Original Article

Language *and* Speech

Language *and* Speech
1–31

DOI: 10.1177/00238309221114433
journals.sagepub.com/home/las
SAGE

Production and Perception Evidence of a Merger: [l] and [n] in Fuzhou Min

Ruoqian Cheng, Allard Jongman and Joan A. Sereno
Department of Linguistics, University of Kansas, USA

Abstract

The current study investigated the merger-in-progress between word-initial nasal and lateral consonants in Fuzhou Min, examining the linguistic and social factors that modulate the merger. First, the acoustic cues to the l-n distinction were examined in Fuzhou Min. Acoustic analyses suggested a collapse of phonemic contrast between prescriptive L and N (phonemes in the unmerged system), with none of the six acoustic cues showing any difference across L and N. Linear discriminant analysis did identify acoustically distinct [l] and [n] tokens, although the mapping onto the phonetic space of prescriptive L and N substantially overlapped. Speakers of all ages and both genders tended to produce [l], and low vowels correlated with more [n]-like classification. In perception, AX discrimination data showed Fuzhou Min listeners confused both prescriptive L and N and acoustic [l] and [n]. Greater sensitivity to the acoustic differences occurred in the context of low vowels and a nasal coda, supported by the acoustics of the stimuli, and younger listeners were more sensitive to the difference between [l] and [n] than older listeners. In two-alternative forced choice (2AFC) identification, Fuzhou Min listeners also identified the merged form as L more frequently than N, with more L responses elicited in the context of low vowels and in the absence of nasal codas. Overall, although Fuzhou Min speakers produced some acoustically distinct [l] and [n] tokens in the context of a sound merger, these productions did not map onto prescriptive L and N. In addition, younger listeners were more sensitive to the acoustic distinction than older listeners, suggesting an emerging acoustic contrast possibly arising due to contact with Mandarin.

Keywords
Nasal, lateral, sound merger, speech production, speech perception

---

Corresponding author:
Ruoqian Cheng, Department of Linguistics, University of Kansas, 1541 Lilac Lane, Blake Hall 427, Lawrence, KS 66045, USA.
Email: rqcheng@outlook.com

1 Introduction

When systematic sound contrasts collapse in speech, a phonological merger appears. Diachronically, the collapse of a phonemic distinction can happen over an individual's lifespan or between generations of individuals (Clark et al., 2013). An ongoing phonological merger may also manifest itself in gradience in speech production and perception, with synchronic variations across linguistic and social contexts. The current paper examines the [l~n] merger-in-progress in Fuzhou Min, investigating the linguistic and social contexts that modulate this gradient phonological merger.

1.1 Phonological mergers in speech production and perception

When evaluating a phonological merger, previous researchers have investigated both speech production and perception (Hay et al., 2013; Labov, 1994; Shen, 1990; Warren et al., 2007; Yao & Chang, 2016).

In production, Labov (1994) examined the phonetic space of the two phonemes, with a merger showing an eventual overlap of the two phonemes in phonetic space. Before a complete overlap, the merger-in-progress goes through a *near-merger* stage in which there is a subtle, but statistically significant, acoustic difference between the two phonemic categories, although the speakers no longer perceive the phonetic difference. In New York English, *SAUCE* (/ɔ/) *and SOURCE* (/ɔɹ/) were merging. Labov (1994) presented F1 and F2 measurements that illustrated a substantial difference between /ɔ/ and /ɔɹ/ for some speakers. Hay et al. (2013) suggested that the acoustic difference reflects distinct abstract representations that are pre-lexical. In a speech production experiment, Hay et al. (2013) used nonsense words (e.g., *fod-fawd*) that had identifying rimes similar to the merging real words (e.g., *cod-cawd*) as a way to elicit the abstract representations. In both the American English [ɑ]-[ɔ] (e.g., *LOT/THOUGHT)* merger and New Zealand English [e]-[a] (e.g., *ELLEN/ALLAN*) merger, the nonsense words were phonetically more distinct than the real words due to speakers' knowledge of the distinct abstract categories. In addition, the use of different speech styles can influence speakers' production. When the strings or words occur as a minimal pair, the speakers may choose a more careful style that increases distinctiveness and reduces mergers. Children in New England produced merging low back vowels in spontaneous speech but made a distinction in overt minimal pairs (D. E. Johnson, 2007). Finally, differences across languages may exist in bilingual populations, providing evidence of language-selective merger. Soo et al. (2021) analyzed the acoustic properties of /l/ and /n/ in spontaneous Cantonese and English speech produced by Cantonese-English bilinguals. At the group level, the acoustic values of /l/ and /n/ (including mid-frequency spectral tilt and F2–F1 spacing) were less distinct in Cantonese than in English, which was considered evidence that /l/ and /n/ were merging in their Cantonese while the contrast was preserved in their English.

