



Accented speech modulates multiple event-related potential components across multiple levels of language processing



Fernando Llanos ^{1,5}✉, Yunan Charles Wu ^{2,5}, Taylor J. Abel ³ & Lori L. Holt ⁴✉

Accents are ubiquitous in spoken communication, and while listeners can rapidly adapt to accented speech, the neural mechanisms supporting this flexibility remain poorly understood. Successful adaptation requires developing new sound representations without compromising the stability of long-term speech norms. This delicate balance between plasticity and stability illustrates a fundamental challenge faced by all cognitive systems. To investigate how the brain manages this trade-off, we recorded electroencephalographic activity from 23 native English speakers as they categorized words produced in either canonical American English or an unfamiliar accent. We contrasted two potential mechanisms: one in which listeners fully restructure their sound-to-category mappings to reflect accent-specific pronunciations, and another in which they downweight the functional relevance of sounds that deviate from long-term expectations. Listeners relied on short-term speech regularities to reduce perceptual weighting of acoustic dimensions that did not conform to the canonical norm. Consistent with this perceptual shift, we observed less robust neural encoding of sound differences along the downweighted dimensions. Notably, these adaptive neural adjustments emerged as early as 100 milliseconds, at latencies associated with subphonemic auditory processing, and persisted through later stages linked to phonological and post-phonological processing. These findings indicate that rapid adaptation to unfamiliar accents involves downweighting the functional relevance of sound cues based on short-term input statistics, rather than fully restructuring native sound-to-category mappings. This mechanism enables flexible adjustment to novel speech inputs while preserving long-term linguistic representations, illustrating how the auditory system negotiates the trade-off between plasticity and representational stability.

Speech accents provide a unique window into how people communicate across different geographic and sociolinguistic boundaries. When listening to a speaker with an unfamiliar accent, we must quickly adjust our expectations to accommodate novel dialectal deviations from our phonetic norm^{1–4}. This perceptual adaptation process involves a delicate balance between auditory plasticity and stability in the human brain. Adult listeners' perception reflects native phonetic norms consistent with neural stability to long-term speech regularities. However, when they are faced with unfamiliar accents, they must prioritize plasticity, rapidly adapting to novel

idiosyncratic pronunciations while preserving the stability of long-term representations. This dynamic interplay between neural stability and plasticity represents a fundamental challenge for all cognitive systems: maintaining stable representations that align with long-term norms while remaining flexible enough to accommodate short-term deviations.

Speech accents are characterized by systematic phonetic departures from a given linguistic norm^{5–7}. Upon encountering an accent, speech comprehension can take a hit. But with exposure, it rebounds and even generalizes to other talkers with similar accents. While the cognitive

¹Department of Linguistics, University of Texas at Austin, Austin, TX, 78712, USA. ²Department of Psychology, Carnegie Mellon University, Pittsburgh, PA, 15213, USA. ³Department of Neurological Surgery, University of Pittsburgh, Pittsburgh, PA, 15213, USA. ⁴Department of Psychology, University of Texas at Austin, Austin, TX, 78712, USA. ⁵These authors contributed equally: Fernando Llanos, Yunan Charles Wu. ✉e-mail: fllanos@utexas.edu; lori.holt@austin.utexas.edu

implementation of this perceptual adaptation process is still a matter of speculation (for review see⁸), there is evidence that listeners take advantage of short-term speech regularities to accommodate speech accents^{9–12}. From a dimension-based learning perspective^{10,12–14}, adaptation to accented speech can be conceptualized as a learning process through which listeners dynamically recalibrate the linguistic relevance of speech dimensions that deviate from their canonical speech norm. For instance, native speakers of American English rely on the onset of laryngeal voicing (*Voice Onset Time*, or VOT) and the fundamental frequency (F0) to distinguish /bir/ 'beer' from /pir/ 'pier', such that short VOT and low F0 are typically perceived as *beer* and long VOT and high F0 as *pier*^{13,15–22}. However, when English listeners are exposed to an artificial accent where *beer* is systematically pronounced with short VOT but high F0 and *pier* is systematically pronounced with long VOT but low F0, they down-weight reliance on F0 to recognize these words^{12,23,24}. This finding indicates that listeners exploit short-term statistical regularities to downplay the linguistic relevance of speech patterns that systematically deviate from their native expectations.

Previous research on speech perception has shown that listeners can adapt to accented speech very quickly after just a few seconds or minutes of exposure depending on the task^{24–27}. Given that the processing of speech patterns spans multiple levels of hierarchical processing that are not directly accessible from discrete behavioural responses, a key question that remains unanswered is how perceptual accommodation of speech accents unfolds across multiple processing latencies in the cortex. Specifically, it is currently uncertain whether perceptual adaptation to accented speech is driven by lower-level adjustments during the early encoding of phonetic features or higher-level structural adjustments in the mapping of these features into phonological and lexical constituents. To address these questions, we investigated the effects of exposure to canonical and accented pronunciations of English minimal pairs (/bir/ 'beer' vs. /pir/ 'pier') on electroencephalographic (EEG) markers of subphonemic processing (N1), phonological processing (Mismatch Negativity), and post-phonemic processing (Late Negativity and Late Positive Component).

The N1, or N100^{28–30} is an early cortical evoked potential generated in the auditory cortex. It is characterized by a negative deflection in the EEG occurring ~100 ms following the onset of sounds, syllables, or words presented with equal or different probabilities. Previous research proposed that changes in N1 amplitude reflect the non-linear mapping from speech sounds to phonological categories^{31–33}. However, recent work has demonstrated a linear relationship between changes in N1 amplitude and concomitant changes in phonetic cues^{30,34,35}. Consequently, the N1 offers a suitable component to investigate the early processing of speech attributes in the auditory cortex.

The Mismatch Negativity (MMN)^{36–39} is generated within a broad neural network of temporal and frontal regions and indexes the brain's automatic detection of sound changes^{28,36,40}. The MMN is characterized by a negative ongoing deflection in the EEG signal occurring ~200 ms after the onset of an oddball or deviant sound. While the MMN can detect sound changes in both speech and non-speech contexts³⁸, previous research has shown that MMN amplitude becomes more negative when sound changes are phonologically relevant in the native language^{28,40}. Consequently, the MMN offers a useful approach to investigate the balance between long-term neural stability and short-term auditory neuroplasticity.

Late Negativity (LN) refers to a negative-going wave typically observed after the MMN. Rather than a unitary component, this wave can reflect several ERP effects associated with higher-order linguistic processing, including the N400⁴¹, late discriminative negativity (LDN)⁴², and reorienting negativity (RON)⁴². Both the N400 and LDN have been linked to lexico-semantic processing. Specifically, the N400 is associated with the automatic processing of semantic content^{43,44}, including the detection and integration of novel word meanings, while the LDN has been linked to lexical processes such as word familiarity and lexical access^{45,46}. In contrast, the RON is typically related to task-driven attentional reorienting and is

generally not elicited during passive exposure designs^{47,48}, where participants are instructed to ignore the sounds and attend to a silent video. Unlike the RON, the N400 and LDN can be elicited in passive word oddball paradigms⁴⁹ and do not require explicit task engagement⁵⁰, although the LDN is less consistently observed in adults than in children. Importantly, in the context of MMN studies, LN provides an index of higher-order linguistic processing beyond automatic deviance detection.

The Late Positive Component (LPC) refers to a late and sustained positive-going deflection observed in paradigms beyond oddball designs, especially in tasks where participants are explicitly asked to categorize or label speech sounds. In these paradigms, the LPC typically emerges several hundred milliseconds after the N1. In contrast to the P3b, which typically peaks earlier (300–500 ms) and is associated with novelty detection in oddball paradigms, the LPC is more robust in paradigms requiring explicit categorization or memory retrieval. The LPC is generally characterized by sustained parietal positivity between 600 and 800 ms^{51,52} and is typically associated with post-lexical processing, including decision-making, memory and semantic retrieval, as well as the integration of complex or ambiguous stimuli^{53–55}.

Research hypotheses

We investigated how exposure to different contextual regularities influenced the neural and perceptual processing of the same, acoustically ambiguous *beer* and *peer* tokens across two conditions. In the canonical condition, listeners were exposed to statistical regularities consistent with English, where *beer* tokens had short VOT and low F0 and *peer* tokens had long VOT and high F0. Our first hypothesis was that, under these canonical mappings, listeners would rely on both VOT and F0 but assign greater weight to VOT, consistent with prior work on English perceptual cue weighting^{24,30}.

In the accented condition, the correlation between VOT and F0 was reversed so that *beer* tokens had short VOT but high F0 and *peer* tokens had long VOT but low F0. Here, we considered two alternative adaptive mechanisms. One possibility is that listeners would fully reverse the correlation between VOT and F0 in canonical English, such that low F0 would bias *peer* responses and high F0 would bias *beer* responses. Alternatively, listeners might downweight the perceptual relevance of F0, treating it as unreliable because it conflicts with long-term English norms. Each strategy offers distinct potential benefits. Reversing the mapping allows listeners to maintain the use of all available cues, including F0, thereby preserving cue redundancy and enhancing robustness at the expense of long-term stability. Downweighting F0, by contrast, prioritizes stability in long-term representations, ensuring that listeners do not adopt mappings that conflict with the long-term norm and thus reducing the cost of re-adaptation when returning to native speech.

Consistent with these behavioural predictions, we expect the ERP markers introduced above to align with listeners' perceptual categorization behaviour. For example, if listeners reverse the canonical correlation between VOT and F0, we should observe similar MMN amplitudes across conditions, reflecting comparable neural sensitivity to the same F0 differences across speech contexts. By contrast, if listeners downweight F0 in the accented condition, MMN amplitude should be attenuated in that condition, reflecting reduced neural sensitivity to F0 differences in the accented speech context.

