

## Research Article

**Cite this article:** Zhang, K. and Peng, G. (2025). Unlocking the barriers to speech normalization in L2: An EEG study on Mandarin L2 learners of Cantonese. *Bilingualism: Language and Cognition* 1–15. <https://doi.org/10.1017/S1366728925100369>

Received: 18 August 2024

Revised: 14 June 2025

Accepted: 27 June 2025


### Keywords:

speech variability; context cues; extrinsic normalization; ERP; L2 immersion

### Corresponding author:

Gang Peng;

Email: [gang.peng@polyu.edu.hk](mailto:gang.peng@polyu.edu.hk)

 This research article was awarded Open Data badge for transparent practices. See the Data Availability Statement for details.

# Unlocking the barriers to speech normalization in L2: An EEG study on Mandarin L2 learners of Cantonese

Kaile Zhang and Gang Peng 

Research Centre for Language, Cognition, and Neuroscience, Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong, China

## Abstract

Understanding high-variability speech is particularly challenging for second-language (L2) learners due to difficulties with extrinsic normalization, a perceptual strategy utilizing contextual cues to overcome speech variability. This study investigates the neural correlates of these difficulties among Mandarin speakers learning Cantonese, using EEG. Behaviorally, Mandarin learners demonstrated a significant yet considerably reduced ability to normalize Cantonese tone variability with contexts compared to native Cantonese speakers. EEG analysis showed that while native speakers engage multiple neural components (N1, P2, and LPC) for acoustic, phonetic/phonological, and cognitive adjustments in extrinsic normalization, Mandarin learners only activated P2, focusing on phonetic/phonological adjustments. This discrepancy underscores the multi-faceted nature of successful extrinsic normalization, which L2 learners fail to fully engage. L2 immersion significantly improves extrinsic normalization, particularly at the cognitive-adjustment stage. Overall, this study illuminates the intricate nature of poor extrinsic normalization in L2 learners and the importance of L2 immersion for effective L2 speech perception.

## 1. Introduction

Speech is fundamental to human communication, yet its inherent variability – arising from anatomical differences, speaking styles, and psychological factors – poses challenges for all language users. These challenges are amplified for L2 learners who must navigate not only new vocabulary and grammar but also the variability inherent in spoken communication. Achieving native-like speech perception requires strategies to adapt to this variability, one of which is extrinsic normalization. This process involves using contextual cues to minimize speech variability (Ainsworth, 1975; Johnson, 2005; Nearey, 1989). While the challenges of perceiving high-variability nonnative speech are well-documented (e.g., Antoniou et al., 2015; Tamati & Pisoni, 2014), the neural mechanisms that underpin L2 learners' utilization of contextual cues to overcome speech variability remain largely unexplored. This study aims to fill this gap by examining the strategies Mandarin learners of Cantonese employ for speech normalization, using both behavioral and EEG methods. The focus of the present study is the perception of Cantonese level tones, which are particularly susceptible to speech variability. The remainder of this introduction will first review relevant studies on extrinsic normalization, especially on lexical tones and its cognitive mechanisms, followed by an outline of the research plan for the present study.

### 1.1. Speech variability in lexical tones and the extrinsic normalization process

Speech signals vary considerably between speakers due to anatomical differences in vocal folds and tracts. Gender-related pitch variations are particularly noticeable: women generally produce higher-pitched speech than men due to shorter and thinner vocal folds. While such differences may have minimal impact on the perception of nontonal languages like English, they become crucial in tonal languages where pitch conveys meaning. In Cantonese, for example, three level tones – high (T55), mid (T33), and low (T22) – differentiate words (Yip, 2002): The same base syllable /ji/ means doctor with T55, meaning with T33, and two with T22. These tones share similar pitch contours, making pitch height the primary distinguishing factor. However, gender-specific pitch variations can confound the identification of these tones. A woman's mid-level tone could have a higher fundamental frequency (F0) than a man's high-level tone, blurring tonal distinctions.

Given that intrinsic speech cues can be unreliable in high-variability situations, listeners often rely on extrinsic cues to aid speech perception. One such strategy is extrinsic normalization, where listeners interpret the target speech cue by referring to the speech cues in external contexts

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use and/or adaptation of the article.

(Ainsworth, 1975; Johnson, 2005; Nearey, 1989). For instance, an ambiguous Cantonese level tone is more likely to be perceived as a low level tone if preceded by a high-pitched context and as a high level tone if preceded by a low-pitched context, showing a *contrastive context effect* (Francis et al., 2006; Wong & Diehl, 2003). This adaptive strategy mitigates the perceptual challenges posed by speaker variability, thereby enhancing perceptual constancy across different speakers. Both Peng et al. (2012) and Wong and Diehl (2003) found that when Cantonese words from multiple speakers were presented in isolation – especially words with level tones – the recognition accuracy was significantly lower. However, embedding these words in speech context (e.g., /ha6 yat1 go3 ji6 hai6/ meaning “The next word is...”) substantially improved identification accuracy.

## 1.2. Cognitive mechanisms underlying the extrinsic normalization process

Understanding the cognitive mechanisms involved in extrinsic normalization is vital for comprehending why L2 learners face difficulties in accommodating speech variability. Speech perception operates hierarchically, beginning with the decoding of acoustic cues and culminating in word selection (McClelland & Elman, 1986). Accordingly, the extrinsic normalization process could theoretically manifest at any stage, and previous research has proposed three main types of mechanisms to explain the extrinsic normalization process (Johnson & Sjerps, 2018; Stilp, 2020; Xie et al., 2023).

### 1.2.1. Acoustic adjustment

Some scholars posit that extrinsic normalization occurs at the stage of acoustic processing. In this framework, the auditory system automatically decodes acoustic cues in the target syllables in contrast to those in the surrounding context (Holt, 2006; Holt et al., 2001; Holt & Lotto, 2002). This notion is bolstered by observations that acoustic contrast between context and target is essential for extrinsic normalization to occur and that even nonspeech contexts could trigger this normalization process (Huang & Holt, 2009). Supporting the acoustic adjustment hypothesis, Sjerps et al. (2011) demonstrated that the N1 component, associated with acoustic analysis of speech signals, was triggered during extrinsic normalization. Further evidence was from neuroimaging studies that highlighted the role of subcortical regions, such as cerebellum, in this process (Guediche et al., 2015).

### 1.2.2. Phonetic/phonological adjustment

The second perspective asserts that extrinsic normalization takes place when acoustic signals are mapped to phonemic representations. For instance, through the Cantonese greeting “早晨” (/zou 25 san 21/, good morning), listeners ascertain the F0 for both the highest (the end point of T25) and lowest tones (the end point of T21) for a specific talker (K. Zhang et al., 2024). Listeners will categorize incoming tones using this talker-specific acoustic-phonemic mapping. Native speech contexts are more effective in this regard than nonnative and nonspeech contexts, implying that linguistic-specific stages involving phonetic and phonological information are at play (Kang et al., 2016; K. Zhang & Peng, 2021). Supporting this phonetic/phonological adjustment perspective, recent Event-Related Potential (ERP) studies identified the presence of the P2 component – a neural marker closely tied to phonetic and phonological processing – during the speech normalization process (K. Zhang, Tao, & Peng, 2023; K. Zhang & Peng, 2021).

### 1.2.3. Cognitive adjustment

A third account posits that normalization happens at the decision-making stage, where a comprehensive set of cues, including acoustic, phonological, and higher-level linguistic knowledge, are considered (Bosker et al., 2017). The evidence was mainly from the neuroimaging research, which reported that speech normalization recruited areas associated with decision-making – such as the inferior frontal gyrus (Myers et al., 2009; Myers & Mesite, 2014). EEG studies also reported the extrinsic normalization of lexical tones triggered a significant late positive component (LPC), an ERP component related to the stimuli evaluation, contextual integration and decision making, and the normalization efficiency at the behavioral level was positively correlated with LPC amplitude (C. Zhang et al., 2013).

To determine the most plausible mechanism for extrinsic normalization, a recent computational model integrated acoustic, phonetic/phonological, and cognitive adjustments and subjected them to empirical tests using real perceptual data (Xie et al., 2023). The study concluded that no single adjustment type alone could adequately explain the complexities of real-world speech perception, and that instead, a combined approach involving the perceptual adjustment at all three stages was the best account for adaptive speech behavior (Bosker et al., 2017; Persson & Jaeger, 2023; Xie et al., 2023). Previous studies about speech normalization in native speakers inform us that a thorough understanding of L2 learners’ normalization strategies necessitates investigating each adjustment stage during the speech normalization process.

## 1.3. The extrinsic normalization process in L2 speakers

L2 learners have challenges in L2 speech perception. This struggle is particularly evident in high-variability conditions, where nonnative speakers lag significantly behind their native counterparts compared to low-variability conditions (Tamati & Pisoni, 2014), and less proficient L2 listeners even demonstrate a more pronounced decline (Antoniou et al., 2015). A main cause of these challenges might be the L2 learners’ inability to adeptly use speech normalization strategies (K. Zhang et al., 2024). For instance, when presented with Mandarin tone continuum T214-T35 that vary in pitch and timing, Mandarin speakers effectively use both pitch and temporal cues from preceding contexts to interpret the target tone. In contrast, English listeners struggle with such interpretation (Jongman & Moore, 2000). Such a pattern also persists in the identification of the Mandarin T35-T55 continuum. Mandarin speakers exhibit a context-dependent interpretation of Mandarin T35-T55 continuum, adjusting responses based on the pitch of the preceding context. Yet, this adaptability is absent among English listeners (Luo & Ashmore, 2014). Even speakers of other tonal languages, who are accustomed to tone normalization in their native language (L1), face difficulties in normalizing L2 tonal categories that do not exist in their L1. For instance, Mandarin has only one level tone (T55), whereas Cantonese has three (T55, T33, and T22). When identifying these Cantonese level tones under high talker variability, Mandarin speakers, who lack a native contrast between level tones, perform worse than native Cantonese speakers (K. Zhang et al., 2024). These studies suggest that extrinsic normalization strategies do not transfer seamlessly from L1 to L2; instead, they develop gradually through L2 learning (K. Zhang et al., 2024).

Talker variability affects different phonological categories to varying degrees. The tone pairs used in previous studies – Mandarin T214-T35 (Jongman & Moore, 2000) and Mandarin T35-T55 (Luo

& Ashmore, 2014) – differ in both pitch height (high vs. low) and pitch slope (falling-rising vs. rising or rising vs. level), making them relatively robust to talker variability. Their identification relies more on intrinsic cues and less on contextual information. In contrast, the three Cantonese level tones (T55, T33, and T22) share a similar pitch slope (i.e., level) and primarily differ in pitch height. Because talker-related pitch differences can obscure intrinsic pitch height cues, listeners must rely on extrinsic pitch height cues from the surrounding context for accurate identification (K. Zhang et al., 2024). Consequently, the effect of normalization is generally stronger for level tones than for contour tones (Chen et al., 2023; K. Zhang, Tao, & Peng, 2023). The high context-dependence of Cantonese level tones makes them particularly suitable for testing normalization processes. Given that L2 learners, especially beginners, generally perform worse than native speakers in L2 tasks, using a more sensitive measure increases the likelihood of detecting performance differences between native and nonnative speakers. Therefore, the present study would employ Cantonese level tones to examine the extrinsic normalization process in L2 learners.

