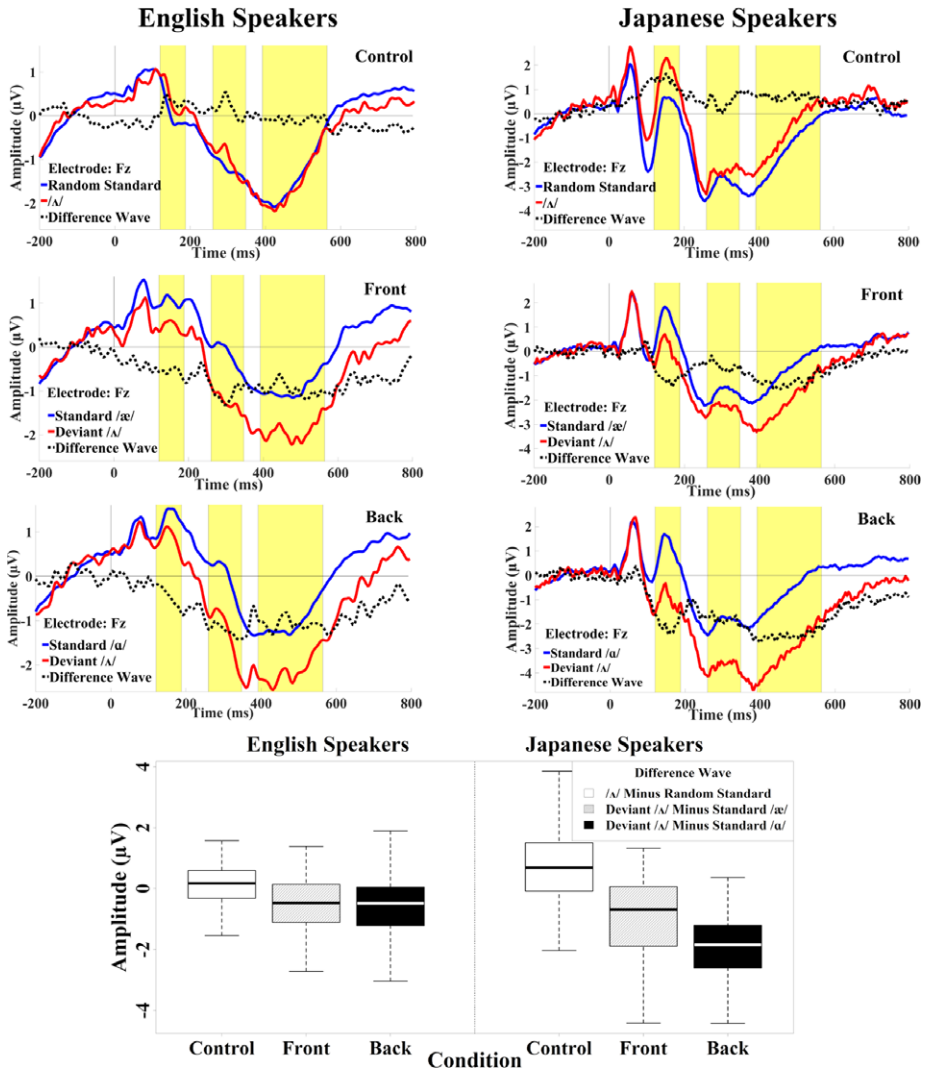


FIGURE 3. Absolute waveforms of the two stimuli (standards and deviants) in the three vowel conditions (control: random vowels vs. /ʌ/, front: /æ/ vs. /ʌ/, back: /ɑ/ vs. /ʌ/) for English and Japanese speakers, and boxplots of the voltage amplitudes of the difference waves (deviants minus standards) in each condition.