In perception, data range from fieldwork interviews to controlled laboratory experiments. In fieldwork, for example, linguists present a listener with two minimal pair strings differing in only one sound (e.g., [pat] and [p^{h}at]) and ask the listener to either judge whether the two strings map onto one word or two words or to pronounce the words. In a community with a merger-in-progress, linguists do not usually obtain clear-cut judgments from speakers. In sociophonetic studies, speakers may be presented with two different words written in standard language orthography. They are asked to pronounce the words out loud and report whether the words sound the same. The sociophonetic approach elicits speech production data for acoustic analysis as well as speakers' introspection on the words (Nycz, 2013). Judgment questions elicit explicit knowledge from the speakers, but actual perception is based on both implicit and explicit knowledge. If the merger is not socially conscious, speakers may not be able to identify it. Meanwhile, distinct orthography may also encourage the speakers to perceive the two words differently.

In perception, a wide range of perceptual tasks has been used to quantify the status of a sound merger. Warren et al. (2007) used a two-alternative forced choice (2AFC) identification task, in which New Zealand English speakers listened to distinct [iə] (e.g., *BEER*) and [eə] (e.g., *BEAR*) in the non-merged system. Higher error rates on [eə] items (12%) than [iə] items (4%) suggested a merger toward the [iə] direction. Hay et al. (2013) additionally used nonsense words, but the effect of lexical status was merger-specific. Nonsense words were less accurately identified than real words in the [ɑ]-[ɔ] (e.g., *LOT/THOUGHT*) merger in American English, while the effect of lexical status was not significant for the New Zealand English [e]-[a] (e.g., *ELLEN/ALLAN*) merger in which listeners showed high accuracy for both real and nonsense words. Using phonetically distinct tokens as stimuli may encourage greater decoupling between the two phonemes, which differs from the naturalistic merging context that engages both phonetically distinct and similar tokens.

Auditory stimuli with ambiguous sounds have also been used in perception experiments. K. Johnson and Song (2016) recruited a Nanjing speaker to produce merging [l] and [n] sounds in Nanjing for a similarity rating task. Nanjing listeners provided higher similarity ratings than native English listeners, suggesting a merger in progress. Cheng (2017) synthesized continua for [l]-[n], [a]-[ŋ], and [m̩] and [ŋ̩] in Cantonese with distinct endpoints using STRAIGHT (Kawahara et al., 2008), and asked Cantonese listeners to select the corresponding lexical items in a 2AFC identification task. Cheng (2017) found that listeners responded dominantly with /l/-words to [l] endpoints, and with /n/ words to [n] endpoints. As the continuum shifted from [l] to [n], the proportion of /l/ responses gradually decreased.

Using data from both speech production and perception helps us understand the complex patterns of merger-in-progress. Acoustic analysis of the production data reflects the phonetic space of the phonemes, and perceptual experiments can reveal whether the merging speakers still contrast the sounds and how they identify the merger. In the current study, we will use acoustic analyses and a 2AFC identification task and extend our investigation by examining AX discrimination, directly testing listeners' contrast sensitivity.

1.2 Factors that influence the phonological mergers

Phonological mergers have been shown to be modulated by linguistic and social factors, as the merger first emerges in a specific context and then diffuses to other linguistic and social contexts.

Phonological environments can differentially affect mergers. In the New Zealand English *NEAR/SQUARE* merger, /eə/ raises into /iə/ when preceding a coronal consonant (Warren et al., 2007), and thus the place of articulation of the consonants following the target diphthong influences the degree of the merger. Specifically, the acoustic distance between the two diphthongs was smaller when the preceding consonants were alveolar (i.e., *really/rarely*) than bilabial (i.e., *beer/bare*). Similarly, the merger between [e] and [a] in New Zealand English occurred only for vowels before /l/ (i.e., *shell/shall*), while the distinction was maintained for vowels preceding obstruents (i.e., *said/sad, dress/trap*).

