JASA ARTICLE

Dichotic training modulates phonetic feature processing in late English learners: Evidence from categorical perception and the mismatch negativity

Margot Bouhon,[1,2,3,4,a)] Anna Fiveash,[5,6] Jade Charron,[7] Agathe Foureix,[7] Barbara Tillmann,[2,6] and Nathalie Bedoin[3,7]

[1]*Laboratoire de Psychologie Sociale et Cognitive (LAPSCO), CNRS-UMR 6024, Université Clermont-Auvergne, Clermont-Ferrand, France*

[2]*Centre de Recherche en Neurosciences de Lyon, Equipe Cognition Auditive et Psychoacoustique, CNRS-UMR 5292, Université Lyon 1, Inserm U 1028, Bron, France*

[3]*Centre de Recherche en Neurosciences de Lyon, Equipe Trajectoires, CNRS-UMR 5292, Université Lyon 1, Inserm U 1028, Bron, France*

[4]*Contentsquare Foundation, Paris, France*

[5]*The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Sydney, Australia*

[6]*LEAD UMR5022, Université Bourgogne Europe, CNRS, 21000, Dijon, France*

[7]*Institut de Psychologie, Université Lyon 2, Lyon, France*

ABSTRACT:
In English, voiceless stop consonants (i.e., /p-t-k/) have long voice onset times (VOT), and voiced stop consonants (i.e., /b-d-g/) have short VOTs. In French, voiceless stop consonants have short VOTs while voiced stop consonants have pre-lag VOTs, which can make late second language (L2) learning of English difficult. The Asymmetric Sampling in Time (AST) model proposes that slow and fast cerebral rhythms dominate in the auditory cortex of the right hemisphere and left hemisphere (LH), respectively, and preferentially support processing of long and short speech events. Our study extends the application of this model to sub-phonemic levels, where short VOTs were presented to the right ear (stimulating the LH), and long VOTs were presented to the left ear (dichotic stimulation). Pre-/post-tests assessed categorical perception performance in adults trained dichotically, binaurally, or without training. Consonant perception improved in dichotic and binaural groups, with generalization of training only in the dichotic group (experiment 1). Following dichotic training, the Mismatch Negativity amplitude increased for voiced English deviants, and hemispheric dominance shifted appropriately for voiceless deviants (experiment 2). These findings demonstrate that perceptual training grounded in an AST framework can induce phonological perception changes in late English learners, suggesting promising directions for L2 perceptual training.
 https://doi.org/10.1121/10.0036725

(Received 7 October 2024; revised 28 April 2025; accepted 30 April 2025; published online 27 May 2025)

[Editor: Sven Mattys] Pages: 3907–3921

I. INTRODUCTION

Speech perception in a second language (L2) can present significant challenges for late learners, necessitating substantial modifications to already developed speech processing circuits (Luk *et al.*, 2020). Recent advancements in neuroscience suggest that the difficulty observed for adults learning an L2 may not be based primarily on a decline in brain plasticity with age. Instead, adult learners' challenges are more likely influenced by motivational factors (DeKeyser and Larson-Hall, 2005), suboptimal learning conditions (Kang *et al.*, 2021), and insufficient input frequency (Ellis, 2002). In the case of French speakers learning English, difficulties may also arise because the assimilation of new phonemes to first language (L1) phonological categories are initially made without clear perception of phonetic differences.

A. French speakers learning English voicing contrasts

For French speakers learning English, two additional sources of phonological errors are anticipated. First, the shared grapheme–phoneme correspondences for consonants between the two languages (e.g., P-/p/, B-/b/) may mask the phonetic variations between the two languages, which include differences in acoustic and articulatory properties. Consequently, when reading English words, French speakers are likely to interpret them using native phonetic contrasts, reinforcing the incorrect assumption that English and French consonants exhibit phonetic similarity. Second, late learners tend to assimilate L2 sounds to the phonological category that is phonetically closest to the related L1 category (Grimaldi *et al.*, 2014), potentially leading to perceptual and production errors.

In this context, the distinction between English voiced and voiceless consonants poses a particular challenge for

[a)]Email: bouhonmargot@gmail.com

French-speaking English learners. The present study focuses on the voicing contrast, which primarily depends on voice-onset-time (VOT) in both languages, albeit in distinct ways. VOT is defined as the duration between the release of a stop consonant (the burst) and the onset of vocal fold vibration (the onset of voicing). In English consonant–vowel (CV) syllables, voiced stop consonants (/b-d-g/) have a short positive VOT, while aspirated voiceless stop consonants (/p-t-k/) are characterized by a long positive VOT (phonetically, [p^h-t^h-k^h]). In contrast, French voiced stop consonants are marked by a pre-lag VOT, while voiceless stop consonants have a short positive VOT ([p-t-k]), similar to the VOT pattern found in English voiced stops.

Consequently, French learners experience what is referred to as "single category assimilation" within the Perceptual Assimilation Model (PAM-L2) (Best and Tyler, 2007) and a "new scenario" in the Second Linguistic Perception model (L2LP) (Escudero, 2009; van Leussen and Escudero, 2015). In these models, the English voicing contrasts (L2) risk being initially assimilated into a single L1 category. For instance, the English sounds [p^h] and [b] may be perceived as atypical and typical exemplars of the L1 /p/ category, respectively. L2 perception has been suggested to improve when learners shift toward a 'two-category assimilation', according to the PAM-L2 model (Tyler, 2019), or a "similar scenario" in the L2LP framework, wherein tokens such as [p^h] and [b] are assimilated into distinct L1 categories (i.e., /p/ and /b/). However, ambiguity arising from differences in the associations between short vs long VOTs and voiced vs voiceless consonants in French and English may hinder learners' progression to this advanced stage of L2 perception. Perceptual training targeting specific phonological contrasts has been put forward as a method to aid learners in developing L2 phonological categories that reflect phonetic differences in the L2.

The revised Speech Learning model (SLM-r) (Flege and Bohn, 2021) posits that the capacity to establish new L2 phonetic–phonemic associations persists throughout the lifespan, contingent upon factors such as the acoustic similarity between L1 and L2 sounds, the precision of L1 categories, and the quantity of authentic L2 input received (Ellis, 2002). The developmental process from naive non-native to native-like L2 speech perception is the object of study of the L2LP model (van Leussen and Escudero, 2015). Similarly, the PAM-L2 (Best and Tyler, 2007) hypothesizes that late learners can attain heightened sensitivity to L2 phonological contrasts, especially following targeted training emphasizing specific phonetic distinctions between L2 phonemes.

The training program employed in our study was grounded in these theoretical frameworks, as well as the neurolinguistic Asymmetric Sampling in Time model (AST) (Poeppel, 2003).

B. Extending the AST model to phonetic features using dichotic training

The AST model posits that speech perception involves bilateral non-primary auditory cortices, with preferential hemispheric processing depending on frequency (Boemio *et al.*, 2005; Teng *et al.*, 2016). Temporal processing of speech is suggested to be governed by the synchronization of rapid, left-hemisphere–dominant endogenous neural oscillations for the processing of short linguistic units (phonemes), and slower, right-hemisphere–dominant neural oscillations for the processing of longer linguistic units (syllables) (Oderbolz *et al.*, 2025).

The current study was based on this hemispheric specialization, with the hypothesis that perceptual training involving dichotic presentation of English stop consonants would improve French learners' ability to identify and discriminate English voiced and voiceless stop consonants. This training method involves presenting stimuli to only one ear to specifically stimulate the auditory cortex in the left hemisphere (LH) or right hemisphere (RH). Functional magnetic resonance imaging (fMRI) data have demonstrated that stimuli presented to the right ear specifically enhances LH activation, while stimuli presented to the left ear specifically enhances RH activation (Stefanatos *et al.*, 2008). Consistently, behavioral dichotic listening experiments have shown that right-handed individuals typically exhibit a right ear advantage (REA): they recall more the words presented to the right ear than those presented to the left ear, when two words were simultaneously presented to the left and right ears. The REA is suggested to reflect both the strong contralateral relationship between the ears and the auditory cortices, and the predominant involvement of the LH in phonological processing (Hugdahl and Wester, 1992; Kimura, 1967).

It is important to note that the contrast between short and long VOTs examined in the current study does not align precisely to the distinction between short events (phonemes) and long events (syllables or prosodic units) described in the AST model. Therefore, our hypothesis entails an additional theoretical implication. Specifically, if dichotic training leads to a greater improvement in the perception of the English voicing contrast compared to binaural training (control condition), this would provide support for extending the AST model to sub-phonemic levels.

This proposed extension is corroborated by previous dichotic training studies, as well as by dichotic listening research testing hemispheric asymmetry in the perception of phonetic features, including temporal features. For example, Bouhon *et al.* (2023) showed improved perception of duration as a phonological cue distinguishing short and long English vowels in a proof-of-concept study. In this study, perceptual training exercises were designed to target rapid or slow sampling skills, depending on vowel duration. Building on this framework, in the current study, it was hypothesized that amplifying hemisphere-specific processing would improve sensitivity to—and/or awareness of—the phonetic features distinguishing English voiced and voiceless stop consonants.

Furthermore, dichotic listening experiments involving concurrent words differing by only one phonetic feature (voicing of their initial consonants) have revealed divergent hemispheric asymmetries between English (Rimol *et al.*, 2006) and French (Bedoin *et al.*, 2010), consistent with the application of the AST framework to sub-phonemic levels for VOT processing. For example, in English-speaking participants, the REA decreased when a voiceless stop consonant (i.e., long VOT) was presented to the left ear and a voiced stop consonant (i.e., short VOT) was presented to the right ear. In contrast, in French speakers, the REA decreased when a voiced stop consonant was presented to the left ear and a voiceless stop consonant was presented to the right ear. For both languages, these findings suggest RH lateralization for long VOTs. Moreover, for word pairs differing by a single phonetic feature, the REA was higher if this feature was associated with brief acoustic cues, which is the case for place of articulation, supporting LH lateralization for short phonetic features (Studdert-Kennedy and Shankweiler, 1970). The aim of the current study was to test whether dichotic training can improve perception at the sub-phonemic level, by investigating the processing of English VOT. This hypothesis is consistent with an extension of the AST model to the sub-phonemic level, in line with the AST prediction that the auditory cortex is particularly sensitive to sound contours, including VOTs (Eggermont, 2001).