*Note*: The time windows analyzed (i.e., 120–188 ms, 260–348 ms, 392–564 ms) are in yellow. Different EEG systems were used to collect English and Japanese speakers' data with different sound presentations (analog speakers and earphones). This causes the overall amplitude difference between the two language groups. The effect of vowel condition (e.g., front vs. back) on the difference waves (deviant minus standard) should be compared between the two language groups.

https://doi.org/10.1017/S0272263121000978 Published online by Cambridge University Press

TABLE 6. Results of a linear mixed-effects model for the voltage data of difference waves (deviant minus standard) in the control (random standards vs. /ʌ/), front (/æ/ vs. /ʌ/) and back (/ɑ/ vs. /ʌ/) conditions for English and Japanese speakers

| Fixed Factor (contrast) | Estimate | *SE* | *t* | *p* | Cohen's *d* |
|---|---|---|---|---|---|
| Intercept | –0.51 | 0.13 | –4.09 | .011 | – |
| Group (English vs. Japanese Speakers) | –0.16 | 0.09 | –1.84 | .072 | –0.53 |
| Condition (Control vs. Front and Back) | –0.48 | 0.03 | –15.56 | < .001 | –1.55 |
| Condition (Front vs. Back) | –0.24 | 0.05 | –4.49 | < .001 | –0.45 |
| Group (English vs. Japanese) and Condition (Control vs. Front and Back) | –0.25 | 0.03 | –8.05 | < .001 | –0.80 |
| Group (English vs. Japanese) and Condition (Front vs. Back) | –0.22 | 0.05 | –4.07 | < .001 | –0.41 |

than that in the front condition across language groups. However, there was a significant two-way interaction of language group (English vs. Japanese) and vowel condition (front vs. back), $\beta = –0.22$, $SE = 0.05$, $t = –4.07$, $p < .001$, suggesting that the MMN effect in each of the front and back conditions was different between English and Japanese speakers.

Further analyses were conducted for each language group using separate linear mixed-effects models. For Japanese speakers, the MMN amplitude was larger for the back (i.e., standard /ɑ/ vs. deviant /ʌ/) than for the front (i.e., standard /æ/ vs. deviant /ʌ/) condition, as shown by a significant vowel condition (front vs. back), $\beta = –0.45$, $SE = 0.08$, $t = –5.58$, $p < .001$. The other vowel condition contrast (control vs. front and back) was also significant, $\beta = –0.72$, $SE = 0.05$, $t = –15.39$, $p < .001$, suggesting that the MMN effect was significant in the MMN-testing conditions compared to the control condition.

However, for the English speakers, a significant MMN amplitude difference between the front (i.e., standard /æ/ vs. deviant /ʌ/) and back (i.e., standard /ɑ/ vs. deviant /ʌ/) conditions was not observed, $\beta = –0.02$, $SE = 0.07$, $t = –0.33$, $p > .05$. This means that the MMN effect was not different between the two conditions, although it was significant in the MMN-testing conditions as demonstrated by a significant vowel condition effect (control vs. front and back), $\beta = –0.23$, $SE = 0.04$, $t = –5.97$, $p < .001$.

These results show that the deviant /ʌ/ in a series of standard /ɑ/ (the back condition) elicited a larger MMN than the deviant /ʌ/ in a series of standard /æ/ (the front condition) for Japanese speakers, but there was no such vowel condition effect on MMN for English speakers.

## DISCUSSION

The present study examined the opposing effects of prototypicality and discriminability of standard and deviant stimuli and investigated which of the two modulated the MMN amplitude more. According to the discriminability account, the front condition (i.e., /ʌ/ deviants in a series of the /æ/ standard stimuli) elicits a larger MMN than does the back condition (i.e., /ʌ/ deviants in a series of the /ɑ/ standard stimuli), as the front contrast (/æ/ vs. /ʌ/) is more discriminable than the back one (/ɑ/ vs. /ʌ/) for Japanese speakers

https://doi.org/10.1017/S0272263121000978 Published online by Cambridge University Press

(Shinohara et al., 2019). However, we hypothesized the opposite, following the prototypicality account, where the back condition elicits a larger MMN than the front one. When Japanese speakers hear the English /ɑ/ as standard stimuli in an oddball paradigm, the prototypical phonetic status of the English /ɑ/ as Japanese /a/ easily generates a short-term memory trace, and their prediction of upcoming stimuli becomes robust. When they hear a deviant /ʌ/, the stronger prediction error occurs, resulting in a larger MMN, compared to when they hear a deviant /ʌ/ in a series of standard /æ/, which is less prototypical for Japanese speakers. The results of this study supported this prototypicality account, indicating that the prototypicality of standard stimuli modulates the MMN amplitude more than the discriminability of standard and deviant stimuli.