A key open question, however, concerns the time course of neural adaptation. If differences between conditions emerge as early as the N1 (~100 ms post stimulus onset), this would indicate that accent adaptation is mediated by subphonemic adjustments unfolding at early auditory processing stages in the cortex. If differences first emerge at the MMN (~200 ms post-stimulus onset), this would suggest that adaptation consolidates at later stages of phonological processing; beyond, but not excluding, the auditory cortex. Finally, if differences first appear at the LN or LPC (>300 ms post-stimulus onset), this would indicate that adaptation is achieved at post-phonological stages of processing involving higher-order lexico-semantic adjustments.

Methods

Participants

Twenty-nine students from Carnegie Mellon University (aged 19–30 years; 12 male, 17 female) participated in the study in exchange for course credit or monetary compensation. Information on participants' sex was self-reported. According to preliminary power analysis (see Zhang and colleagues work⁵⁶ for details) this sample size exceeded the estimated requirement of 16 participants needed to achieve 80% power at a significance level of 0.05. Power analysis was informed by categorization responses to F0-differentiated stimuli collected in a prior study⁵⁶. All participants reported normal hearing and American English as the primary language used at home before age two. Six participants were removed from the study due to EEG recording issues, and thus also excluded from the behavioural analyses. Participants provided informed consent in accordance with protocols established by the Institutional Review Board of Carnegie Mellon University. There are no study preregistrations to be disclosed. Data on race and ethnicity are not presented because they are unrelated to the research questions.

Stimulus grid

The stimulus grid was created following the procedures specified in a previous study²⁴. It included 49 exemplars of *beer* or *pier* ranging from 0 ms VOT to 30 ms VOT across seven steps of 5 ms, and from 200 Hz to 320 Hz across seven F0 steps of 20 Hz. Grid exemplars were derived from one *pier* exemplar naturally produced by a female native-English speaker. This exemplar was acoustically manipulated in Praat⁵⁷ to match the VOT and F0 values included in the grid. VOT was operationalized as the time elapsed between the oral release of the stop consonant and onset of laryngeal voicing at the following vowel. VOT duration was manipulated by removing 5 ms segments at zero-crossings from the naturally produced *pier* exemplar. F0 was operationalized as the value of the fundamental frequency at the onset of the vowel. F0 onsets were adjusted manually in Praat. They remained constant through the first 80 ms of the vowel and decreased to 180 Hz during the next 150 ms.

Baseline block

The entire experiment was conducted in a double-walled sound-attenuated and electrically shielded booth. Participants were first instructed to categorize as *beer* or *pier* five repetitions of 25 grid exemplars (125 trials total). Baseline exemplars included all possible combinations of VOT and F0 values between 5 and 25 ms and 220 and 300 Hz. Stimuli (44.1 kHz sampling rate) were presented in random order at 75 dB sound pressure level (SPL) by an RME UXF+ Audio Interface (RME, Haimhausen, Germany) via Etymotic ER-1 (Etymotic, Elk Grove Village, IL) linear headphones. The RME sent triggers routed through a S/PDIF-to-TTL converter (Electronics Designs Facility, Boston University) for compatibility with the BioSemi EEG system (Amsterdam, The Netherlands). Participants were instructed to listen to each trial and use their mouse to click one of two alternative responses presented on a screen while the experimenters prepared the EEG cap.

Participants' responses were coded as 0 (response = *beer*) and 1 (response = *pier*). To investigate the effects of VOT duration and F0 height on participants' behaviour, we fitted a generalized logistic regression model to binary responses collapsed across all participants. The model was coded in MATLAB 2024b⁵⁸ using the following equation: $response \sim VOT + F0$ (binomial distribution, logit link). The significance of VOT and F0 was determined by the *p*-values of the first (VOT) and second (F0) beta coefficients of the model. Responses from three participants were excluded from the logistic model to avoid convergence problems caused by having too few 0 or 1 responses.

To assess individual perceptual weights of VOT and F0, we fitted separate generalized logistic regression models to each participant's binary responses. Individual models were coded in MATLAB following the parameters specified above and individual perceptual weights were determined

by the alpha (VOT) and beta (F0) coefficients of the models. We conducted a two-sample *t*-test analysis to compare the distributional means of individual VOT and F0 weights. Cohen's *d* was computed as the mean difference divided by the pooled standard deviation.

Active exposure blocks

Following the baseline block, participants were instructed to follow the same task procedures to categorize 10 exposure exemplars and two test exemplars presented multiple times at random across 26 blocks of 60 trials per condition. Each active exposure block lasted ~3 min and was followed by a passive exposure block described in the next section. In the canonical speech condition, exposure stimuli included the following combinations of VOT and F0 values: [0 ms, 220 Hz], [5 ms, 220 ms], [10 ms, 220 Hz], [5 ms, 200 Hz], [5 ms, 240 Hz], [20 ms, 300 Hz], [25 ms, 300 Hz], [30 ms, 300 Hz], [25 ms, 320 Hz], [25 ms, 280 Hz]. In the accented speech condition, exposure stimuli included the following combinations of VOT and F0 values: [0 ms, 300 Hz], [5 ms, 300 Hz], [10 ms, 300 Hz], [5 ms, 320 Hz], [5 ms, 280 Hz], [20 ms, 220 Hz], [25 ms, 220 Hz], [30 ms, 220 Hz], [25 ms, 240 Hz], [25 ms, 200 Hz]. Test stimuli remained constant across conditions: [15 ms, 220 Hz], [15 ms, 300 Hz]. Participants were presented with the active exposure blocks from one condition, followed by the ones from the other condition. The order of the two conditions was counterbalanced across participants.

Participants' behavioural responses across test exemplars were coded as 0 (response = *beer*) and 1 (response = *pier*). To evaluate the effect of each condition on the perception of F0 contrasts, we fitted a generalized logistic mixed-effect regression model. The model was coded in R⁵⁹ using the following *glmer*⁶⁰ equation: $response \sim F0 * condition + (1 | participant)$. This equation incorporates fixed effects by F0 (higher F0, lower F0) and condition (canonical, accented), their interaction, and random intercepts by participant. Model assumptions were evaluated through visual inspection of standard diagnostic plots. The normality of random effects was assessed using a *Q-Q plot* of the subject-level random intercepts, which showed no substantial deviation from normality. Linearity of the logit was assessed using simulation-based residual plots generated with the *DHARMA* package, which revealed no systematic patterns suggestive of nonlinearity.

Differences between F0 levels by condition were determined via Tukey-adjusted post-hoc analysis. Pairwise comparisons were coded in R using the following *emmeans*⁶¹ equation: $pairwise \sim F0 | condition$. This equation compares the distributional means of *pier* responses in higher- vs. lower-F0 test exemplars within each condition.

Overt categorization trials during active exposure blocks were further used to collect N1 components, as in previous related work^{30,35,62}. N1 is most reliably observed during active tasks, where attention enhances its amplitude and functional relevance. In passive listening paradigms, such as the oddball paradigm, this ERP component can be masked by other ERPs like the MMN³⁸, which makes it more difficult to interpret.

Passive exposure blocks

To collect MMN and LN components, participants were exposed to an oddball sequence made of test exemplars while watching a silent video right after each active exposure block (i.e., every 60 overt categorization trials). Each oddball sequence lasted ~30 s. The combination of VOT and F0 values remained constant across conditions: test stimulus 1 = [15 ms, 300 Hz], test stimulus 2 = [15 ms, 220 Hz]. Test stimuli were presented with an 85:15 standard-to-oddball ratio to elicit robust MMN waves. Within each sequence, standard and oddball stimuli were presented 17 and 3 times, respectively. Stimulus presentation was pseudorandomized, so each oddball sound was preceded by at least three standard sounds. Interstimulus interval was fixed at 700 ms and each sequence began and ended with 1 s of silence. Standard and oddball stimuli were counterbalanced across blocks within each participant. Participants were told to pay attention to a silent movie and ignore the sounds during the presentation of oddball sequences.

EEG acquisition and preprocessing

During both active exposure and passive listening blocks, continuous EEG signals were digitized at a sampling rate of 1024 Hz using a BioSemi ActiveTwo system. EEG signals were acquired through 32 Ag/AgCl sintered electrodes embedded in a Biosemi Headcap (10-20 system) and left/right mastoid (M1, M2) electrodes. A pair of electrodes placed on the outer canthus of each eye allowed for calculation of the horizontal electro-oculogram (EOG) and an additional electrode placed on the left cheek bone allowed for detection of vertical EOG. An experimenter encouraged participants to remain as still as possible to minimize muscle artifacts. Messages on the screen encouraged participants to take brief breaks between blocks.

EEG preprocessing was conducted using MNE-PYTHON⁶³ open-source software. EEG signals were resampled to 128 Hz to reduce pre-processing time and band-pass filtered between 0.1 Hz and 32 Hz with a zero-phase finite impulse response (FIR) filter. Next, we performed independent component analysis (ICA), visually inspecting components to remove those generated by eye blinks, saccades, heartbeats, and muscle movements.

Mismatch negativity procedures

Preprocessed EEG signals from the passive listening blocks were re-referenced to the average reference⁶⁴, segmented into epochs ranging from 200 ms before stimulus onset to 800 ms after stimulus onset, and baseline-corrected. Individual EEG epochs were averaged by channel ($N = 32$), oddball status (standard, deviant), and condition (canonical, accented). Individual MMN waves were computed by subtracting the standard wave from the oddball (or deviant) wave across channels and conditions. Individual MMN amplitudes were calculated as the mean amplitude of the MMN wave between 150 ms and 250 ms.

