

Article

Not peer-reviewed version

Language Experience Shapes Neural Grouping of Speech by Accent: EEG Evidence from Native, L2, and Heritage Listeners

Lauren Hong , Chao Han , [Philip J. Monahan](#) *

Posted Date: 25 December 2025

doi: 10.20944/preprints202512.2352.v1

Keywords: mismatch negativity (MMN); accent processing; second language (L2); talker identity; speech perception; electroencephalography (EEG)



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Language Experience Shapes Neural Grouping of Speech by Accent: EEG Evidence from Native, L2, and Heritage Listeners

Lauren Hong^{1,2,3}, Chao Han⁴ and Philip J. Monahan^{1,4,5,*}

¹ Department of Psychology, University of Toronto Scarborough, Toronto M1C 1A4, Canada

² KITE Research Institute, Toronto Rehabilitation Institute, University Health Network, Toronto M5G 2A2, Canada

³ Department of Psychology, University of Toronto, Toronto M5S 3G3, Canada

⁴ Department of Language Studies, University of Toronto Scarborough, Toronto M1C 1A4, Canada

⁵ Department of Linguistics, University of Toronto, Toronto M5S 3G3, Canada

* Correspondence: philip.monahan@utoronto.ca

Abstract

Accented speech contains talker-indexical cues that listeners can use to infer social group membership, yet it remains unclear how the auditory system categorizes accent variability and how this process depends on language experience. The current study used EEG and the MMN oddball paradigm to test pre-attentive neural sensitivity to accent changes of English words stopped produced by Canadian English or Mandarin Chinese accented English talkers. Three participant groups were tested: Native English listeners, L1-Mandarin listeners, and Heritage Mandarin listeners. In the Native English and L1-Mandarin groups, we observed MMNs to the Canadian accented English deviant, indicating that the brain can group speech by accent despite substantive inter-talker variation and is consistent with an experience-dependence sensitivity to accent. Exposure to Mandarin Chinese accented English modulated MMN magnitude. Time-frequency analyses suggested that α and low- β power during accent encoding varied with language background, with Native English listeners showing stronger activity when presented with Mandarin Chinese accented English. Finally, the neurophysiological response in the Heritage Mandarin group reflected a broader phonological space encompassing both Canadian English and Mandarin-accented English, and its magnitude was predicted by Chinese proficiency. These findings provide brain-based evidence that automatic accent categorization is not uniform across listeners but interacts with native phonology and second-language experience.

Keywords: mismatch negativity (MMN); accent processing; second language (L2); talker identity; speech perception; electroencephalography (EEG)

Introduction

Speech transmits both the linguistic message and talker-indexical characteristics [1]. From the acoustic signal alone—often with sparse exposure to a given talker—listeners reliably assess talker age [2–5], sex and gender [6–9], race and ethnicity [10–13], as well as physical attributes, such as height and weight [14,15]. Listeners also identify non-native accents with as little acoustic information as a single phoneme [16,17]. Such cues to talker accent and dialect are used to socially categorize individuals or establish group membership [18,19]. The current study examines whether the brain categorizes speech based on talker accent and whether language experience affects this categorization.

Spatial activity patterns in superior temporal cortex highlight a functional dissociation between linguistic content and talker-indexical cues [20]. Phonetic information is encoded in overlapping superior temporal regions, including Heschl's gyrus, planum temporale (PT), superior temporal

gyrus (STG), superior temporal sulcus (STS), and middle temporal gyrus [21–27]. Conversely, repeated talker exposure drives right anterior STS activation patterns [28], although some spatial overlap is observed between speech recognition regions in left STS/STG and talker-specific vocal tract parameters [29]. Functional imaging also reveals enhanced activation of auditory and memory-related networks for familiar compared with unfamiliar voices, leading to improved performance in listening and word memory tasks [30,31]; using electrophysiology, familiar voices elicit larger amplitude event-related potentials compared to unfamiliar talkers [32,33] and voice familiarity differences are observed in α -band desynchronization [34], while θ -band and γ -band activity dominate linguistic encoding [35,36]. Together, the evidence indicates at least partially distinct mechanisms for the neural processing of talker-indexical and linguistic cues.

Studies of forensic talker profiling indicate that nonnative accent and regional dialect are easily identifiable features of an unknown voice and have been used in legal proceedings [37]. In instances where dialect and accent are probabilistically linked with marginalized groups, profiling leads to discrimination [38,39], the denial of access to housing [19,40,41], and triggers exclusionary behavior [42]. Linguistic profiling persists even in putatively socially conscious, younger individuals [43]. Despite these social and economic implications, the neurophysiological mechanisms for the grouping and categorization of individuals based on accent and dialect remain underexplored.

A powerful tool to investigate the neurophysiological categorization and grouping of speech is the mismatch negativity (MMN), an automatic and pre-attentive change detection event-related potential [44–46]. The MMN localizes to supratemporal auditory cortex [45,47–50], making it suitable for testing talker-based categorization. Oddball paradigms, which present frequent standard stimuli interspersed with infrequent deviants, are commonly used to elicit the MMN. Detection of a deviant along perceptual or representational dimensions evokes a negative deflection in the event-related potential, typically peaking 150–350 ms after stimulus onset and largest over fronto-central electrode sites. While traditional oddball paradigms involved the repeated presentation of physically identical standard stimuli [51–53], the MMN is also observed when within-category physical variation is introduced into the standard stimuli [54–61]. Relevant to the current design, changes in voice are sufficient to elicit a MMN [32,62], and the linguistic background of participants also impacts MMN response characteristics [63–72].

MMN latency is influenced by accent familiarity. An earlier MMN was observed when Standard American English listeners were exposed to acoustically variable African American English tokens of *hello* as standards and a Standard American English *hello* as the deviant; conversely, a later MMN was observed when the standard was in Standard American English and the deviant in African American English [73]. The presence of MMNs in both configurations indicated that accent variation is categorized during early cortical phonetic processing, while latency differences reflect the influence of accent familiarity on the temporal and spatial dynamics of the MMN. Specifically, deviant stimuli produced in a familiar accent elicited an earlier MMN relative to deviants produced in an unfamiliar accent, suggesting that the familiar accent was psycho-acoustically more salient. In contrast, deviants produced in the less familiar accent elicited larger MMN responses in Standard German and Swiss-German participants, which was interpreted as the unfamiliar accent imposing greater processing costs [64]. That said, the MMN latency reported in Bühler et al. [64] was consistent with the MMN latency to the unfamiliar deviant reported in Scharinger et al. [73]. Furthermore, changes in talker-sex and talker-accent in an oddball paradigm result in the elicitation of the MMN response in native and nonnative participants alike [74]; that said, the stimuli consisted of isolated vowels, with only one token per category, making it unclear whether listeners perceived standard-deviant differences as meaningful accent variation or merely low-level acoustic contrasts, especially as native and non-native listeners showed similar MMN responses. Overall, the extent to which language experience, particularly accent familiarity and phonological background, shapes neurophysiological processing of accent variability remains an open question.

Previous MMN tests assessing accent processing have either used a single stimulus [32,64,74] or multiple stimuli produced by the same talker [73]. While this approach isolates accent information

from other talker-specific cues, it leaves unaddressed the question of whether the brain generalizes across talkers to extract accent. The MMN has been employed to show that the brain posits generalizations on the basis of abstract phonological features despite substantial inter-category phonetic variation [54,57,59] and on the basis of vowel category despite substantial inter-talker variation [61]. The current study introduces inter-talker variability in the oddball paradigm to determine whether the auditory system can generalize an accent beyond an individual talker.

Using an inter-talker varying oddball paradigm, the current study tested whether the brain automatically categorizes speech based on accent across individual talkers and whether the magnitude or timing of the brain response—as indexed by the MMN—is modulated by language experience. Following Scharinger et al. [73], we used whole-word English stimuli. In an oddball paradigm, we presented native English listeners, heritage Mandarin listeners, and L1-Mandarin learners of English with the word *stopped* produced in two accents (i.e., Standard Canadian English (SCE), Mandarin-accented English (MAE)) while their electroencephalogram (EEG) was recorded. Crucially, standard and deviant tokens were sampled from ten different talkers of each accent. Overall, an accent change should elicit an MMN response if the brain perceptually groups multiple talkers from the same accent together. Both Scharinger et al. [73] and Bühler et al. [64] report an accent-change induced MMN. That said, the reported latencies and amplitudes differed depending on deviant familiarity. As such, there are two possible outcomes. First, we might predict a larger and earlier MMN to the familiar accent due to its psychoacoustic salience. That is, we predict a larger MMN to the SCE deviant in native English listeners and a larger MMN in the L1 Mandarin participants when the MAE tokens are the deviant. Alternatively, an unfamiliar accent may impose greater processing costs and hence result in a more robust MMN. Given their language background, heritage Mandarin listeners were expected to be familiar with both Standard Canadian English and Mandarin-accented English, thus exhibiting greater tolerance to accent variability and potentially, a reduced MMN.

Methods

Participants

Sixty-nine participants (47 females; mean age: 19.1 years, range: 17-24 years) were recruited from the University of Toronto community and received course credit for participation. All participants self-reported no known hearing, language, neurological, or visual deficits. Prior to the experiment, participants provided written informed consent and completed a language background questionnaire. Four participants were excluded due to technical errors during recording. Based on self-reported language background, the remaining 65 participants were categorized into the following groups: Native English listeners ($n = 22$) with no formal or informal experience learning Mandarin Chinese; Heritage Mandarin listeners ($n = 17$), who were raised in an English-speaking community, but whose parent(s) spoke Mandarin; and advanced L1-Mandarin learners of English ($n = 26$), who received formal English instruction and were enrolled at the University of Toronto. Participants self-reported their linguistic background and self-assessed their familiarity with and likelihood of exposure to Mandarin-accented English on a 10-point Likert scale (1: minimal level of familiarity/exposure; 10: highest level). Table 1 provides participant demographics. As expected, Native English listeners reported the lowest familiarity with MAE, whereas Heritage Mandarin listeners and L1-Mandarin listeners showed comparable MAE familiarity. For likelihood of exposure, English listeners also reported the lower rating, while L1-Mandarin listeners reported the highest rating. The experiment was approved by the Research Ethics Board of the University of Toronto.