The results of this study showed that generating a short-term memory trace by hearing a series of standard stimuli that is easily mapped onto a listener's L1 phonological category is more important in eliciting a larger MMN than the discriminability between standard and deviant stimuli. This interpretation is supported by the results of previous studies. For example, Shafer et al. (2004) found that no MMN was elicited when English speakers heard a nonnative phone as standard and a native phone as deviant, although an MMN was observed when they heard a native phone as standard and a nonnative phone as deviant. This is because hearing a nonnative phone repeatedly does not generate a robust short-term memory trace, whereas hearing repetitions of a native phone stably generates a memory trace. Thus, the phonetic status in listeners' phonological representations, namely the prototypicality as listeners' L1 phoneme, affects the prediction of upcoming stimuli, resulting in a modulation of the MMN amplitude.

A more recent study testing Japanese speakers' perception of English vowels also demonstrated similar results. Shafer et al. (2021) found that a larger MMN was elicited when a prototype stimulus (e.g., a nonnative stimulus sharing phonetic features of a Japanese phoneme) was used as standard and a nonprototype stimulus was used as deviant, compared to the reversed order. In Shafer et al. (2021), naturally recorded stimuli that vary in both spectral and durational cues were used and its prototypicality in L1 phonology was determined by the feature-based analysis. The current study controlled the stimuli more carefully, and the difference between the English vowel stimuli was set only in F2 frequency. Japanese speakers' goodness-rating scores of the English stimuli were measured in Shinohara et al. (2019) and were statistically compared to confirm the perceptual assimilation patterns and their prototypicality as the Japanese vowel /a/. Even after careful stimulus control, the current study found similar results. Although there are contradictory results in the literature (e.g., Aaltonen et al., 1997), given that Shafer et al. (2021) and the present study obtained similar results, it is plausible to conclude that the phonetic status of standard stimuli (prototypicality/familiarity) in an oddball paradigm has a significant effect on the MMN amplitude, at least for the perception of the English /æ/, /ʌ/, and /ɑ/ by Japanese speakers.

The findings in the present study show that the MMN amplitude can be used as an index of the phonetic status in listeners' phonological representations. Newborns statistically learn the acoustic-phonetic features of an ambient language and develop the discrimination of frequently perceived speech sounds (i.e., native phonemes), but decline that of infrequently perceived ones (i.e., nonnative phones) (Kuhl, 2010; Kuhl et al., 2006, 2008). When such frequently perceived speech sounds (e.g., two native phonemes) are used as standards and deviants in an oddball paradigm, a large MMN is elicited

https://doi.org/10.1017/S0272263121000978 Published online by Cambridge University Press

(Näätänen et al., 1997; Peltola et al., 2003). In addition, a phonetic training study demonstrated that when two sounds that have been learned in a bimodal distribution are used as standard and deviant stimuli, a larger MMN is elicited, compared to when two sounds that have been learned in a unimodal distribution are used as standards and deviants (Wanrooij et al., 2014). These results in both the current and previous studies suggest that the size of MMN amplitude indicates statistical phonetic learning. If a standard stimulus used in an oddball paradigm has been learned by listeners as a prototype in a phoneme distribution, a larger MMN is elicited when they hear a deviant stimulus. If the standard stimulus is not a prototype in a phoneme distribution, a smaller MMN is elicited. Thus, the MMN amplitude can be used as an index of the phonetic status in listeners' phonological representations.