Lexical frequency was also found to influence phonological mergers. The gradience of a phonological merger along lexical items was initially proposed in *merger by transfer* by Labov (1994). When discussing sound changes, Wang (1969) initially postulated that abrupt categorical change does not apply to all lexical entries that contain the target phoneme at the same time. Instead, some lexical items illustrate the merger earlier than other lexical items. Recent research suggests that lexical frequency correlates with the order of application of phonological mergers. Interestingly, the suggested role of lexical frequency has exhibited contradictory data patterns in the literature. In an exemplar-based model, Todd et al. (2019) found that in phonological mergers, high-frequency words change at a faster rate than low-frequency words. However, in a study of the Shanghainese /ã/ and /ɑ̃/ merger, high-frequency words entered the merging process later than low-frequency words (Shen, 1990).

In addition to linguistic and lexical factors, social factors including age, gender, and socioeconomic status have also been shown to affect phonological mergers (Labov, 2001). Labov (1994) showed that when a merger is ongoing, younger speakers are more likely to use the merged forms than older speakers because younger speakers are more advanced in the sound change. In the current paper, we elaborate on age and gender. Instead of conducting a longitudinal study to capture the dynamic sound change of a merger over time, we recruit participants from different age groups to observe the synchronic variations in *apparent time.*

Interestingly, mergers can reverse due to contact with another language that maintains the contrast between the original phonemes (Yao & Chang, 2016). In Shanghainese, high-mid /e/ and low-mid /ɛ/ have merged into /ɛ/. Despite merged pronunciation in Shanghainese, these words map onto distinct pronunciations in Mandarin, including /aj/, /an/, and /ej/. Older (mean age = 59.2 years) and younger (mean age = 29.8 years) Shanghainese-Mandarin bilingual speakers participated in a production experiment in which they produced Shanghainese words with the merged targets. Acoustic analyses of the Shanghainese production data showed that younger Shanghainese speakers showed greater distinctiveness across the items (i.e., items that mapped onto different phonetic representations in Mandarin). In addition, younger speakers' Shanghainese production of the word with /ej/ sounds in Mandarin had shifted toward [i] as compared to the Shanghainese production of older speakers. It was concluded that exposure to Mandarin made younger Shanghainese speakers reverse the merger, although they were meant to be the leaders of the merger if the sound change only happened within the Shanghainese system. A similar reversal was discovered when Kang and Nagy (2016) compared the VOT merger across Seoul Korean speakers and Toronto Korean speakers who were born in the 1940s, 1960s, and 1980s. In Seoul Korean, the younger the speakers, the smaller the VOT difference observed between lenis and aspirated stops and the more speakers relied on F0 as the primary cue. In Toronto Korean, the VOT difference decreased from the 1940s group to the 1960s groups, but it increased from the 1960s group to the 1980s group. The VOT merger is internal to Korean and the Seoul and Toronto Korean speakers differ in their degree of exposure to English. The researchers attributed the reversal of the VOT merger in younger Toronto Korean speakers to contact with English because English utilizes VOT as the primary cue to distinguish voiced and voiceless stops.

Gender has also been shown to play a role in the process of sound change (Labov, 2001). Specifically, in a sound change connected to social stereotypes, females tend to be more conservative in speech than males as a reflection of sociological insecurity triggered by socioeconomic weakness. Hence, females tend to use prestige variants more than males as a source of symbolic power to balance their lower economic power (Wolfram and Schilling-Estes, 2006). For instance, in Philadelphia English, women used the prestigious *-ing* form more often than men and the stigmatized *-in* form less often (Labov, 2001). Similar to lexical frequency, contrary data patterns about sound change and gender have been observed. Chambers (1995) observed that, if a sound change does not relate to social stereotypes, females are likely to use more innovative variants, due to their advantage of using a larger repertoire of variants and commanding more styles than males. The /ʒ/ devoicing in Buenos Aires (Wolf & Jiménez, 1979) showed no difference across formal and casual styles and hence did not link to social stereotypes, yet females were leaders in this sound change. Specifically, females used more devoiced items than males, especially among speakers who were younger than 55 years old.