Table 1. Participant demographics obtained from the language background questionnaire. The median is provided for the self-reported proficiency ratings. One standard deviation of the central tendency measure is presented in parentheses.

	Age of acquisition (English)	English proficiency		Chinese proficiency		MAE exposure	
		Listening	Speaking	Listening	Speaking	Familiarity	Exposure Likelihood
Native English Listeners (<i>n</i> = 22)	NA	10 (0.2)	10 (0)	NA	NA	4 (2.3)	2 (2)
Heritage Mandarin Listeners (<i>n</i> = 17)	2.5 (3)	10 (0.4)	10 (0.9)	8 (2.2)	7 (2.7)	8 (2.2)	6 (2.4)
L1-Mandarin Listeners (<i>n</i> = 26)	6.7 (2.3)	8 (1.1)	7 (1.3)	10 (0.4)	10 (0.2)	8 (1.2)	7 (2.1)

Materials

Experimental stimuli were obtained from the recordings by 10 native Standard Canadian English (SCE) talkers and 10 L1 Mandarin Chinese accented English (MCE) talkers. L1 Mandarin talkers were all born in China (mean length of residence in Canada: 1.83 years, range: 5 months–2 years). SCE talkers were of Chinese descent but were either born in Canada or arrived in Canada before 5 years of age [38,39]. Acoustic recordings were made with a Røde NT5 condenser microphone in a sound-attenuated cabin. Sound files were recorded with 44.1 kHz sampling rate and 16-bit depth encoding on a Mix-Pre 3 (Sound Devices, USA) soundcard. Talkers read a list of seventeen words, repeated six times. Words were selected to contain phonetic properties (e.g., syllable structures, phonemes) that are commonly difficult for L1 Mandarin Chinese English learners to produce and had been previously reported to be familiar to L2 East Asian English speakers [75]. All items were monosyllabic real words of English, included a complex coda [e.g., [p^hakt] “packed”, [ɡlɪmps] “glimpse” 76], and contained one of the following vowels: [æ ʌ ai] [77].

The English word *stopped* was selected as the test item, as the authors deemed it to have the most consistent pronunciation across talkers of the same accent from the word list. Only recordings from female talkers that were clear, produced at a comfortable speech rate, and did not deviate from the common pronunciation among this group were included (*n* = 36). Then, the most tonally neutral repetition per talker was extracted as assessed by the first Author. All stimuli were digitally scaled to have an equal root mean square (RMS) intensity in *Praat* [78]. A seven-point auditory Likert rating task (1 = little accent relative to SCE; 7 = very accented relative to SCE) was administered online to naïve, native Standard Canadian English participants (*n* = 17). One token of the word *stopped* for each talker was presented in pseudorandom order. Participants completed the accentedness rating task on an online form. The ten SCE talkers that were judged to be least accented relative to SCE (median = 1, *SD* = 0) and the ten L1 Mandarin talkers that were judged to be most accented (median = 6; *SD* = 0.95) were selected for inclusion in the main experiment. See the Appendix for the ratings of each talker. Individuals that recorded experimental items were excluded from participation in the EEG experiment.

Procedure

Participants were seated in a quiet, dimly-lit room and passively listened to the stimuli, while watching a silent movie to maintain an awake state and reduce excessive ocular movements [79]. An auditory oddball paradigm was used to elicit the MMN. Auditory stimuli were delivered via Beyerdynamic DT 770 PRO headphones calibrated for 70 dB SLP auditory playback; the sound presentation level was constant across participants. The experiment consisted of two blocks. In the MAE-deviant block, participants were presented with the English word *stopped* produced by SCE talkers as the standard. The deviant was also the English word *stopped* but produced by MAE talkers. In the SCE-deviant block, participants were presented with the English word *stopped* produced by MAE talkers as the standard. The deviant was also the English word *stopped* but produced by SCE

talkers. In both blocks, the experimental tokens were sampled from the twenty different female talkers selected for use in the EEG experiment, ten from each accent group. We opted to present multiple talkers for each accent group, as the selection of just one talker per accent group would have confounded accent change with voice change, and voice changes have been known to elicit MMN responses [62]. The current study presented multiple talkers to ensure that neurophysiological responses reflected abstract categorizations based on accent rather than individual acoustic properties [58,80]. Block order was randomized across participants.

Each deviant was preceded by approximately eight standard stimuli (range: 4–10) for the elicitation of an MMN response [65,81]. The number of standards preceding each deviant was randomly drawn from a uniform distribution. Seventy deviants were presented in each block. Individual tokens were randomly sampled from the ten tokens that belonged to the standard accent or from the ten tokens that belonged to the deviant accent. The interstimulus interval duration was randomly sampled from a uniform distribution between 0.7–0.95 seconds. The experimental design was a mixed 3 (Group: Native English Listeners, Heritage Mandarin Listeners, Mandarin Learners of English) $\times$ 2 (Block: SCE-deviant, MAE-deviant) $\times$ 2 (Stimulus: standard, deviant) design, with Group being a between-participants factor and both Block and Stimulus being within-participants factors.

EEG Recording and Analysis

Continuous EEG recordings were acquired from 32 actiCAP active electrodes connected to an actiCHamp amplifier (Brain Products GmbH). The EEG signal was digitized at a 500 Hz sampling frequency with a 200 Hz on-line low-pass filter. Electrodes were positioned on the scalp according to the International 10–20 system. Positions included Fp1/2, F3/4, F7/8, FC1/2, FC5/6, FT9/10, C3/4, T7/8, CP1/2, CP5/6, TP9/10, P3/4, P7/8, O1/2, Oz, Fz, Cz, and Pz. A ground electrode was placed at Fpz. The EEG signal was referenced to the right mastoid (TP10) on-line. Impedances were reduced to below 10 k Ω at each electrode site prior to recording. The experiment was deployed using PsychoPy [82]. In addition to the EEG channels, the auditory signal was also sent to the amplifier using the StimTrak device (Brain Products GmbH). This allows for off-line correction of temporal delays between delivery of the auditory stimulus and the digital trigger marker [83]

EEG recordings were preprocessed using MNE-Python [Version 1.6.1] [84]. A band-pass filter from 1 to 100 Hz was applied to the continuous EEG signal, which was then down sampled to 250 Hz. Bad channels were identified and interpolated using the PyPREP toolbox [85]. Independent Component Analysis (ICA) was performed to identify and remove artifacts. To run the ICA, data were average re-referenced and segmented into epochs ranging from –100 ms–900 ms relative to stimulus onset. ICA was performed using the extended infomax algorithm, and components were automatically classified using the ICLabel classifier [86]. Components not classified as *brain* or *other* within the first 15 components (ranked by variance explained) were excluded. All classifications were manually inspected to ensure that salient artifacts (e.g., eye blinks, saccades) were rejected (mean = 6 components per participant, SD = 2). The ICA solution was then applied to the continuous data, which were subsequently re-referenced to the average linked mastoid [87].

EEG Analysis

Event-Related Potentials

The preprocessed continuous EEG signal was segmented into epochs time-locked to stimulus onset, spanning –100 ms pre-stimulus to 900 ms post-stimulus. Epochs with peak-to-peak amplitudes greater than 100 μ V were rejected (accounting for 3% of total trials). To ensure that the ERP to standard tokens reflected an established memory trace of the abstract accentual category, the first two standards in each standard train were excluded. For each participant, standards and deviants were averaged separately in each block. The identity MMN (iMMN) was computed for each accent by subtracting the ERP to the accent serving as standards in one block from the ERP to the same accent serving as deviants in the other block [88]. This was done to isolate the MMN response from

ERP differences intrinsic to physical property difference between stimuli. For example, the SCE iMMN was calculated by subtracting the ERP to SCE standards in the MAE deviant block from the ERP to SCE deviants in the SCE deviant block. The presence of an iMMN suggests that observed differences are attributable to accent status rather than physical stimulus properties. Inferential statistics were conducted using the spatiotemporal cluster-based permutation test applied to the 0–900 ms time interval. A two-tailed t -statistic threshold corresponding to $p = 0.05$ identified time samples and electrodes showing significant deviation from zero. Cluster-based correction for multiple comparisons was applied with a $p = 0.05$ significance threshold.

Event-Related Spectral Perturbation

To further explore the mechanisms underlying accent processing, we computed the Event-Related Spectral (ERSP) to determine how accent modulates the oscillatory dynamics in each participant Group (Native English listeners, Heritage Mandarin listeners, advanced English learners). The ERSP epochs were extracted from the preprocessed continuous EEG signal using a time window from –1.0 seconds to 2.0 seconds relative to stimulus onset. Artifact rejection was applied using a peak-to-peak threshold of 100 μ V (7% of trials). The time-frequency decomposition was performed using a Morlet wavelet convolution. Frequencies ranged from 3 to 30 Hz, linearly spaced across 28 steps. The number of wavelet cycles increased linearly with frequency, starting at 3 cycles and increasing by 0.8 steps per frequency bin. This approach balances temporal and spectral resolution. ERSPs were computed for each trial and averaged across epochs and all EEG channels for each standard and deviant condition in each group. Power estimates were baseline corrected using log-ratio normalization, with a baseline window of –450 to –300 ms pre-stimulus. Group-level statistical comparisons were conducted using a cluster-based permutation test [89], with cluster adjacency defined across spatial (channel), spectral (frequencies), and temporal (time samples) dimensions. The threshold for initial cluster formation was set to the t -value corresponding to $p = 0.01$ (two-tailed), and statistical significance was assessed using a cluster-level permutation at $p < 0.05$. Analyses were restricted to the 3–30 Hz frequency range and the –300 to 700 ms post-stimulus interval.