MMN asymmetry is affected by many factors, as the perceptual salience is attributed to both language-universal (e.g., favoring focal vowels: Masapollo et al., 2015, 2017; Polka & Bohn, 2003, 2011; Schwartz et al., 2005) and language-specific perception bias (e.g., favoring native phoneme prototypes: Iverson & Kuhl, 1995; Kuhl, 1991; Kuhl et al., 1992; and underspecification: Eulitz & Lahiri, 2004; Hestvik & Durvasula, 2016). Although the current study demonstrated that the prototypicality of the standard stimuli in listeners' phonological representations modulates the MMN amplitude more robustly than does the conscious discriminability between standard and deviant stimuli, it was not possible to isolate the prototypicality and discriminability effects from other factors or to confirm that the same result is seen in the perception of other nonnative phones. Future studies should conduct another experiment with other pairs of stimuli to test additional predictions. The ways in which the factors intervene with each other and the MMN amplitude gets affected by the interaction of those factors must be investigated.

In conclusion, the auditory ERP experiment showed that MMN is not a mere reflection of the discriminability of speech sounds. The prototypicality of the standard stimulus modulates the MMN amplitude more than the discriminability.

## REFERENCES

Aaltonen, O., Eerola, O., Hellström, Å., Uusipaikka, E., & Lang, A. H. (1997). Perceptual magnet effect in the light of behavioral and psychophysiological data. *The Journal of the Acoustical Society of America*, *101*, 1090–1105. https://doi.org/10.1121/1.418031

Amenedo, E., & Escera, C. (2000). The accuracy of sound duration representation in the human brain determines the accuracy of behavioural perception. *European Journal of Neuroscience*, *12*, 2570–2574. https://doi.org/10.1046/j.1460-9568.2000.00114.x

Atienza, M., & Cantero, J. L. (2001). Complex sound processing during human REM sleep by recovering information from long-term memory as revealed by the mismatch negativity (MMN). *Brain Research*, *901*, 151–160. https://doi.org/10.1016/S0006-8993(01)02340-X

Atienza, M., Cantero, J. L., & Escera, C. (2001). Auditory information processing during human sleep as revealed by event-related brain potentials. *Clinical Neurophysiology*, *112*, 2031–2045. https://doi.org/10.1016/S1388-2457(01)00650-2

Atienza, M., Cantero, J. L., & Gómez, C. M. (1997). The mismatch negativity component reveals the sensory memory during REM sleep in humans. *Neuroscience Letters*, *237*, 21–24. https://doi.org/10.1016/S0304-3940(97)00798-2

Atienza, M., Cantero, J. L., & Gómez, C. M. (2000). Decay time of the auditory sensory memory trace during wakefulness and REM sleep. *Psychophysiology*, *37*, S0048577200980697. https://doi.org/10.1017/S0048577200980697

Atienza, M., Cantero, J. L., & Quian Quiroga, R. (2005). Precise timing accounts for posttraining sleep-dependent enhancements of the auditory mismatch negativity. *NeuroImage*, *26*, 628–634. https://doi.org/10.1016/j.neuroimage.2005.02.014

Atienza, M., Cantero, J. L., & Stickgold, R. (2004). Posttraining sleep enhances automaticity in perceptual discrimination. *Journal of Cognitive Neuroscience*, *16*, 53–64. https://doi.org/10.1162/089892904322755557

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

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models Using lme4. *Journal of Statistical Software*, *67*, 1–51. https://doi.org/10.18637/jss.v067.i01

Best, C. T. (1994a). Learning to perceive the sound pattern of English. In C. Rovee-Collie & L. P. Lipsitt (Eds.), *Advances in infancy research* (pp. 217–304). Ablex Publishers.

Best, C. T. (1994b). The emergence of native-language phonological influences in infants: A perceptual assimilation model. *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*, 167, *167*–224. https://doi.org/10.1007/978-94-015-8234-6_24

Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). York Press.