To evaluate the effects of condition (canonical, accented) on the amplitude of the MMN, individual MMN amplitudes were modelled in R (*lmer* library) with the following linear mixed-effects equation: $MMN\ amplitude \sim condition * channel + (1|participant)$. This equation incorporates fixed effects by EEG channel (32 levels) and condition (canonical, accented), their interaction, and random intercepts by participant. The following model assumptions were evaluated via visual inspection: linearity and homoscedasticity of residuals (using residuals vs. fitted plots), and normality of residuals (using Q-Q plots). Visual diagnostics indicated no substantial deviations from these assumptions, suggesting that model fit was appropriate for psychological research.

Differences between condition levels across channels and between condition levels by channel were determined via Tukey-adjusted (2 levels) or FDR-adjusted (32 levels) post-hoc comparisons. Pairwise comparisons were coded in R using the two following *emmeans* equations: (1) *pairwise ~ condition*, (2) *pairwise ~ condition|channel*. These equations compare the distributional means of the conditions across all channels and channel-by-channel. For this and all subsequent linear mixed-effects models, Cohen's *d* was computed as the model estimate divided by the model's estimated noise, defined as the residual standard deviation. Similarly, due to the large number of contrasts, the results of the condition-by-channel pairwise comparisons are presented in the Supplementary Materials.

LN procedures

Preprocessed EEG signals from the passive listening blocks were re-referenced, segmented, baseline-corrected, and averaged following the MMN procedures specified above⁶⁵. Individual LN waves were computed by subtracting the standard wave from the oddball (or deviant) wave across channels and conditions. Individual LN amplitudes were calculated as the mean amplitude of the LN wave between 300 ms and 600 ms. To evaluate the effects of condition on the amplitude of the LN, individual LN amplitudes were modelled in R with the following linear mixed-effects equation: $LN\ amplitude \sim condition * channel + (1|participant)$. This equation incorporates fixed effects by EEG channel (32 levels) and condition (canonical, accented), their interaction, and random intercepts by participant. The following model assumptions were evaluated via visual inspection:

linearity and homoscedasticity of residuals and normality of residuals. Visual diagnostics indicated no substantial deviations from these assumptions, suggesting that model fit was appropriate for psychological research.

Differences between condition levels across channels and between condition levels by channel were determined via Tukey-adjusted (2 levels) or FDR-adjusted (32 levels) post-hoc comparisons. Pairwise comparisons were coded in R using the two following *emmeans* equations: (1) *pairwise ~ condition*, (2) *pairwise ~ condition|channel*. These equations compare the distributional means of the conditions across all channels and channel-by-channel.

N1 procedures

Preprocessed EEG signals from test exemplars in active exposure blocks were re-referenced to the average mastoid³⁰. Re-referenced signals were segmented into epochs ranging from 200 ms before stimulus onset to 800 ms after stimulus onset and baseline-corrected. Individual EEG epochs were averaged by channel and condition. Individual difference waves were computed by subtracting the wave of the lower-F0 test exemplar from the wave of the higher-F0 test exemplar across channels and conditions. Individual differences in N1 amplitude were calculated as the mean amplitude of the difference wave between 75 ms and 125 ms.

To evaluate the effects of condition on the amplitude of the difference wave, individual differences in N1 amplitude were modelled in R with the following linear mixed-effects equation: $difference\ wave\ amplitude \sim condition * channel + (1|participant)$. This equation incorporates fixed effects by EEG channel (32 levels) and condition (canonical, accented), their interaction, and random intercepts by participant. The following model assumptions were evaluated via visual inspection: linearity and homoscedasticity of residuals, and normality of residuals. Visual diagnostics indicated no substantial deviations from these assumptions, suggesting that model fit was appropriate for psychological research.

Differences between condition levels across channels and between condition levels by channel were determined via Tukey-adjusted (2 levels) or FDR-adjusted (32 levels) post-hoc comparisons. Pairwise comparisons were coded in R using the two following *emmeans* equations: (1) *pairwise ~ condition*, (2) *pairwise ~ condition|channel*. These equations compare the distributional means of the conditions across all channels and channel-by-channel.

LPC procedures

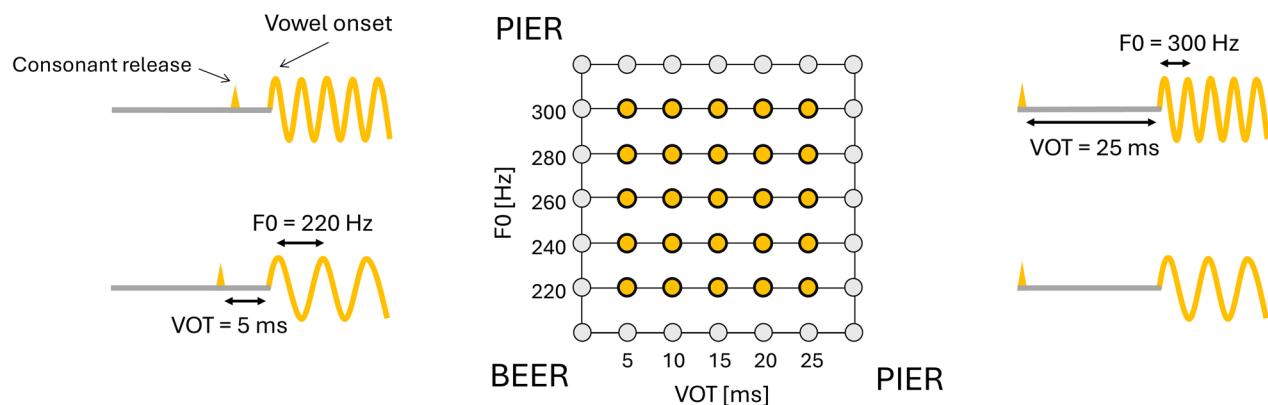
Preprocessed EEG signals from test exemplars in active exposure blocks were re-referenced, segmented, baseline-corrected, and averaged following the N1 procedures specified above. Individual differences in LPC amplitude were calculated as the mean amplitude of the difference wave between 600 ms and 800 ms. To evaluate the effects of condition on the amplitude of the LPC, individual differences in LPC amplitude were modelled in R with the following linear mixed-effects equation: $LPC\ amplitude \sim condition * channel * F0\ height + (1|participant)$. This equation incorporates fixed effects by EEG channel (32 levels), condition (canonical, accented), F0 height (higher, lower), their interaction, and random intercepts by participant. The following model assumptions were evaluated via visual inspection: linearity and homoscedasticity of residuals, and normality of residuals. Visual diagnostics indicated no substantial deviations from these assumptions, suggesting that model fit was appropriate for psychological research.

Differences between condition levels across channels and between condition levels by channel were determined via Tukey-adjusted (2 levels) or FDR-adjusted (32 levels) post-hoc comparisons. Pairwise comparisons were coded in R using the two following *emmeans* equations: (1) *pairwise ~ condition*, (2) *pairwise ~ condition|channel*. These equations compare the distributional means of the conditions across all channels and channel-by-channel.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

A. Baseline Block Structure



B. Group-level Logistic Function

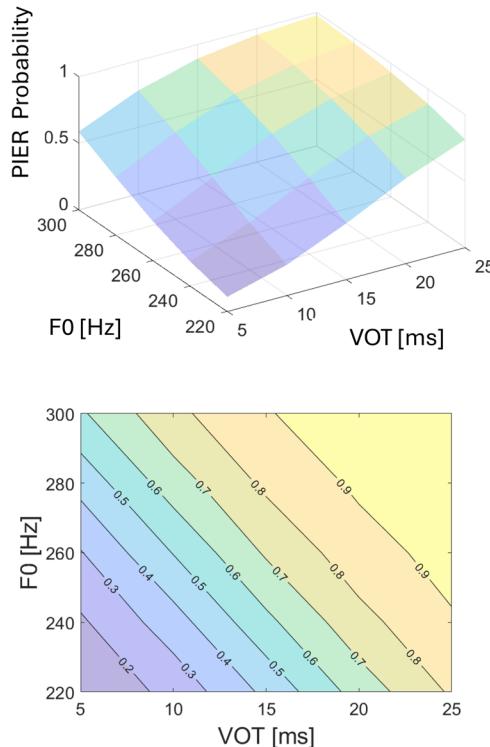
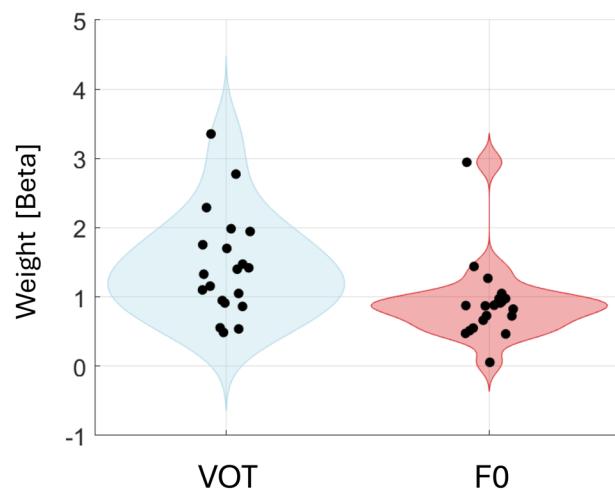


Fig. 1 | Baseline categorization block. A Participants ($n = 23$) were first instructed to categorize as *beer* or *pier* a series of baseline speech exemplars across five VOT and F0 levels. VOT was defined as the time elapsed between the consonant release and the onset of the vowel. F0 was defined as the fundamental frequency of the speech

C. Individual Perceptual Weights



signal at the onset of the vowel. B The proportion of *pier* responses across participants increased with longer VOT and higher F0 values. C While both VOT and F0 were perceptually relevant, the perceptual weight of VOT was slightly higher than the perceptual weight of F0.