Generalized Additive Modeling

To assess whether brain responses to MAE are modulated by individual experience with Mandarin, we implemented Generalized Additive Models (GAMs) following Meulman et al. [90]. GAMs were built using the `bam()` function from the `mgcv` package [91] in R [92] to capture nonlinear temporal dynamics in EEG signals, resulting in the following model:

$$\mu V \sim s(\text{Time}, \text{by} = \text{Condition}) + s(\text{Time}, \text{Item}, \text{bs} = "fs", m = 1)$$

Condition represents the Block (SCE-deviant, MAE-deviant) $\times$ Stimulus (standard, deviant) interaction. *Item* refers to the specific auditory stimulus. The term $s(\text{Time}, \text{Item}, \text{bs} = "fs", m = 1)$ allows for a separate smooth over time for each stimulus, accounting for item-level variability and penalize overfitting. The model was applied to single-trial EEG data, averaged over nine frontocentral electrodes (i.e., Cz, Fz, FC1, FC2, CP1, CP2, C3, C4, Pz), where the MMN response is observed [50]. For each participant and each accent, a difference ERP (i.e., deviant ERP minus standard ERP, for the same accent across blocks) was derived from the model-fitted values using the `difference_smooth()` function from the `gratia` package [93]. Based on the model-derived difference waveform, the MMN component was identified as the negative deflection containing the global negative peak within the 0–900 ms post-stimulus interval. For each participant and each accent condition, two individual-level measures were extracted. First, the normalized modeled peak was calculated. This is the peak amplitude of the identified MMN component, divided by 1.96 times the standard error of the fitted value at that time point. This measure indicates response robustness and has been reported to reliably reflect language background [90]. Second, the half-area latency was calculated. This is the time point at which 50% of the total area under the negative deflection has been reached, representing the

temporal dynamics of the brain response [94]. The latter measure was motivated by the findings reported in Scharinger et al. [73], who observed MMN timing differences due to dialect familiarity.

In addition to the GAM-derived measures, we also extracted a traditional ERP amplitude measure for each participant: the mean ERP amplitude within the time window and electrode region identified by the group-level spatiotemporal permutation test. These three individual-level neural measures were then correlated with participant responses to experience with MAE, including the familiarity and the likelihood of exposure, as well as their Chinese language proficiency (for advanced English learners only).

Results

Event Related Potentials

Figure 1 displays the ERP waveforms for each group, averaged over the eight frontocentral electrodes (Fz, FC1, FC2, Cz, C3, C4, CP1, CP2). For SCE (upper panel), a clear negative deflection for SCE deviants relative to SCE standards was observed in both Native English listeners (Figure 1A) and L1-Mandarin listeners (Figure 1C).

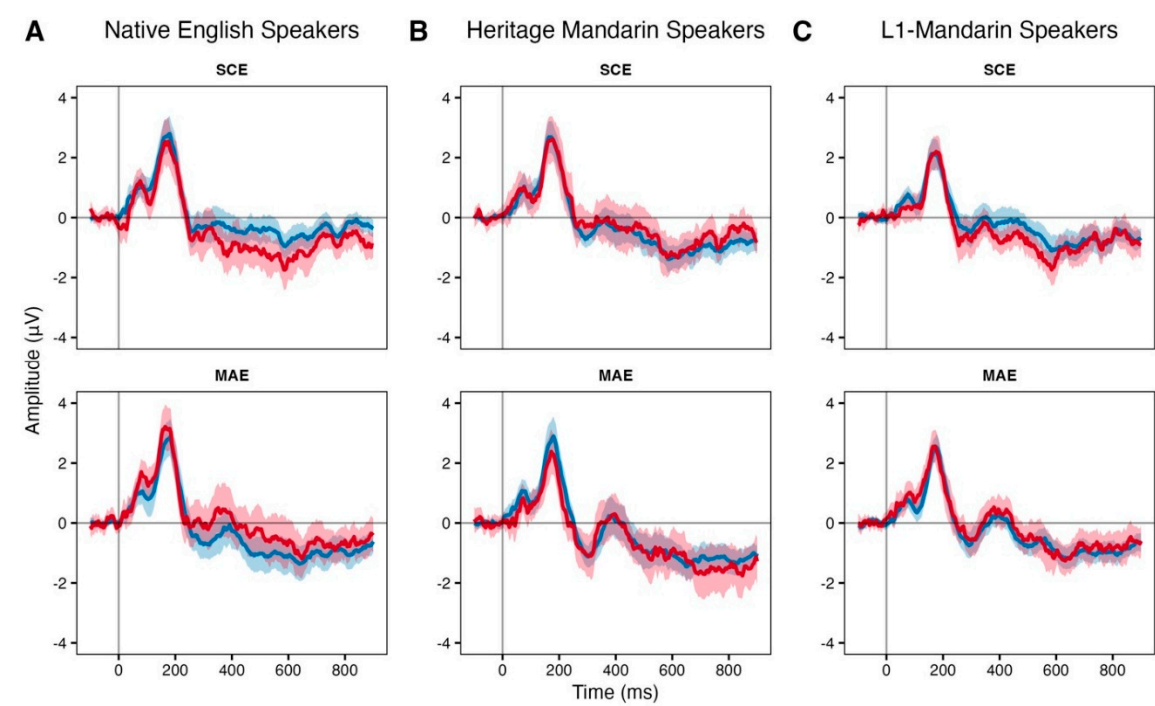


Figure 1. Event-related potentials (ERP) to standards (blue) and deviants (red) averaged over frontocentral electrodes for Native English listeners (A), Heritage Mandarin listeners (B), and L1-Mandarin listeners (C). For each group, the upper panel shows the ERPs to SCE as the standard (from MAE-deviant block) and deviant (from SCE-deviant block); the lower panel shows the ERPs to MAE as the standard (from the SCE-deviant block) and deviant (from the MAE-deviant block). Shaded regions represent the 95% confidence interval of the mean ERP waveform.

To determine whether accent deviation elicited a robust iMMN response regardless of talker variability, we conducted cluster-based permutation tests on the difference ERP waveforms, derived from comparing each accent when presented as deviants against the same accent when presented as standards. This approach offers a direct comparison to Scharinger et al. [73], who also computed the iMMN. Figure 2 and Figure 3 display the waveforms and topographies, respectively, corresponding to the permutation test results. Waveforms were averaged over the electrodes contributing to the significant spatiotemporal clusters, and the topographies reflect ERP activity averaged over

significant time windows. For Native English listeners, two clusters were identified. The first cluster (456–592 ms) showed a frontocentral negativity for the difference between the SCE deviant (from SCE-deviant block) and the SCE standard (from MAE-deviant block). The waveform revealed that the SCE deviant elicited a more negative response, and the frontocentral distribution of this difference aligns with classic MMN topographies [45]; The second cluster (252–460 ms) showed a frontocentral positivity for the MAE deviants compared to the MAE standards. The reversed polarity is not typical of MMN responses but aligns with previous finding using inter-category many-to-one oddball paradigm, particularly when the standard is the unmarked category [54,57]. For L1-Mandarin listeners, one cluster was found (380–604 ms). The topography showed a frontocentral negativity for the SCE deviant relative to the SCE standard, mirroring the pattern in Native English listeners and supporting the presence of a late MMN. No clusters were observed for the MAE deviant. In contrast, the Heritage Mandarin listeners showed no clusters for either accent contrast, suggesting an attenuated iMMN response.

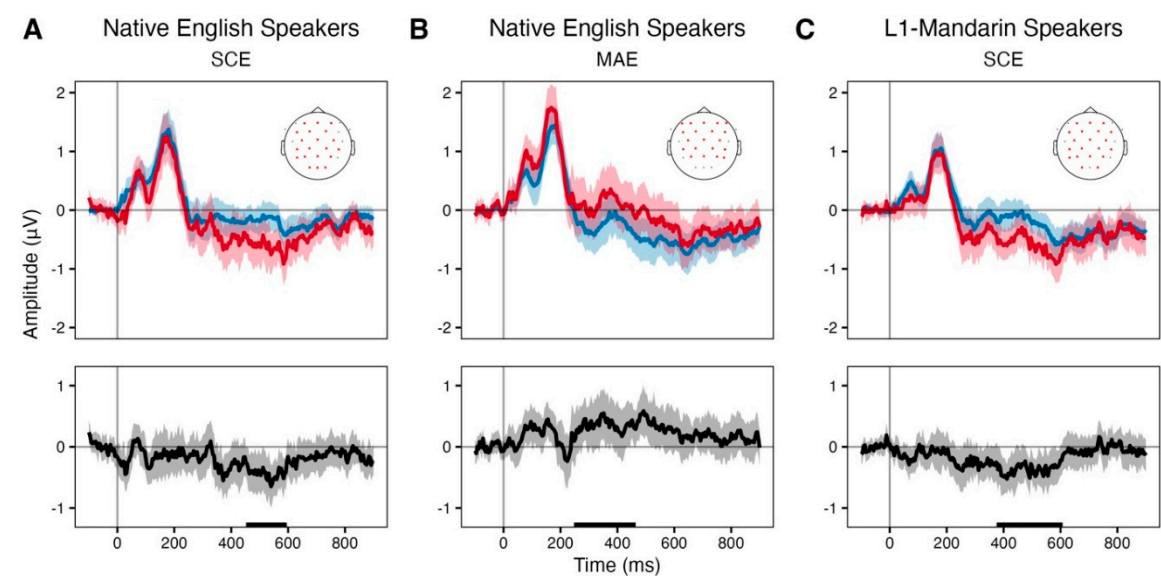


Figure 2. Event-related potentials to the standard (blue) and deviant (red) stimuli, as well as the corresponding difference waveform (deviant – standard, black) averaged over electrodes within the significant cluster (highlighted in red in the topography). (A) ERP to SCE serving as standard (from MAE-deviant block) and deviant (from SCE-deviant block) in Native English listeners. (B) ERPs to MAE serving as the standard (from SCE-deviant block) and deviant (from MAE-deviant block) in Native English listeners. (C) ERPs to SCE serving as the standard (from MAE-deviant block) and deviant (from SCE-deviant block) in L1-Mandarin learners. Shaded areas represent 95% confidence interval of the ERP waveform. Rug plots along x-axes mark time samples within the significant cluster. Only conditions associated with significant clusters are shown.