C. A training program targeting the learning of sub-phonemic rules

To optimize hemispheric recruitment for phonetic analysis, the LH and RH were trained separately using the current perceptual program. This training harnessed hemispheric specialization to enable better processing of voiced and voiceless English stop consonants based on their VOT duration. It was hypothesized that dichotic training would lead to greater improvement than binaural training in the identification and discrimination of English CVs (as determined by their voicing features).

The perceptual training program used in the previous proof-of-concept study resulted in greater accuracy for the production of length differences between short /ɪ/ versus long /iː/ vowels after dichotic training compared to binaural training, for both dyslexic and non-dyslexic French-speaking adults learning English (Bouhon *et al.*, 2023). This outcome was interpreted as evidence that trained perception skills transferred to L2 production.

In the present study, trained consonants (/p-b/, /k-g/) were compared against untrained consonants (/t-d/), to determine whether sub-phonemic rules were acquired and generalized to untrained items. This hypothesis aligns with the featural approach (de Jong *et al.*, 2009; Olson, 2019), which suggests that L2 learners can master new sub-phonemic components and rules. Evidence of generalization to untrained L2 phonemes would support this approach, in contrast to the segmental perspective (Flege and Bohn, 2021), which asserts that each phoneme must be specifically trained to be learned.

The goal of the current study was to investigate whether L2 perception of sub-phonemic features could be enhanced using principles outlined in the AST model. Three groups were recruited to investigate this question: (a) a control group who received no training, (b) a group who completed the training with binaural stimulation, and (c) a group who completed the training with dichotic stimulation. Categorical perception tests were administered twice to all groups. To ensure test–retest reliability and establish a baseline, the control group completed the categorical perception tests at two time points without training in between. It was predicted that both the binaural and dichotic stimulation groups would show an improvement in English voicing perception (measured through the categorical perception tests) after the perceptual training program. However, it was hypothesized that dichotic stimulation (i.e., lateralized presentation) of voiced and voiceless English consonants would facilitate sub-phonemic rule learning, resulting in enhanced categorical perception of both trained and untrained consonants (experiment 1). To further investigate the effect of dichotic training on phonological perception, phoneme discrimination was assessed in the dichotic listening group with the mismatch-negativity (MMN) response in an oddball electroencephalographic (EEG) paradigm (experiment 2). The MMN amplitude was predicted to increase following dichotic training, indicating that listeners developed preattentional sensitivity to the English voicing contrast.

II. EXPERIMENT 1: BEHAVIORAL EVALUATION OF THE TRAINING PROGRAM

A. Methods

1. Participants

Sixty monolingual native French speakers recruited from French universities participated: 20 were enrolled in the Dichotic Group [$M_{age} = 21.50$ years; standard deviation (SD) = 2.60; 18 females), 20 in the Binaural Group ($M_{age} = 21.60$ years; $SD = 2.80$; 16 females), and 20 in the Untrained Group ($M_{age} = 22.30$ years; $SD = 1.90$; 16 females). The sample size was consistent with previous studies in this field (e.g., Grenon *et al.*, 2019; Tamminen *et al.*, 2015; Ylinen *et al.*, 2006). Participants were selected based on their low-intermediate proficiency in English, not exceeding the B2 level [Common European Framework of Reference for Languages (CECRL)]. All were right-handed ($\geq 80\%$ "right hand" responses on the Edinburgh Handedness Inventory) and exhibited a REA for processing short phonetic features.

REA was signaled by more recalled words from the right ear than from the left ear (e.g., /tas/, for /pas/-/tas/), in a dichotic listening test where the two presented words only differed by place of articulation (Bedoin *et al.*, 2010). A REA was interpreted as indicative of LH dominance for processing short phonetic cues. All participants reported normal hearing, normal or corrected-to-normal vision, and no history of neurological, psychiatric, or language-related disorders, nor any childhood learning disabilities. The study was approved by the French national ethics committee. Participants gave informed consent after the procedure had been explained to them. As compensation, they received books and board games.

2. Stimuli

a. Training. A total of 172 English words beginning with /p/, /t/, /k/, or /g/ were used to create minimal (e.g., pad-bad, 69%) and near-minimal (e.g., gaze-case, 31%) pairs. Two female and two male speakers with a standard British accent produced the words and one version of each was randomly selected. Each word was illustrated with five different Creative Commons licensed images (860 images in total); one for each presentation during the training.

b. Testing sessions. English continua (from /ki/ to /gi/, and from /ti/ to /di/) were constructed using Praat software (Phonetic Sciences, Amsterdam, the Netherlands), by systematically varying the VOT of consonants from 0 to +60 in 10 ms increments, resulting in seven stimuli per continuum (Fig. 1). Two consonant pairs that differed by place of articulation were included, as such variations may influence the categorization of voicing contrasts (Benki, 2001). The syllables were produced by one of the two males described above. The vowel /i:/ was selected instead of the more traditional /a/ in CV syllables, because it provided a clearer context for isolating the effect of VOT duration from other voicing cues. The /i:/ vowel is characterized by relatively stable formants over time (Hillenbrand *et al.*, 1995), which facilitates the examination of purely temporal processing (VOT). The /ki/-/gi/ continuum was utilized to assess the learning effect, while the /ti/-/di/ continuum was utilized to assess potential transfer effects.

3. Procedure

a. General procedure. All three groups participated in a pre-test/post-test design over seven consecutive days (see Fig. 2). During the testing sessions (days 1 and 7), the categorical perception of English consonants /k/, /g/, /t/, and /d/ was assessed using stimuli presented binaurally to participants who were seated in a sound attenuated room and wore headphones (Beyerdynamic DT 770 Pro, Beyerdynamic, Heilbronn, Germany). They responded by pressing designated computer keys. From days 2–6, the participants in the Dichotic and Binaural training groups were instructed to complete the training program independently at home, ensuring that the right earphone was placed in the right ear and the left earphone in the left ear. To monitor compliance and ensure accurate training, participants were contacted daily. In the Dichotic group, participants were not told of the lateralized presentation of stimuli. The program was structured so that English words beginning with /g/ or /b/ (i.e., voiced stop consonants, short VOT) were presented to the right ear, while English words beginning with /k/ or /p/ (i.e., voiceless stop consonants, long VOT) were presented to the left ear, with white noise competing in the other ear. In the binaural group, the same word was presented in stereo to both ears. The Untrained group did not undergo any training. As shown in Fig. 2, categorical perception was also evaluated for French phonemes. The expected lack of impact of English training on French phoneme categorization was observed (pre-/post-test comparison), and this result will not be further discussed for the sake of brevity.

b. Training. The training program focused on the English voicing contrast in /k-g/ and /p-b/, and was developed in accordance with the principles of high variability phonetic training (HVPT) (e.g., Bradlow *et al.*, 1999; Grenon *et al.*, 2019; Saito *et al.*, 2022; Strange and Dittman, 1984). Based on these principles, the program incorporated the following elements: (1) various exercises, lexical and vocalic contexts, and speaker voices, (2) the presentation of minimal and near-minimal pairs, (3) trial-by-trial feedback, (4) simultaneous auditory and visual stimulation, (5) incentives aimed at increasing the participants' vocabulary, and (6) engaging content through the use of numerous words and appealing images. Furthermore, each training day focused on a single pair of consonants, with /k-g/ targeted on days 2 and 6 and /p-b/ on days 3 and 5. The two words of (near-)minimal pairs were processed within the same day, facilitating the learners' experience of phonetic features in L2 phonemes, as proposed by Tyler (2019). Each word was presented twice per session, and new words were introduced daily.

Three types of exercises were proposed from days 2–6 (Fig. 3). Every training day (except day 4) began with a lexical learning exercise: participants listened to an English word accompanied by its spelling, French translation, and a picture illustrating its meaning. Then, a spelling judgement task was proposed on these words: participants listened to a word and chose the correct spelling from two options on the screen, with a picture aiding their decision. On day 4, two judgement tasks were administered, one involving a forced-

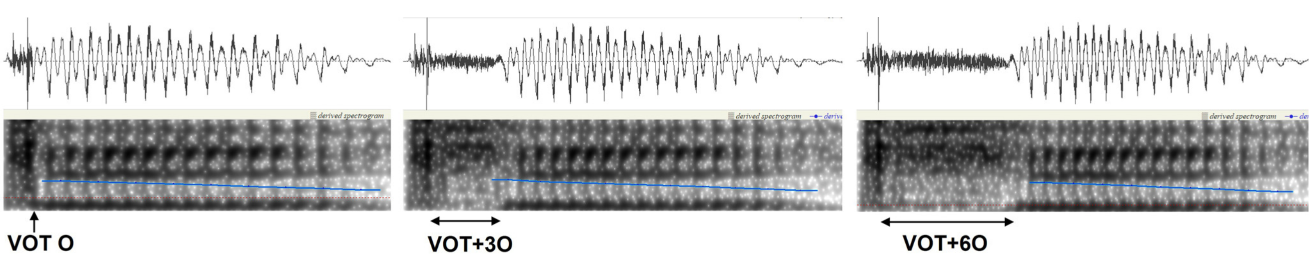

FIG. 1. Oscillograms and spectrograms of three examples of syllables along the English /di/-/ti/ continuum, and the duration of their VOT in ms, from typical /di/ (left), to typical /ti/ (right).

https://doi.org/10.1121/10.0036725

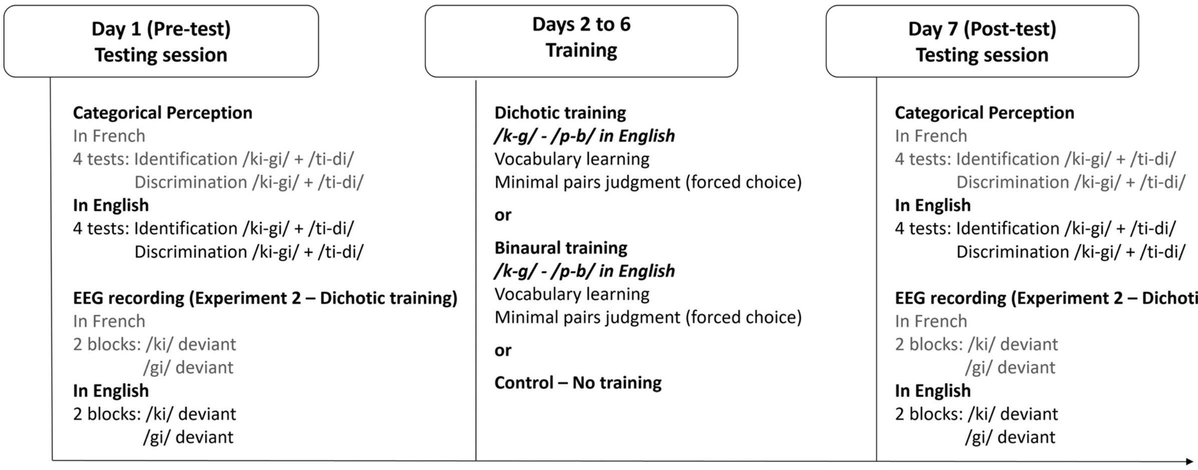

FIG. 2. General procedure of experiment 1 and experiment 2, with pretest-training-posttest design. Results for the French phonemes (in gray) are not presented in this article.

choice between two written words and the other between two images, both based on the words learned on days 2 and 3.

c. Testing sessions. In both pre- and post-tests, categorical perception was evaluated using a discrimination task followed by an identification task, as outlined by Fugate (2013). Both tasks commenced with stimuli from the /ki-gi/ continuum, followed by stimuli from the /ti-di/ continuum, in order to minimize potential contamination of the trained pair. Each task began with ten familiarization trials. No feedback was provided.