Despite these observed challenges, little is known about the neural mechanisms underlying L2 learners' extrinsic normalization. As discussed in Section 1.2, normalization likely involves multiple stages of perceptual adjustment. However, behavioral studies alone cannot pinpoint which of these stages are compromised in L2 learners. Unraveling this could illuminate neural underpinnings of perceptual normalization in L2 acquisition and shape targeted training interventions. With its high temporal resolution, EEG offers the potential to explore the normalization process at these distinct perceptual adjustment stages. The primary goal of the present study is to harness EEG to decipher the neural mechanisms of L2 learners' normalization strategies.

The extrinsic normalization process improves along with L2 learning. Antoniou et al. (2015) demonstrated that Mandarin speakers who resided in the United States for several years outperformed those with no exposure to English-speaking environments when identifying English words with high speaker variability, indicating that extensive language immersion facilitates L2 normalization process. K. Zhang et al. (2024) found that among overall Cantonese proficiency, Cantonese immersion, and Cantonese tone proficiency, only the Cantonese immersion significantly enhanced the extrinsic normalization process of Cantonese tones in Mandarin learners. Although these studies revealed the pivotal role of language immersion in facilitating the L2 normalization, the specific mechanisms through which immersion exerts its influence remain unclear due to the multifaceted nature of immersion experiences.

L2 immersion sharpens learners' sensitivity to subtle phonetic contrasts crucial for distinguishing L2 phonemes (Casillas, 2020; Winkler et al., 1999), and fosters a more native-like speech cue weighting (Ylinen et al., 2010), which in turn empowers learners to develop more robust and precise cognitive representations of L2 phonemes. Additionally, immersion broadens learners' L2 expertise, notably in areas like vocabulary and grammar, which are also important for the extrinsic normalization process, as meaningful linguistic contexts elicited an 11% greater normalization effect than meaningless word sequences (C. Zhang et al., 2015). This heightened proficiency provides L2 learners with extra time and cognitive resources to continuously evaluate and refine their linguistic performance. It's plausible that language immersion positively impacts all three stages of the extrinsic normalization process (i.e., acoustic, phonological, and cognitive adjustments), although its magnitude may differ across these stages. A deeper dive into this dynamic

would shed light on how language immersion and the development of perceptual normalization strategies intertwine in the realm of L2 acquisition. EEG, which can probe various depths of language processing, seems a suitable approach to this inquiry. Hence, building on the findings from K. Zhang et al. (2024), the present study would use EEG techniques to explore the ways in which L2 immersion enhances learners' extrinsic normalization capabilities (the second objective of the present study).

#### 1.4. The present study

The first research question of this study is: What are the neural mechanisms underlying the extrinsic normalization processes in L2 learners? This question would be addressed by asking Mandarin learners of Cantonese to perceive Cantonese level tones spoken by various speakers in contexts with different pitch heights. The high context-dependence of Cantonese level tones makes them an ideal candidate for testing normalization processes. Given that Mandarin lacks level tone contrast, Mandarin learners have no prior experience in normalizing multiple level tones in L1. Therefore, Mandarin learners' acquisition of Cantonese level tones would allow us to examine how tonal-language speakers acquire a more sophisticated tone normalization strategy when learning a L2 with a more complex tonal system.

The normalization process would be measured at both the behavioral and cortical levels. According to the definition of extrinsic normalization where listeners interpret the target speech according to the speech cues in context, at the behavioral level, the normalization process is quantified by examining whether participants' perceptions change according to the pitch height of the context. Specifically, the normalization process would lead subjects to give more high level tone responses (T55) in a low-F0 context, more mid level tone responses (T33) in a mid-F0 context, and more low level tone responses (T22) in a high-F0 context (i.e., the contrastive context effect) (Francis et al., 2006; Wong & Diehl, 2003). By statistically testing whether the expected responses in each pitch-height condition are significantly greater than the alternatives, we could determine if participants were engaging in normalization process. Moreover, by comparing these expected responses between Mandarin and Cantonese subject groups in each condition, we could assess whether Mandarin learners have achieved proficiency comparable to native Cantonese speakers in employing these strategies.

At the cortical level, the quantification strategy for normalization process differs from the one used at the behavioral level. Participants did the normalization process regardless of whether the context pitch is high, mid, or low. In such cases, contrasting EEG responses following contexts with different pitch heights would likely cancel out, or at least substantially diminish, the effects attributed to normalization process. Moreover, because of normalization, listeners would identify the same target words (i.e., /ji33/) as different words (e.g., /ji22/, /ji33/, and /ji55/) according to the context pitch heights. Thus, the contrast of pitch heights in such conditions might reflect the retrieval of different lexical tones or words (e.g., /T22/ vs. /T55/) rather than the normalization process itself. Besides, based on previous studies with a similar design, the scalp EEG might not be sensitive enough to capture the subtle differences caused by the retrieval of different words (K. Zhang & Peng, 2025). To better capture the normalization process at the cortical level, we compared EEG responses elicited by speech and nonspeech contexts. Numerous studies have consistently demonstrated that a nonspeech context does not trigger the normalization



process during Cantonese tone perception, whereas a speech context does (e.g., Francis et al., 2006; K. Zhang et al., 2017). Therefore, we believe that the differences in EEG responses between speech and nonspeech contexts should include, if not be entirely attributable to, the normalization process, which could be an effective way to detect the online normalization process. Previous studies (e.g., C. Zhang et al., 2013; K. Zhang & Peng, 2021) have effectively detected online extrinsic normalization processes using this speech–nonspeech contrast method. Thus, the present study would adopt this comparative approach between speech and nonspeech to observe the extrinsic normalization process at the cortical level, with participants perceiving Cantonese level tones in both speech and nonspeech contexts.

Given the likelihood of extrinsic normalization being a multi-stage process, speech–nonspeech discrepancies in EEG signals (i.e., the online extrinsic normalization process) might be observable across various time windows. Three ERP components, previously identified in extrinsic normalization studies, would be tested in the present study: N1, reflecting acoustic adjustment of perceived signals (Sjerps et al., 2011); P2, indicative of phonetic/phonological adjustment (Zhang & Peng, 2021); and LPC, representing cognitive adjustment during the final decision stage (C. Zhang et al., 2013). We hypothesized that all three ERP components would emerge during native Cantonese speaker's Cantonese tone perception. However, we hypothesized that Mandarin learners might display normalization processes predominantly in the N1 time window, associated with acoustic adjustment at the auditory level, with lesser to no impact during the P2 and LPC stages, which were heavily reliant on language-specific knowledge.

A secondary research question is: How does language immersion influence the extrinsic normalization process in L2 learners? Specifically, we aim to examine the relationship between different stages of normalization and language immersion scores to identify how L2 immersion enhances specific facets of L2 knowledge or cognitive resources. We hypothesized that L2 immersion score would positively affect all stages of normalization processing in Mandarin learners of Cantonese, ranging from the acoustic decoding to the cognitive adjustment.

## 2. Methods

### 2.1. Participants

Initially, the experiment engaged 30 Mandarin learners of Cantonese and 30 native Cantonese speakers. However, the final dataset considered only 29 Mandarin participants (Mean age = 20.52; 15 females) and 25 Cantonese participants (Mean age = 21.66; 13 females) due to the exclusion of participants with poor EEG signal quality – comprising one Mandarin and five Cantonese participants (see Section 2.4 for details). All Mandarin participants were born in Mainland China and recognized Mandarin as their L1 and predominantly communicated in it. Among any languages or Chinese dialects they know, none have more than one level tone, as validated by the first author. At the time of the experiment, they were enrolled as undergraduate or postgraduate students at The Hong Kong Polytechnic University and were learning Cantonese to aid their acclimatization to life and academic pursuits in Hong Kong. Conversely, the Cantonese participants were born in Hong Kong and were also students at The Hong Kong Polytechnic University during the experiment's timeframe. All participants had minimal exposure to professional music training, capping at 3 years, and self-reported normal hearing and language abilities. All

participants were right-handed, as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971). Comprehensive insights into the experimental procedures were provided to all participants, and informed consent was acquired prior to the commencement of the experiment. Participants received appropriate compensation for their time. The study was conducted in adherence to the ethical guidelines approved by the Human Subjects Ethics Sub-committee of The Hong Kong Polytechnic University.

Mandarin participants provided basic information about their Cantonese background during the recruitment phase. Additionally, during the experiment, they completed the Language History Questionnaire (LHQ3) to offer a more comprehensive account of their Cantonese experiences (Li et al., 2020). The LHQ3 is a web-based tool for assessing the linguistic profile of users, which was developed based on the most commonly asked questions in published studies. It generates aggregate scores based on the participant's answer to each question to quantify the overall proficiency, dominance, and immersion levels for each language the participant has learned. Table 1 summarizes the Cantonese background information for each Mandarin participant. Specifically, the duration of residence in Cantonese-speaking regions was obtained from the recruitment form, while age of acquisition (AOA), years of use, Cantonese proficiency, Cantonese immersion, Cantonese dominance, and the Cantonese-to-Mandarin dominance ratio were derived from the LHQ3 data. For the calculation formulas of these aggregate scores, please refer to <https://lhq-blclab.org/static/docs/aggregate-scores.html>. In this study, Mandarin learners provided details regarding their acquisition and usage of Mandarin and Cantonese exclusively, excluding other L2s such as English. Note that two participants (IDs #9 and #25) did not complete the LHQ3.