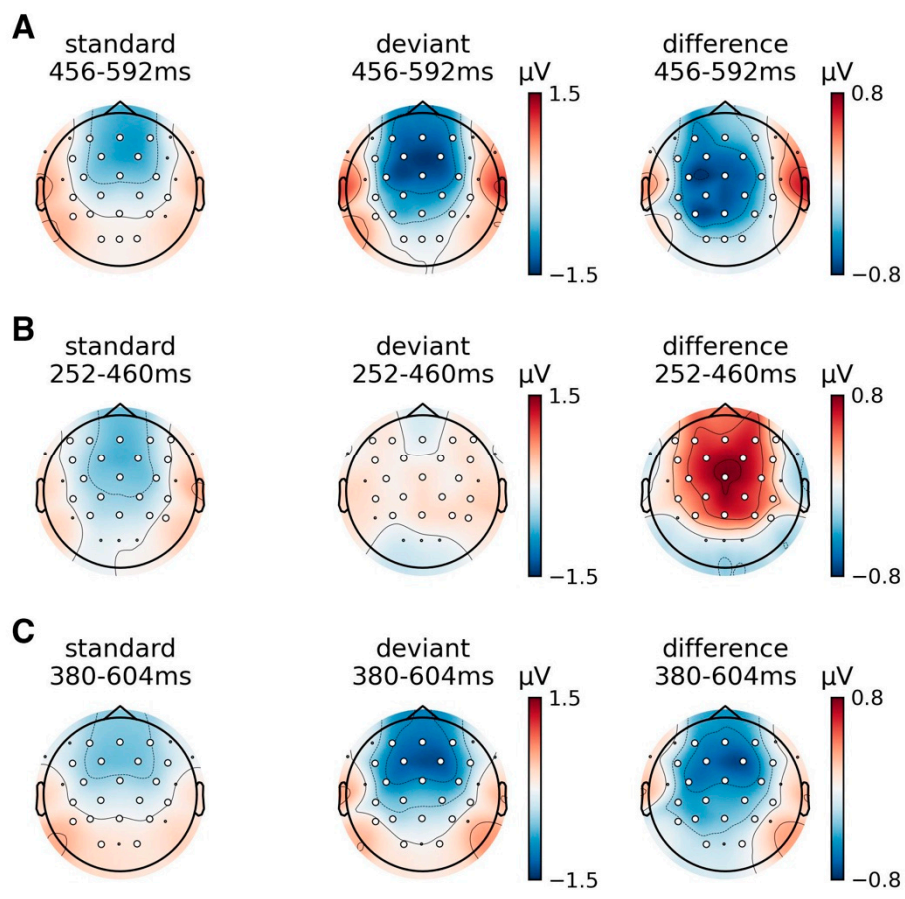


Figure 3. Scalp ERP topographies averaged over time samples identified by the spatiotemporal permutation test for the standard (left), deviant (middle) and their difference (deviant – standard, right). (A) SCE in Native English listeners. (B) MAE in Native English listeners. (C) SCE in L1-Mandarin listeners. Only conditions associated with significant clusters are shown. Electrode sites marked with white circles indicate where there was a significant difference identified within the time window.

To summarize, iMMN responses to SCE accent were observed in both Native English listeners and L1-Mandarin learners of English, suggesting that in both populations, the auditory cortex robustly detected accent-based deviations. In contrast, this accent-based neural discrimination was absent in Heritage Mandarin listeners, suggesting that these participants may have treated SCE and MAE as belonging to the same category or as insufficiently distinct to trigger an automatic deviance response.

Event-Related Spectral Perturbation

To assess whether neural activity underlying the accent difference detection varied across groups, we analyzed the oscillatory power of the EEG responses to the standard tokens. Given the lack of specific hypothesis regarding relevant frequency bands, we conducted a time-frequency analysis spanning 3–30 Hz and extending to 900 ms post-stimulus onset. Permutation tests comparing the two standard ERSPs in Native English listeners revealed a cluster in which the MAE standard elicited increased oscillatory power relative to the SCE standard (Figure 4). The cluster showed increased power in the low- β band (12–13 Hz) approximately -150 ms pre-stimulus to 250 ms post-stimulus, and in the α -band (8–12 Hz) sustaining up to 750 ms post-stimulus. The increase in low β -power aligns with prior work suggesting that β -activity reflects top-down predictive coding [57,95], suggesting that the MAE standard was perceived as less expected. The sustained enhancement in α -power likely reflects greater processing difficulty [96] or top-down inhibition [97] when encoding the less familiar MAE speech. In contrast, no ERSP differences between SCE and MAE

standards were observed for either Heritage Mandarin listeners or L1-Mandarin listeners, suggesting comparable oscillatory dynamics for the two accent types in bilingual listeners. This may reflect increased experience with both accents and hence a reduced neural distinction between them at the level of predictive and attention-related oscillatory activity.

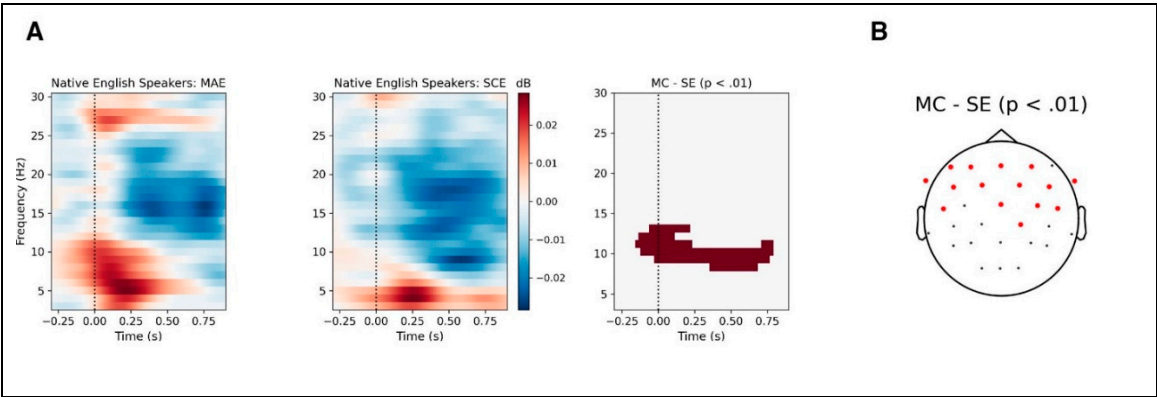


Figure 4. Cluster-based permutation test comparing the SCE standard and the MAE standard in Native English listeners. (A) ERSP for SCE standards (left) and MAE standards (center). Right panel shows the time-frequency bins for the significant cluster. The ERSP response to the MAE standard shows significantly higher oscillatory power compared to the SCE standard in the α -band (8–12 Hz) from -150 to 750 ms, and in the low β -band (12–13 Hz) from -150 ms to 250 ms. (B) Topographic distribution of electrodes in the significant cluster at one or more time-frequency bins marked in red.

Generalized Additive Models

To examine how neurophysiological responses were modulated by individual language experience and whether such relationships differed across language groups, we extracted individual-level brain measures from fitted GAM models. These measures included normalized modeled peak and half-area latency, which index the magnitude and timing of the MMN response, respectively. For comparison with previous studies, we also computed a traditional amplitude measure of the MMN, defined as the mean amplitude averaged over the 400–600 ms time window and eight frontocentral electrodes (i.e., Fz, FC1, FC2, CP1, CP2, Cz, C3, C4). For each language experience variable (i.e., MAE familiarity, MAE likelihood) and for each MMN condition (i.e., iMMN to MAE deviant, iMMN to SCE deviant), we fit linear regression models regressing the MMN measure to Group (Native English, Heritage Mandarin, and L1-Mandarin), Rating, and their interaction. Interactions were followed up using the *emmeans* package [98] to estimate slopes within each group and to test whether the slopes differed across groups.

For the normalized modeled peak measure, the model revealed a trending Group $\times$ MAE-likelihood Rating interaction: $F(2, 56) = 3.08, p = 0.05$. Individual slopes did not reach significance. Follow-up pairwise slope comparisons indicated that the interaction was driven by opposite patterns in Native English listeners and L1-Mandarin listeners: $t(56) = 2.46, p = 0.04$. Native English listeners exhibited a positive correlation where greater likelihood of hearing MAE was associated with a larger MMN response to the SCE deviant (Figure 5A). In contrast, L1-Mandarin listeners showed a negative slope: Greater exposure to MAE correlated with reduced MMN responses to the SCE deviants. No correlations were observed for half-area latency or for the traditional amplitude measure.

For Heritage Mandarin listeners, we examined whether the MMN was predicted by Chinese proficiency by correlating each MMN measure (i.e., normalized modeled peak, half-area latency, traditional amplitude measure) with participant’s self-rated proficiency scores. Since Chinese speaking and listening ratings correlated ($r = 0.57, p < 0.001$), we averaged the two to derive a single composite proficiency score. Only the traditional amplitude measure showed an association with proficiency (Figure 5B): Higher Chinese proficiency predicted reduced (more positive) MMN

responses to the SCE deviant ($r = 0.75, p = 0.001$) but was also associated with more negative MMN responses to the MAE deviant ($r = -0.7, p = 0.004$).