Boersma, P., & Weenink, D. (2017). *Praat: doing phonetics by* computer [Computer program]. Version 6.0.33. http://www.praat.org/

Boersma, P., & Weenink, D. (2019). *Praat: doing phonetics by computer* [Computer program]. Version 6.0.48. http://www.praat.org/

Bühler, J. C., Schmid, S., & Maurer, U. (2017). Influence of dialect use on speech perception: A mismatch negativity study. *Language, Cognition and Neuroscience*, *32*, 757–775. https://doi.org/10.1080/23273798.2016.1272704

Dehaene-Lambertz, G., & Baillet, S. (1998). A phonological representation in the infant brain. *NeuroReport*, *9*, 1885–1888. https://doi.org/10.1097/00001756-199806010-00040

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

Dien, J. (2010). The ERP PCA Toolkit: An open source program for advanced statistical analysis of event-related potential data. *Journal of Neuroscience Methods*, *187*, 138–145. https://doi.org/10.1016/j.jneumeth.2009.12.009

Eulitz, C., & Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience*, *16*, 577–583. https://doi.org/10.1162/089892904323057308

Friston, K. (2002). Functional integration and inference in the brain. *Progress in Neurobiology*, *68*, 113–143. https://doi.org/10.1016/S0301-0082(02)00076-X

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*, 815–836. https://doi.org/10.1098/rstb.2005.1622

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology*, *120*, 453–463. https://doi.org/10.1016/j.clinph.2008.11.029

Grimaldi, M., Sisinni, B., Gili Fivela, B., Invitto, S., Resta, D., Alku, P., & Brattico, E. (2014). Assimilation of L2 vowels to L1 phonemes governs L2 learning in adulthood: A behavioral and ERP study. *Frontiers in Human Neuroscience*, *8*, 1–14. https://doi.org/10.3389/fnhum.2014.00279

Hestvik, A., & Durvasula, K. (2016). Neurobiological evidence for voicing underspecification in English. *Brain and Language*, *152*, 28–43. https://doi.org/10.1016/j.bandl.2015.10.007

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, *97*, 3099–3111. https://doi.org/10.1121/1.411872

Horváth, J., Czigler, I., Jacobsen, T., Maess, B., Schröger, E., & Winkler, I. (2008). MMN or no MMN: No magnitude of deviance effect on the MMN amplitude. *Psychophysiology*, *45*, 60–69. https://doi.org/10.1111/j.1469-8986.2007.00599.x

Imada, T., Hari, R., Loveless, N., McEvoy, L., & Sams, M. (1993). Determinants of the auditory mismatch response. *Electroencephalography and Clinical Neurophysiology*, *87*, 144–153. https://doi.org/10.1016/0013-4694(93)90120-k

Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America*, *97*, 553–562. https://doi.org/10.1121/1.412280

Jacobsen, T., Horváth, J., Schröger, E., Lattner, S., Widmann, A., & Winkler, I. (2004). Pre-attentive auditory processing of lexicality. *Brain and Language*, *88*, 54–67. https://doi.org/10.1016/s0093-934x(03)00156-1

Javitt, D. C., Grochowski, S., Shelley, A. M., & Ritter, W. (1998). Impaired mismatch negativity (MMN) generation in schizophrenia as a function of stimulus deviance, probability, and interstimulus/interdeviant interval. *Electroencephalography and Clinical Neurophysiology*, *108*, 143–153. https://doi.org/10.1016/s0168-5597(97)00073-7

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, *50*, 93–107. https://doi.org/10.3758/BF03212211

Kuhl, P. K. (2010). Brain mechanisms in early language acquisition. *Neuron*, *67*, 713–727. https://doi.org/10.1016/j.neuron.2010.08.038

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*, 979–1000. https://doi.org/10.1098/rstb.2007.2154

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*, F13–F21. https://doi.org/10.1111/j.1467-7687.2006.00468.x

Kuhl, P. K., Williams, K., Lacerda, F., Stevens, K., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*, 606–608. https://doi.org/10.1126/science.1736364

Lang, H., Nyrke, T., Ek, M., Aaltonen, O. K., Raimo, I., & Näätänen, R. (1990). Pitch discrimination performance and auditive event-related potentials. *Psychophysiological Brain Research*, *1*, 294–298.