### 2.2. Stimuli

The stimuli for the Cantonese tone identification task replicated those used in Tao et al. (2021) and K. Zhang et al. (2017). Each trial incorporated a context (either speech or nonspeech) and a speech target. The speech context was the Cantonese phrase 呢個字係 (/li55 ko33 tsi22 hvi22/, meaning “this word is ...”), articulated by four native Cantonese speakers with varying pitch ranges [female high (FH), female low (FL), male high (MH), and male low (ML)]. To introduce intra-talker variability, the F0 trajectories of the original recordings were adjusted three semitones up or down in Praat (Boersma & Weenink, 2023), forming three distinct pitch-height contexts: high-F0, mid-F0, and low-F0 speech contexts. To maintain naturalness, context stimuli durations remained unaltered. The intensities of the speech contexts were set at 55 dB using Scale intensity function in Praat (Boersma & Weenink, 2023). The F0 trajectory of each speech context is illustrated in Figure 1A, showing its F0 value and duration. Nonspeech contexts, created from triangle waves, were manipulated to mirror the pitch trajectories and durations of the speech counterparts, resulting in 12 nonspeech contexts (4 speakers × 3 pitch heights). Specifically, consecutive triangular waveforms were used to model the F0 trajectory. Each upward triangle was immediately followed by its downward counterpart to form one complete cycle of the F0 pattern, with the intensity of each cycle matching that of the corresponding speech period. For unvoiced consonants – where no F0 data were available – linear interpolation between the immediately preceding and following F0 values was applied. This approach ensured that both the intensity profile and the duration of the nonspeech waveform closely aligned with those of the corresponding speech contexts. Furthermore, the intensities of

**Table 1.** The Cantonese background of Mandarin participants

Subject ID	Residence in Cantonese-speaking regions (months)	AOA	Years of use	Cantonese Proficiency	Cantonese Immersion	Cantonese Dominance	Cantonese to Mandarin Dominance Ratio
1	5	18	1	.21	.05	.12	.18
2	26	19	1	.29	.05	.14	.22
3	5	18	1	.14	.05	.07	.12
4	10	18	1	.43	.06	.22	.43
5	13	18	2	.43	.09	.25	.45
6	14	17	1	.29	.08	.16	.23
7	38	18	3	.64	.13	.34	.56
8	8	19	1	.14	.03	.09	.14
9	25	NA	NA	NA	NA	NA	NA
10	11	18	1	.21	.05	.12	.22
11	8	17	1	.43	.04	.26	.33
12	4	18	2	.14	.08	.07	.11
13	15	18	3	.43	.13	.31	.22
14	8	19	1	.43	.05	.22	.47
15	6	18	1	.21	.05	.17	.19
16	26	18	2	.5	.09	.25	.37
17	35	28	3	.43	.08	.25	.27
18	3	18	1	.36	.04	.19	.28
19	4	18	1	.21	.03	.11	.2
20	6	18	1	.43	.05	.25	.32
21	29	19	2	.29	.06	.17	.23
22	5	19	2	.71	.1	.45	.87
23	6	18	1	.43	.05	.26	.49
24	31	20	1	.14	.05	.09	.15
25	20	NA	NA	NA	NA	NA	NA
26	11	19	1	.29	.05	.17	.35
27	4	18	1	.21	.03	.11	.18
28	48	28	4	.5	.13	.29	.72
29	8	26	1	.21	.03	.11	.19
Mean	14.9	19.17	1.55	.34	.06	.19	.31
SE	2.26	.54	.15	.03	.01	.02	.03

the nonspeech contexts were set at 75 dB to match the perceived loudness of the speech contexts. This adjustment was based on C. Zhang et al. (2012), which employed a similar method to generate both speech and nonspeech stimuli. In that study, two native Cantonese speakers and the first author evaluated the synthesized nonspeech stimuli. Their feedback indicated that when the nonspeech intensities were set 20 dB higher than those of the speech stimuli, the nonspeech stimuli were perceived as having a similar loudness to the corresponding speech stimuli.

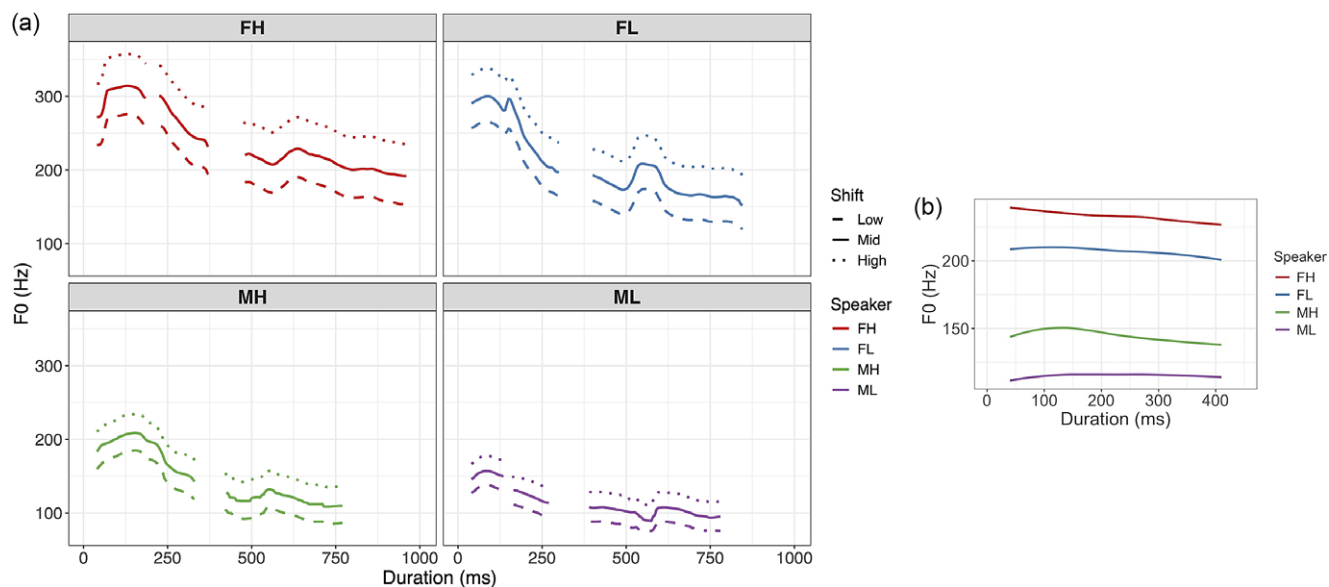
The speech targets consistently utilized the Cantonese syllable 意 (/ji33/, meaning), which were recorded by the same four speakers in a sound-proofed booth. The F0 of each target stimulus was kept unchanged. Target stimuli durations were normalized to 450 ms using the Lengthen (overlap-add) function in Praat to accommodate EEG signal processing. The F0 trajectory of the

speech target from each speaker is illustrated in Figure 1B. The intensities of the speech targets were set at 55 dB using the Scale intensity function in Praat. Each trial ensured that the speaker of the context and target stimuli matched.

Additional fillers, undergoing similar pitch manipulation as the test stimuli, were also included. The context fillers were Cantonese phrases: 我而家讀 (/ŋo23 ji21 ka55 tuk2/, now I will read...) and 請留心聽 (/tshɿŋ25 ləu21 səm55 thɿŋ55/, please listen to... carefully). The target fillers were Cantonese 意 (/ji33/, meaning) or 二 (/ji22/, two).

### 2.3. Experimental procedure

Participants took a Cantonese tone identification task, executed in a soundproof booth, consisting of a speech-context block and a



**Figure 1.** The F0 trajectory of each speech context (a) and speech target (b).

nonspeech-context block. Each block contained 108 test trials (4 speakers  $\times$  3 pitch heights  $\times$  9 repetitions) and 36 fillers. Audio stimuli were bilaterally delivered through inserted earphones. Each trial initiated with a 500 ms fixation, followed by a context stimulus. A target stimulus was then played after a silent interval varying between 300 and 500 ms. A question mark appeared on the screen 350–550 ms post the target stimulus, prompting participants to press the corresponding keys to indicate the last word heard: 醫 (/ji55/, doctor), 意 (/ji33/, meaning), or 二 (/ji22/, two). The subsequent trial was presented 2000 ms post the appearance of the question mark.

#### 2.4. EEG signal recording and preprocessing

EEG signals were recorded using a SynAmps 2 amplifier (NeuroScan, Charlotte, NC, USA) with a cap carrying 64 Ag/AgCl electrodes placed on the scalp surface at the standard locations according to the international 10–20 system. Offline reference channels were located at the left and right mastoids. Horizontal and vertical eye movements were monitored by two bipolar channels for electrooculography (EOG). Impedance between the online reference electrode (placed between Cz and CPz) and any recording electrode was kept below 5 k $\Omega$ . EEG signals were continuously recorded during the Cantonese tone identification task at a 1000 Hz sampling rate.

EEG data were processed using custom scripts in MATLAB, utilizing EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014) functions. The EEG signals were filtered offline with a 0.1 Hz high-pass and a 30 Hz low-pass filter (both slopes = 12 dB/Oct) and re-referenced offline to the average mastoid recordings. Epochs from –100 to 800 ms (time locked to the onset of target stimulus) were extracted and baseline-corrected based on the –100–0 ms pre-target stimulus activity. Any epochs exceeding  $\pm 100$   $\mu$ V at any scalp channels were discarded. Eye blinks were detected automatically by a moving window peak-to-peak threshold criterion on the VEOG data with the threshold of 100  $\mu$ V, a window size of 200 ms, and a window step of 50 ms. Horizontal eye movements were detected automatically by a step-

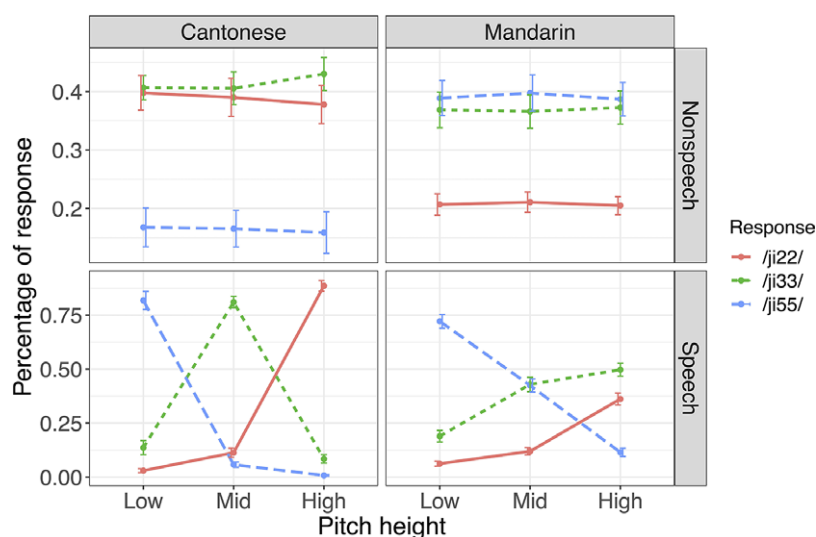
like threshold criterion on the HEOG data with a threshold of 40  $\mu$ V, a window size of 400 ms, and a window step of 10 ms. Participants with <70% accepted epochs were excluded, resulting in 29 Mandarin and 25 Cantonese participants for analysis. Overall, acceptance rates were 91.3% (SD = 7.4%) in nonspeech and 90.6% (SD = 8.7%) in speech contexts.

### 3. Results

To ensure consistency in the behavior and EEG data analysis, participants excluded from the EEG data analysis due to poor EEG signal quality were also excluded from the behavioral data analysis. Consequently, the final analyses were conducted on 29 Mandarin and 25 Cantonese participants. Although the sample sizes were slightly unbalanced in two groups, we employed mixed-effects models, which are well-known for their robustness to such designs. In addition, we subsampled 25 Mandarin subjects to match 25 Cantonese subjects and re-ran the analyses using this balanced dataset (see the [Supplementary Material](#) for details). The results from two datasets were qualitatively consistent. Therefore, in the main text, we reported the results based on 29 Mandarin and 25 Cantonese participants.

#### 3.1. The behavioral results of the Cantonese tone identification task

Figure 2 presents participants' responses under each experimental condition. To assess the emergence of the extrinsic normalization process within each group and any differences therein between the two groups, a multinomial regression model was applied to responses from the Cantonese tone identification task using the *met* package (Venables & Ripley, 2002) in R, with *response* (three levels: /ji55/, /ji33/, and /ji22/) as the dependent variable and *pitch height* (three levels: high, mid, and low), *speechness* (indicating whether the nature of context is speech or nonspeech; two levels: speech context and nonspeech context), *group* (two levels: Cantonese and Mandarin), and their possible two-way and three-way interactions as the fixed factors. Only by-participants intercept



**Figure 2.** The percentage of each response in different experiment conditions. The error bar represents the standard error of the mean.

and by-speaker (FH, FL, MH, ML) intercept were included as the random factors due to the convergence issue. The sum coding scheme was applied to all the categorical variables (*pitch height*: −1 for low, 0 for mid, and 1 for high; *speechness*: −1 for nonspeech context and 1 for speech context; *group*: −1 for Cantonese and 1 for Mandarin). The reference level for *response* was set to /ji33/.