In the discrimination task, participants were required to determine whether the two stimuli in each trial, separated with an interstimulus interval (ISI) of 24 ms, were perceived as "same" or "different," as quickly as possible. Each block contained 130 trials (50 "different," 70 "same"). The task lasted 10 min.

In the identification task, participants listened to and categorized each of the 140 stimuli as /ki/ or /gi/ in the first block and as /ti/ or /di/ in the second block, by pressing one of two response keys. Each block included ten repetitions of each of the seven stimuli, presented in a randomized order. The task lasted 5 min.

4. Data analyses

Discrimination responses were converted into a d' sensitivity index following the procedure described in Macmillan and Creelman (2005). The category boundary was defined by the discrimination peak along the VOT continuum for each participant (e.g., VOT + 10 if it occurred at

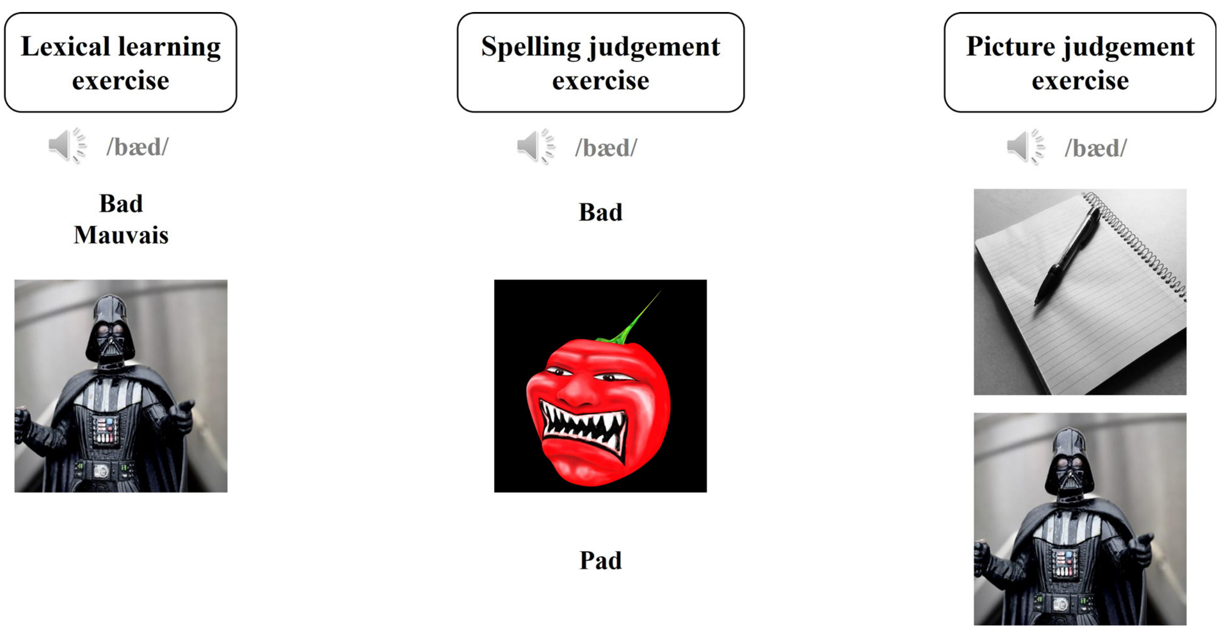

FIG. 3. Examples of stimuli in the perceptual training program. Reprinted with permission from PxHere (bottom left and right); Piotr Siedlecki via PublicDomainPictures.net (bottom middle); ShannonShort via Pixabay (top right).

the 0/+20 ms VOT pair). We chose to evaluate only the peak as it provides information on both the position of the boundary and its accuracy. The rest of the discrimination curve seemed less relevant if no comparison was made between observed and predicted perception, or if we are looking for an allophonic perception. Consistent with previous research (Theodore *et al.*, 2009), this boundary varied according to the place of articulation, and, on average, was observed for each group at VOT + 10 for /ki-gi/ and VOT + 20 for /ti-di/ (Figs. 3–5).

The identification scores for each participant were calculated as the mean of /ki/ or /ti/ responses for each stimulus. Categorical precision was quantified by the slope of the psychometric function using the Palamedes toolbox from MATLAB (Prins and Kingdom, 2018) and the PAL_PFML_Fit function. The latter adjusts the threshold (alpha) and slope (beta) parameters of a logistic function to the identification data. The VOT corresponding to 50% /ki/ or /ti/ responses was determined for each participant and regarded as the categorical boundary.

For statistical analysis, a significance threshold of $p \leq 0.05$ was applied and effect sizes were calculated using partial eta-squared (η_p^2) or Cohen's d (Cohen, 1988).

B. Results

1. Pre-training results

Three analyses of variance (ANOVAs) were conducted to assess initial group differences and establish baseline performance, with Pair (/ki-gi/, /ti-di/) as a within-subject factor and Group (untrained, dichotic, binaural) as a between-subjects factor. Analyses focused on the magnitude of the discrimination peaks (d'), identification slopes, and boundaries in the pre-test data.

The analysis revealed a significant main effect of Group on d'[$F(2, 114) = 3.96$; $p = 0.022$; $\eta_p^2 = 0.06$], indicating significantly higher sensitivity in the untrained group ($M = 1.89$; $SD = 1.00$) compared to the binaural group [$M = 1.29$; $SD = 1.06$); $t(77.72) = 2.61$; $p = 0.011$; $d = 0.58$] at pre-test. The untrained group and the dichotic group ($M = 1.68$; $SD = 0.80$) did not significantly differ [$t(74.28) = 1.14$; $p = 0.258$]. There was no significant difference between the two trained groups at baseline [$t(72.32) = 1.77$; $p = 0.080$]. There was no main effect of Pair ($F < 1$), nor Group × Pair interaction ($F < 1$).

The identification boundary was higher for /ti-di/ ($M_{VOT} = +16.99$ ms) than for /ki-gi/ ($M_{VOT} = +12.52$ ms), as indicated by a significant main effect of Pair [$F(1, 114) = 14.00$;

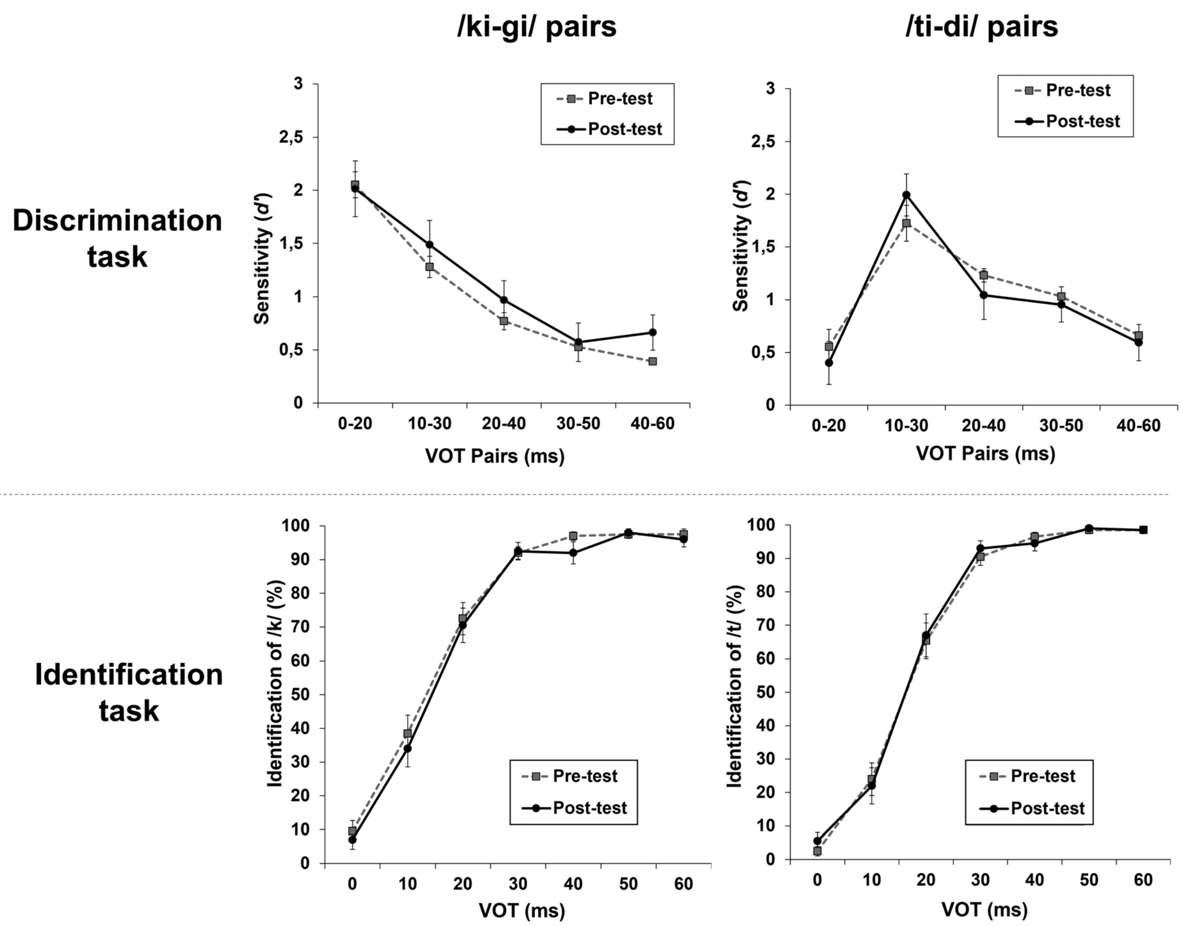

FIG. 4. Mean sensitivity to differences (discrimination task) for /ki-gi/ and /ti-di/ continua (top), and mean percentages of /k/ and /t/ identification (bottom), at pre- (dotted line) and post-test (solid line) in the untrained group. Error bars represent standard errors.