Results

VOT duration and F0 height are perceptually relevant in baseline categorization

First, we examined the perceptual relevance of the two phonetic dimensions (VOT and F0) whose acoustic relationship were systematically manipulated to create canonical and accented speech stimuli. Twenty-three native speakers of English were instructed to categorize as ‘beer’ or ‘pier’ multiple repetitions of 25 baseline exemplars drawn from a two-dimensional phonetic grid perceptually varying from *beer* to *pier* across seven VOT and

F0 steps (see Fig. 1A). Consistent with the results of previous work^{11,12,17,19,66}, the number of *pier* responses increased for longer VOT and higher F0 values and decreased for shorter VOT and lower F0 values (VOT: $\beta = 0.89$, $ci = [0.80, 0.98]$, $p < 0.001$; F0: $\beta = 0.61$, $ci = [0.52, 0.69]$, $p < 0.001$; see Fig. 1B). Individual β coefficients were higher for VOT compared to F0 (two-sample t -test: $t_{38} = 2.60$, $ci = [0.12, 0.97]$, $p = 0.013$, Cohen’s $d = 0.82$). This finding indicates that, while both speech dimensions are perceptually relevant, VOT has stronger perceptual weight than F0 in the decision-making process (see Fig. 1C). Consequently, the influence of F0 on word recognition was

expected to be stronger for acoustically ambiguous VOT values (e.g., 15 ms VOT).

Reduced perceptual reliance on F0 in the accented condition

Having established the perceptual relevance of VOT duration and F0 height for the recognition of *beer* and *pier*, we proceeded to investigate the effects of experience with canonical and accented speech regularities on the recognition of the same words (see Fig. 2A). Following the baseline block (see Fig. 2B), participants were instructed to categorize a new subset of *beer* and *pier* exemplars across two conditions. In the canonical speech condition, they categorized canonical exemplars of *beer* (short VOT and low F0 values) and *pier* (long VOT and high F0 values). In the accented speech condition, they categorized accented pronunciations of the same words created by reversing the canonical correlation between VOT and F0 in English. Thus, stimuli in this condition consisted of a subset of *beer* exemplars with short VOT but high F0 values and *pier* exemplars with long VOT but low F0 values.

To assess the impact of canonical and accented speech regularities on participants' reliance on F0 during overt categorization, we focused on the categorization of two test exemplars with the same ambiguous VOT duration (15 ms) but differing in F0 height (220 Hz vs. 300 Hz). The results of the generalized logistic mixed-effects model fitted to the number of *peer* responses across test exemplars revealed a significant interaction between F0 and condition ($p < 0.001$; see Table 1 for the full statistical report), indicating that the effect of F0 was greater in the canonical speech context compared to the accented speech context.

In the canonical condition (see Fig. 2C), higher F0 increased the number of *peer* responses, whereas lower F0 increased the number of *beer* responses (post-hoc pairwise comparison: $\beta = 3.18$, $ci = [2.99 3.37]$, $SE = 0.09$, $z = 33.55$, $p < 0.001$, odds ratio = 24.2). In the accented speech condition, by contrast, the effect of F0 on the number of *beer* and *peer* responses was not statistically significant (post-hoc pairwise comparison: $\beta = 0.01$, $ci = [-0.13 0.16]$, $SE = 0.07$, $z = 0.22$, $p = 0.82$, odds ratio = 1.01). Together, these findings indicate that exposure to accented speech downweighted listeners' perceptual reliance on F0.

Mismatch negativity is modulated by accent

To investigate the neural mechanisms underlying the phonological processing of F0 contrasts in each condition, participants were exposed to oddball sequences of test exemplars after every 60 overt categorization trials (see Fig. 2B and Fig. 3A). Oddball sequences consisted of a frequently repeated standard sound (e.g., the test exemplar with ambiguous VOT and lower F0) randomly interspersed with an infrequent oddball sound (e.g., the test exemplar with ambiguous VOT and higher F0). We recorded EEGs while participants watched a silent video, and modelled the effects of condition and channel on individual MMN amplitudes using a linear mixed-effects model.

The amplitude of the MMN was stronger (i.e., more negative) in the canonical speech condition ($M = -0.34 \mu\text{V}$, $SEM = 0.03 \mu\text{V}$) compared to the accented speech condition ($M = 0.27 \mu\text{V}$, $SEM = 0.03 \mu\text{V}$; post-hoc pairwise comparison: $\beta = 0.62$, $ci = [0.54 0.84]$, $SE = 0.03$, $df = 1364$, $t = 16.03$, $p < 0.001$, Cohen's $d = 0.84$). In the canonical speech condition, the MMN showed a negative deflection peaking between 150 ms and 250 ms at fronto-central channels (see Fig. 3B). In the accented speech condition, by contrast, MMN deflection from baseline amplitudes (0–150 ms) was not detectable. Together, these findings indicate that the neural encoding of phonological contrasts by F0 was hindered in the accented speech condition, relative to the canonical condition.

Late negativity is modulated by accent

Following the MMN, we observed an LN component in the difference wave of each condition peaking around 400 ms (see Fig. 3C, bottom panels). To investigate the effects of accented speech on this component, we fitted a linear mixed-effects model to individual LN amplitudes extracted from the difference wave (see Fig. 3C, top left panel). We found stronger (i.e., more

negative) LN amplitudes in the canonical speech condition ($M = -0.38 \mu\text{V}$, $SEM = 0.03 \mu\text{V}$) compared to the accented speech condition ($M = 0.13 \mu\text{V}$, $SEM = 0.04 \mu\text{V}$) (post-hoc pairwise comparisons: $\beta = 0.52$, $ci = [0.42 0.62]$, $SE = 0.04$, $df = 1364$, $t = 10.63$, $p < 0.001$, Cohen's $d = 0.56$). This suggests that in the canonical condition, the processing of lexico-semantic differences in ambiguous VOT tokens was facilitated by the encoding of F0 cues, as reflected in the LN difference between standard and deviant trials contrasting in F0. In contrast, the LN difference in the accented speech condition was significantly smaller, indicating that accented speech hindered the encoding of lexico-semantic differences for the same speech exemplars.

N1 is modulated by accent

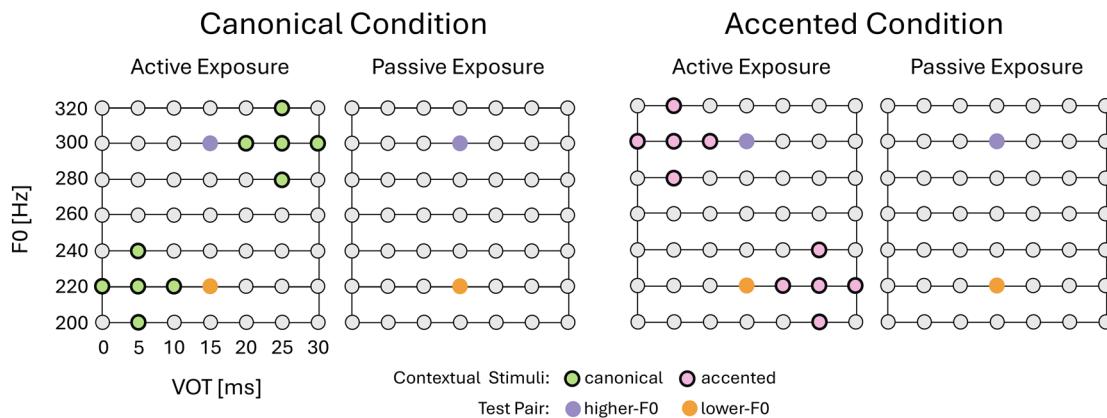
To further investigate the temporal dynamics underlying the MMN and LN differences documented above, we examined the amplitude of the N1 component during the overt categorization of test exemplars in active exposure blocks. We fitted a linear mixed-effects model to evaluate the effects of condition and channel on F0-level differences in N1 amplitude, and found stronger subphonemic encoding of F0 differences in the canonical speech condition (N1 difference: $M = 0.53 \mu\text{V}$, $SEM = 0.04 \mu\text{V}$) relative to the accented speech condition (N1 difference: $M = -0.20 \mu\text{V}$, $SEM = 0.03 \mu\text{V}$) (post-hoc pairwise comparison: $\beta = 0.74$, $ci = [0.65 0.83]$, $SE = 0.04$, $df = 1364$, $t = 16.09$, $p < 0.001$, Cohen's $d = 0.84$; see Fig. 4B, top-left). Condition differences were more pronounced at frontal sensors (see Fig. 4B, top-right). In the canonical speech condition, lower-F0 exemplars elicited more negative N1 peaks than higher-F0 exemplars (see Fig. 4B, bottom-left). In the accented speech condition, however, higher- and lower-F0 exemplars were associated with similar N1 amplitudes (Fig. 4B, bottom-right). Together, these findings suggest that the subphonemic encoding of F0 differences was severely disrupted by the accent at early cortical latencies linked to the processing of sound patterns in the auditory cortex.