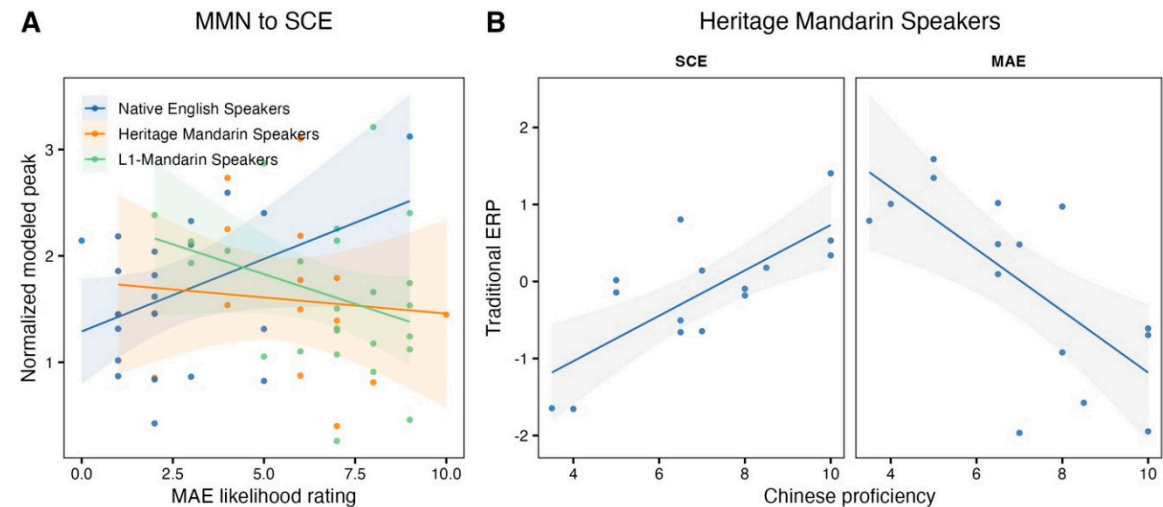


Figure 5. (A) Relationship between MAE likelihood ratings and the normalized modeled peak (more positive values indicating stronger MMNs) for each language group. (B) Scatterplots showing the relationship between self-reported Chinese proficiency and the traditional amplitude MMN measure (more negative values indicating stronger MMNs) for Heritage Mandarin listeners.

Discussion

The present study tested whether the auditory system pre-attentively generalizes accent beyond inter-talker variability, and whether such neurophysiological categorization is shaped by language experience. Using a multi-talker oddball design, Native English listeners and L1-Mandarin learners showed an MMN when SCE served as the deviant accent, indicating that accent change can be detected at an abstract level beyond talker-specific acoustics. Native English listeners exhibited an accent-familiarity asymmetry, consistent with prior accent MMN studies showing an earlier or more robust MMN for the more familiar accent deviant [64,99]. Time-frequency results provide converging evidence: MAE standards elicited increased α (8–12 Hz) and low- β (12–13 Hz) power relative to the SCE standard in Native English listeners. Oscillatory responses in the β -band have been linked to top-down signaling and prediction maintenance in speech perception [57,97,100]. The relative low- β power increase for MAE standards is consistent with the idea that Native English listeners engaged stronger top-down predictive control in the categorization of an unfamiliar accent. In parallel, enhanced α -power indexes greater cognitive demand and effortful suppression of competing information in challenging auditory processing contexts [96,97]. For Native English listeners, MAE may have led to greater processing difficulty and lower intelligibility, as it was less familiar, thereby increasing the need for sustained control (β -power) and effort/inhibitory regulation (α -power) during neurophysiological encoding of the standard.

In contrast, L1-Mandarin learners did not show a familiarity-driven MMN when MAE served as the deviant, despite showing an MMN to the SCE deviant. One explanation to this unexpected pattern, that is, a familiarity effect in Native English listeners but not in L1-Mandarin listeners, is that the observed MMN was driven by low-level auditory processing rather than accent-level categorization. Specifically, when the same contrast is tested in opposite directions (e.g., A deviants among B standards versus B deviants among A standards), the MMN amplitude can exhibit a deviance-direction asymmetry that arises from the interaction between the physical change (e.g., duration increment versus decrement) and stimulus role (standard versus deviant) [101–103]. Under this view, if the SCE deviant was more acoustically salient when embedded among MAE standards than in the reverse configuration, then a larger MMN for the SCE deviant could emerge without requiring accent-level categorization; however, a purely acoustic account cannot explain why

Heritage Mandarin listeners did not show a similar asymmetry (i.e., MMN to SCE but not to MAE) and no reliable MMN to either accent, albeit the smaller sample size. Furthermore, the oscillatory pattern found in Native English listeners (i.e., increased low- β/α power during standard encoding to MAE relative to SCE) was absent in L1-Mandarin learners.

Although the L1-Mandarin listeners had substantial exposure to MAE, this does not entail that SCE was an unfamiliar accent, nor that MAE should be more familiar than SCE. The current L1-Mandarin listeners were advanced English users living in an English-speaking community, attended university classes taught in English, and routinely communicated in English with English-speaking peers. Therefore, even if they encountered MAE more often and might understand MAE more easily relative to Native English listeners, immersion in an English-speaking environment likely led to greater SCE familiarity, supporting robust SCE perception and lexical access. Our stimuli were also English words, and as such, L2 lexical items for the L1-Mandarin participants. Previous work on L2 speech perception suggests that as L2 experience and proficiency increase, neural indices of L2 speech processing tend to be more native-like [104–106]. Extensive L2 experience with English lexical items may place learners in a perceptual state that is more tuned for SCE relative to MAE. Evidence from a sentence acceptability judgement task [107] suggests that L2 listeners show increased sensitivity to morphosyntactic violations in the non-native language produced in the target-language accent, relative to a nonnative-accented version. Specifically, Gosselin and colleagues reported that English learners of Spanish were better at detecting the Spanish gender and number errors pronounced in a native Spanish accent than in English-accented Spanish. Applied to the present context, L1-Mandarin listeners might be sensitive to words produced in an SCE accent, but this sensitivity does not necessarily result in increased lexical processing effort relative to MAE. In this light, the absence of an ERSP difference between MAE and SCE standards in L1-Mandarin learners may reflect reduced accent-dependent processing difficulty relative to Native English listeners. Summarizing this account, familiarity may still contribute to the observation of an MMN to SCE in both Native English listeners and L1-Mandarin learners, but the additional processing challenges indexed by sustained β/α -differences may be specific to Native English listeners, rather than a uniform consequence of accent familiarity across groups.

Regarding Heritage Mandarin listeners, we found no spatiotemporal clusters for either accent deviant, and no ERSP difference between SCE and MAE standards. These null results suggest that, for Heritage Mandarin listeners who began acquiring both Mandarin and English early in life, MAE and SCE may be accommodated within an expanded phonological space. In other words, the deviant accent may be perceived as part of the expected realizations compatible with the accent representation of the standard accent and therefore, does not trigger a pre-attentive change-detection response under high talker variability. Overall, the MMN and ERSP results suggest that the auditory system can abstract an accent-level template that generalizes over inter-talker variability. Moreover, the MMN group differences are consistent with experience-dependent familiarity effects, aligning with prior findings of enhanced sensitivity to familiar voices [32,108], more efficient lexical access [109], and grammatical error detection in speech produced in a native accent [110,111].

GAM-based individual-difference analyses indicated that individual accent experience may reshape the accent perceptual space in a language-specific manner. Specifically, for Native English listeners, a higher likelihood of hearing MAE was associated with a larger MMN to SCE deviants, whereas for L1-Mandarin listeners, greater MAE exposure was associated with a reduced MMN to SCE deviants. It is possible that, for Native English listeners, increased contact with MAE sharpens the accent contrast such that MAE becomes a more clearly delimited accent category, yielding stronger predictions based on the MAE standards, hence a stronger prediction error elicited by SCE deviants. For L1-Mandarin listeners, increased MAE exposure may instead increase overlap between MAE and SCE within a broadened English phonological space, reducing sensitivity to SCE deviants. In this sense, exposure can either increase contrast (category sharpening) or increase tolerance (category blending), depending on the learner's starting point and the functional role of MAE in daily

input; that said, these individual-level effects should be interpreted cautiously given the marginal level of statistical significance.

For Heritage Mandarin listeners, higher Chinese proficiency may shift processing toward an L1-Mandarin-like pattern, in which MAE is perceived as relatively unmarked, yielding a reduced MMN when MAE serves as the standard, but an enhanced MMN when MAE serves as the deviant. Such a proficiency effect would imply that heritage experience is not categorical; rather, it may place listeners along a continuum of accent representations that re-weight expectations. In this sense, increased Chinese proficiency may bias Heritage Mandarin listeners toward an L1-Mandarin-like MAE weighting, consistent with the reduced MMN in the MAE-standard configuration and the enhanced MMN in the MAE-deviant configuration.

Conclusion

The current study provides neurophysiological evidence that the auditory system can extract accent-level regularities that generalize across talker variability. This is reflected in the MMN to the SCE deviant in both Native English and L1-Mandarin listeners. Accent processing is also shaped by language experience: familiarity and language background influence accent differences both in deviant detection and in oscillatory processing dynamics. Specifically, Native English listeners showed a familiarity-driven MMN, as well as the ERSP difference, during standard processing, consistent with greater processing demands for MAE relative to SCE, whereas L1-Mandarin listeners showed an MMN to SCE deviants but no corresponding ERSP difference, consistent with more comparable processing efficiency across accents. Finally, heritage experience appears to potentially reorganize the phonological space in qualitatively different ways. With greater exposure to MAE, Heritage Mandarin performance was more L1-Mandarin-like, highlighting that accent familiarity is not a unitary mechanism, but interacts with native phonology and the dynamics of the language learning.

Supplemental Materials: Stimuli and EEG datasets are publicly available in the Open Science Framework: https://osf.io/8du7s/overview?view_only=864869e97b3b466487e378bbc1648c27.

Author Contributions: Conceptualization, design: LH, PJM; Running the experiment: LH; Analysis: CH, LH; Writing: LH, CH, PJM.:

Funding: This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada, grant number: RGPIN-2024-06584.

Institutional Review Board: The study was conducted in accordance with the Declaration. of Helsinki and approved by the Research Ethics Board of the University of Toronto (Protocol #: 31245; approved on 11 December 2023).