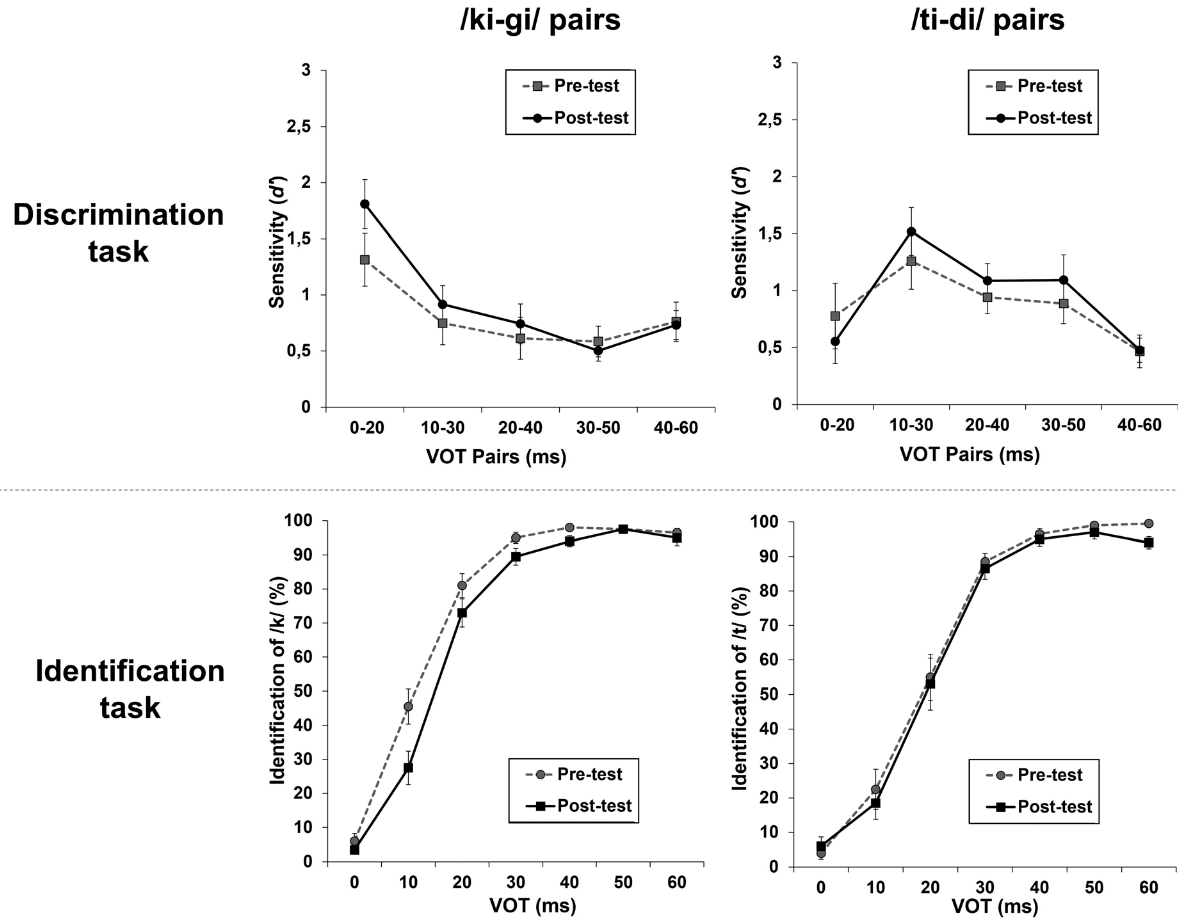

FIG. 5. Mean performance in the binaural group. Sensitivity to differences in the discrimination task (top) and percentages of identification (bottom) for the trained /ki-gi/ (left) and untrained /ti-di/ (right) syllables for the pre-test (dotted line) and post-test (solid line). Error bars represent standard errors.

$p < 0.001$; $\eta_p^2 = 0.11$]. This result was consistent with the location of the discrimination peaks.

Given the higher initial d' for discrimination observed in the untrained group, we refrained from conducting further ANOVAs including this group as a control.[1] There were no significant differences between the two trained groups at baseline for the discrimination peak magnitude, identification slope, or boundary values (all $p > 0.47$). However, the greater visual variability observed in the binaural group's performance compared to the dichotic group (Figs. 5 and 6), combined with distinct hypotheses for each training condition, led to the decision to perform separate statistical analyses for the binaural and dichotic groups, respectively.

2. Test-retest effects–untrained group

To exclude the possibility that pre-/post-test performance differences were due to test–retest effects, rather than training effects, we conducted analyses to assess the reliability of the categorical perception tests when tested at two timepoints. Three repeated-measures ANOVAs were performed with Session (pre-, post-test) and Pair (/ki-gi/, /ti-di/) as within-subject factors, examining the untrained group's performance on the magnitude of the discrimination peaks (d') (Fig. 4, top), identification slopes, and boundaries (Fig. 4, bottom).

No significant main effect of Session was observed for any measure [d': $F < 1$; slope: $F(1, 19) = 1.91$, $p = 0.182$; boundary: $F < 1$]. Additionally, no significant Session × Pair interaction was found [d': $F(1, 19) = 1.54$; $p = 0.229$; slope: $F(1, 19) = 1.96$; $p = 0.177$; boundary: $F < 1$].

The results confirm the reliability of the categorical perception tests, validating their use in assessing the impact of the training conditions.

3. Binaural training program: Pre- versus post-tests

To evaluate the training effect on the binaural group (pre-, post-test), we analyzed the magnitude of discrimination peaks (d') (Fig. 5, top), and the identification slopes and boundaries (Fig. 5, bottom) separately for the trained /ki-gi/ and untrained /ti-di/ syllables, using two-tailed paired-sample t-tests with Bonferroni corrections.

For the discrimination task, the analysis revealed a significant increase in d' following binaural training for the trained pair /ki-gi/ [$t(19) = 2.15$; $p = 0.045$; $d = 0.49$], but not for the untrained pair [/ti-di/, $t(19) = 1.27$; $p = 0.220$; $d = 0.25$]. In the identification task, no significant change was observed in the slope for either pair ($t_s < 1$), but the

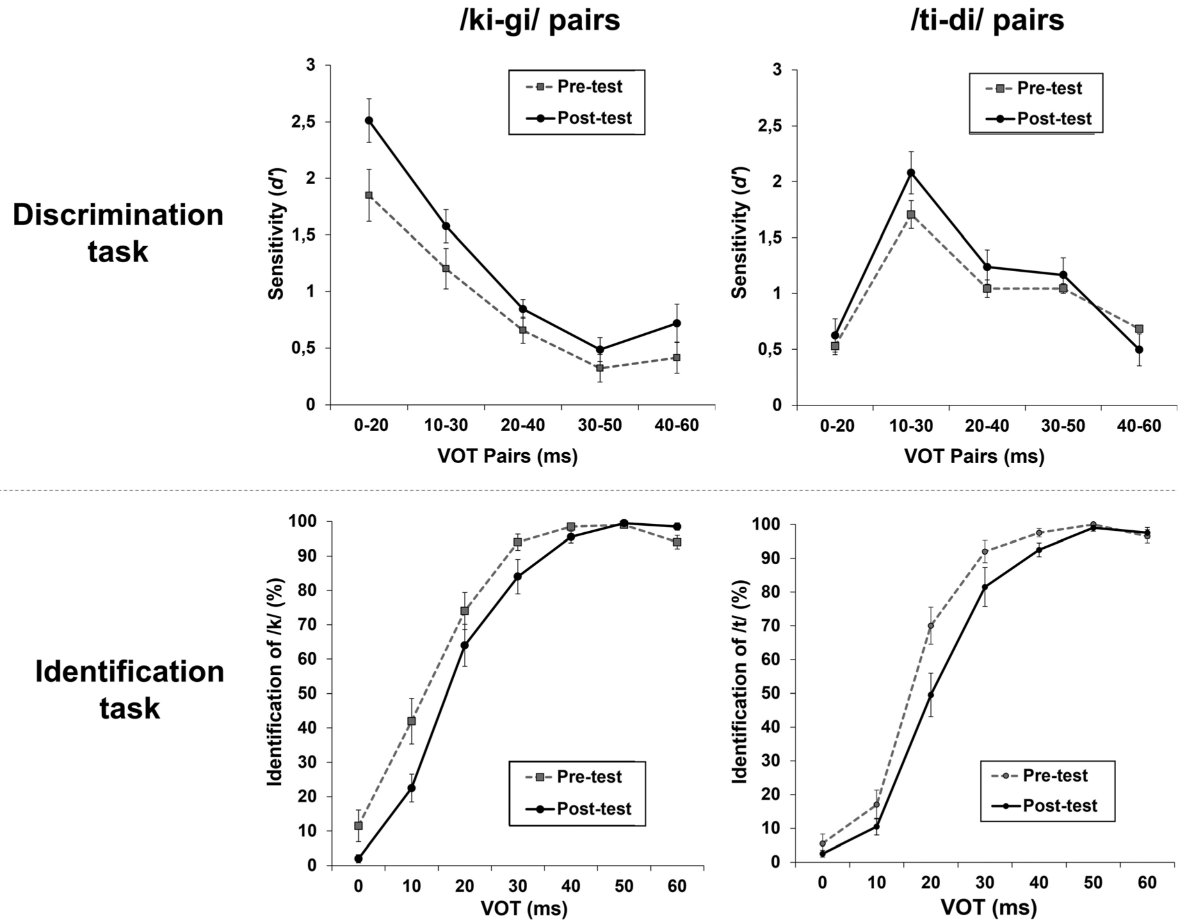

FIG. 6. Mean performance in the dichotic group. Sensitivity to differences in the discrimination task (top) and percentages of identification (bottom) for the trained /ki-gi/ (left) and untrained /ti-di/ (right) syllables for the pre-test (dotted line) and post-test (solid line). Error bars represent standard errors.

boundary for the trained pair shifted significantly to the right [$t(19) = 3.50$; $p = 0.002$; $d = 0.74$], primarily due to an increased perception of /g/ as voiced at VOT + 10 [$t(19) = 4.41$; $p < 0.001$; $d = 0.76$]. The training factor had no significant impact on the boundary for the untrained pair ($t < 1$).