Labov (2001) highlighted the interaction between gender and age of the speakers in the context of sound change (pp. 294–322). In the /ʒ/ devoicing case, while the overall trend showed that younger listeners devoiced /ʒ/ more frequently than older listeners (Wolf & Jiménez, 1979), the difference between females and males varied across different age groups. While both males and females older than 55 had similar devoicing rates, the gender effect increased among speakers aged between 26 and 55 and further increased among teenagers. Moreover, females were found to be one generation faster than males. In the /ʒ/ devoicing case, females in the 36 to 55 age group and males around 15 years old had a similar devoicing rate (approximately 20%) and females at 15 years old had a similar devoicing rate to males at 9 years old (around 65%). This gender difference was also found in Seoul Korean /o/ raising (Chae, 1995), in which the gender difference was larger in the second generation, relative to the first and third generations. In Labov's (2001) scheme, the lag between males and females will be the most salient at the temporal midpoint of the sound change, and be reduced at the beginning of the merger, or when the merger nears completion.

Interactions can also appear between linguistic factors and social factors. When revisiting the department store /ɹ/ study in New York City (Labov, 1966/2006, 1972). Mather (2012) observed that across all the speech styles and social stratifications, /ɹ/ consistently occurred less in the preobstruent context (e.g., *fourth*) than in the word-final context (e.g., *floor*). However, the increase in /ɹ/ usage from word-final to preobstruent position manifested differently across speech styles and social stratification. The emphatic style elicited a larger increase in /ɹ/ usage than the casual style. Meanwhile, the increase was larger in a more prestigious store than a less prestigious store. The interplay between linguistic and social factors was also identified in an /iː/-/ei/ merger in Swabian (Beaman & Tomaschek, 2019). For high-frequency words, speakers with a low sense of social identity reduced the phonetic distance between the phonemes more than speakers with a high sense of social identity. The speakers with a low sense of identity desired to mimic the standard form of the dialect, so they produced the merger with a greater reduction of phonetic distance over time than the speakers with a high sense of identity.

These data demonstrate that both linguistic and social factors affect phonological mergers. The current study will examine a phonological merger in progress in a Chinese language, Fuzhou Min, taking into account linguistic context, age, and gender. In the next section, we focus on previous research on the l-n merger in Chinese.

1.3 The merger between [l] and [n] in Chinese languages

The contrast between word-initial lateral [l] and nasal [n] consonants is disappearing in many Chinese languages (Cheng, 2017; K. Johnson & Song, 2016; Shi, 2015; Shi & Zhang, 2017; Shi & Liang, 2017; Shi & Xiang, 2010; Tian, 2009; Zhang & Levis, 2021). Previous studies have made claims about the [l~n] merger based on impressionistic transcription, measurement of nasal and oral air flow, and perceptual experiments.

Based on the impressionistic documentation of 40 varieties of 10 Chinese languages, Tian (2009) summarized two effects of the phonological environment on the merger. First, the [l~n] merger is more likely to occur in low-vowel contexts (i.e., [la]-[na]) than high-vowel contexts (i.e., [lu]-[nu]). Second, the merger co-occurs more frequently with rimes with a nasal coda or a nasalized vowel than rimes without a nasal coda and usually tends toward [n] in these contexts.

The effects of vowel height and nasal coda can be manifested in nasalance data. Reflecting the degree of velopharyngeal opening, nasalance is expressed as the ratio of nasal flow to overall airflow in the nasal and oral cavities (Sebastian et al., 2015).

Studies of Mandarin and Tianjin, two Chinese languages with the [l] versus [n] contrast, consistently found that the nasality of [n] is above 90%, and the nasality of [l] is between 20% and 40%

(Shi et al., 2010). However, for Chinese languages with the [l~n] merger, nasalance data revealed three distinct patterns of the [l~n] merger. In Chengdu, both variants showed more nasal airflow than a typical lateral variant, with nasalance of 93.5% and 61%, and there was no effect of vowel quality (Shi, 2015). In Nanjing, Shi and Liang (2017) found two variants, a variant with nasal airflow closer to a nasal variant preceding /i/ and /y/, and a lateral variant in other vowel contexts. In Wuhan, Shi and Xiang (2010) identified three variants, including a strong nasal consonant, an intermediate variant that was acoustically between [l] and [n], and a lateral consonant.