The summarized model statistics are presented in Table 2. A significant main effect of pitch height was observed, revealing that when the context changed from low to high, participants

demonstrated a general contrastive context effect by giving significantly fewer /ji55/ responses ( $\beta = -1.317$ , OR = .268,  $p < .001$ ) and more /ji22/ responses ( $\beta = .983$ , OR = 2.672,  $p < .001$ ), relative to /ji33/. However, nuanced differences were noticed between the two groups. The significant interaction between pitch height and group indicated variations in context effects. Specifically, as the context pitch height ascended from low to high, Mandarin learners exhibited a propensity to give more /ji55/ responses ( $\beta = .64$ , OR = 1.897,  $p < .001$ ) and fewer /ji22/ responses ( $\beta = -.702$ , OR = .495,  $p < .001$ )

**Table 2.** The multinomial logistic regression modeling on the Cantonese tone normalization task

	/ji55/ (relative to /ji33/)					
	$\beta$	SE	95% CI	$z$	$p$	OR
Intercept	−.268	.019	[−.293–.243]	−20.995	< .001	.765
pitch height	−1.317	.047	[−1.409–1.226]	−28.138	< .001	.268
context	−.37	.038	[−.445–.295]	−9.667	< .001	.691
group	.816	.038	[.741 .891]	21.314	< .001	2.262
pitch height: context	−1.286	.047	[−1.378–1.194]	−27.467	< .001	.276
pitch height: group	.64	.047	[.548 .732]	13.673	< .001	1.897
context: group	.324	.038	[.249 .4]	8.472	< .001	1.383
pitch height: context: group	.616	.047	[.525 .708]	13.166	< .001	1.852
	/ji22/ (relative to /ji33/)					
	$\beta$	SE	95% CI	$z$	$p$	OR
Intercept	−.255	.016	[−.276–.234]	−24.16	< .001	.775
pitch height	.983	.043	[.9 1.068]	22.665	< .001	2.672
context	−.446	.032	[−.508–.383]	−14.067	< .001	.64
group	−.002	.032	[−.064 .06]	−.0564	.955	.998
pitch height: context	1.014	.043	[.93 1.1]	23.398	< .001	2.758
pitch height: group	−.702	.043	[−.79–.62]	−16.195	< .001	.495
context: group	.254	.032	[.192 .316]	8.027	< .001	1.29
pitch height: context: group	−.724	.043	[−.809–.639]	−16.702	< .001	.485

Note: OR refers to odds ratio.

compared to native Cantonese speakers, indicative of a diminished contrastive context effect in Mandarin learners. Furthermore, a significant interaction between pitch height and speechness unveiled disparities in context effects between speech and nonspeech contexts. As pitch height in contexts increased from low to high, participants provided fewer /ji55/ responses ( $\beta = -1.286$ ,  $OR = .276$ ,  $p < .001$ ) and more /ji22/ responses ( $\beta = 1.014$ ,  $OR = 2.758$ ,  $p < .001$ ) in speech contexts compared to nonspeech contexts, delineating a more pronounced contrastive context effect within speech contexts.

The regression model also revealed a significant *pitch height* by *speechness* by *group* interaction. The post-hoc analysis on this three-way interaction was conducted in two directions. First, we tested if the extrinsic normalization process of lexical tones emerged in each group-speechness condition (i.e., Cantonese-speech, Cantonese-nonspeech, Mandarin-nonspeech, and Mandarin-nonspeech) by comparing each response in three pitch heights. As outlined in Section 1.4, if subjects showed the extrinsic normalization process, their response should change according to the pitch height of the context. In such conditions, more /ji55/ responses were expected when pitch height of context changed from high to low, and in contrast, more /ji22/ responses were expected when pitch height of context changed from low to high. Second, we tested if two groups showed differences in the Cantonese tone normalization process by comparing their responses in each speechness-pitch height condition (i.e., speech-high, speech-mid, speech-low, nonspeech-high, nonspeech-mid, and nonspeech-low).

The first post-hoc analysis revealed that both groups showed a typical context-dependent perception of target tones in speech contexts. The statistical results for each pairwise contrast were summarized in Table 3. Specifically, native Cantonese speakers gave the most /ji55/ responses in the low-F0 speech contexts (low:  $prob = .831$ ,  $SE = .013$ ; mid: .059, .008; high: .008, .003), they gave the most /ji33/ responses in mid-F0 speech contexts (low: .139, .012; mid: .826, .013; high: .086, .009), and they also gave most /ji22/ responses in high-F0 speech contexts (low: .03, .006; mid: .116, .011; high: .905, .01). In speech contexts, Mandarin learners gave the most /ji55/ responses in low-F0 contexts (low: .741, .014; mid: .436, .016; high: .118, .01), and they also gave the most /ji22/ responses in high-F0 speech contexts (low: .064, .008; mid: .123, .01; high: .377,

.015;  $ps < .001$ ). Although Mandarin learners gave more /ji33/ responses (.441, .016) than /ji22/ responses (.123, .01) in mid-F0, their /ji33/ responses were comparable to /ji55/ responses (.436, .016), the only pair not displaying context-dependent perception in the speech-context condition. In nonspeech contexts, there were no significant variations in responses between low and high contexts ( $ps > .05$ ) for both groups, reaffirming the findings of Francis et al. (2006), C. Zhang et al. (2012), Zhang and Peng (2021) about the unequal effects of speech and nonspeech contexts on Cantonese tone normalization process.

The second post-hoc analysis revealed significant group differences. In speech contexts, Cantonese speakers showed a more typical normalization process, as they gave more expected responses in each pitch height condition (i.e., /ji55/ in high-F0 context, /ji33/ in mid-F0 context, and /ji22/ in low-F0 context). Specifically, they gave more /ji55/ responses than Mandarin learners in the low-F0 speech contexts (.831 vs. .741;  $\beta = -.09$ , 95%  $CI = [-.168 \text{ } -.011]$ ,  $SE = .02$ ,  $d = -.05$ ,  $t = -4.805$ ,  $p = .01$ ), more /ji33/ responses in the mid-F0 speech contexts (.826 vs. .441;  $\beta = -.385$ , 95%  $CI = [-.47 \text{ } -.3]$ ,  $SE = .02$ ,  $d = -.212$ ,  $t = -19.1$ ,  $p < .001$ ), and more /ji22/ responses in the high-F0 speech contexts (.906 vs. .371;  $\beta = -.534$ , 95%  $CI = [-.61 \text{ } -.46]$ ,  $SE = .02$ ,  $d = -.294$ ,  $t = -29.544$ ,  $p < .001$ ). Noticeable perceptual differences between the two groups were also observed in nonspeech contexts. As visualized in the upper panels of Figure 2, Cantonese subjects were more likely to perceive the target tone as /ji22/ and /ji33/, but Mandarin subjects were more likely to perceive it as /ji33/ and /ji55/, showing an influence from Mandarin tone system which comprises only one level tone, that is, Mandarin T55.

### 3.2. The ERP results

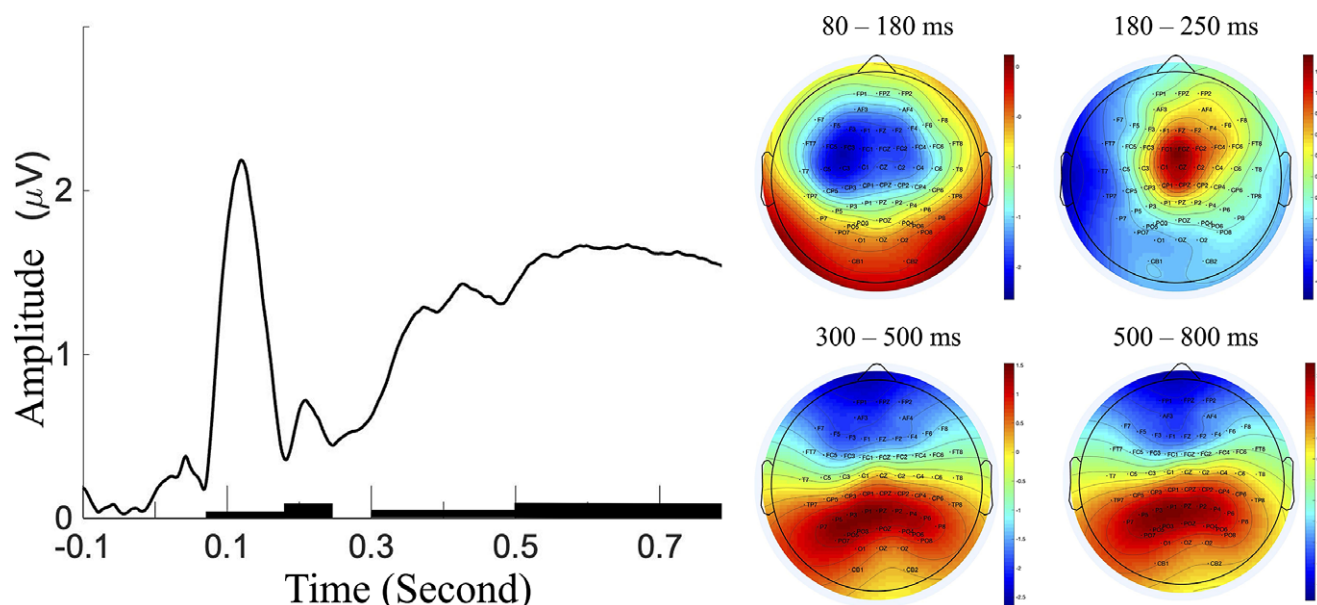
The global field power over time in Figure 3 (left) represents the root mean square of the ERP value at each time point averaged across 64 electrodes, six contexts (2 speechness  $\times$  3 pitch heights), nine repetitions, four speakers (FH, FL, MH, and ML), and 54 participants (29 Mandarin and 25 Cantonese participants). The global field power (Figure 3) and the ERP waves (averaged across repetitions, speakers, and participants; Figure 4) indicated that N1, P2, and LPC reported in the previous speech normalization studies

**Table 3.** The statistical result for each pairwise contrast in the speech context condition

Group	Contrast	PH	$\beta$	SE	95% CI	$t$	$p$	$d$
Can	/ji33/ - /ji22/	H	-.819	.019	[-.868 -.77]	-42.914	< .001	-.452
Can	/ji55/ - /ji22/	H	-.898	.011	[-.926 -.869]	-81.141	< .001	-.495
Can	/ji33/ - /ji55/	L	-.692	.024	[-.752 -.631]	-29.396	< .001	-.381
Can	/ji55/ - /ji22/	L	.8	.016	[.76 .841]	50.69	< .001	.441
Can	/ji33/ - /ji55/	M	.767	.018	[.72 .814]	41.833	< .001	.423
Can	/ji33/ - /ji22/	M	.71	.022	[.653 .767]	31.926	< .001	.392
Man	/ji33/ - /ji22/	H	.139	.029	[.064 .214]	4.76	< .001	.076
Man	/ji55/ - /ji22/	H	-.254	.02	[-.306 -.201]	-12.405	< .001	-.14
Man	/ji33/ - /ji55/	L	-.546	.025	[-.611 -.482]	-21.805	< .001	-.301
Man	/ji55/ - /ji22/	L	.677	.018	[.63 .725]	36.659	< .001	.373
Man	/ji33/ - /ji55/	M	.005	.029	[-.071 .081]	.167	1	.003
Man	/ji33/ - /ji22/	M	.318	.021	[.263 .373]	14.894	< .001	.175

Note: PH refers to pitch height.