Lengeris, A. (2009). Perceptual assimilation and L2 learning: Evidence from the perception of Southern British English vowels by native speakers of Greek and Japanese. *Phonetica*, *66*, 169–187. https://doi.org/10.1159/000235659

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. *Clinical Neurophysiology*, *120*, 916–921. https://doi.org/10.1016/j.clinph.2009.03.010

Luck, S. J. (2005). *An introduction to the event-related potential technique*. MIT Press.

Masapollo, M., Polka, L., & Ménard, L. (2015). Asymmetries in vowel perception: Effects of formant convergence and category "goodness." *The Journal of the Acoustical Society of America*, *137*, 2385–2385. https://doi.org/10.1121/1.4920678

Masapollo, M., Polka, L., Molnar, M., & Ménard, L. (2017). Directional asymmetries reveal a universal bias in adult vowel perception. *The Journal of the Acoustical Society of America*, *141*, 2857–2869. https://doi.org/10.1121/1.4981006

May, P., Tiitinen, H., Ilmoniemi, R. J., Nyman, G., Taylor, J. G., & Näätänen, R. (1999). Frequency change detection in human auditory cortex. *Journal of Computational Neuroscience*, *6*, 99–120. https://doi.org/10.1023/a:1008896417606

Näätänen, R. (1992). *Attention and brain function*. Routledge. https://doi.org/10.4324/9780429487354

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. http://www.ncbi.nlm.nih.gov/pubmed/11321610

Näätänen, R., & Alho, K. (1997). Mismatch Negativity: The measure for central sound representation accuracy. *Audiology and Neurotology*, *2*, 341–353. https://doi.org/10.1159/000259255

Näätänen, R., Kujala, T., & Light, G. (2019). *The mismatch negativity*. Oxford University Press.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (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. *Clinical Neurophysiology*, *118*), 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026

Näätänen, R., Schröger, E., Karakas, S., Tervaniemi, M., & Paavilainen, P. (1993). Development of a memory trace for a complex sound in the human brain. *NeuroReport*, *4*, 503–506. https://doi.org/10.1097/00001756-199305000-00010

Nashida, T., Yabe, H., Sato, Y., Hiruma, T., Sutoh, T., Shinozaki, N., & Kaneko, S. (2000). Automatic auditory information processing in sleep. *Sleep*, *23*, 821–828. http://www.ncbi.nlm.nih.gov/pubmed/11007449

Pakarinen, S., Lovio, R., Huotilainen, 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. *Biological Psychology*, *82*, 219–226. https://doi.org/10.1016/j.biopsycho.2009.07.008

Pakarinen, S., Takegata, R., Rinne, T., Huotilainen, M., & Näätänen, R. (2007). Measurement of extensive auditory discrimination profiles using the mismatch negativity (MMN) of the auditory event-related potential (ERP). *Clinical Neurophysiology*, *118*, 177–185. https://doi.org/10.1016/j.clinph.2006.09.001

Peltola, M. S., Kujala, T., Tuomainen, J., Ek, M., Aaltonen, O., & Näätänen, R. (2003). Native and foreign vowel discrimination as indexed by the mismatch negativity (MMN) response. *Neuroscience Letters*, *352*, 25–28. https://doi.org/10.1016/j.neulet.2003.08.013

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., McGinnis, M., & Roberts, T. (2000). Auditory cortex accesses phonological categories: An MEG mismatch study. *Journal of Cognitive Neuroscience*, *12*, 1038–1055. https://doi.org/doi:10.1162/08989290051137567

Polka, L., & Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication*, *41*, 221–231. https://doi.org/10.1016/S0167-6393(02)00105-X

Polka, L., & Bohn, O.-S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics*, *39*, 467–478. https://doi.org/10.1016/j.wocn.2010.08.007

Rinker, T., Alku, P., Brosch, S., & Kiefer, M. (2010). Discrimination of native and non-native vowel contrasts in bilingual Turkish-German and monolingual German children: Insight from the Mismatch Negativity ERP component. *Brain and Language*, *113*, 90–95. https://doi.org/10.1016/j.bandl.2010.01.007