These results align with previous research providing evidence for the effectiveness of HVPT-based training programs for enhancing perception of trained L2 phonemes; however, progress was restricted to the trained consonants and did not transfer to other phoneme pairs (as also shown in Maye and Gerken, 2001; Olson, 2021; Tamminen *et al.*, 2015).

4. Dichotic training program: Pre- versus post-tests

Following the demonstration that the categorical perception tasks did not show test–retest effects, and the observed benefits of the binaural training program for trained pairs, we next assessed whether dichotic stimulation led to improvements in categorical perception for both trained and untrained consonants. Training effects were analyzed in the dichotic group using two-tailed paired-sample *t*-tests (Bonferroni corrected), separately for trained and untrained pairs.

In the discrimination task, the sensitivity (d') significantly increased after dichotic training for the trained pair [$t(19) = 2.42$; $p = 0.026$; $d = 0.57$], and marginally for the untrained pair, with a moderate effect size [$t(19) = 2.06$; $p = 0.053$; $d = 0.47$] (Fig. 6, top).

In the identification task, there was no significant change in the slope for either the trained pair ($t < 1$) or the untrained pair [$t(19) = 1.53$; $p = 0.142$]. However, the boundary shifted significantly rightward for both the trained pair [$t(19) = 3.53$; $p = 0.002$; $d = 0.71$] and the untrained pair [$t(19) = 3.29$; $p = 0.004$; $d = 0.75$]. This shift toward longer positive VOTs for voiceless stop consonants was more consistent with English phonetic rules, suggesting a perceptual change. As illustrated in Fig. 6, these shifts were particularly driven by increased perception of /g/ as voiced at VOT0 [$t(19) = 2.17$; $p = 0.043$; $d = 0.63$] and VOT + 10 ms [$t(19) = 4.03$; $p < 0.001$; $d = 0.79$], as well as increased perception of /d/ as voiced at VOT + 20 [$t(19) = 5.30$; $p < 0.001$; $d = 0.76$]. These improvements occurred for VOTs of the English voiceless consonants, which were particularly ambiguous for French speakers if they used French phonetic rules to interpret them.

Taken together, these results suggest that both binaural and dichotic perceptual training enhance perception of trained English phonetic contrasts. However, dichotic training resulted in a transfer also to untrained material, suggesting generalization of learned rules. These transfer effects were evident in improved identification and discrimination of untrained consonants, with a medium effect size. These results highlight the benefit of perceptual training in general, as well as the additional benefit of dichotic presentation for enhancing perception of untrained material. Therefore, these findings support the AST model (Poeppel, 2003) and the featural approach to L2 learning (de Jong *et al.*, 2009; Olson, 2019), emphasizing the potential of dichotic training for learning and generalizing subphonemic rules.

III. EXPERIMENT 2: ELECTROPHYSIOLOGICAL EVALUATION OF DICHOTIC TRAINING

Experiment 1 yielded three key insights: (1) categorical perception tests are reliable, (2) binaural training based on HVPT principles improves categorical perception in late learners of L2, but the benefit did not transfer to untrained phonemes, and (3) dichotic training, designed to engage the LH or RH based on VOT durations, improves performance for trained phonemes and leads to transfer of skills to untrained phonemes. Based on these behavioral changes, it is particularly interesting to investigate how dichotic training may influence underlying neural responses.

While categorical perception tasks assess explicit processing of attended speech, they do not capture pre-attentional mechanisms or neural responses. The MMN component in EEG recordings is automatically elicited by perceptual deviants (Näätänen *et al.*, 1989), and can be used as a measure of pre-attentive perception. The MMN provides a reliable neural index of automatic change detection, independent of attention and task-related effects, such as motor preparation (Kujala *et al.*, 2007; Näätänen *et al.*, 2007). The amplitude of the MMN has been shown to increase when the deviants belong to the listener's phonological system (Sams *et al.*, 1985). Further, the MMN has been shown to differentiate native from non-native contrasts in both children (Shestakova *et al.*, 2003) and adults (Ylinen *et al.*, 2006), making it a reliable tool to measure changes in pre-attentive phonetic processing.

Experiment 2 employed the passive oddball paradigm, using /ki/ and /gi/ as standard and deviant stimuli (and *vice versa*), to test whether dichotic training induces a change in the neural response to phonetic features in L2. Following dichotic training, an increase in MMN amplitude was hypothesized, along with changes in brain asymmetry when processing short vs long VOTs, consistent with the LH and RH involvement predicted by the AST model for short- vs long-duration events. Only the dichotic training group was assessed, as the hypothesis specifically targeted hemispheric changes associated with this condition.

A. Method

1. Participants

Sixteen participants from the dichotic group in experiment 1 ($M_{\text{age}} = 21.60$; $SD = 2.90$; 14 females) agreed to participate in experiment 2.

2. Materials

The experiment presented two blocks of English CV syllables, using /ki/ and /gi/ stimuli selected from the /ki-gi/ continua in experiment 1 for their typical VOTs. The voiced consonant /g/ had a VOT + 10 and the voiceless /k/ had a VOT + 60.

3. Procedure

Participants underwent EEG recording during a passive listening task (oddball paradigm) at pre- and post-test. Testing occurred individually in a sound-proof room, with participants seated in a comfortable armchair in front of a screen. They selected a movie from a list of visually engaging, muted cartoons without subtitles. Sounds, normalized for intensity, were presented binaurally through Beyerdynamics DT 770 Pro 250# headphones. Participants were instructed to focus on the movie and disregard the sounds, to prevent attention-dependent event-related potential (ERP) components that might interfere with the MMN response (Näätänen *et al.*, 2007). Stimuli were administered using Presentation software (v.14.9, Neurobehavioral Systems, Albany, CA).

The experiment included 506 items divided into two blocks: one with /gi/ as the standard and the other with /ki/, presented in counterbalanced order across participants. Each block consisted of 85% standards and 15% deviants. Stimuli were pseudo-randomized, with at least four standards preceding each deviant. Standards following deviants were excluded from analyses. Each block lasted for 5 min, with a rest period between blocks, resulting in a total recording time of 10 min. This experiment followed the categorical perception tasks conducted at the pre- and post-test stages.

4. EEG data acquisition and preprocessing

EEG was recorded continuously from 32 scalp sites (Electro-Cap International, Eaton, OH, according to the international 10–20 system) using the Biosemi ActiveTwo EEG system (version 5.36, 02–06-2006, BioSemi B.V., Amsterdam, Netherland), with a sampling rate of 512 Hz and a 24 bit resolution. Seven active external electrodes were placed on the face: four around the eyes to monitor ocular movements (vertical and horizontal electro-oculograms), one on the nose, and two on the left and right mastoids. Individual electrode offsets were adjusted to remain stable below ± 20 mV.

Data were analyzed offline with the ERPLAB (Lopez–Calderon and Luck, 2014) module (v.1.0.0.42) within the EEGLAB (Delorme and Makeig, 2004) toolbox (v.9.0.2.3.b) for MATLAB (v.7.0.9.R2009b, The MathWorks, Natick, MA). The EEG signal was bandpass filtered between 0.1 and 30 Hz (12 dB/oct). Data were epoched from –200 to 800 ms relative to stimulus onset. Baseline was corrected by subtracting the mean voltage in the 200 ms prestimulus interval from each sample in the epoch. Automatic artifact rejection excluded epochs with a trace exceeding ± 100 μV.

For each condition, signals were averaged per participant and re-referenced to the average of the two mastoid channels. ERPs time-locked to the syllable were averaged offline for each participant at each electrode site across experimental conditions. MMN responses were extracted by subtracting the average standard waveform from the average deviant waveform for each participant in each experimental block.

B. Data analyses

The MMN was identified for each participant as the most negative peak occurring between 100 and 300 ms post-stimulus onset at electrode Fz (Kujala *et al.*, 2007), and MMN mean amplitude was calculated. Time windows were selected based on MMN latency over Fz for all participants, separately for each deviant type. For the /ki/ deviant, a 150–250 ms window was used, while for the /gi/ deviant, a 200–300 ms window was used. Visual inspection confirmed that the MMN for each participant peaked within these windows.

Electrodes F3, Fz, and F4 were grouped to represent the Frontal scalp region of interest; FC1 and FC2 for the frontocentral region; C3, Cz, and C4 for the central region; CP1 and CP2 for the centroparietal region; and P3, Pz, P4 for the parietal region. This electrode grouping is standard in MMN analyses, as it helps to reveal the front-back distribution of brain responses (Kujala *et al.*, 2007).

C. Results

1. Pre-training results

To assess the presence of significant MMNs prior to training, two-tailed *t*-tests (Bonferroni corrected) compared the mean amplitude within the MMN window for each of the two experimental blocks against a test value of zero.

In the pre-test, *t*-tests revealed that the mean MMN amplitude differed significantly from zero for /ki/ [$t(15) = 3.06$; $p = 0.008$; $d = 1.08$], but not for /gi/ [$t(15) = 1.11$; $p = 0.285$] (Fig. 7, upper panel). This result may be attributed to the long aspiration in /ki/, which is perceived as a more salient nonrelevant acoustic event than the brief silence following the burst in English voiced stop consonants (Flege, 1991). Thus, the pre-training MMN likely reflects automatic orientation toward basic physical changes (Sams *et al.*, 1985).