**Figure 3.** The global field power (left) and topographies in different times windows (right).

also emerged during target tone perception in the present study. Additionally, a noticeable N400 associated with semantic processing (Kutas & Federmeier, 2011), was discerned during target tone perception. C. Zhang et al. (2013) reported that extrinsic normalization of Cantonese tones interacted with semantic retrieval during the N400 time window. Consequently, the current study incorporated N400 in the data analysis to investigate its interaction with extrinsic normalization and to explore variations in semantic processing between the two groups.

The selected time window and electrodes for each ERP component are detailed in Table 4. The time window for each ERP component was identified based on the timeframe within which the ERP component appeared, as discerned through visual inspection of the global field power (refer to the black bar in Figure 3, left). The electrodes corresponding to each ERP component were chosen based on the topographies where the ERP amplitudes were anticipated to peak (see Figure 3, right).

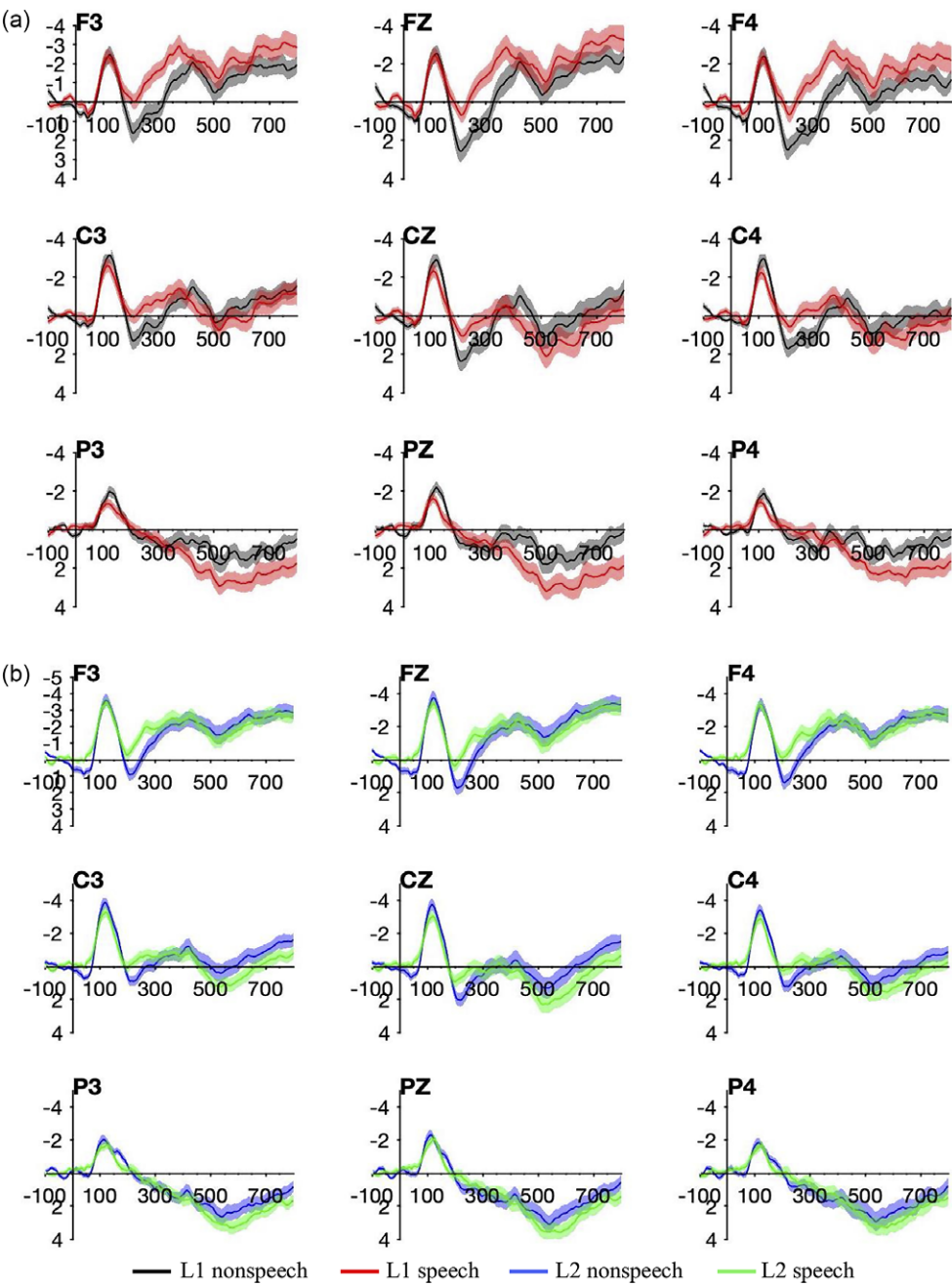
### 3.2.1. ERP amplitude results

A linear mixed-effects regression model was applied to the mean ERP amplitude, utilizing the 'lme4' package (Bates et al., 2015) in R. The ERP wave for each condition was calculated by averaging the EEG epochs over nine repetitions and four speakers in MATLAB. Given that nine repetitions are insufficient to acquire reliable ERPs, the EEG epochs were further averaged across four speakers (FH, FL, MH, and ML), as the study did not intend to examine the response of participants to individual speakers. The mean ERP amplitudes, which were used for model fitting, were further averaged over electrodes for two key reasons. First, models with electrodes as a random effect failed to converge. Second, given the EEG signals' low spatial resolution, the analyses on each electrode might not yield meaningful insights. Consequently, the model was designed to include only *ERP component* (four levels: N1, P2, N400, and LPC; dummy coded), *speechness* (two levels: speech context and non-speech context; dummy coded), *pitch height* (two levels: high, mid, low; dummy coded), *group* (two levels: Cantonese and Mandarin; dummy coded), and any potential two-way, three-way, and four-way interactions as fixed effects. Due to convergence issues, only the

by-subject intercept was included as the random effect. The model explained 37.8% of the variance ( $R^2_m = .378$ ) due to the fixed effects, and 55.5% of the variance ( $R^2_c = .555$ ) due to both fixed and random effects.

The  $p$ -values for each fixed factor were derived using Anova function in the car package. Partial  $\eta^2$  was calculated to represent the effect sizes of each predictor. After identifying significant main effects or interactions, post hoc pairwise comparisons were conducted using the 'emmeans' package (Lenth, 2019), applying Bonferroni adjustment for multiple comparisons. Cohen's  $d$  was calculated for each pairwise comparison in the post hoc analysis to measure the effect size. As outlined in Section 1.4, the cortical normalization process was quantified by contrasting EEG responses in speech and nonspeech conditions. Thus, a significant main effect of speechness or interaction involving speechness (i.e., speech–nonspeech difference) indicates the engagement of the normalization process. The analysis revealed a significant main effect of *ERP component* [ $\chi^2(3) = 982.333, p < .001, \eta^2_p = .436$ ]. The amplitudes of four ERPs differ significantly from each other (N1:  $-1.964$ , P2:  $.886$ , N400:  $-1.656$ , LPC:  $1.815$ ). There is a significant main effect of *speechness* [ $\chi^2(1) = 4.683, p = .031, \eta^2_p = .004$ ]. The ERP amplitude is generally more negative in speech contexts ( $-.347$ ) than in nonspeech contexts ( $-.113$ ). The analysis also revealed a significant *group by ERP component* interaction [ $\chi^2(3) = 23.184, p < .001, \eta^2_p = .018$ ], a significant *speechness by ERP component* interaction [ $\chi^2(3) = 54.053, p < .001, \eta^2_p = .043$ ], and a significant *speechness by group by ERP component* interaction [ $\chi^2(3) = 8.358, p = .039, \eta^2_p = .007$ ]. Pitch height and its interaction with other factors were not significant ( $ps > .05$ ). This finding suggests that scalp EEG may lack sufficient sensitivity to detect the subtle cortical changes associated with pitch height variations, which are primarily related to the retrieval of different lexical items.

Since two significant two-way interactions were subsumed within the three-way interaction, post hoc analysis was conducted only on the three-way interaction. The post hoc analysis aimed to investigate how ERP amplitudes differed between speech and non-speech conditions within each group, for each Group-ERP combination (e.g., Cantonese-N1). The results suggested that ERP



**Figure 4.** The ERP waves at nine electrodes during the target tone perception for Cantonese (a) and Mandarin (b) subjects.

**Table 4.** The time window and electrodes for each ERP component

ERP	Time windows	Electrodes
N1	80–180 ms	F3, F1, FC5, FC3, FC1, C5, C3, C1, FZ, FCZ, F2, FC2
P2	180–250 ms	FC1, C1, CP1, FZ, FCZ, CZ, CPZ, F2, FC2, C2, CP2, F4, FC4, C4
N400	300–500 ms	FC5, FC3, F7, F5, F3, F1, AF3, FZ, F2, AF4, F4, F6
LPC	500–800 ms	P5, P3, P1, PO7, PO5, PO3, PZ, POZ, P2, P4

amplitudes in nonspeech and speech contexts differ significantly from each other for P2 (speech: .414, nonspeech: 1.8;  $\beta = 1.386$ ,  $SE = .312$ , 95% CI = [.774, 1.999],  $t = 4.439$ ,  $p < .001$ ,  $d = .739$ ), N400 (speech:  $-1.958$ , nonspeech:  $-.726$ ;  $\beta = 1.231$ ,  $SE = .312$ , 95%

CI = [.619, 1.844],  $t = 3.944$ ,  $p < .001$ ,  $d = .656$ ), and LPC (speech: 2.127, nonspeech: .956;  $\beta = -1.171$ ,  $SE = .312$ , 95% CI = [−1.784, −.559],  $t = -3.751$ ,  $p < .001$ ,  $d = -.624$ ) in the Cantonese group. However, for the Mandarin group, only in P2 time window, the ERP amplitude in speech contexts is significantly different from that in nonspeech contexts (speech: .159, nonspeech: 1.171;  $\beta = 1.012$ ,  $SE = .29$ , 95% CI = [.443, 1.581],  $t = 3.49$ ,  $p < .001$ ,  $d = .54$ ).