Late positive component is modulated by accent

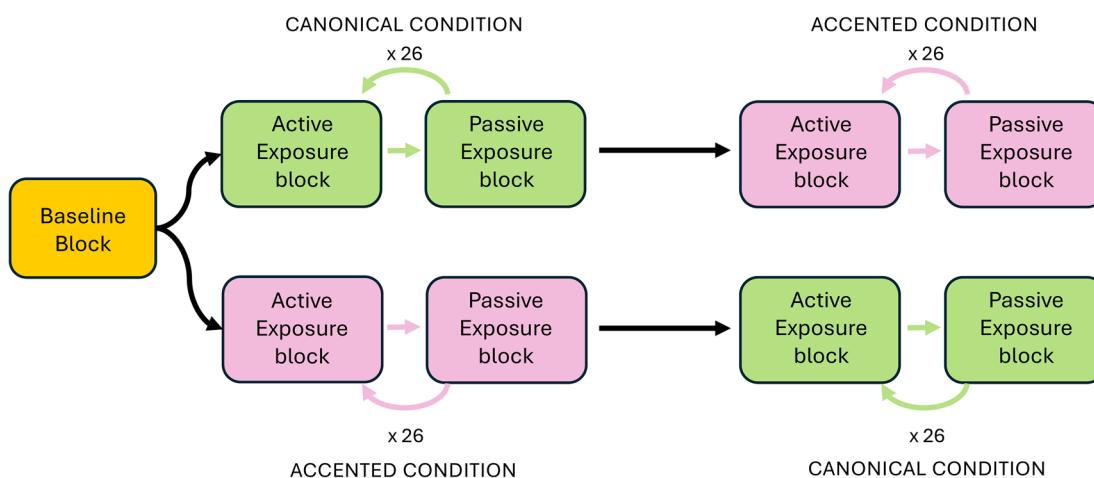
Finally, following Toscano and colleagues³⁰, we examined the effects of experience with canonical and accented speech regularities on post-phonological processing latencies following the N1 (Fig. 4C). In their study, Toscano and colleagues observed a parietal positivity corresponding to the P3 component. The P3 provides a complementary measure to N1 by indexing higher-order decision-making processes involved in post-phonemic categorization^{67,68}. This component is typically indexed by a positive-going deflection peaking between 300 and 800 milliseconds after the stimulus onset. In overt speech categorization tasks, stimulus-specific changes in the amplitude of the P3 have been shown to reflect the mapping between acoustic exemplars and phonological categories, with between-category tokens eliciting larger amplitude shifts than within-category ones³⁰. In contrast to the findings of Toscano and colleagues, we observed a later and more sustained parietal positivity, which is more consistent with an LPC component.

To examine the effect of F0 on LPC amplitude within each condition, we fitted a linear mixed-effects model predicting individual LPC amplitudes from test stimulus type (higher F0 vs. lower F0), speech condition (canonical vs. accented), and their interaction. We found a significant main effect of condition, with higher overall LPC amplitudes in the accented condition ($M = 0.37 \mu\text{V}$, $SEM = 0.03 \mu\text{V}$) compared to the canonical condition ($M = 0.19 \mu\text{V}$, $SEM = 0.04 \mu\text{V}$), suggesting slightly greater processing demand in response to accented speech ($\beta = 0.19$, $ci = [0.10 0.27]$, $SE = 0.04$, $df = 2750$, $t = 4.35$, $p < 0.001$, Cohen's $d = 0.16$). Crucially, we also observed a significant interaction between condition and F0 height ($\beta = 1.13$, $ci = [0.22 2.14]$, $SE = 0.52$, $df = 2750$, $t = 2.20$, $p = 0.028$), indicating that the effect of F0 on LPC amplitude was stronger in the canonical condition than in the accented condition. In the canonical condition, the effect of F0 was significant across several channels (Fig. 4C, top left), with higher F0 stimuli eliciting significantly greater LPC amplitudes over left posterior and centroparietal electrodes. This left-lateralized LPC enhancement is consistent with increased post-lexical processing demands during the recognition of acoustically ambiguous or less frequent words, potentially reflecting greater

A. Exposure Blocks by Condition



B. Exposure Blocks Control



C. F0 Weight in Active Exposure Blocks

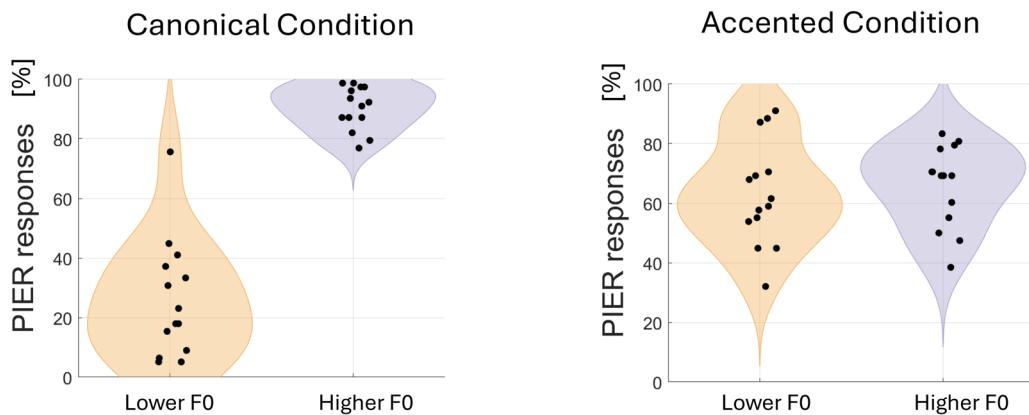


Fig. 2 | Active exposure categorization. A, B Following the baseline block, participants ($n = 23$) were instructed to categorize as *beer* or *pier* speech exemplars conveying canonical (green symbols) or accented (pink symbols) VOT \times F0 correlations. Participants were also instructed to categorize a subset of two F0-differentiated test exemplars with perceptually ambiguous VOT (active exposure), and passively exposed to oddball sequences of them while watching a silent video

(passive exposure). Overt categorization trials across active exposure blocks were used to elicit behavioural responses, as well as N1 and LPC components (see Fig. 4). Passive exposure trials were used to elicit MMN and LN components related to the neural encoding of voicing contrasts and lexico-semantic features (see Fig. 3). C In the canonical speech condition, the overt categorization of test exemplars was strongly influenced by stimulus differences in F0.

Table 1 | Logistic Mixed-Effects Model

Term	β	CI	SE	z	p
(Intercept)	-1.10	[-1.55 -0.65]	0.22	-4.83	1.35e-06
Higher-F0	3.18	[3 3.37]	0.09	33.56	<2e-16
Accented	1.76	[1.61 1.97]	0.08	21.90	<2e-16
Higher-F0:accented	-3.20	[-3.44 -2.96]	0.12	-26.46	<2e-16

lexical retrieval effort or categorization difficulty^{53,54}. In contrast, no statistically significant differences in LPC magnitude were found between higher and lower F0 stimuli in the accented speech condition (see Fig. 4C, top-left), underscoring a sharp divergence from the canonical condition, where higher F0 stimuli elicited stronger left-lateralized LPC responses. This null effect suggests that the accented speech context may have neutralized the perception of the F0 contrast.

Discussion

We investigated the neural mechanisms that support rapid adaptation to accented speech, focusing on how listeners adjust to short-term variability without destabilizing long-term phonological representations. We contrasted two alternative mechanisms: a restructuring of sound-to-category mapping versus a context-sensitive downweighting of acoustic-phonetic cues that deviate from canonical norms. Under the restructuring account, accent adaptation should show no differences in neural processing between canonical and accented speech. In contrast, the downweighting hypothesis predicts less robust neural encoding of non-canonical cues. Consistent with the downweighting hypothesis, we found that participants relied on both VOT and F0 cues in the baseline (canonical) block, with greater weighting on VOT. However, reliance on F0 decreased significantly in the accented block, indicating that listeners selectively downweighted the cue that deviated from long-term expectations. A parallel pattern emerged in the neural dataset: across all examined latencies (subphonemic, phonemic, and post-phonemic) neural encoding of F0 contrasts was less robust in the accented speech condition. These converging behavioural and neural results support the downweighting account, suggesting that rapid accent adaptation is achieved through flexible modulation of cue weighting rather than through a fundamental restructuring of native phonological categories.

Distinct cortical responses were elicited by the same speech exemplars, modulated by short-term speech regularities. This effect was particularly evident during passive exposure, which operates outside the focus of voluntary attention. Exposure to accented speech leads to the phonological neutralization of contrasts that deviate from the canonical speech norm. Rather than reversing the canonical mapping between speech cues (e.g., lower-F0 and higher-F0) and speech categories (e.g., voiced and voiceless sounds) to mirror the accent, listeners downplay the linguistic relevance of accented cues that depart from long-term expectations. This mechanism provides a functional balance between neural stability and plasticity, enabling adaptation to novel accents without compromising the stability of long-term representations. By selectively downweighting the linguistic relevance of noncanonical speech patterns, listeners optimize the alignment between short- and long-term representations of words. This accommodation strategy can potentially facilitate the segregation of linguistic (contrastive) from extralinguistic features (e.g., speaker's indexical features) in non-canonical auditory landscapes.

We found that adaptation to accented speech is facilitated by a cascade of rapid neural adjustments operating across multiple levels of hierarchical processing. These adjustments are first observed at short cortical latencies (~100 ms post-stimulus) associated with the subphonemic encoding of speech patterns in the auditory cortex. Our results show that already at this early stage of processing, acoustic speech dimensions are encoded in manner dependent upon local short-term speech regularities. In canonical speech contexts, lower-F0 exemplars elicit more robust N1 peaks than higher-F0 exemplars. This pattern aligns with prior research documenting

stronger N1 peaks for voiced consonants (e.g., /b/, short VOT) compared to voiceless consonants (e.g., /p/, long VOT). Similarly, low-frequency tones evoke stronger early cortical responses than high-frequency tones, likely due to cochlear dynamics, where low-frequency sounds activate a larger neuronal population⁶⁹. These results suggest that learning across short-term speech regularities blurs bottom-up differences in the encoding of speech patterns that deviate from the canonical norm. Notably, this is the case even across the 80 Hz differentiating test stimuli. The fact that accented speech can disrupt the subphonemic encoding of such a large acoustic difference indicates that short-term statistical regularities can drastically alter the internal state of the auditory system at early pre-attentive stages of processing.