2. Dichotic training effects

A repeated-measures ANOVA assessed the impact of dichotic training on the MMN amplitude, with Training (pre-, post-test), Voicing (/ki/, /gi/), and Position (frontal, frontocentral, central, centroparietal, parietal) as within-participant factors. Significant interactions were investigated using two-tailed paired-sample *t*-tests (Bonferroni correction). A significance level of $p \leq 0.05$ was used, and effect sizes were evaluated.

The ANOVA showed no significant MMN amplitude difference between frontal and frontocentral electrodes [$t(15) = 1.20$; $p = 0.249$]. However, MMN amplitude was higher at frontocentral electrodes compared to more posterior sites (Table I), as indicated by a significant main effect of Position [$F(4, 60) = 6.41$; $p < 0.001$; $\eta_p^2 = 0.30$]. This finding replicated the literature and confirmed the presence

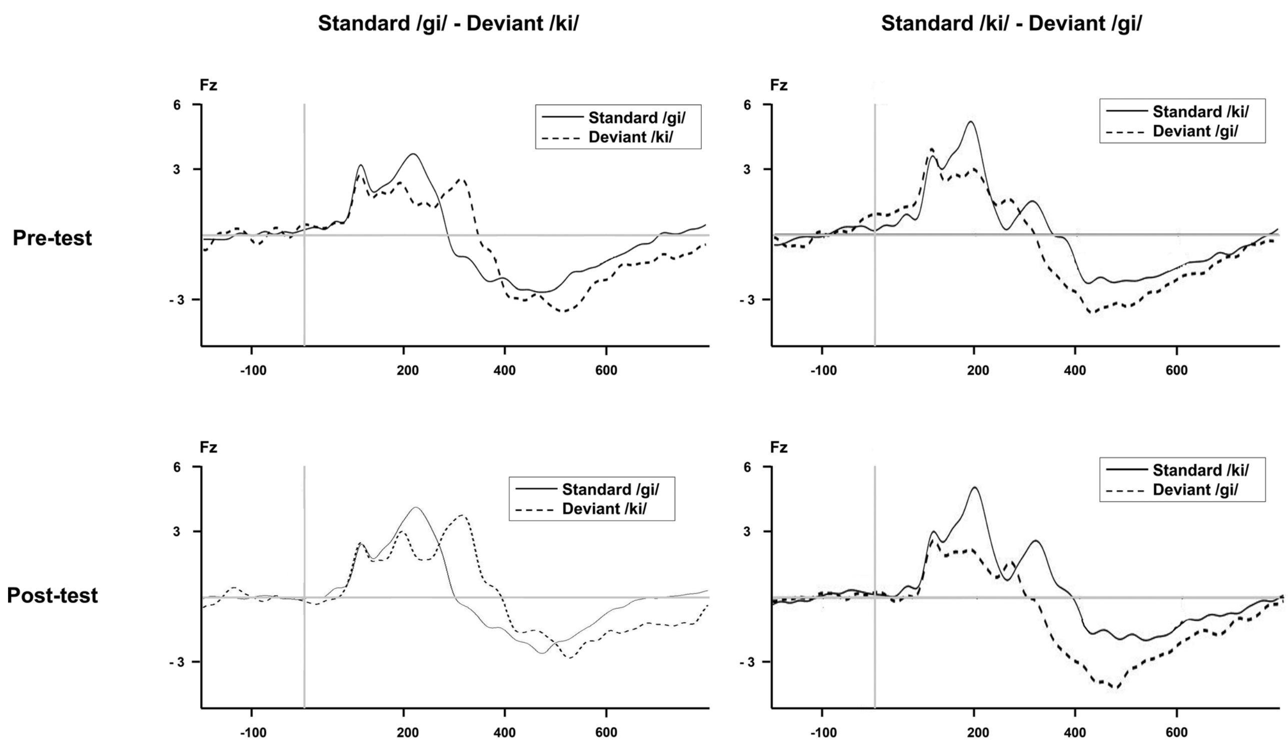

FIG. 7. Grand averages (μV) of the ERPs elicited by the standard (solid line) and deviant (dotted line) stimuli at Fz, presented before (upper panel) and after (lower panel) dichotic training, with /ki/ (left panel) or /gi/ (right panel) as the deviant.

TABLE I. Results of two-tailed *t*-tests comparing MMN amplitudes between frontocentral electrodes and more posterior sites.

Electrodes positions (mean MMN amplitude)	*t* test (df = 15)	*p*	Cohen's *d*
Frontocentral (−1.30 μV)			
Central (−1.14 μV)	−2.68	0.017	0.03
Centroparietal (−1.07 μV)	−2.25	0.040	0.14
Parietal (−0.82 μV)	−4.30	<0.001	0.32

of an MMN. There was no significant main effect of Training or Voicing (all $ps > 0.201$).

The Training × Voicing interaction [$F(1, 15) = 6.23$; $p = 0.025$;, $\eta_p^2 = 0.29$], revealed that the MMN increased after dichotic training for the English /gi/ deviant (pre-training: $M = -0.42$, $SD = 1.50$; post-training: $M = -1.53$; $SD = 1.32$) [$t(15) = 2.40$; $p = 0.030$; $d = 0.26$], but no change occurred for the /ki/ deviant. All other interactions were non-significant (all $p > 0.579$).

Additional analyses examined the expected changes in hemispheric asymmetry for VOT duration processing post-training. Following training, the involvement of the LH was hypothesized to increase for /gi/, while the involvement of the RH was anticipated to increase for /ki/.

As shown in Fig. 8, during pre-test, MMN amplitude for /ki/ was significantly higher at left than right anterior electrodes in both frontal and frontocentral areas, respectively [$t(15) = 2.34$; $p = 0.034$; $d = 0.85$] and frontocentral [$t(15) = 2.52$; $p = 0.024$; $d = 0.78$]. This asymmetry disappeared in the post-test: there was no difference between the left and right electrodes in frontal: $t(15) = -0.16$; $p = 0.873$; or frontocentral: $t(15) = -0.23$; $p = 0.822$ areas.

In summary, following dichotic training, the MMN increased for the English voiced deviant /gi/, indicating that its short positive VOT was processed as a phonetic feature characterizing a phonological category, which was distinct from the category associated with the standard /ki/. This result supports the expected change in the relationship

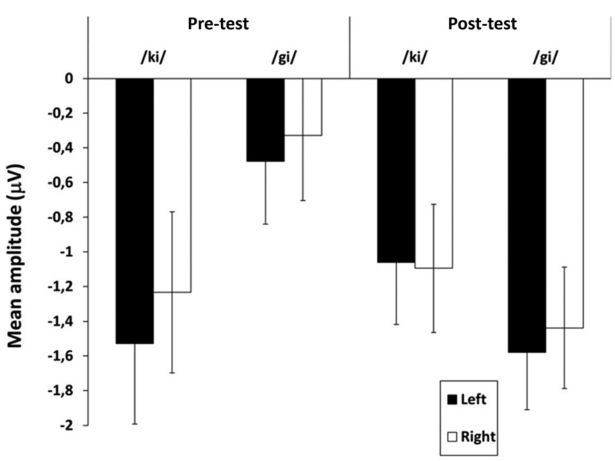

FIG. 8. Mean MMN amplitude (μV) for English stimuli before (left panel) and after (right panel) dichotic training, for the left (black) and right electrodes (white). Error bars represent standard errors.

between phonetic features and phonological categories, aligning with the "two-category assimilation" process (PAM-L2) or a "similar scenario" (L2LP). Previous research has linked larger MMNs to changes in the perception of abstract features (Tervaniemi *et al.*, 1994) and increased linguistic expertise (Näätänen, 2001). The increased MMN for the /gi/ deviant post-training may reflect enhanced English-like pre-attentive categorization, including at the neural level (Kujala *et al.*, 2007; Näätänen *et al.*, 2007).

Additionally, the post-training decrease in LH involvement for deviants with long VOTs (i.e., /ki/) suggests a change in appropriate cerebral asymmetry for event durations, as proposed by the AST model (Poeppel, 2003; Poeppel *et al.*, 2008), extending the model to sub-phonemic features.

IV. GENERAL DISCUSSION

The current study investigated whether a brief perceptual training program could facilitate the acquisition of the voicing contrast for English consonants among adult French speakers. It was hypothesized that dichotic presentation, stimulating the LH with voiced stop consonants and the RH with voiceless stop consonants, would enhance learning compared to training with binaural presentation. This hypothesis was grounded in the AST model suggesting complementary hemispheric dominance for short and long events. The hypothesis was tested using categorical perception tasks and pre-attentional EEG measures. Results indicated that both binaural and dichotic training improved behavioral categorical perception for trained phonemes, and that this benefit extended to untrained phonemes only with dichotic training, indicating generalization of sub-phonemic rules. Additionally, after dichotic training, an MMN response emerged for English voiced stop consonants, with a change in hemispheric lateralization aligning with the long VOT of these English consonants. These findings support the hypothesis that phonetic perception can be enhanced in late learners with both binaural and dichotic stimulation. Further, dichotic stimulation training can support the processing of untrained phonemes, improving featural-level associations with phonemes, neural sensitivity to the English voicing contrast, and hemispheric specialization for short and long VOTs.

A. Enhancing English voicing perception using a brief training program

Prior to training, all participant groups exhibited categorical perception of voiced vs voiceless stop consonants in English, as evidenced by the shape of their identification and discrimination curves, but the categorical boundary was more appropriate to English following perceptual training than at this initial step. This result is consistent with production data demonstrating that adult Spanish-speaking learners of English produce English voiceless stop consonants with intermediate VOT values that are longer than Spanish voiceless stop consonants, but shorter than those produced by native English speakers (Flege, 1991). This result is relevant, as Spanish voiced consonants are associated with short positive VOTs, like in French. Relatedly, in the present study, post-training results indicated a perceptual shift toward more English-like VOT perception. Specifically, an increase in the peak of discrimination suggested improved boundary precision (Medina *et al.*, 2010). Moreover, the identification boundary shifted from VOT + 12 to VOT + 19, aligning more closely with the phonemic boundary of VOT + 20 captured by the discrimination task, signaling that perception was based on sub-phonemic rules more in line with English phonology.