**3.2.2. ERP latency results**

The peak latency of each component, for every subject, was determined as the time point corresponding to either the minimal (for N1 and N400) or maximal (for P2 and LPC) point of the 2nd-order polynomial curve fitted to the ERP wave. The statistical approach employed for analyzing ERP latency mirrored the one used for the ERP amplitude, incorporating *ERP component*, *speechness*, *pitch*

height, group, and their potential two-way, three-way, and four-way interactions as the fixed effects, and by-subject intercept as the random effect within the linear mixed-effects model. The model explained 92.2% of the variance ( $R^2_m = .922$ ) due to the fixed effects, and 92.4% of the variance ( $R^2_c = .924$ ) due to both fixed and random effects.

The analysis revealed a significant main effect of *ERP component* [ $\chi^2(3) = 15699.201, p < .001, \eta^2_p = .436$ ]. The latencies of the four ERPs differ significantly (N1: .133, P2: .216, N400: .403; LPC: .641). There is also a significant main effect of *group* [ $\chi^2(1) = 4.199, p = .041, \eta^2_p = .015$ ]. In general, Cantonese participants ( $M = .353, SE = .0031$ ) showed significantly later ERPs than Mandarin participants ( $M = .344, SE = .0029$ ). More importantly, there is a significant *group by ERP component* interaction [ $\chi^2(3) = 27.735, p < .001, \eta^2_p = .018$ ]. Post hoc analysis was conducted to examine differences in ERP latencies between the two groups, with Bonferroni adjustment applied to control for multiple comparisons. The results suggested that Cantonese and Mandarin groups showed significant difference only in the LPC latency ( $\beta = .037, SE = .007, 95\% \text{ CI} = [.023, .051], t = 5.31, p < .001, d = .659$ ), with Cantonese participants ( $M = .659, SE = .0051$ ) showing a significantly later LPC than Mandarin participant ( $M = .623, SE = .0047$ ). Pitch height and its interactions with other factors were not significant ( $ps > .05$ ).

### 3.3. The relationship between behavioural and cortical responses in the Cantonese tone normalization process

Linear regression models were employed to examine how the extrinsic normalization process at different stages at the cortical level affects the normalization performance at the behavioral level. As outlined in Section 1.4, at the behavioral level, if participants did the normalization process, they would give more expected responses than two alternatives according to the pitch heights of contexts (e.g., more /ji55/ in the low-F0 contexts). Therefore, in the regression models, the normalization performance at the behavioral level was indexed by the proportion of expected responses. Since normalization process only emerged in speech but not in nonspeech contexts, the normalization process at the cortical level was indexed by the speech–nonspeech difference wave, which was calculated by subtracting the EEG signals in nonspeech contexts from those in speech contexts. The amplitudes and latencies of four ERP components (i.e., N1, P2, N400, and LPC) for the difference wave were included in the regression models to represent different stages of the cortical-level normalization process.

The first linear regression model was fitted on the expected response rate using the *lm* function in R with the following predictors: *N1amplitude*, *P2amplitude*, *N400amplitude*, *LPCamplitude*, *group* (two levels: Mandarin and Cantonese; dummy coded), and interaction terms between each ERP amplitude and *group*. The significance of each predictor was assessed by Anova function from the *car* package. The analysis revealed a significant main effect of *group* [ $F(1, 44) = 27.28, p < .001, \eta^2_p = .687$ ], with Cantonese speakers ( $M = .844, SE = .026$ ) outperforming Mandarin speakers ( $M = .511, SE = .023$ ) in the expected response rate. However, none of the ERP components nor their interactions with the group were significant ( $ps > .05$ ).

A similar regression model was then fitted on the expected response rate, but with the latency of each ERP as predictors. The analysis revealed a significant main effect of *LPC latency* [ $F(1, 44) = 9.47, p = .003, \eta^2_p = .498$ ]. Group, other ERP latencies, and ERP latency by group interactions were not significant

( $ps > .05$ ). The positive significant main effect of LPC latency suggested that participants who exhibited higher proficiency in the extrinsic normalization process of Cantonese tones at the behavioral level also demonstrated later LPC at the cortical level, and, more importantly, this trends held for both Mandarin and Cantonese participants.

### 3.4. The relationship between L2 immersion and different ERPs

Linear regression models were fitted to test how the L2 immersion affects the different stages of normalization process at the cortical level. This analysis focused on 27 Mandarin participants who completed the LHQ3. A linear regression model was first fitted to the ERP amplitude of difference waves using *lmer* function in *lme4* package in R with *ERP component* (four levels: N1, P2, N400, and LPC, dummy coded), *L2 immersion*, and *their interaction* as the fixed effects, and by-subject intercept as the random effects. The model explained 16% of the variance ( $R^2_m = .16$ ) due to the fixed effects, and 54% of the variance ( $R^2_c = .54$ ) due to both fixed and random effects. The Anova function from the *car* package was used to test the significance of each predictor. The analysis revealed a significant main effect of *ERP component* [ $\chi^2(3) = 16.704, p < .001, \eta^2_p = .18$ ], suggesting that ERP amplitudes vary across different ERP components. L2 immersion and its interaction with ERP component were not significant ( $ps > .05$ ).

Another regression model was fitted to the ERP latencies of difference waves with the same setting. The model explained 93% of the variance ( $R^2_m = .93$ ) due to the fixed effects, and 93% of the variance ( $R^2_c = .93$ ) due to both fixed and random effects. The analysis revealed a significant main effect of ERP component [ $\chi^2(3) = 160.909, p < .001, \eta^2_p = .68$ ], L2 immersion [ $\chi^2(1) = 15.626, p < .001, \eta^2_p = .11$ ], and ERP components by L2 immersion interaction [ $\chi^2(3) = 13.386, p = .004, \eta^2_p = .15$ ]. To better understand the significant ERP components by L2 immersion interaction, the *emtrend* function from the *emmeans* package was used to calculate the estimated slope of L2 immersion for each ERP component. The analysis revealed that the L2 immersion significantly affect the ERP latency only at the LPC time window ( $\beta = 1.386, SE = .351, 95\% \text{ CI} = [.69, 2.082], t = 3.953, p < .001, d = .78$ ), indicating that Mandarin participants with longer L2 immersion also exhibited later LPC difference waves while normalizing Cantonese tones at the cortical level.

## 4. Discussion

### 4.1. Successful extrinsic normalization: A joint effort of multiple adjustments

To understand the neural mechanisms underlying L2 learners' extrinsic normalization process, this study invited Mandarin learners of Cantonese and native Cantonese speakers to perceive Cantonese level tones in different contexts, and concurrently recorded their EEG signals. Before contrasting the outcomes between Mandarin learners and native Cantonese speakers, it is crucial to first address the multi-stage nature of the extrinsic normalization process as unveiled by the ERP results, which would aid in a deeper understanding of the L2 learners' performance.

The unequal effect of speech and nonspeech contexts on the extrinsic normalization process was reduplicated at the behavioral level. The pitch heights in nonspeech contexts did not significantly impact the tone perception of either group, denoting the absence of a reliable normalization process in nonspeech contexts. Conversely,



within speech contexts, the alteration in context pitch heights led both groups to perceive the same target as distinct tones, thereby illuminating a pronounced extrinsic normalization process. The different impacts of speech and nonspeech contexts for both groups imply that the extrinsic normalization process at the cortical level is most discernible when we contrast the ERPs in speech and nonspeech contexts during the perception of the target Cantonese level tones (C. Zhang et al., 2013; Zhang & Peng, 2021).

This comparative approach between speech and nonspeech indeed revealed, among native Cantonese speakers, that the perception of the target tone within speech contexts manifested diminished P2 and amplified LPC compared to nonspeech contexts. This implicates a continuous implementation of the extrinsic normalization process in the P2 and LPC time windows for native speakers of Cantonese. The P2, originating from the planum temporale and BA 22, is related to phonetic and phonological processes (Crowley & Colrain, 2004; Godey et al., 2001), while the LPC correlates with advanced cognitive processes like the integration of context information, evaluation of stimuli and response options, and decision making (Finnigan et al., 2002; Q. Zhang et al., 2023). These two ERP components, each with distinct functional implications, consequently signify different adjustment stages of the extrinsic normalization process, such as P2 for phonetic/phonological adjustment (Zhang & Peng, 2021) and LPC for cognitive adjustment at the decision-making stage (C. Zhang et al., 2013).

Two adjustments appear to play a distinctive role in the extrinsic normalization process. The phonetic/phonological adjustment likely serves a pivotal role in the extrinsic normalization process, since Mandarin learners who showed significant extrinsic normalization at the behavioral level only demonstrated speech–nonspeech differences at the P2 time window. The tone perception in both groups prompted a reduced P2 amplitude in speech compared to nonspeech contexts, suggesting that successful phonetic/phonological adjustments may facilitate easier mapping of phonetic and phonological representations in speech contexts, thereby resulting in a smaller P2. The cognitive adjustment is intricately related to normalization accuracy. A later and larger LPC was observed in speech than in nonspeech contexts during Cantonese speakers' tone perception, and more importantly, Cantonese speakers who showed better extrinsic normalization at the behavioral level elicited later LPCs. The cognitive adjustment seems to necessitate both time (evidenced by larger ERP latency) and cortical resources (indicated by larger ERP amplitudes). As indicated by the positive relationship between LPC latency and normalization performance at the behavioral level in the Cantonese group, the longer the subjects engage in cognitive adjustment during decision-making, the more successful the extrinsic normalization appears to be.

In contrast to previous studies (e.g., Sjerps et al., 2011), we did not observe a significant speech–nonspeech difference in the N1 time window for either the Cantonese or Mandarin groups. Considering that the N1 component – originating in the primary auditory cortex – is closely associated with decoding the acoustic properties of auditory signals (Näätänen & Picton, 1987), the absence of a speech–nonspeech difference at N1 suggests that listeners may not engage in perceptual adjustments during this early acoustic processing stage. In this time window, the primary role of the auditory system might be to decode the target tone's acoustic features. It is only when the system interprets these acoustic signals into higher-level linguistic representations (e.g., phonetic features or phonological representations) that contextual cues are employed to recalibrate perception (Zhang & Peng, 2021). Another

possibility was that listeners also involved in the perceptual adjustment at the acoustic processing stage in both speech and nonspeech condition, as both speech and nonspeech contexts provide acoustic contrasts relative to the target tone. However, the speech and nonspeech contexts in this study were matched in terms of F0 trajectories, duration, and perceived loudness, resulting in comparable acoustic contrasts between the contexts and the target. Consequently, this balancing likely contributed to the lack of observed differences in the N1 component across speech and nonspeech conditions. Future studies could further elucidate this question by employing complementary neuroimaging techniques, such as magnetoencephalography or intracranial EEG, which can provide higher spatiotemporal resolution to capture subtle auditory processing differences and to determine whether the normalization process is also present in the primary auditory cortex (i.e., an index of normalization at the acoustic processing stage).