Informed Consent Statement: All participants provided informed consent prior to taking part in the Experiment.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix

Table A1. Average and median ratings of each talker sample, and the overall averages, medians, and standard deviation of the medians of each accent group. Talkers are listed according to recruitment order per accent.

Talker	Median Rating	Standard Deviation
Chinese Accented		
1	5	1.57
2	7	1.45
3	4	1.18

4	6	1.2
5	7	0.75
6	5	1.6
7	5	1.28
8	6	1.59
9	6	1.33
10	6	1.58
Overall	6	0.95 (median)
Canadian English		
11	1	1.17
12	1	0.92
13	1	1.17
14	1	1.22
15	1	1.23
16	1	1.41
17	1	1.2
18	1	0.93
19	1	2.1
20	1	1.44
Overall	1	0 (median)

References

1. Bent, T.; Holt, R.F. Representation of Speech Variability. *WIREs Cogn. Sci.* **2017**, *8*, e1434, doi:10.1002/wcs.1434.

2. Hartman, D.E. The Perceptual Identity and Characteristics of Aging in Normal Male Adult Speakers. *J. Commun. Disord.* **1979**, *12*, 53–61, doi:10.1016/0021-9924(79)90021-2.

3. Hartman, D.E.; Danhauer, J.L. Perceptual Features of Speech for Males in Four Perceived Age Decades. *J. Acoust. Soc. Am.* **1976**, *59*, 713–715, doi:10.1121/1.380894.

4. Ptacek, P.H.; Sander, E.K. Age Recognition from Voice. *J. Speech Hear. Res.* **1966**, *9*, 273–277, doi:10.1044/jshr.0902.273.

5. Shipp, T.; Hollien, H. Perception of the Aging Male Voice. *J. Speech Hear. Res.* **1969**, *12*, 703–710, doi:10.1044/jshr.1204.703.

6. Hillenbrand, J.M.; Clark, M.J. The Role of F0 and Formant Frequencies in Distinguishing the Voices of Men and Women. *Atten. Percept. Psychophys.* **2009**, *71*, 1150–1166, doi:10.3758/APP.71.5.1150.

7. Lass, N.J.; Hughes, K.R.; Bowyer, M.D.; Waters, L.T.; Bourne, V.T. Speaker Sex Identification from Voiced, Whispered, and Filtered Isolated Vowels. *J. Acoust. Soc. Am.* **1976**, *59*, 675–678, doi:10.1121/1.380917.

8. Mullennix, J.W.; Johnson, K.A.; Topcu-Durgun, M.; Farnsworth, L.M. The Perceptual Representation of Voice Gender. *J. Acoust. Soc. Am.* **1995**, *98*, 3080–3095, doi:10.1121/1.413832.

9. Remez, R.E.; Fellowes, J.M.; Rubin, P.E. Talker Identification Based on Phonetic Information. *J. Exp. Psychol. Hum. Percept. Perform.* **1997**, *23*, 651–666, doi:10.1037/0096-1523.23.3.651.

10. Babel, M.; Russell, J. Expectations and Speech Intelligibility. *J. Acoust. Soc. Am.* **2015**, *137*, 2823–2833, doi:10.1121/1.4919317.

11. Baugh, J. Racial Identification by Speech. *Am. Speech* **2000**, *75*, 362–364, doi:10.1215/00031283-75-4-362.

12. Perrachione, T.K.; Chiao, J.Y.; Wong, P.C.M. Asymmetric Cultural Effects on Perceptual Expertise Underlie an Own-Race Bias for Voices. *Cognition* **2010**, *114*, 42–55, doi:10.1016/j.cognition.2009.08.012.

13. Thomas, E.R.; Reaser, J. Delimiting Perceptual Cues Used for the Ethnic Labeling of African American and European American Voices. *J. Socioling.* **2004**, *8*, 54–87, doi:10.1111/j.1467-9841.2004.00251.x.

14. Van Dommelen, W.A.; Moxness, B.H. Acoustic Parameters in Speaker Height and Weight Identification: Sex-Specific Behaviour. *Lang. Speech* **1995**, *38*, 267–287, doi:10.1177/002383099503800304.

15. Smith, D.R.R.; Patterson, R.D.; Turner, R.; Kawahara, H.; Irino, T. The Processing and Perception of Size Information in Speech Sounds. *J. Acoust. Soc. Am.* **2005**, *117*, 305–318, doi:10.1121/1.1828637.

16. Flege, J.E. The Detection of French Accent by American Listeners. *J. Acoust. Soc. Am.* **1984**, *76*, 692–707, doi:10.1121/1.391256.
17. Park, H. Detecting Foreign Accent in Monosyllables: The Role of L1 Phonotactics. *J. Phon.* **2013**, *41*, 78–87, doi:10.1016/j.wocn.2012.11.001.
18. Kinzler, K.D. Language as a Social Cue. *Annu. Rev. Psychol.* **2021**, *72*, 241–264, doi:10.1146/annurev-psych-010418-103034.
19. Purnell, T.; Idsardi, W.; Baugh, J. Perceptual and Phonetic Experiments on American English Dialect Identification. *J. Lang. Soc. Psychol.* **1999**, *18*, 10–30, doi:10.1177/0261927X99018001002.
20. Formisano, E.; De Martino, F.; Bonte, M.; Goebel, R. “Who” Is Saying “What”? Brain-Based Decoding of Human Voice and Speech. *Science* **2008**, *322*, 970–973, doi:10.1126/science.1164318.
21. Binder, J. The New Neuroanatomy of Speech Perception. *Brain* **2000**, *123*, 2371–2372, doi:10.1093/brain/123.12.2371.
22. Hickok, G.; Poeppel, D. The Cortical Organization of Speech Processing. *Nat. Rev. Neurosci.* **2007**, *8*, 393–402, doi:10.1038/nrn2113.
23. Liebenthal, E.; Binder, J.R.; Spitzer, S.M.; Possing, E.T.; Medler, D.A. Neural Substrates of Phonemic Perception. *Cereb. Cortex* **2005**, *15*, 1621–1631, doi:10.1093/cercor/bhi040.
24. Mesgarani, N.; Cheung, C.; Johnson, K.; Chang, E.F. Phonetic Feature Encoding in Human Superior Temporal Gyrus. *Science* **2014**, *343*, 1006–1010, doi:10.1126/science.1245994.
25. Obleser, J.; Eisner, F. Pre-Lexical Abstraction of Speech in the Auditory Cortex. *Trends Cogn. Sci.* **2009**, *13*, 14–19, doi:10.1016/j.tics.2008.09.005.
26. Scott, S.K.; Johnsrude, I.S. The Neuroanatomical and Functional Organization of Speech Perception. *Trends Neurosci.* **2003**, *26*, 100–107, doi:10.1016/S0166-2236(02)00037-1.
27. Yi, H.G.; Leonard, M.K.; Chang, E.F. The Encoding of Speech Sounds in the Superior Temporal Gyrus. *Neuron* **2019**, *102*, 1096–1110, doi:10.1016/j.neuron.2019.04.023.
28. Belin, P.; Zatorre, R.J. Adaptation to Speaker’s Voice in Right Anterior Temporal Lobe: *NeuroReport* **2003**, *14*, 2105–2109, doi:10.1097/00001756-200311140-00019.
29. Kriegstein, K. von; Smith, D.R.R.; Patterson, R.D.; Kiebel, S.J.; Griffiths, T.D. How the Human Brain Recognizes Speech in the Context of Changing Speakers. *J. Neurosci.* **2010**, *30*, 629–638, doi:10.1523/JNEUROSCI.2742-09.2010.
30. Drozdova, P.; Van Hout, R.; Scharenborg, O. Talker-Familiarity Benefit in Non-Native Recognition Memory and Word Identification: The Role of Listening Conditions and Proficiency. *Atten. Percept. Psychophys.* **2019**, *81*, 1675–1697, doi:10.3758/s13414-018-01657-5.
31. Johnsrude, I.S.; Mackey, A.; Hakyemez, H.; Alexander, E.; Trang, H.P.; Carlyon, R.P. Swinging at a Cocktail Party: Voice Familiarity Aids Speech Perception in the Presence of a Competing Voice. *Psychol. Sci.* **2013**, *24*, 1995–2004, doi:10.1177/0956797613482467.
32. Beauchemin, M.; De Beaumont, L.; Vannasing, P.; Turcotte, A.; Arcand, C.; Belin, P.; Lassonde, M. Electrophysiological Markers of Voice Familiarity. *Eur. J. Neurosci.* **2006**, *23*, 3081–3086, doi:10.1111/j.1460-9568.2006.04856.x.
33. Plante-Hébert, J.; Boucher, V.J.; Jemel, B. The Processing of Intimately Familiar and Unfamiliar Voices: Specific Neural Responses of Speaker Recognition and Identification. *PLOS ONE* **2021**, *16*, e0250214, doi:10.1371/journal.pone.0250214.
34. del Giudice, R.; Lechinger, J.; Wislowska, M.; Heib, D.P.J.; Hoedlmoser, K.; Schabus, M. Oscillatory Brain Responses to Own Names Uttered by Unfamiliar and Familiar Voices. *Brain Res.* **2014**, *1591*, 63–73, doi:10.1016/j.brainres.2014.09.074.
35. Giraud, A.-L.; Poeppel, D. Cortical Oscillations and Speech Processing: Emerging Computational Principles and Operations. *Nat. Neurosci.* **2012**, *15*, 511–517, doi:10.1038/nn.3063.
36. Poeppel, D.; Assaneo, M.F. Speech Rhythms and Their Neural Foundations. *Nat. Rev. Neurosci.* **2020**, *21*, 322–334, doi:10.1038/s41583-020-0304-4.
37. Schilling, N.; Marsters, A. Unmasking Identity: Speaker Profiling for Forensic Linguistic Purposes. *Annu. Rev. Appl. Linguist.* **2015**, *35*, 195–214, doi:10.1017/S0267190514000282.