After downweighting the functional relevance of noncanonical speech patterns, listeners no longer rely on them to discriminate phonemes at mid-cortical latencies associated with the MMN. The MMN is generated in frontal and temporal regions. Contemporary MMN models, particularly those grounded in predictive coding^{37,70}, propose that MMN amplitude reflects the strength of an error signal transmitted to frontal regions when an unexpected sound change is detected in temporal regions. Specifically⁷¹, MMN peaks become more negative when sound changes are linguistically relevant^{36,38}. In our study, the phonological neutralization of noncanonical phonetic contrasts at MMN processing latencies highlights the importance of neural stability for long-term linguistic norms while processing speech accents.

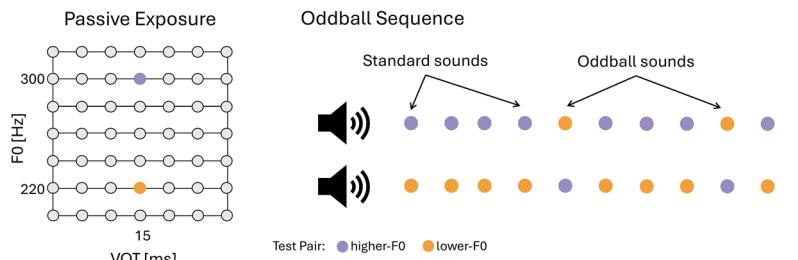
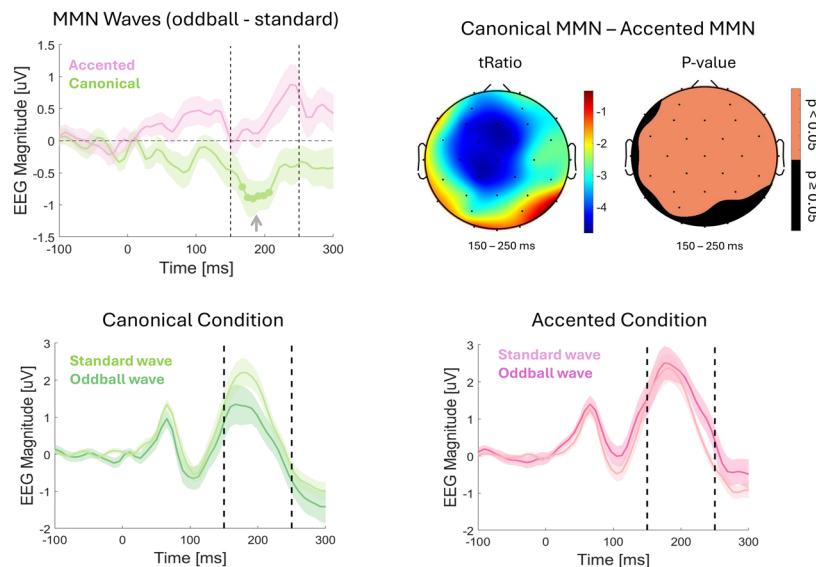
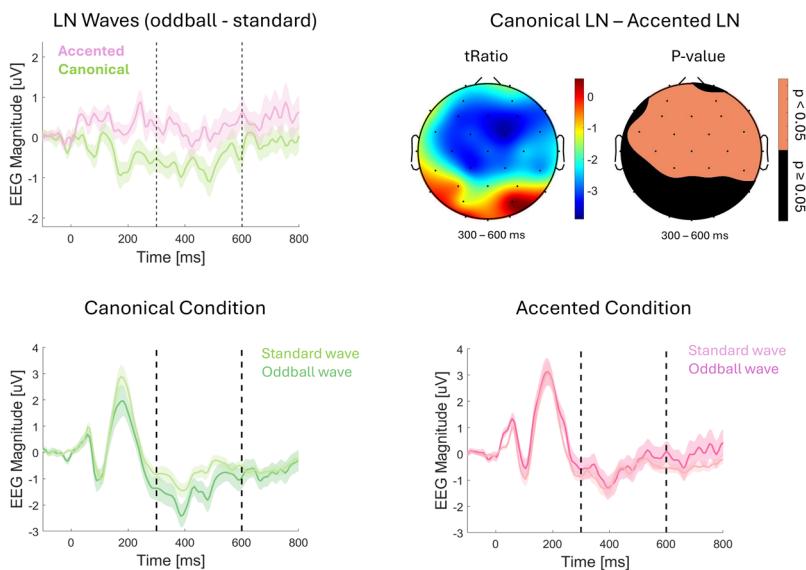
The interleaved statistical regularities that influenced the MMN indicate that the effects of dimension-based statistical learning extend beyond overt categorization. This finding demonstrates the utility of our experimental design. MMN elicitation depends on first-order statistical regularities conveyed by the contrast between a *standard* sound and a lower-probability *deviant* sound. However, the oddball paradigm used to elicit this contrast is unsuitable for examining the effects of second-order statistical regularities that are not directly tied to stimulus marginal probabilities. To address this limitation, we interleaved brief passive listening blocks of oddball sequences with blocks of overt categorization trials, exposing participants to canonical and noncanonical correlations of phonetic cues. The carryover effects from active categorization to passive listening highlights the persistent influence of dimension-based statistical learning on the perceptual accommodation of speech accents over time and across tasks.

The perceptual downweighting of noncanonical speech features influences post-phonological processing at longer neural processing latencies linked to lexico-semantic processing. In contrast to the canonical speech condition, in the accented speech condition LN magnitude was not modulated by F0 differences. This finding suggests that short-term experience with noncanonical speech regularities may result in the under-specification of lexical items. Thus, rather than restructuring the canonical mapping between sounds and meanings, our findings suggest that the temporary conflict between canonical and accented pronunciations of words is solved by adopting more flexible lexical representations.

The effects of contextual statistics on higher-order post-phonological processing are further supported by the LPC results, as the effect of F0 level on LPC was only significant in the canonical speech context. Unlike the findings reported by Toscano and colleagues³⁰, who observed a parietal positivity peaking at ~500 ms, our results revealed a later and more sustained posterior positivity between 600–800 ms, which is more consistent with an LPC component. The explanation for this divergence may lie in the nature of our test stimuli. Whereas Toscano and colleagues used auditory words with comparable lexical frequencies and contrasting along spanning multiple VOT values, our design focused on auditory words with VOT ambiguous and different lexical frequencies (*beer* is more frequent than *pier*). These stimulus differences may have increased categorization difficulty and shift processing demands toward mechanisms involved in lexical resolution and decision-making. Taken together, these findings underscore the sensitivity of the LPC to post-lexical ambiguity and highlight how listeners dynamically adapt their categorization strategies depending on the phonological clarity and lexical properties of the input.

Fig. 3 | Neural results for passive exposure.

A During passive exposure blocks, participants ($n = 23$) watched a silent video while being exposed to oddball sequences of two test exemplars differing in F0 height (high vs. low frequency). **B** Participants showed a more robust neural encoding of F0 contrasts in the canonical speech condition relative to the accented speech condition, approximately 200 ms following the onset of the oddball sound. Dots in the difference wave (top-left) mark time points where the wave significantly deviated from zero ($p < 0.05$), as determined by FDR-corrected one-sample t -tests. **C** Participants also showed a more robust neural encoding of lexico-semantic differences, as indexed by their LN component, in the canonical speech condition relative to the accented speech condition, approximately 400 ms following the onset of the oddball sound. Brain waves (mean and standard error mean) are shown for the following representative channels: Fz (MMN), Cz (LN).

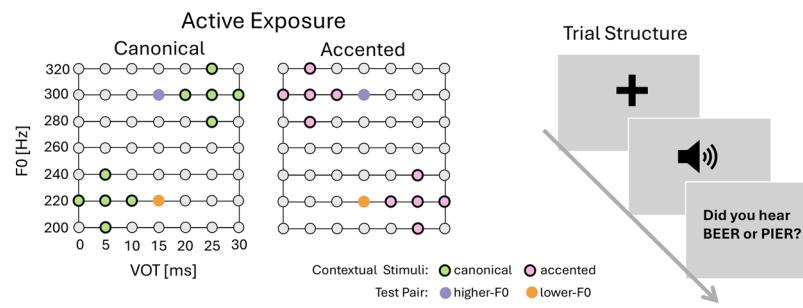
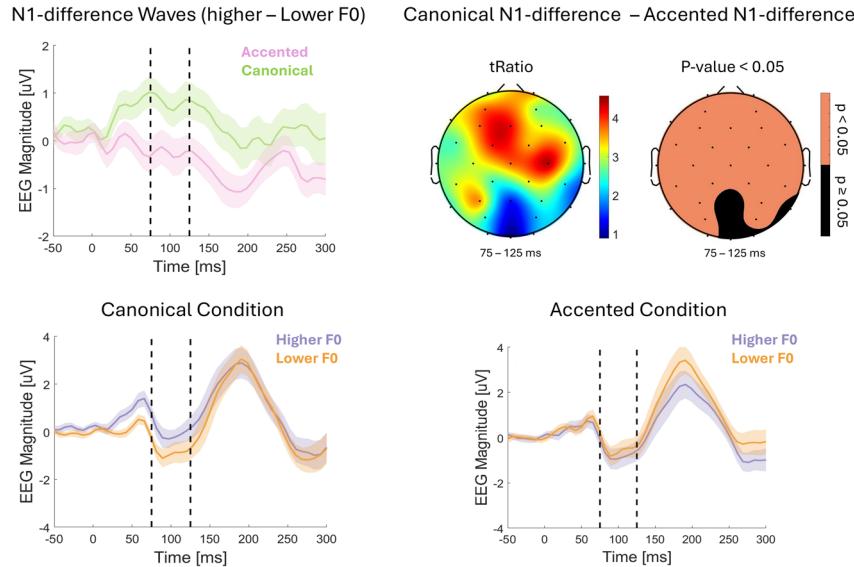
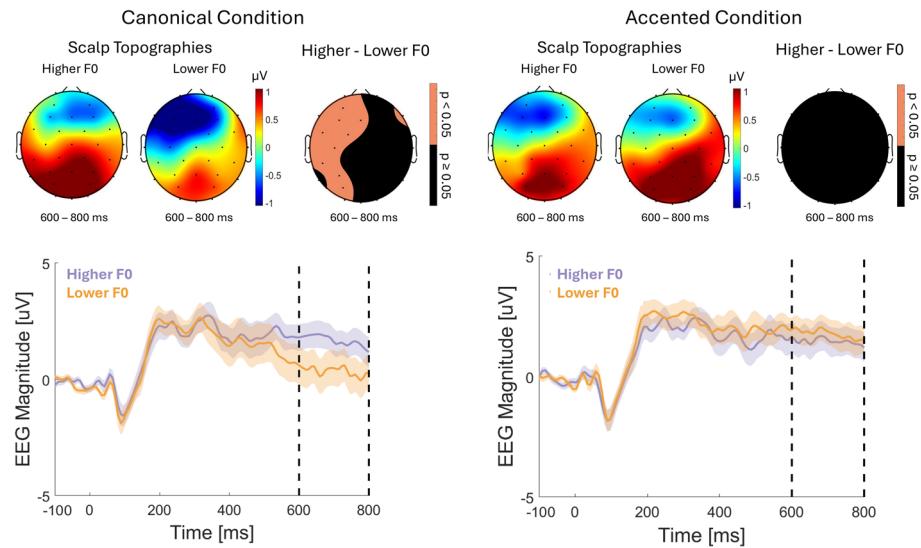
A. Mismatch Negativity (MMN) and Late Negativity (LN) Elicitation**B. MMN Results****C. LN Results**