Most previous research has linked extrinsic normalization predominantly to a single phase of speech processing; this could be the decoding of acoustic cues as suggested by Holt and Lotto (2002), the mapping of phonetic/phonological representation proposed by Magnuson and Nusbaum (2007), or the cognitive adjustment of the final decision in Bosker et al. (2017). Similarly, past ERP studies have interpreted their results within the framework of a single-stage normalization process, localizing the extrinsic normalization to the earliest ERPs – like N1 in Sjerps et al. (2011) and P2 in K. Zhang and Peng (2021) – or to the most reliable ERPs, such as LPC in C. Zhang et al. (2013). However, a recent computational modelling study by Xie et al. (2023) challenged these isolated perspectives, revealing that a single account could not fully elucidate the intricacies of human speech perception. Instead, the synthesis of these diverse processes provided a more coherent explanation, thereby offering computational simulation evidence supporting a continuous and multi-stage normalization process. Partially aligned with Xie et al.'s findings, our study also uncovered a multi-stage extrinsic normalization process utilizing EEG techniques. We observed that the native speakers' extrinsic normalization processes activated significant P2 and LPC components, encompassing the phonetic/phonological and cognitive adjustment processes, respectively. To our understanding, this is the first study to furnish direct neurological evidence corroborating a continuous extrinsic normalization process. However, based on the experiment design of the present study, we did not identify a perceptual adjustment at the acoustic processing stage. Future intracranial EEG studies which are more sensitive to the subtle cortical activities might be carried out to clarify whether perceptual adjustment occurs at the acoustic processing stage as well.

#### **4.2. L2 learners' extrinsic normalization deficiencies: the impact of absent comprehensive adjustments**

L2 learners are worse at perceiving high-variability L2 speech compared to native speakers (Antoniou et al., 2015; Tamati & Pisoni, 2014), which is potentially due to their suboptimal extrinsic normalization strategies (Jongman & Moore, 2000; Luo & Ashmore, 2014). By comparing the perception of Cantonese tones between Mandarin learners and native Cantonese speakers, the present study revealed that while Mandarin learners did exhibit significant context-dependent interpretations of Cantonese lexical tones at the behavioral level, the effect size was considerably smaller compared to native Cantonese speakers, reduplicating the diminished efficacy of the extrinsic normalization process in L2 learners (K. Zhang et al., 2024).



To explore the reasons behind the inability of L2 learners to utilize contextual cues as effectively as native speakers in normalizing nonnative speech variability, this study used EEG techniques to probe the neural mechanisms underlying the extrinsic normalization process of Mandarin learners for Cantonese level tones. The EEG data highlighted pronounced disparities between Mandarin learners and native Cantonese speakers in their cortical normalization processes. Mandarin learners only exhibited speech–nonspeech differences in the P2 component, and no group differences was shown in P2, indicating that the observed significant normalization process at the behavioral level in Mandarin learners predominantly stems from phonetic/phonological adjustments. Conversely, native Cantonese speakers, exhibiting more extensive extrinsic normalization effects at the behavioral level, manifested a sustained online extrinsic normalization process spanning the P2 and LPC time windows. The ERP patterns of the two groups suggest that the attenuated extrinsic normalization effect in Mandarin learners is predominantly attributed to deficits in the cognitive adjustments.

The ERP results were partially distinct from the hypothesis outlined in Section 1.4. First of all, we did not detect a normalization process at the acoustic processing stage at both Mandarin and Cantonese groups. Probably, only when listeners need to interpret the perceived acoustic signals into a meaningful linguistic label, they started to use the context cues to adjust the perceived target signals (i.e., normalization process). Moreover, the observed similarity in EEG responses in the P2 time window between two groups also diverged from the initial hypothesis. Given that phonetic/phonological adjustment was influenced by L2 language knowledge, we anticipated a diminished or nonexistent speech–nonspeech difference in P2 among Mandarin learners. However, Mandarin learners of Cantonese mirrored the P2s of native Cantonese speakers, suggesting equivalent phonetic/phonological adjustments in the normalization process across two groups. To ensure participants could successfully complete the Cantonese tone identification task, all Mandarin participants were subjected to a screening test wherein they were required to translate five Cantonese sentences (including the context sentences used in the experiment) into Mandarin and read five Cantonese words (including three target words in the experiment). It is plausible that Mandarin participants who passed this screening had acquired sufficient phonetic/phonological knowledge to execute the phonetic/phonological adjustment effectively during the extrinsic normalization process.

The EEG results about LPC were congruent with our hypothesis. The speech–nonspeech comparison revealed significant differences in both LPC amplitude and latency in the Cantonese group, but no such distinctions were found in the Mandarin group, suggesting a lack of significant cognitive adjustment in Mandarin learners. Additionally, the Mandarin group demonstrated earlier LPC, suggesting a possible shallower engagement in cognitive adjustment compared to native Cantonese speakers, given the positive relationship between LPC latency and extrinsic normalization performance at the behavioral level. Cognitive adjustment represents the apex of processing in speech perception, integrating diverse information to form coherent interpretations of speech signals. L2 learners may not have the same level of linguistic knowledge in Cantonese as native speakers, which hinders their ability to access and apply the relevant information effectively. Aside the limitation in L2 linguistic knowledge, the processing effort is another factor to consider. L2 learners may need to invest more cognitive effort in processing individual speech sounds and phonetic details, leaving

fewer cognitive resources available for higher-level cognitive adjustments (Morishima, 2013).

Beyond the less effective cognitive adjustments, the lack of semantic integration may also impair the extrinsic normalization process in Mandarin learners. The Cantonese group displayed a significant speech–nonspeech difference in N400, indicating a higher-order semantic integration (Kutas & Federmeier, 2011; Van Berkum et al., 1999). The amplified N400 in speech contexts implies that native Cantonese speakers tried to integrate the semantic information between perceived words and preceding contexts, regardless of the neutral semantic content of the speech contexts. Conversely, Mandarin learners did not manifest this semantic integration, showing comparable N400s in speech and nonspeech contexts. The top-down semantic information constraints interpretations of instant auditory input (Davis & Johnsrude, 2007). The insignificant context-target semantic integration in Mandarin group could potentially affect the extrinsic normalization accuracy, and this effect might be more prominent in semantically biased contexts.

#### 4.3. L2 immersion and extrinsic normalization: Unlocking enhancement through cognitive adjustment

Prior research has indicated that L2 learners, with extensive L2 immersion, are more adept at employing the extrinsic normalization strategy in L2 speech perception (Antonious et al., 2015; K. Zhang et al., 2024). L2 immersion, however, influences many aspects of L2 speech processing, from sensitivity to L2 speech cues to an advanced grasp of L2 linguistic knowledge, each contributing to the enhancement of the extrinsic normalization process. To delineate the mechanisms underlying these benefits, the present study examined multiple ERPs representing various processing stages in the extrinsic normalization process and explored their relationship with L2 immersion scores with linear regressions. The results revealed that L2 immersion scores significantly influenced the ERP latency of the LPC component. Mandarin learners with more extensive Cantonese immersion exhibited delayed LPC difference waves when perceiving ambiguous Cantonese tones with context cues. Given that the LPC component is associated with cognitive adjustment, and that longer LPC latencies are linked to better extrinsic normalization performance at the behavioral level (as indicated by the regression analysis in Section 3.3), our findings suggest that L2 immersion helps Mandarin learners' cognitive adjustment in normalizing Cantonese tones approach the proficiency level of native Cantonese speakers.

The cognitive adjustment stage, as reflected by LPC, represents the highest-level processing in speech perception. This stage involves integrating multiple sources of information, making contextually appropriate adjustments, and arriving at a final perceptual decision. L2 immersion, with its exposure to real-world language use, provides opportunities to learn how to consider the broader linguistic and situational context when interpreting speech cues. As a result, the influence of L2 immersion may be more pronounced at this speech process stage, leading to a significant correlation between L2 immersion and LPC. Hence, LPC could serve as a neurological indicator that reflects how L2 learners' brains undergo adaptation with increased exposure and training (i.e., the neuroplasticity in L2 learning).

However, there was no significant correlation found between L2 immersion score and N1 and P2. In this study, the N1 difference wave represents the acoustic adjustments in the extrinsic normalization process, involving the initial decoding and processing of

acoustic features. The lack of correlation possibly implies that acoustic adjustments, being highly automatic, might not manifest substantial changes even with increased language exposure. Alternatively, the correlation between L2 immersion and acoustic sensitivity in L2 might not be linear; learners may initially experience improvements in acoustic sensitivity to L2 speech, but such improvements may plateau once a certain proficiency level is attained, requiring additional facilitating factors for further advancements. P2 difference wave in the present study represents phonetic and phonological adjustment in the extrinsic normalization process, which requires a deeper understanding of the L2's sound structure and patterns. The acquisition of L2 phonetic and phonological knowledge might require more focused training or explicit instruction, and thus shows an insignificant correlation with L2 immersion, which is more related to implicit learning.

## 5. Conclusion

The present study investigated the extrinsic normalization process of Cantonese tones in Mandarin learners and native Cantonese speakers. Our findings suggest that L2 learners utilize context cues to normalize L2 speech variability less effectively than native speakers, demonstrating significant adjustments of perceived speech signals primarily at the phonetic/phonological stage. Native speakers exhibited a multi-stage normalization process, invoking P2 and LPC components, reflecting phonetic/phonological and cognitive adjustments in the extrinsic normalization process. The limitations observed in L2 learners are predominantly attributed to the lack of substantial cognitive adjustments in the final decision-making stage. Moreover, the study illustrated the pivotal role of L2 immersion in enhancing the extrinsic normalization process in L2 learners. It was found to significantly bolster high-level cognitive adjustments, indicating the potential of L2 immersion as a strategic tool to refine high-level linguistic knowledge in L2 learners and to bridge perceptual gaps in L2 speech perception.

**Supplementary material.** The supplementary material for this article can be found at <http://doi.org/10.1017/S1366728925100369>.

**Data availability statement.** The data that support the findings of this study are openly available at <https://osf.io/skwnb/>.

**Funding statement.** This study was supported by the National Natural Science Foundation of China (NSFC: 12304526) and the Research Grants Council of Hong Kong (GRF: 15607518).