38. Newman, M.; Wu, A. "Do You Sound Asian When You Speak English?" Racial Identification and Voice in Chinese and Korean Americans' English. *Am. Speech* **2011**, *86*, 152–178, doi:10.1215/00031283-1336992.
39. Xue, S.An.; Hao, J.G. Normative Standards for Vocal Tract Dimensions by Race as Measured by Acoustic Pharyngometry. *J. Voice* **2006**, *20*, 391–400, doi:10.1016/j.jvoice.2005.05.001.
40. Squires, G.D.; Chadwick, J. Linguistic Profiling: A Continuing Tradition of Discrimination in the Home Insurance Industry? *Urban Aff. Rev.* **2006**, *41*, 400–415, doi:10.1177/1078087405281064.
41. Wright, K.E. Housing Policy and Linguistic Profiling: An Audit Study of Three American Dialects. *Language* **2023**, *99*, e58–e85.
42. Grey, K.L. Deviant Bodies, Stigmatized Identities, and Racist Acts: Examining the Experiences of African-American Gamers in Xbox Live. *New Rev. Hypermedia Multimed.* **2012**, *18*, 261–276, doi:10.1080/13614568.2012.746740.
43. Berry, J. Linguistic Profiling: Its Prevalence and Consequence. *Mich. Acad.* **2021**, *47*, 78.
44. Näätänen, R. *Attention and Brain Function*; Routledge, 1992; ISBN 978-0-429-48735-4.
45. Näätänen, R.; Paavilainen, P.; Rinne, T.; Alho, K. The Mismatch Negativity (MMN) in Basic Research of Central Auditory Processing: A Review. *Clin. Neurophysiol.* **2007**, *118*, 2544–2590, doi:10.1016/j.clinph.2007.04.026.
46. Titova, N.; Näätänen, R. Preattentive Voice Discrimination by the Human Brain as Indexed by the Mismatch Negativity. *Neurosci. Lett.* **2001**, *308*, 63–65, doi:10.1016/S0304-3940(01)01970-X.
47. Alho, K. Cerebral Generators of Mismatch Negativity (MMN) and Its Magnetic Counterpart (MMNm) Elicited by Sound Changes. *Ear Hear.* **1995**, *16*, 38–51.
48. Hari, R.; Hämäläinen, M.; Ilmoniemi, R.; Kaukoranta, E.; Reinikainen, K.; Salminen, J.; Alho, K.; Näätänen, R.; Sams, M. Responses of the Primary Auditory Cortex to Pitch Changes in a Sequence of Tone Pips: Neuromagnetic Recordings in Man. *Neurosci. Lett.* **1984**, *50*, 127–132, doi:10.1016/0304-3940(84)90474-9.
49. Javitt, D.C.; Steinschneider, M.; Schroeder, C.E.; Vaughan, H.G.; Arezzo, J.C. Detection of Stimulus Deviance within Primate Primary Auditory Cortex: Intracortical Mechanisms of Mismatch Negativity (MMN) Generation. *Brain Res.* **1994**, *667*, 192–200, doi:10.1016/0006-8993(94)91496-6.
50. Näätänen, R.; Alho, K. Mismatch Negativity – A Unique Measure of Sensory Processing in Audition. *Int. J. Neurosci.* **1995**, *80*, 317–337, doi:10.3109/00207459508986107.
51. Näätänen, R.; Lehtokoski, A.; Lennes, M.; Cheour, M.; Huutilainen, M.; Iivonen, A.; Vainio, M.; Alku, P.; Ilmoniemi, R.J.; Luuk, A.; et al. Language-Specific Phoneme Representations Revealed by Electric and Magnetic Brain Responses. *Nature* **1997**, *385*, 432–434, doi:10.1038/385432a0.
52. Sams, M.; Paavilainen, P.; Alho, K.; Näätänen, R. Auditory Frequency Discrimination and Event-Related Potentials. *Electroencephalogr. Clin. Neurophysiol. Potentials Sect.* **1985**, *62*, 437–448, doi:10.1016/0168-5597(85)90054-1.
53. Sharma, A.; Dorman, M.F. Cortical Auditory Evoked Potential Correlates of Categorical Perception of Voice-Onset Time. *J. Acoust. Soc. Am.* **1999**, *106*, 1078–1083, doi:10.1121/1.428048.
54. Fu, Z.; Monahan, P.J. Extracting Phonetic Features From Natural Classes: A Mismatch Negativity Study of Mandarin Chinese Retroflex Consonants. *Front. Hum. Neurosci.* **2021**, *15*, 1–15, doi:10.3389/fnhum.2021.609898.
55. Gomes, H.; Ritter, W.; Vaughan, H.G. The Nature of Preattentive Storage in the Auditory System. *J. Cogn. Neurosci.* **1995**, *7*, 81–94, doi:10.1162/jocn.1995.7.1.81.
56. Hu, A.; Gu, F.; Wong, L.L.N.; Tong, X.; Zhang, X. Visual Mismatch Negativity Elicited by Semantic Violations in Visual Words. *Brain Res.* **2020**, *1746*, 147010, doi:10.1016/j.brainres.2020.147010.
57. Monahan, P.J.; Schertz, J.; Fu, Z.; Pérez, A. Unified Coding of Spectral and Temporal Phonetic Cues: Electrophysiological Evidence for Abstract Phonological Features. *J. Cogn. Neurosci.* **2022**, *34*, 618–638, doi:10.1162/jocn_a_01817.
58. Phillips, C.; Pellathy, T.; Marantz, A.; Yellin, E.; Wexler, K.; Poeppel, D.; McGinnis, M.; Roberts, T. Auditory Cortex Accesses Phonological Categories: An MEG Mismatch Study. *J. Cogn. Neurosci.* **2000**, *12*, 1038–1055, doi:10.1162/08989290051137567.
59. Politzer-Ahles, S.; Jap, B.A.J. Can the Mismatch Negativity Really Be Elicited by Abstract Linguistic Contrasts? *Neurobiol. Lang.* **2024**, 1–26, doi:10.1162/nol_a_00147.

60. Schröger, E.; Paavilainen, P.; Näätänen, R. Mismatch Negativity to Changes in a Continuous Tone with Regularly Varying Frequencies. *Electroencephalogr. Clin. Neurophysiol. Potentials Sect.* **1994**, *92*, 140–147, doi:10.1016/0168-5597(94)90054-X.
61. Shestakova, A.; Brattico, E.; Huottilainen, M.; Galunov, V.; Soloviev, A.; Sams, M.; Ilmoniemi, R.J.; Näätänen, R. Abstract Phoneme Representations in the Left Temporal Cortex: Magnetic Mismatch Negativity Study: *NeuroReport* **2002**, *13*, 1813–1816, doi:10.1097/00001756-200210070-00025.
62. Knösche, T.R.; Lattner, S.; Maess, B.; Schauer, M.; Friederici, A.D. Early Parallel Processing of Auditory Word and Voice Information. *NeuroImage* **2002**, *17*, 1493–1503, doi:10.1006/nimg.2002.1262.
63. Allen, J.; Kraus, N.; Bradlow, A. Neural Representation of Consciously Imperceptible Speech Sound Differences. *Percept. Psychophys.* **2000**, *62*, 1383–1393, doi:10.3758/BF03212140.
64. Bühler, J.C.; Schmid, S.; Maurer, U. Influence of Dialect Use on Speech Perception: A Mismatch Negativity Study. *Lang. Cogn. Neurosci.* **2017**, *32*, 757–775, doi:10.1080/23273798.2016.1272704.
65. Dehaene-Lambertz, G.; Dupoux, E.; Gout, A. Electrophysiological Correlates of Phonological Processing: A Cross-Linguistic Study. *J. Cogn. Neurosci.* **2000**, *12*, 635–647, doi:10.1162/089892900562390.
66. Hacquard, V.; Walter, M.A.; Marantz, A. The Effects of Inventory on Vowel Perception in French and Spanish: An MEG Study. *Brain Lang.* **2007**, *100*, 295–300, doi:10.1016/j.bandl.2006.04.009.
67. Kazanina, N.; Phillips, C.; Idsardi, W.J. The Influence of Meaning on the Perception of Speech Sounds. *Proc. Natl. Acad. Sci.* **2006**, *103*, 11381–11386, doi:10.1073/pnas.0604821103.
68. Lipski, S.C.; Escudero, P.; Benders, T. Language Experience Modulates Weighting of Acoustic Cues for Vowel Perception: An Event-related Potential Study. *Psychophysiology* **2012**, *49*, 638–650, doi:10.1111/j.1469-8986.2011.01347.x.
69. Miglietta, S.; Grimaldi, M.; Calabrese, A. Conditioned Allophony in Speech Perception: An ERP Study. *Brain Lang.* **2013**, *126*, 285–290, doi:10.1016/j.bandl.2013.06.001.
70. Sharma, A.; Dorman, M.F. Neurophysiologic Correlates of Cross-Language Phonetic Perception. *J. Acoust. Soc. Am.* **2000**, *107*, 2697–2703, doi:10.1121/1.428655.
71. Winkler, I.; Lehtokoski, A.; Alku, P.; Vainio, M.; Czigler, I.; Csépe, V.; Aaltonen, O.; Raimo, I.; Alho, K.; Lang, H.; et al. Pre-Attentive Detection of Vowel Contrasts Utilizes Both Phonetic and Auditory Memory Representations. *Cogn. Brain Res.* **1999**, *7*, 357–369, doi:10.1016/S0926-6410(98)00039-1.
72. Winkler, I.; Kujala, T.; Alku, P.; Näätänen, R. Language Context and Phonetic Change Detection. *Cogn. Brain Res.* **2003**, *17*, 833–844, doi:10.1016/S0926-6410(03)00205-2.
73. Scharinger, M.; Monahan, P.J.; Idsardi, W.J. You Had Me at “Hello”: Rapid Extraction of Dialect Information from Spoken Words. *NeuroImage* **2011**, *56*, 2329–2338, doi:10.1016/j.neuroimage.2011.04.007.
74. Tuninetti, A.; Chládková, K.; Peter, V.; Schiller, N.O.; Escudero, P. When Speaker Identity Is Unavoidable: Neural Processing of Speaker Identity Cues in Natural Speech. *Brain Lang.* **2017**, *174*, 42–49, doi:10.1016/j.bandl.2017.07.001.
75. Hardman, J. Accentedness and Intelligibility of Mandarin-Accented English for Chinese, Koreans, and Americans. In Proceedings of the In Proceedings of the International Symposium on the Acquisition of Second Language Speech; 2014; Vol. Vol. 5, pp. 240–260.
76. Hansen Edwards, J.G. Sociolinguistic Variation in Asian Englishes: The Case of Coronal Stop Deletion. *Engl. World-Wide J. Var. Engl.* **2016**, *37*, 138–167, doi:10.1075/eww.37.2.02han.
77. Cheng, L.S.P.; Kramer, M.A. Exploring Asian North American English: A YouTube Corpus-Based Approach. Presented at the LSA 2022 Annual Meeting, Washington, DC, United States (virtual), 2022.
78. Boersma, P.; Weenink, D. Praat: Doing Phonetics by Computer 2020.
79. Tervaniemi, M.; Kujala, A.; Alho, K.; Virtanen, J.; Ilmoniemi, R.J.; Näätänen, R. Functional Specialization of the Human Auditory Cortex in Processing Phonetic and Musical Sounds: A Magnetoencephalographic (MEG) Study. *NeuroImage* **1999**, *9*, 330–336, doi:10.1006/nimg.1999.0405.
80. Monahan, P.J. Phonological Knowledge and Speech Comprehension. *Annu. Rev. Linguist.* **2018**, *4*, 21–47, doi:10.1146/annurev-linguistics-011817-045537.
81. Cowan, N.; Winkler, I.; Teder, W.; Näätänen, R. Memory Prerequisites of Mismatch Negativity in the Auditory Event-Related Potential (ERP). *J. Exp. Psychol. Learn. Mem. Cogn.* **1993**, *19*, 909–921, doi:10.1037/0278-7393.19.4.909.