Additionally, the VOT boundary shifted from a French-like location (approximately VOT0) (Caramazza and Yeni-Komshian, 1974) to a more English-like one (around VOT + 30) (Lisker *et al.*, 1977). Despite the English voiced [g] containing a positive VOT of 10 or 20 ms that is prone, together with the voiceless [k^h], to "single-category assimilation" to /k/, it was better identified as /g/ following the two trainings, suggesting a shift toward what would be considered a "two-category assimilation" process in PAM-L2 (Best and Tyler, 2007) or a "similar scenario" in L2LP (Escudero, 2009; van Leussen and Escudero, 2015). In summary, the training program (with both binaural and dichotic presentation) grounded in HVPT principles (Grenon *et al.*, 2019; Saito *et al.*, 2022; Strange and Dittman, 1984) successfully enhanced adult learners' perception of phonetic features and their phonological roles for trained English phonemes, without explicit instruction. These results support existing research suggesting that adult L2 learners can improve their phonological processing of temporal features (Cebrian and Carlet, 2014; Menning *et al.*, 2002; Olson, 2019; Tamminen *et al.*, 2015), as predicted by the SLM-r and PAM-L2 models.

B. Enhancing attentional categorical perception through dichotic stimulation

Dichotic stimulation proved more effective than binaural stimulation in enhancing the phonetic perception of French speakers learning English, resulting in improved perception of both trained and untrained phonemes. This finding suggests that dichotic training, which engages asymmetric hemispheric recruitment for processing short vs long events, facilitates the formation of new sub-phonemic associations between VOT and phonological voicing values for English stop consonants. The current finding aligns with previous research showing an advantage of this same dichotic training program applied to vowel length difference learning (Bouhon *et al.*, 2023). This previous training was also grounded in the AST model, and successfully improved the perception of the non-native duration differences between English /ɪ/ and /iː/ for French learners, with skill transfer observed from perception to production (Bouhon *et al.*, 2023).

Another tentative explanation for the generalization of dichotic training to untrained contrasts is that it facilitates learning by distributing processing resources adequately across both hemispheres. Alternatively, one might argue that the novel dichotic presentation may have heightened learners' curiosity and attention, promoting deeper processing of stimuli. To further investigate the role of hemispheric asymmetry in processing phonetic feature durations, as suggested by the AST model, future studies could implement a reversed dichotic condition, with voiceless English stop consonants presented to the right ear and voiced stop consonants to the left. Based on the current findings and hypotheses, performance in this condition would likely be lower than in the two trained groups of the current study. This reversed condition was avoided in the present study to prevent disrupting brain processing and potentially hindering the appropriate engagement of the LH and RH in processing short and long acoustic events, impacting L2 speech processing.

While improvement after binaural training could reflect learning of individual phonemes in line with segmental models (SLM, PAM), dichotic training may work at the sub-phonemic level, aligning with the featural approach to L2 learning (de Jong *et al.*, 2009). This perspective is neglected in favor of segmental models in educational contexts, but it is supported by the current results, encouraging development of sub-phonemic learning tools. Combined, these results support the use of dichotic stimulation for training English perception skills in adult French learners.

C. Extension of the AST to phonetic features

The benefits of dichotic training observed in this study align with the predictions of the AST model. As outlined above, the hemispheric asymmetry proposed by this model has not previously been applied to phonetic features. AST links rapid gamma oscillations in the left non-primary auditory cortex with phoneme processing, while slow theta oscillations in the right non-primary auditory cortex are associated with syllable processing (Boemio *et al.*, 2005; Poeppel, 2003). The VOTs analyzed in this study are at the sub-phonemic level, suggesting that the AST model can be extended to encompass phonetic features.

Gamma oscillations, with frequencies above 30.5 Hz, are sufficiently rapid to process VOTs in English voiced stop consonants, which last under 30 ms (Cho and Ladefoged, 1999). In the current study, dichotic training effectiveness in learning a generalizable association between short phonetic features (VOT) and voiced English consonants may be facilitated by the dominance of gamma oscillations in the LH.

Furthermore, dichotic training may enhance the processing of longer VOTs in English voiceless stop consonants by engaging the RH, associated with slower oscillations. The AST model suggests that theta and alpha oscillations in the RH process longer acoustic events, such as syllables and prosody within the 4–10 Hz frequency range (Giraud and Poeppel, 2012). These oscillations may aid in processing long VOTs (100–250 ms).

https://doi.org/10.1121/10.0036725

However, VOTs in English voiceless stop consonants can range from 30–150 ms (Docherty, 1992; Kupske, 2017), potentially corresponding to frequencies in the 10–33 Hz range, overlapping with beta oscillations (13–30 Hz) for VOTs between 30 and 76 ms and part of the alpha band (10–13 Hz) for VOTs between 76 and 100 ms. Although evidence for alpha and beta oscillations in the RH related to sound durations is limited, studies have shown that theta and alpha activity is enhanced in the RH when non-verbal sounds are presented to the left ear (Jiwani *et al.*, 2021), suggesting the availability of these oscillations in the RH.

Beta oscillations have also been observed to dominate in response to 30–500 ms acoustic transients (in the right associative auditory cortex) (Giroud *et al.*, 2020). Additionally, higher magnetoencephalographic responses to amplitude modulations at 2, 4, but also 10 Hz (alpha) have been noted in the RH compared to the LH in typical-level adult readers, but not in dyslexic adults (Hämäläinen *et al.*, 2012), suggesting that the RH can support long VOT processing. Finally, auditory evoked cortical magnetic field recordings indicate that, in the auditory cortex, the LH processes periodic noise repetition rates within a 25 ms temporal window, while the RH processes them within a 50–200 ms window. This window in the RH may be well-suited to processing most of the VOTs in English voiceless stop consonants (Keceli *et al.*, 2015). These findings, combined with the current results, collectively support the proposition to extend the AST model to phonetic features, particularly VOTs.

D. Enhancing pre-attentional phonological processing in L2 *via* dichotic training

The finding that dichotic training enhanced the MMN amplitude in response to the English voiced deviant /gi/ aligns with the behavioral data from experiment 1, providing evidence for improved perception of short positive VOTs as cues to voiced stop consonants at both behavioral and neural levels. Prior to training, the MMN also revealed greater LH involvement as compared to the RH in the processing of the English voiceless stop consonant /k/ by French-speaking English learners, contradicting predictions from the AST. This finding suggests an overreliance on abstract phonological representation in the L1, associating voiceless consonants with LH processing due to short VOTs in French. This cerebral misalignment with long acoustic cues is consistent with magnetoencephalography findings showing higher LH phase locking to long noises in dyslexic adults (Hämäläinen *et al.*, 2012). The unexpected LH dominance in long VOT processing indicates that French learners of English require support in optimizing brain asymmetry for temporal processing in English (Bouhon *et al.*, 2023). Dichotic training in the current study effectively facilitated this asymmetric hemispheric processing of the English voicing contrast.

However, it should be noted that the current study measured the MMN only in the dichotic training group, and the EEG results should be considered preliminary, as the results cannot yet be directly interpreted as linked to dichotic stimulation. Future research should examine the effects of binaural training on the MMN to compare its effectiveness with that of dichotic training, including comparisons with a passive control group. The comparisons with a control group showing similar baseline performance to the trained groups are also required to confirm the benefits of the training. Nevertheless, two neural changes were observed following dichotic training: enhanced preattentional sensitivity to English /gi/ (short VOT) and reduced LH involvement in processing English /ki/ (long VOT). These findings suggest changes at the encoding featural level in L2 learning, consistent with de Jong *et al.* (2009) and the current behavioral results.

V. CONCLUSION

This study provides new evidence supporting the potential for improvements in late L2 learning through brief perceptual training, and advocates for the use of neurologically inspired methodologies, such as dichotic stimulation, to facilitate this training. The findings suggest that adult French-speaking learners of English can enhance their phonological perception of phonetic features in L2, refine representations of L2 phoneme boundaries, and strengthen appropriate phonetic-phonological associations, thereby increasing pre-attentional sensitivity to L2 features. These results underscore the value of targeted training programs that focus on sub-phonemic contrasts without explicit instructions. However, further investigation of potential long-term effects is necessary. Considering the crucial role of phonetic accuracy in lexical identification, and previous research on vowel perception (Bouhon *et al.*, 2023), these findings encourage further exploration of dichotic stimulation as a tool for enhancing L2 learning.

ACKNOWLEDGMENTS

This study was funded by a doctoral grant to M.B. supported by Contentsquare–Contentsquare Foundation.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Ethics Approval

The experiment was approved by the French Ethics Committee (Comité de Protection des Personnes, n°21-AMBU-01) and participants gave informed consent after the procedures had been explained to them.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

[1]Two participants across the full sample exhibited performance exceeding 2 *SD* from the mean; however, removing them did not change the pattern of results, so they were kept in the analysis.