Sallinen, M., Kaartinen, J., & Lyytinen, H. (1994). Is the appearance of mismatch negativity during stage 2 sleep related to the elicitation of K-complex? *Electroencephalography and Clinical Neurophysiology*, *91*, 140–148. https://doi.org/10.1016/0013-4694(94)90035-3

Sallinen, M., Kaartinen, J., & Lyytinen, H. (1996). Processing of auditory stimuli during tonic and phasic periods of REM sleep as revealed by event-related brain potentials. *Journal of Sleep Research*, *5*, 220–228. https://doi.org/10.1111/j.1365-2869.1996.00220.x

Sams, M., Alho, K., & Näätänen, R. (1983). Sequential effects on the ERP in discriminating two stimuli. *Biological Psychology*, *17*, 41–58. https://doi.org/10.1016/0301-0511(83)90065-0

Schwartz, J.-L., Abry, C., Boë, L.-J., Ménard, L., & Vallée, N. (2005). Asymmetries in vowel perception, in the context of the Dispersion-Focalisation Theory. *Speech Communication*, *45*, 425–434. https://doi.org/10.1016/j.specom.2004.12.001

Sculthorpe, L. D., Ouellet, D. R., & Campbell, K. B. (2009). MMN elicitation during natural sleep to violations of an auditory pattern. *Brain Research*, 1290, 52–62. https://doi.org/10.1016/j.brainres.2009.06.013

Shafer, V. L., Kresh, S., Ito, K., Hisagi, M., Vidal, N., Higby, E., Castillo, D., & Strange, W. (2021). *The neural timecourse of American English vowel discrimination by Japanese, Russian and Spanish second-language learners of English*. Bilingualism: Language and Cognition, 1–14. https://doi.org/10.1017/S1366728921000201

Shafer, V. L., Schwartz, R. G., & Kurtzberg, D. (2004). Language-specific memory traces of consonants in the brain. *Cognitive Brain Research*, *18*, 242–254. https://doi.org/10.1016/j.cogbrainres.2003.10.007

Shinohara, Y., Han, C., & Hestvik, A. (2019). Effects of perceptual assimilation: The perception of English /æ/, /ʌ/, and /ɑ/ by Japanese speakers. In S. Calhoun, P. Escudero, M. Tabain, and P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 2344–2348). Australasian Speech Science and Technology Association Inc.

Stanton, A. (Director). (2008). *Wall-E* [Film]. Walt Disney Home Entertainment.

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., Nishi, K., & Jenkins, J. J. (1998). Perceptual assimilation of American English vowels by Japanese listeners. *Journal of Phonetics*, *26*, 311–344. https://doi.org/10.1006/jpho.1998.0078

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. *Neuroscience Letters*, *226*, 1–4. https://doi.org/10.1016/S0304-3940(97)00217-6

Wacongne, C., Changeux, J.-P., & Dehaene, S. (2012). A neuronal model of predictive coding accounting for the mismatch negativity. *Journal of Neuroscience*, *32*), 3665–3678. https://doi.org/10.1523/JNEUROSCI.5003-11.2012

Wanrooij, K., Boersma, P., & van Zuijen, T. L. (2014). Fast phonetic learning occurs already in 2-to-3-month old infants: An ERP study. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00077

Winkler, I., & Czigler, I. (2012). Evidence from auditory and visual event-related potential (ERP) studies of deviance detection (MMN and vMMN) linking predictive coding theories and perceptual object representations. *International Journal of Psychophysiology*, *83*, 132–143. https://doi.org/10.1016/j.ijpsycho.2011.10.001

Zhang, Yang, Kuhl, P. K., Imada, T., Kotani, M., & Tohkura, Y. (2005). Effects of language experience: Neural commitment to language-specific auditory patterns. *NeuroImage*, *26*, 703–720. https://doi.org/10.1016/j.neuroimage.2005.02.040

Zhang, Yun, Yan, F., Wang, L., Wang, Y., Wang, C., Wang, Q., & Huang, L. (2018). Cortical areas associated with mismatch negativity: A connectivity study using propofol anesthesia. *Frontiers in Human Neuroscience*, *12*, 392. https://doi.org/10.3389/fnhum.2018.00392