**Competing interests.** The authors declare no competing interests.

## References

- Ainsworth, W. A. (1975). Intrinsic and extrinsic factors in vowel judgments. In G. Fant & M. A. A. Tatham (Eds.), *Auditory analysis and perception of speech* (pp. 103–113). Academic Press. <https://doi.org/10.1016/j.jml.2016.12.002>.
- Antoniu, M., Wong, P. C. M., & Wang, S. (2015). The effect of intensified language exposure on accommodating talker variability. *Journal of Speech, Language, and Hearing Research*, 58(3), 722–727. [https://doi.org/10.1044/2015\\_JSLHR-S-14-0259](https://doi.org/10.1044/2015_JSLHR-S-14-0259).
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Boersma, P., & Weenink, D. (2023). Praat: Doing phonetics by computer [computer program]. Version 6.3.09, retrieved 2 March 2023 from <http://www.praat.org/>.
- Bosker, H. R., Reinisch, E., & Sjerps, M. J. (2017). Cognitive load makes speech sound fast, but does not modulate acoustic context effects. *Journal of Memory and Language*, 94, 166–176. <https://doi.org/10.1016/j.jml.2016.12.002>.
- Casillas, J. V. (2020). The longitudinal development of fine-phonetic detail: Stop production in a domestic immersion program. *Language Learning*, 70(3), 768–806. <https://doi.org/10.1111/lang.12392>.
- Chen, F., Zhang, K., Guo, Q., & Lv, J. (2023). Development of achieving onstancy in lexical tone identification with contextual cues. *Journal of Speech, Language, and Hearing Research*, 1–17. [https://doi.org/10.1044/2022\\_JSLHR-22-00257](https://doi.org/10.1044/2022_JSLHR-22-00257).
- Crowley, K. E., & Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: Age, sleep and modality. *Clinical Neurophysiology*, 115(4), 732–744. <https://doi.org/10.1016/j.clinph.2003.11.021>.
- Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research*, 229(1–2), 132–147. <https://doi.org/10.1016/j.heares.2007.01.014>.
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>.
- Finnigan, S., Humphreys, M. S., Dennis, S., & Geffen, G. (2002). ERP “old/new” effects: Memory strength and decisional factor(s). *Neuropsychologia*, 40(13), 2288–2304. [https://doi.org/10.1016/S0028-3932\(02\)00113-6](https://doi.org/10.1016/S0028-3932(02)00113-6).
- Francis, A. L., Ciocca, V., Wong, N. K. Y., Leung, W. H. Y., & Chu, P. C. Y. (2006). Extrinsic context affects perceptual normalization of lexical tone. *The Journal of the Acoustical Society of America*, 119(3), 1712–1726. <https://doi.org/10.1121/1.2149768>.
- Godey, B., Schwartz, D., De Graaf, J. B., Chauvel, P., & Liégeois-Chauvel, C. (2001). Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: A comparison of data in the same patients. *Clinical Neurophysiology*, 112(10), 1850–1859. [https://doi.org/10.1016/S1388-2457\(01\)00636-8](https://doi.org/10.1016/S1388-2457(01)00636-8).
- Guediche, S., Holt, L. L., Laurent, P., Lim, S. J., & Fiez, J. A. (2015). Evidence for cerebellar contributions to adaptive plasticity in speech perception. *Cerebral Cortex*, 25(7), 1867–1877. <https://doi.org/10.1093/cercor/bht428>.
- Holt, L. L. (2006). The mean matters: Effects of statistically defined nonspeech spectral distributions on speech categorization. *The Journal of the Acoustical Society of America*, 120(5), 2801–2817. <https://doi.org/10.1121/1.2354071>.
- Holt, L. L., & Lotto, A. J. (2002). Behavioral examinations of the level of auditory processing of speech context effects. *Hearing Research*, 167(1), 156–169. [https://doi.org/10.1016/S0378-5955\(02\)00383-0](https://doi.org/10.1016/S0378-5955(02)00383-0).
- Holt, L. L., Lotto, A. J., & Kluender, K. R. (2001). Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement? *The Journal of the Acoustical Society of America*, 109(2), 764–774. <https://doi.org/10.1121/1.1339825>.
- Huang, J., & Holt, L. L. (2009). General perceptual contributions to lexical tone normalization. *The Journal of the Acoustical Society of America*, 125(6), 3983–3994. <https://doi.org/10.1121/1.3125342>.
- Johnson, K. (2005). Speaker normalization in speech perception. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 363–389). Blackwell Publishing. <https://doi.org/10.1002/9780470757024.ch15>.
- Johnson, K., & Sjerps, M. (2018). Speaker normalization in speech perception. *UC Berkeley Phonetics and Phonology Lab Annual Report*, 2018.
- Jongman, A., & Moore, C. (2000). The role of language experience in speaker and rate normalization processes. In *Proceedings of the sixth International Conference on Spoken Language Processing*, Vol. 1, pp. 62–65.
- Kang, S., Johnson, K., & Finley, G. (2016). Effects of native language on compensation for coarticulation. *Speech Communication*, 77, 84–100. <https://doi.org/10.1016/j.specom.2015.12.005>.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62(1), 621–647. <https://doi.org/10.1146/annurev.psych.093008.131123>.
- Lenth, R. (2019). Emmeans: Estimated marginal means. In *R package version 1.4.2*. <https://cran.r-project.org/package=emmeans>.
- Li, P., Zhang, F., Yu, A., & Zhao, X. (2020). Language history questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism*, 23(5), 938–944. <https://doi.org/10.1017/S1366728918001153>.

- Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, 8(1 APR), 1–14. <https://doi.org/10.3389/fnhum.2014.00213>
- Luo, X., & Ashmore, K. B. (2014). The effect of context duration on mandarin listeners' tone normalization. *The Journal of the Acoustical Society of America*, 136(2), EL109–EL115. <https://doi.org/10.1121/1.4885483>.
- Magnuson, J. S., & Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance*, 33(2), 391–409. <https://doi.org/10.1037/0096-1523.33.2.391>.
- McClelland, J. L., & Elman, J. L. (1986). Exploiting lawful variability in the speech wave. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 360–385). Erlbaum.
- Morishima, Y. (2013). Allocation of limited cognitive resources during text comprehension in a second language. *Discourse Processes*, 50(8), 577–597. <https://doi.org/10.1080/0163853X.2013.846964>.
- Myers, E. B., Blumstein, S. E., Walsh, E., & Eliassen, J. (2009). Inferior frontal regions underlie the perception of phonetic category invariance. *Psychological Science*, 20(7), 895–903. <https://doi.org/10.1111/j.1467-9280.2009.02380.x>.
- Myers, E. B., & Mesite, L. M. (2014). Neural systems underlying perceptual adjustment to non-standard speech tokens. *Journal of Memory and Language*, 76(2), 80–93. <https://doi.org/10.1016/j.jml.2014.06.007>.
- Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology*, 24(4), 375–425.
- Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *The Journal of the Acoustical Society of America*, 85(5), 2088–2113. <https://doi.org/10.1121/1.397861>.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9(1), 97–113. [https://doi.org/10.1016/0028-3932\(71\)90067-4](https://doi.org/10.1016/0028-3932(71)90067-4).
- Peng, G., Zhang, C., Zheng, H., Minett, J. W., & Wang, W. S.-Y. (2012). The effect of intertalker variations on acoustic – perceptual mapping in Cantonese. *Journal of Speech, Language, and Hearing Research*, 55, 579–596. [https://doi.org/10.1044/1092-4388\(2011/11-0025\)](https://doi.org/10.1044/1092-4388(2011/11-0025))
- Persson, A., & Jaeger, T. F. (2023). Evaluating normalization accounts against the dense vowel space of central Swedish. *Frontiers in Psychology*, 14(June). <https://doi.org/10.3389/fpsyg.2023.1165742>
- Sjerps, M. J., Mitterer, H., & McQueen, J. M. (2011). Listening to different speakers: On the time-course of perceptual compensation for vocal-tract characteristics. *Neuropsychologia*, 49(14), 3831–3846. <https://doi.org/10.1016/j.neuropsychologia.2011.09.044>.
- Stilp, C. E. (2020). Evaluating peripheral versus central contributions to spectral context effects in speech perception. *Hearing Research*, 392, 107983. <https://doi.org/10.1016/j.heares.2020.107983>.
- Tamati, T. N., & Pisoni, D. B. (2014). Non-native listeners' recognition of high-variability speech using PRESTO. *Journal of the American Academy of Audiology*, 25(09), 869–892. <https://doi.org/10.3766/jaaa.25.9.9>.
- Tao, R., Zhang, K., & Peng, G. (2021). Music does not facilitate lexical tone normalization: A speech-specific perceptual process. *Frontiers in Psychology*, 12(October), 1–14. <https://doi.org/10.3389/fpsyg.2021.717110>.
- Van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, 11(6), 657–671. <https://doi.org/10.1162/089929999563724>.
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S*, Fourth Edition (Springer).
- Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., Czigler, I., Csépe, V., Ilmoniemi, R. J., & Näätänen, R. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology*, 36(5), 638–642. <https://doi.org/10.1017/S0048577299981908>.
- Wong, P. C. M., & Diehl, R. L. (2003). Perceptual normalization for inter- and intratalker variation in cantonese level tones. *Journal of Speech, Language, and Hearing Research*, 46(2), 413–421. [https://doi.org/10.1044/1092-4388\(2003\)034](https://doi.org/10.1044/1092-4388(2003)034).
- Xie, X., Jaeger, T. F., & Kurumada, C. (2023). What we do (not) know about the mechanisms underlying adaptive speech perception: A computational framework and review. *Cortex*, 166, 377–424. <https://doi.org/10.1016/j.cortex.2023.05.003>.
- Yip, M. (2002). *Tone*. Cambridge : Cambridge University Press.
- Ylinen, S., Uther, M., Latvala, A., Vepsäläinen, S., Iverson, P., Akahane-Yamada, R., & Näätänen, R. (2010). Training the brain to weight speech cues differently: A study of Finnish second-language users of English. *Journal of Cognitive Neuroscience*, 22(6), 1319–1332. <https://doi.org/10.1162/jocn.2009.21272>.
- Zhang, C., Peng, G., & Wang, W. S.-Y. (2012). Unequal effects of speech and nonspeech contexts on the perceptual normalization of Cantonese level tones. *The Journal of the Acoustical Society of America*, 132(2), 1088–1099. <https://doi.org/10.1121/1.4731470>.
- Zhang, C., Peng, G., & Wang, W. S. Y. (2013). Achieving constancy in spoken word identification: Time course of talker normalization. *Brain and Language*, 126(2), 193–202. <https://doi.org/10.1016/j.bandl.2013.05.010>.
- Zhang, C., Peng, G., Wang, X., & Wang, W. S. (2015). Cumulative effects of phonetic context on speech perception. In *The Scottish Consortium for ICPhS 2015* (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow. ISBN 978-0-85261-941-4. Paper number 0085.1-5 retrieved from <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPhS0085.pdf>
- Zhang, K., Li, D., & Peng, G. (2024). Achieving perceptual constancy with context cues in second language speech perception. *Journal of Phonetics*, 103, 101299. <https://doi.org/10.1016/j.wocn.2024.101299>.
- Zhang, K., & Peng, G. (2021). The time course of normalizing speech variability in vowels. *Brain and Language*, 222(July), 105028. <https://doi.org/10.1016/j.bandl.2021.105028>.
- Zhang, K., & Peng, G. (2025). The modulation of cognitive load on speech normalization: A neurophysiological perspective. *Brain and Language*, 266-(April), 105579. <https://doi.org/10.1016/j.bandl.2025.105579>.
- Zhang, K., Tao, R., & Peng, G. (2023). The advantage of the music-enabled brain in accommodating lexical tone variabilities. *Brain and Language*, 247(October), 105348. <https://doi.org/10.1016/j.bandl.2023.105348>
- Zhang, K., Wang, X., & Peng, G. (2017). Normalization of lexical tones and nonlinguistic pitch contours: Implications for speech-specific processing mechanism. *The Journal of the Acoustical Society of America*, 141(1), 38–49. <https://doi.org/10.1121/1.4973414>.
- Zhang, Q., Mou, C., Yang, X., Yang, Y., & Li, L. (2023). The effect of contextual arousal on the integration of emotional words during discourse comprehension. *Quarterly Journal of Experimental Psychology*, 76(4), 850–861. <https://doi.org/10.1177/17470218221098838>.