82. Peirce, J.; Gray, J.R.; Simpson, S.; MacAskill, M.; Höchenberger, R.; Sogo, H.; Kastman, E.; Lindeløv, J.K. PsychoPy2: Experiments in Behavior Made Easy. *Behav. Res. Methods* **2019**, *51*, 195–203, doi:10.3758/s13428-018-01193-y.
83. Pérez, A.; Monahan, P.J.; Lambon Ralph, M.A. Joint Recording of EEG and Audio Signals in Hyperscanning and Pseudo-Hyperscanning Experiments. *MethodsX* **2021**, *8*, 101347, doi:10.1016/j.mex.2021.101347.
84. Gramfort, A. MEG and EEG Data Analysis with MNE-Python. *Front. Neurosci.* **2013**, *7*, doi:10.3389/fnins.2013.00267.
85. Bigdely-Shamlo, N.; Mullen, T.; Kothe, C.; Su, K.-M.; Robbins, K.A. The PREP Pipeline: Standardized Preprocessing for Large-Scale EEG Analysis. *Front. Neuroinformatics* **2015**, *9*, doi:10.3389/fninf.2015.00016.
86. Pion-Tonachini, L.; Kreutz-Delgado, K.; Makeig, S. ICLabel: An Automated Electroencephalographic Independent Component Classifier, Dataset, and Website. *NeuroImage* **2019**, *198*, 181–197, doi:10.1016/j.neuroimage.2019.05.026.
87. Mahajan, Y.; Peter, V.; Sharma, M. Effect of EEG Referencing Methods on Auditory Mismatch Negativity. *Front. Neurosci.* **2017**, *11*, 560, doi:10.3389/fnins.2017.00560.
88. Jacobsen, T.; Horenkamp, T.; Schröger, E. Preattentive Memory-Based Comparison of Sound Intensity. *Audiol. Neurotol.* **2003**, *8*, 338–346, doi:10.1159/000073518.
89. Maris, E.; Oostenveld, R. Nonparametric Statistical Testing of EEG- and MEG-Data. *J. Neurosci. Methods* **2007**, *164*, 177–190, doi:10.1016/j.jneumeth.2007.03.024.
90. Meulman, N.; Sprenger, S.A.; Schmid, M.S.; Wieling, M. GAM-Based Individual Difference Measures for L2 ERP Studies. *Res. Methods Appl. Linguist.* **2023**, *2*, 100079, doi:10.1016/j.rmal.2023.100079.
91. Wood, S.; Wood, M.S. Package ‘Mgcv’ 2015.
92. R Core Team R: A Language and Environment for Statistical Computing 2021.
93. Simpson, G.L. Gratia: An R Package for Exploring Generalized Additive Models. *J. Open Source Softw.* **2024**, *9*, 6962, doi:10.21105/joss.06962.
94. Hansen, J.C.; Hillyard, S.A. Endogeneous Brain Potentials Associated with Selective Auditory Attention. *Electroencephalogr. Clin. Neurophysiol.* **1980**, *49*, 277–290, doi:10.1016/0013-4694(80)90222-9.
95. Arnal, L.H.; Giraud, A.-L. Cortical Oscillations and Sensory Predictions. *Trends Cogn. Sci.* **2012**, *16*, 390–398, doi:10.1016/j.tics.2012.05.003.
96. McMahon, C.M.; Boisvert, I.; de Lissa, P.; Granger, L.; Ibrahim, R.; Lo, C.Y.; Miles, K.; Graham, P.L. Monitoring Alpha Oscillations and Pupil Dilation across a Performance-Intensity Function. *Front. Psychol.* **2016**, *7*, doi:10.3389/fpsyg.2016.00745.
97. Strauß, A.; Wöstmann, M.; Obleser, J. Cortical Alpha Oscillations as a Tool for Auditory Selective Inhibition. *Front. Hum. Neurosci.* **2014**, *8*, doi:10.3389/fnhum.2014.00350.
98. Lenth, R.V.; Piskowski, J. Emmeans: Estimated Marginal Means, Aka Least-Squares Means; 2025;
99. Scharinger, M.; Monahan, P.J.; Idsardi, W.J. You Had Me at “Hello”: Rapid Extraction of Dialect Information from Spoken Words. *NeuroImage* **2011**, *56*, 2329–2338, doi:10.1016/j.neuroimage.2011.04.007.
100. Scharinger, M.; Monahan, P.J.; Idsardi, W.J. Linguistic Category Structure Influences Early Auditory Processing: Converging Evidence from Mismatch Responses and Cortical Oscillations. *NeuroImage* **2016**, *128*, 293–301, doi:10.1016/j.neuroimage.2016.01.003.
101. Jaramillo, M.; Alku, P.; Paavilainen, P. An Event-Related Potential (ERP) Study of Duration Changes in Speech and Non-Speech Sounds. *NeuroReport* **1999**, *10*, 3301, doi:10.1097/00001756-199911080-00010.
102. Takegata, R.; Tervaniemi, M.; Alku, P.; Ylinen, S.; Näätänen, R. Parameter-Specific Modulation of the Mismatch Negativity to Duration Decrement and Increment: Evidence for Asymmetric Processes. *Clin. Neurophysiol.* **2008**, *119*, 1515–1523, doi:10.1016/j.clinph.2008.03.025.
103. Peter, V.; McArthur, G.; Thompson, W.F. Effect of Deviance Direction and Calculation Method on Duration and Frequency Mismatch Negativity (MMN). *Neurosci. Lett.* **2010**, *482*, 71–75, doi:10.1016/j.neulet.2010.07.010.
104. White, E.J.; Titone, D.; Genesee, F.; Steinhauer, K. Phonological Processing in Late Second Language Learners: The Effects of Proficiency and Task. *Biling. Lang. Cogn.* **2017**, *20*, 162–183, doi:10.1017/S1366728915000620.

105. Winkler, I.; Kujala, T.; Tiitinen, H.; Sivonen, P.; Alku, P.; Lehtokoski, A.; Czigler, I.; Csépe, V.; Ilmoniemi, R.J.; Näätänen, R. Brain Responses Reveal the Learning of Foreign Language Phonemes. *Psychophysiology* **1999**, *36*, 638–642, doi:10.1111/1469-8986.3650638.
106. Liberto, G.M.D.; Nie, J.; Yeaton, J.; Khalighinejad, B.; Shamma, S.A.; Mesgarani, N. Neural Representation of Linguistic Feature Hierarchy Reflects Second-Language Proficiency. *NeuroImage* **2021**, *227*, 117586, doi:10.1016/j.neuroimage.2020.117586.
107. Gosselin, L.; Martin, C.D.; González Martín, A.; Caffarra, S. When a Nonnative Accent Lets You Spot All the Errors: Examining the Syntactic Interlanguage Benefit. *J. Cogn. Neurosci.* **2022**, *34*, 1650–1669, doi:10.1162/jocn_a_01886.
108. Gustavsson, L.; Kallioinen, P.; Klintfors, E.; Lindh, J. Neural Processing of Voices—Familiarity. *J. Acoust. Soc. Am.* **2013**, *133*, 3569, doi:10.1121/1.4806527.
109. Brunellière, A.; Soto-Faraco, S. The Speakers' Accent Shapes the Listeners' Phonological Predictions during Speech Perception. *Brain Lang.* **2013**, *125*, 82–93, doi:10.1016/j.bandl.2013.01.007.
110. Hanulíková, A.; van Alphen, P.M.; van Goch, M.M.; Weber, A. When One Person's Mistake Is Another's Standard Usage: The Effect of Foreign Accent on Syntactic Processing. *J. Cogn. Neurosci.* **2012**, *24*, 878–887, doi:10.1162/jocn_a_00103.
111. Caffarra, S.; Martin, C.D. Not All Errors Are the Same: ERP Sensitivity to Error Typicality in Foreign Accented Speech Perception. *Cortex* **2019**, *116*, 308–320, doi:10.1016/j.cortex.2018.03.007.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.