Bedoin, N., Ferragne, E., and Marsico, E. (**2010**). "Hemispheric asymmetries depend on the phonetic feature: A dichotic study of place of articulation and voicing in French stops," Brain Lang. **115**, 133–140.
Benki, J. R. (**2001**). "Place of articulation and first formant transition pattern both affect perception of voicing in English," J. Phon. **29**, 1–22.
Best, C. T., and Tyler, M. D. (**2007**). "Nonnative and second-language speech perception: Commonalities and complementarities," in *Second Language Speech Learning: The Role of Language Experience in Speech Perception and Production*, edited by M. J. Munro and O.-S. Bohn (John Benjamins, Amsterdam, the Netherlands), pp. 13–34.
Boemio, A., Fromm, S., Braun, A., and Poeppel, D. (**2005**). "Hierarchical and asymmetric temporal sensitivity in human auditory cortices," Nat. Neurosci. **8**, 389–395.
Bouhon, M., Ferreira, C., Bahuon, S., Tillmann, B., and Bedoin, N. (**2023**). "Improving non-native duration contrast with dichotic training in dyslexic and non-dyslexic individuals," Dyslexia **29**, 151–158.
Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., and Tohkura, Y. (**1999**). "Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production," Percept. Psychophys. **61**, 977–985.
Caramazza, A., and Yeni-Komshian, G. (**1974**). "Voice onset time in two French dialects," J. Phon. **2**, 239–245.
Cebrian, J., and Carlet, A. (**2014**). "Second-language learners' identification of target-language phonemes: A short-term phonetic training study," Can. Mod. Lang. Rev. **70**, 474–499.
Cho, T., and Ladefoged, P. (**1999**). "Variation and universals in VOT: Evidence from 18 languages," J. Phon. **27**, 207–229.
Cohen, J. (**1988**). *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. (Lawrence Erlbaum Associates, Hillside, NJ).
de Jong, K. J., Silbert, N. H., and Park, H. (**2009**). "Generalization across segments in second language consonant identification," Lang. Learn. **59**, 1–31.
DeKeyser, R., and Larson-Hall, J. (**2005**). "What does the critical period really mean," in *Handbook of Bilingualism: Psycholinguistic Approaches*, edited by J. F. Kroll and M. B. de Groot Annette (Oxford University Press, New York), pp. 88–108.
Delorme, A., and Makeig, S. (**2004**). "EEGLAB: an open-source toolbox for analysis of single-trial EEG dynamics," J. Neurosci. Methods **134**, 9–21.
Docherty, G. J. (**1992**). "*The Timing of Voicing in British English Obstruents*," (Foris, New York).
Eggermont, J. J. (**2001**). "Between sound and perception: Reviewing the search for a neural code," Hear. Res. **157**, 1–42.
Ellis, N. C. (**2002**). "Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition," Stud. Second Lang. Acquis. **24**, 143–188.
Escudero, P. (**2009**). "Linguistic perception of 'similar' L2 sounds," in *Phonology and Perception*, edited by P. Boersma and S. Hamann (Mouton de Gruyter, Berlin), pp. 151–190.
Flege, J. E. (**1991**). "Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language," J. Acoust. Soc. Am. **89**, 395–411.
Flege, J. E., and Bohn, O.-S. (**2021**). "The revised speech learning model (SLM-r)," in *Second Language Speech Learning: Theoretical and Empirical Progress)*, edited by R. Wayland (Cambridge University Press, Cambridge, UK), pp. 3–83.
Fugate, J. M. (**2013**). "Categorical perception for emotional faces," Emot. Rev. **5**, 84–89.
Giraud, A.-L., and Poeppel, D. (**2012**). "Cortical oscillations and speech processing: Emerging computational principles and operations," Nat. Neurosci. **15**, 511–517.
Giroud, J., Trébuchon, A., Schön, D., Marquis, P., Liegeois-Chauvel, C., Poeppel, D., and Moillon, B. (**2020**). "Asymmetric sampling in human auditory cortex reveals spectral processing hierarchy," PLoS Biol. **18**, e3000207.
Grenon, I., Kubota, M., and Sheppard, C. (**2019**). "The creation of a new vowel category by adult learners after adaptive phonetic training," J. Phon. **72**, 17–34.
Grimaldi, M., Sisinni, B., Gili Fivela, B., Invitto, S., Resta, D., Alku, P., and Brattico, E. (**2014**). "Assimilation of L2 vowels to L1 phonemes governs L2 learning in adulthood: A behavioral and ERP study," Front. Hum. Neurosci. **8**, 279.
Hämäläinen, J. A., Rupp, A., Soltész, F., Szücs, D., and Goswami, U. (**2012**). "Reduced phase locking to slow amplitude modulation in adults with dyslexia: An MEG study," Neuroimage **59**, 2952–2961.
Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.
Hugdahl, K., and Wester, K. (**1992**). "Dichotic listening studies of hemispheric asymmetry in brain damaged patients," Int. J. Neurosci. **63**, 17–29.
Jiwani, S., Doesburg, S. M., Papsin, B. C., and Gordon, K. A. (**2021**). "Effects of long-term unilateral cochlear implant use on large-scale network synchronization in adolescents," Hear. Res. **409**, 108308.
Kang, X., Matthews, S., Yip, V., and Wong, P. C. (**2021**). "Language and nonlanguage factors in foreign language learning: Evidence for the learning condition hypothesis," NPJ Sci. Learn. **6**, 28.
Keceli, S., Okamoto, H., and Kakigi, R. (**2015**). "Hierarchical neural encoding of temporal regularity in the human auditory cortex," Brain Topogr. **28**, 459–470.
Kimura, D. (**1967**). "Functional asymmetry of the brain in dichotic listening," Cortex **3**, 163–168.
Kujala, T., Tervaniemi, M., and Schröger, E. (**2007**). "The mismatch negativity in cognitive and clinical neuroscience: Theoretical and methodological considerations," Biol. Psychol. **74**, 1–19.
Kupske, F. (**2017**). "A complex approach on integrated late bilinguals' English VOT production: A study on south Brazilian immigrants in London," Ilha Desterro **70**, 81–93.
Lisker, L., Liberman, A. M., Erickson, D. M., and Dechovitz, D. (**1977**). "On pushing the voice-onset-time (Vot) boundary about," Lang. Speech **20**, 209–216.
Lopez-Calderon, J., and Luck, S. J. (**2014**). "ERPLAB: An open-source toolbox for the analysis of event-related potentials," Front. Hum. Neurosci. **8**, 213.
Luk, G., Pliatsikas, C., and Rossi, E. (**2020**). "Brain changes associated with language development and learning: A primer on methodology and applications," System **89**, 102209.
Macmillan, N. A., and Creelman, C. D. (**2005**). *Detection Theory: A User's Guide*, 2nd ed. (Lawrence Erlbaum Associates, New York).
Maye, J., and Gerken, L. (**2001**). "Learning phonemes: How far can the input take us?," in *Proceedings of the 25th Annual Boston University Conference on Language Development, 2001*, Somerville, MA, pp. 480–490.
Medina, V., Hoonhorst, I., Bogliotti, C., and Serniclaes, W. (**2010**). "Development of voicing perception in French: Comparing adults, adolescents, and children," J. Phon. **38**, 493–503.
Menning, H., Imaizumi, S., Zwitserlood, P., and Pantev, C. (**2002**). "Plasticity of the human auditory cortex induced by discrimination learning of non-native, mora-timed contrasts of the Japanese language," Learn. Mem. **9**, 253–267.
Näätänen, R. (**2001**). "The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm)," Psychophysiology **38**, 1–21.
Näätänen, R., Paavilainen, P., Alho, K., Reinikainen, K., and Sams, M. (**1989**). "Do event-related potentials reveal the mechanism of the auditory sensory memory in the human brain?," Neurosci. Lett. **98**, 217–221.
Näätänen, R., Paavilainen, P., Rinne, T., and Alho, K. (**2007**). "The mismatch negativity (MMN) in basic research of central auditory processing: A review," Clin. Neurophysiol. **118**, 2544–2590.
Oderbolz, C., Poeppel, D., and Meyer, M. (**2025**). "Asymmetric sampling in time: Evidence and perspectives," Neurosci. Biobehav. Rev. **171**, 106082.
Olson, D. J. (**2019**). "Feature acquisition in second language phonetic development: Evidence from phonetic training," Lang. Learn. **69**, 366–404.
Olson, D. J. (**2021**). "Phonetic feature size in second language acquisition: Examining VOT in voiceless and voiced stops," Sec. Lang. Res. **38**, 913–940.
Poeppel, D. (**2003**). "The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time,' " Speech Commun. **41**, 245–255.
Poeppel, D., Idsardi, W. J., and van Wassenhove, V. (**2008**). "Speech perception at the interface of neurobiology and linguistics," Philos. Trans. R. Soc. B **363**, 1071–1086.
Prins, N., and Kingdom, F. A. A. (**2018**). "Applying the model-comparison approach to test specific research hypotheses in psychophysical research using the Palamedes Toolbox," Front. Psycho. **9**, 1–14.

Rimol, L. M., Eichele, T., and Hugdahl, K. (**2006**). "The effect of voice-onset-time on dichotic listening with consonant–vowel syllables," Neuropsychologia **44**, 191–196.
Saito, K., Hanzawa, K., Petrova, K., Kachlicka, M., Suzukida, Y., and Tierney, A. (**2022**). "Incidental and multimodal high variability phonetic training: Potential, limits, and future directions," Lang. Learn. **72**, 1049–1091.
Sams, M., Paavilainen, P., Alho, K., and Näätänen, R. (**1985**). "Auditory frequency discrimination and event-related potentials," Electroencephalogr. Clin. Neurophysiol. **62**, 437–448.
Shestakova, A., Huotilainen, M., Čeponien, R., and Cheour, M. (**2003**). "Event-related potentials associated with second language learning in children," Clin. Neurophysiol. **114**, 1507–1512.
Stefanatos, G. A., Joe, W. Q., Aguirre, G. K., Detre, J. A., and Wetmore, G. (**2008**). "Activation of human auditory cortex during speech perception: Effects of monaural, binaural, and dichotic presentation," Neuropsychologia **46**, 301–315.
Strange, W., and Dittmann, S. (**1984**). "Effects of discrimination training on the perception of /rl/ by Japanese adults learning English," Percept. Psychophys. **36**, 131–145.
Studdert-Kennedy, M., and Schankweiler, D. (**1970**). "Hemispheric specialization for speech perception," J. Acoust. Soc. Am. **48**, 579–594.
Tamminen, H., Peltola, M. S., Kujala, T., and Näätänen, R. (**2015**). "Phonetic training and non-native speech perception–new memory traces evolve in just three days as indexed by the misMatch negativity (MMN) and behavioural measures," Int. J. Psychophysiol. **97**, 23–29.
Teng, X., Tian, X., and Poeppel, D. (**2016**). "Testing multi-scale processing in the auditory system," Sci. Rep. **6**, 34390.
Tervaniemi, M., Maury, S., and Näätänen, R. (**1994**). "Neural representations of abstract stimulus features in the human brain as reflected by the mismatch negativity," NeuroReport **5**, 844–846.
Theodore, R. M., Miller, J. L., and DeSteno, D. (**2009**). "Individual talker differences in voice-onset-time: Contextual influences," J. Acoust. Soc. Am. **125**, 3974–3982.
Tyler, M. D. (**2019**). "PAM-L2 and phonological category acquisition in the foreign language classroom," in *A Sound Approach to Language Matters*, edited by A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen, and M. H. Sørensen (Aarhus University), pp. 607–630.
van Leussen, J. W., and Escudero, P. (**2015**). "Learning to perceive and recognize a second language: The L2LP model revised," Front. Psychol. **6**, 1000.
Ylinen, S., Shestakova, A., Huotilainen, M., Alku, P., and Näätänen, R. (**2006**). "Mismatch negativity (MMN) elicited by changes in phoneme length: A cross-linguistic study," Brain Res. **1072**, 175–185.