In summary, our findings indicate that adaptation to accented speech is regulated by a trade-off between neural stability and flexibility. This challenges the traditional view that listeners adapt by fully recalibrating the canonical mapping between speech features and categories. In English, the canonical relationship between F0 and voicing categorization is such that higher F0 values are typically associated with voiceless responses (e.g., *pier*),

while lower F0 values are associated with voiced responses (e.g., *beer*). If listeners had fully recalibrated this mapping to reflect the reversed F0 distributions presented in the accented speech condition, we would expect a reversal in these associations; namely, increased categorization of high-F0 tokens as *beer* and low-F0 tokens as *pier*. However, this pattern was not observed. Rather than adopting a reversed mapping, listeners reduced their

Fig. 4 | Neural results for active exposure.

A During active exposure blocks, participants ($n = 23$) were instructed to categorize as *beer* or *pier* ten condition-specific and two test exemplars. B Participants showed a more robust neural encoding of subphonemic F0 differences in the canonical speech condition relative to the accented speech condition, approximately 100 ms following the stimulus onset. C They also showed a more robust neural encoding of lexical differences (*beer/pier*) in the canonical speech condition relative to the accented speech condition, approximately 650 ms following the onset of the stimulus. Brain waves (mean and standard error mean) are shown for the following representative channels: Fz (N1), Pz (LPC).

A. N1 and Late Positive Component (LPC) Elicitation**B. N1 Results****C. LPC Results**

reliance on F0 when it no longer conformed to the long-term statistical norm, consistent with a cue down-weighting strategy. This adaptive mechanism can be traced to early cortical latencies associated with the processing of fine-grained sound features in the auditory cortex, and results in the lexical under-specification of accented words at later stages of linguistic processing. Although the lexical under-specification of accented

speech may increase listening effort and hinder word recognition in acoustically ambiguous contexts, this strategy is far more economical than disrupting the linguistic relationships among native phonetic cues. Additionally, this adaptive mechanism provides a neural basis for the functional segregation between speech features that convey linguistically relevant information and those that signal speaker identity.

Limitations

One limitation of the current study concerns the selection of lexical stimuli. Although the *beer-pier* contrast offers a clean minimal pair and has been widely used in prior work, the two words differ in important lexical properties, most notably frequency. The higher frequency of *beer* relative to *pier* may have influenced lexical processing and reduced sensitivity to ERP components such as the P3, which have been observed in studies of cue encoding under conditions of uncertainty or conflict^{21,35}. Future studies could better isolate the time course of accent adaptation by using stimulus sets that control for lexical frequency and neighbourhood density across conditions.

Another limitation relates to the duration and ecological validity of the exposure. Participants were exposed to an artificial accent over the course of a single experimental session lasting only a few minutes. While this design isolates early neural markers of adaptation, it may not reflect the full range of mechanisms engaged over extended exposure. It remains possible that with sufficient experience, listeners may shift from downweighting inconsistent cues to fully restructuring phonological categories to match the accent. Longitudinal studies tracking this transition would help clarify the time course and stability of such category-level changes.

A third limitation involves the interpretability of EEG signals, particularly for components associated with post-phonological processing. While early components linked to auditory encoding and phonemic categorization are well-characterized in the ERP literature, later effects, especially those reflecting lexical or decisional processes, are more variable and harder to localize. Combining EEG with complementary methods such as MEG or fMRI could help disambiguate the cortical sources and functional roles of late-stage accent adaptation effects.

Finally, the present study focused on F0, a secondary cue to voicing in English. While this choice allowed us to probe subtle adjustments in cue weighting, it remains unclear how adaptation proceeds when primary cues are fully disrupted. Listeners may rely on qualitatively different mechanisms when canonical cues are unavailable or misleading, potentially recruiting visual or lexical information to resolve ambiguity. Future research should explore adaptation in contexts where primary cue reliability is manipulated, and assess how listeners integrate alternative sources of information during speech processing.

Data availability

Anonymized data and code supporting the findings of this study are publicly available at the following OSF repository <https://osf.io/n7pky/>.

Code availability

Anonymized data and code supporting the findings of this study are publicly available at the following OSF repository <https://osf.io/n7pky/>.

Received: 27 February 2025; Accepted: 10 October 2025;

Published online: 21 December 2025

References

1. Baese-Berk, M. M., Bradlow, A. R. & Wright, B. A. Accent-independent adaptation to foreign accented speech. *J. Acoust. Soc. Am.* **133**, EL174–EL180 (2013).
2. Cristia, A. et al. Linguistic processing of accented speech across the lifespan. *Front. Psychol.* **3**, 479 (2012).
3. Munro, M. J. & Dervling, T. M. Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Lang Speech* **38**, 289–306 (1995).
4. Van Engen, K. J. & Peele, J. E. Listening effort and accented speech. *Front. Hum. Neurosci.* <https://doi.org/10.3389/fnhum.2014.00577> (2014).
5. Bent, T., Atagi, E., Akbik, A. & Bonifield, E. Classification of regional dialects, international dialects, and nonnative accents. *J. Phonetics* **58**, 104–117 (2016).
6. Bent, T. & Holt, R. F. Representation of speech variability. *WIREs Cogn. Sci.* **8**, e1434 (2017).
7. Bradlow, A. R. & Bent, T. Perceptual adaptation to non-native speech. *Cognition* **106**, 707–729 (2008).
8. Ullas, S., Bonte, M., Formisano, E. & Vroomen, J. Adaptive plasticity in perceiving speech sounds. In *Speech Perception. Springer Handbook of Auditory Research*, (eds Holt, L. L., Peele, J. E., Coffin, A. B., Popper, A. N. & Fay, R. R.) 74 (Springer, 2022).
9. Jasmin, K., Tierney, A., Obasih, C. & Holt, L. Short-term perceptual reweighting in suprasegmental categorization. *Psychon. Bull. Rev.* **30**, 373–382 (2021).
10. Lehet, M. & Holt, L. L. Dimension-based statistical learning affects both speech perception and production. *Cogn. Sci.* **41**, 885–912 (2017).
11. Schertz, J., Cho, T., Lotto, A. & Warner, N. Individual differences in perceptual adaptability of foreign sound categories. *Atten Percept. Psychophys.* **78**, 355–367 (2016).
12. Wu, Y. C. & Holt, L. L. Phonetic category activation predicts the direction and magnitude of perceptual adaptation to accented speech. *J. Exp. Psychol. Hum. Percept. Perform.* **48**, 913–925 (2022).
13. Idemaru, K. & Holt, L. L. Generalization of dimension-based statistical learning. *Attention Percept. Psychophys.* **82**, 1744–1762 (2020).
14. Liu, R. & Holt, L. L. Dimension-based statistical learning of vowels. *J. Exp. Psychol. Hum. Percept. Perform.* **41**, 1783–1798 (2015).
15. Abramson, A. S. & Whalen, D. H. Voice onset time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *J. Phonetics* **63**, 75–86 (2017).
16. Chodroff, E. & Wilson, C. Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *J. Phonetics* **61**, 30–47 (2017).
17. Dmitrieva, O., Llanos, F., Shultz, A. A. & Francis, A. L. Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *J. Phonetics* **49**, 77–95 (2015).
18. Kingston, J., Diehl, R. L., Kirk, C. J. & Castleman, W. A. On the internal perceptual structure of distinctive features: The [voice] contrast. *J. Phonetics* **36**, 28–54 (2008).
19. Llanos, F., Dmitrieva, O., Shultz, A. & Francis, A. L. Auditory enhancement and second language experience in Spanish and English weighting of secondary voicing cues. *J. Acoust. Soc. Am.* **134**, 2213–2224 (2013).
20. Shultz, A. A., Francis, A. L. & Llanos, F. Differential cue weighting in perception and production of consonant voicing. *J. Acoust. Soc. Am.* **132**, EL95–EL101 (2012).
21. Toscano, J. C. & McMurray, B. Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cogn. Sci.* **34**, 434–464 (2010).
22. Winn, M. B., Chatterjee, M. & Idsardi, W. J. Roles of voice onset time and F0 in stop consonant voicing perception: effects of masking noise and low-pass filtering. *J. Speech Lang. Hearing Res.* **56**, 1097–1107 (2013).
23. Hodson, A. J., Shinn-Cunningham, B. G. & Holt, L. L. Statistical learning across passive listening adjusts perceptual weights of speech input dimensions. *Cognition* **238**, 105473 (2023).
24. Idemaru, K. & Holt, L. L. Word recognition reflects dimension-based statistical learning. *J. Exp. Psychol. Hum. Percept. Perform.* **37**, 1939 (2011).
25. Brown, V. A., McLaughlin, D. J., Strand, J. F. & Van Engen, K. J. Rapid adaptation to fully intelligible nonnative-accented speech reduces listening effort. *Quart. J. Exp. Psychol.* **73**, 1431–1443 (2020).
26. Clarke, C. M. & Garrett, M. F. Rapid adaptation to foreign-accented English. *J. Acoust. Soc. Am.* **116**, 3647–3658 (2004).
27. Floccia, C., Butler, J., Goslin, J. & Ellis, L. Regional and foreign accent processing in English: can listeners adapt? *J. Psychol. Res.* **38**, 379–412 (2009).
28. Näätänen, R. et al. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* **385**, 432–434 (1997).

29. Näätänen, R. & Picton, T. The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology* **24**, 375–425 (1987).

30. Toscano, J. C., McMurray, B., Denhardt, J. & Luck, S. J. Continuous perception and graded categorization: electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychol. Sci.* **21**, 1532–1540 (2010).

31. Sharma, A., Marsh, C. M. & Dorman, M. F. Relationship between N1 evoked potential morphology and the perception of voicing. *J. Acoust. Soc. Am.* **108**, 3030–3035 (2000).

32. Sharma, A. & Dorman, M. F. Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *J. Acoust. Soc. Am.* **106**, 1078–1083 (1999).

33. Steinschneider, M., Volkov, I. O., Noh, M. D., Garell, P. C. & Howard, M. A. Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. *J. Neurophysiol.* **82**, 2346–2357 (1999).

34. Getz, L. M. & Toscano, J. C. The time-course of speech perception revealed by temporally-sensitive neural measures. *WIREs Cogn. Sci.* **12**, e1541 (2021).

35. Pereira, O., Gao, Y. A. & Toscano, J. C. Perceptual encoding of natural speech sounds revealed by the N1 event-related potential response. *Auditory Percep. Cogn.* **1**, 112–130 (2018).

36. García-Sierra, A., Ramírez-Esparza, N., Silva-Pereyra, J., Siard, J. & Champlin, C. A. Assessing the double phonemic representation in bilingual speakers of Spanish and English: An electrophysiological study. *Brain Lang.* **121**, 194–205 (2012).

37. Garrido, M. I. et al. The functional anatomy of the MMN: A DCM study of the roving paradigm. *NeuroImage* **42**, 936–944 (2008).

38. Näätänen, R., Paavilainen, P., Rinne, T. & Alho, K. The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin. Neurophysiol.* **118**, 2544–2590 (2007).

39. Wig, N. & García-Sierra, A. Matching the Mismatch: The interaction between perceptual and conceptual cues in bilinguals' speech perception. *Bilingualism* **24**, 467–480 (2021).

40. Dehaene-Lambertz, G., Dupoux, E. & Gout, A. Electrophysiological correlates of phonological processing: a cross-linguistic study. *J. Cogn. Neurosci.* **12**, 635–647 (2000).

41. Kutas, M. & Federmeier, K. D. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. R. Psychol.* **62**, 621–647 (2011).

42. Wetzel, N. & Schröger, E. On the development of auditory distraction: a review. *PsyCh J.* **3**, 72–91 (2014).

43. Chwilla, D. J., Brown, C. M. & Hagoort, P. The N400 as a function of the level of processing. *Psychophysiology* **32**, 274–285 (1995).

44. Rhodes, S. M. & Donaldson, D. I. Association and not semantic relationships elicit the N400 effect: electrophysiological evidence from an explicit language comprehension task. *Psychophysiology* **45**, 50–59 (2008).

45. Čepenienė, R. et al. Event-related potentials associated with sound discrimination versus novelty detection in children. *Psychophysiology* **41**, 130–141 (2004).

46. Kuuluvainen, S. *Cortical Processing Of Sublexical Speech And Nonspeech Sounds In Children And Adults*. <https://helda.helsinki.fi/items/ecd95d5f-11dc-4c15-8ce3-6c41789118ca> (2016).

47. Schröger, E. & Wolff, C. Attentional orienting and reorienting is indicated by human event-related brain potentials. *Neuroreport* **9**, 3355–3358 (1998).

48. Justo-Guillén, E. et al. Auditory mismatch detection, distraction, and attentional reorientation (MMN-P3a-RON) in neurological and psychiatric disorders: a review. *Int. J. Psychophysiol.* **146**, 85–100 (2019).

49. Lindborg, A., Musiolek, L., Ostwald, D. & Rabovsky, M. Semantic surprise predicts the N400 brain potential. *Neuroimage. Rep.* **3**, 100161 (2023).

50. Jamison, C. et al. Preliminary investigation of the passively evoked n400 as a tool for estimating speech-in-noise thresholds. *Am. J. Audiol.* **25**, 344–358 (2016).

51. Petten, C. V., Kutas, M., Kluender, R., Mitchiner, M. & McIsaac, H. Fractionating the word repetition effect with event-related potentials. *J. Cognit. Neurosci.* **3**, 131–150 (1991).

52. Swaab, T. Y., Brown, C. & Hagoort, P. Understanding ambiguous words in sentence contexts: Electrophysiological evidence for delayed contextual selection in Broca's aphasia. *Neuropsychologia* **36**, 737–761 (1998).

53. Evans, K. M. & Federmeier, K. D. The memory that's right and the memory that's left: Event-related potentials reveal hemispheric asymmetries in the encoding and retention of verbal information. *Neuropsychologia* **45**, 1777–1790 (2007).

54. Kandhadai, P. & Federmeier, K. D. Automatic and controlled aspects of lexical associative processing in the two cerebral hemispheres. *Psychophysiology* **47**, 774–785 (2010).

55. Stuss, D. T., Picton, T. W., Cerri, A. M., Leech, E. E. & Stethem, L. L. Perceptual closure and object identification: electrophysiological responses to incomplete pictures. *Brain Cogn.* **19**, 253–266 (1992).

56. Zhang, X. & Holt, L. L. Simultaneous tracking of coevolving distributional regularities in speech. *J. Exp. Psychol. Hum. Percept. Perform.* **44**, 1760 (2018).

57. Boersma, P. & Weenink, D. Praat: doing phonetics by computer [Computer program]. Version 6.0. 37. Retrieved March **14**, 2018 (2018).

58. The MathWorks Inc. *MATLAB Version: 9.13.0 (R2022b)*, Natick, Massachusetts. <https://www.mathworks.com> (2020).

59. Team, R. C. *RA Language And Environment For Statistical Computing, R Foundation for Statistical*. <https://www.r-project.org> (2020).

60. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. *lmerTest* package: tests in linear mixed effects models. *J. Stat. Softw.* **82**, 1–26 (2017).

61. Lenth, R., Singmann, H., Love, J., Buerkner, P. & Herve, M. *Emmeans: Estimated marginal means, aka least-squares means*. *R Package Version 1*, 3 (2018).

62. Bidelman, G. M. & Walker, B. S. Attentional modulation and domain-specificity underlying the neural organization of auditory categorical perception. *Eur. J. Neurosci.* **45**, 690–699 (2017).

63. Gramfort, A. et al. MNE software for processing MEG and EEG data. *neuroimage* **86**, 446–460 (2014).

64. Mahajan, Y., Peter, V. & Sharma, M. Effect of EEG referencing methods on auditory mismatch negativity. *Front. Neurosci.* **11**, 560 (2017).

65. Šoškić, A., Jovanović, V., Styles, S. J., Kappenman, E. S. & Ković, V. How to do better N400 studies: reproducibility, consistency and adherence to research standards in the existing literature. *Neuropsychol. Rev.* **32**, 577–600 (2022).

66. Llanos, F. & Francis, A. L. The effects of language experience and speech context on the phonetic accommodation of english-accented spanish voicing. *Lang Speech* **60**, 3–26 (2017).

67. Picton, T. W. The P300 wave of the human event-related potential. *J. Clin. Neurophysiol.* **9**, 456–456 (1992).

68. Polich, J. Updating P300: An integrative theory of P3a and P3b. *Clin. Neurophysiol.* **118**, 2128–2148 (2007).

69. Picton, T. W., Woods, D. L. & Proulx, G. B. Human auditory sustained potentials. II. *Stimulus Relationships*. *Electroencephalogr. Clin. Neurophysiol.* **45**, 198–210 (1978).

70. Garrido, M. I., Kilner, J. M., Stephan, K. E. & Friston, K. J. The mismatch negativity: a review of underlying mechanisms. *Clin. Neurophysiol.* **120**, 453–463 (2009).

71. Naatanen, R. Mismatch negativity (MMN): perspectives for application. *Int. J. Psychophysiol.* **37**, 3–10 (2000).

Acknowledgements

This research was supported by a doctoral dissertation grant to LH and YW from the National Science Foundation (BCS2420979; BCS2346989). The

funders (NSF) had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

Conceptualization, Y.W. and L.H.; Methodology, Y.W., L.H. and F.L.; Software, Y.W. and F.L.; Data curation, F.L.; Visualization, F.L.; Writing—original draft, F.L.; Writing—review and editing, T.A., Y.W. and L.H.; Funding acquisition, L.H. and Y.W.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s44271-025-00345-z>.

Correspondence and requests for materials should be addressed to Fernando Llanos or Lori L. Holt.

Peer review information *Communications Psychology* thanks Urs Maurer and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Jixing Li and Troy Ka-Yan Lui. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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

© The Author(s) 2025