In perception, K. Johnson and Song (2016) tested the Nanjing [l~n] merger with a similarity rating task. Younger Nanjing listeners (<45 years), older Nanjing listeners (>45 years), and native American English listeners listened to Nanjing stimuli produced by a Nanjing speaker. American English listeners rated [l] and [n] as more different (average = 3.18; on a scale from 1 as similar to 5 as different) than both the younger (average = 2.21) and the older Nanjing listeners (average = 1.47). The results provided perceptual evidence for the merger of [l~n] in Nanjing. Older Nanjing listeners' lower difference rating was explained as a result of reduced exposure to Mandarin which has a l-n contrast. National mandatory Mandarin education started in the 1980s in China, so older speakers were not required to learn Mandarin in primary school and therefore have less exposure to Mandarin than younger speakers.

A similar age effect also occurred among Cantonese-English bilinguals. In Cantonese, recent transcription data showed that more than 90% of the participants replaced [n] by [l], showing a merger with most identifications tending toward [l] (To et al., 2015). Moreover, in both Hong Kong and Vancouver, younger Cantonese listeners tended to perceive a synthesized [l~n] continuum more categorically than older Cantonese listeners (Cheng, 2017), due to more exposure to English, a language with the [l] versus [n] contrast.

Gender also influenced the production of the [l~n] merger by Cantonese speakers from Hong Kong and Vancouver (Cheng et al., 2019). The transcribed production data (transcribed by a native Cantonese speaker) showed that bilingual Cantonese speakers pronounced prescriptive nasal onsets as [l]. Averaging across age and location, females used the innovative merged form [l] more frequently than males. The older group in Vancouver also showed the most salient gender difference, with females using the [l] variant much more frequently than males.

In summary, while some Chinese languages retain the [l~n] contrast, most show a merging of the distinction. Two linguistic factors (vowel height and nasal coda) and two social factors (age and gender) have been found to modulate the [l~n] merger in Chinese languages. Vowel height as well as rimes with a nasal coda influences the direction of the merger. And older speakers and female speakers showed a greater prevalence of the merger than younger male speakers.

The current study systematically investigates the production and perception of the [l~n] merger in a single Chinese language, Fuzhou Min. While Fuzhou Min has been claimed to exhibit an ongoing merger, empirical evidence has yet to be provided. Previous studies on the production and perception of the [l~n] merger have shown that different methods and tasks can tap into different aspects of the knowledge that speakers have about their language. The current study will complement previous research on the [l~n] merger by providing both acoustic and perceptual data from a single language.

1.4 The present study: the [l~n] merger in Fuzhou Min

Fuzhou Min is a variety of the Min language family, spoken in the eastern region of Fujian province in China. The syllable structure in Fuzhou Min is (C)(G)V(V)(C), including consonants, glides, and vowels. The segments in parentheses are optional. While all three nasal consonants /m, n, ŋ/ can occur in the onset position, only /n/ can occur in the coda position. There is one lateral consonant, /l/, and it can occur in onset position but not in coda position. In the city of Fuzhou, Fuzhou Min is the only local language, but most speakers are also taught Mandarin. In Chinese, monosyllabic words are written as Chinese characters, *hanzi* 汉字. In Middle Chinese, words with [l] and [n] onsets were phonologically distinct and represented by distinct Chinese characters. For example, "南" codes a syllable with the [n] onset (i.e., [nan]), and "蓝" represents a syllable with the [l] onset (i.e., [lan]) in Middle Chinese. Mandarin and Fuzhou Min, as modern varieties of Chinese, both derived from Middle Chinese and inherited *hanzi*, which nowadays maps onto the same morphemes in both languages. These morphemes are pronounced differently in Mandarin and Fuzhou Min. In Mandarin, the word-initial [l] and [n] are still distinct in both production and perception. It should be noted that, due to exposure to Mandarin, when Fuzhou Min speakers read *hanzi*, they can be aware of the prescriptive pronunciation of the word based on the words' Mandarin sounds. However, impressionistic analysis (Chen, 1998; Tao, 1956) suggests that Fuzhou Min speakers do not produce distinct word-initial [l] and [n] when speaking Fuzhou Min. In other words, Fuzhou has been claimed to have merged the word-initial [l] and [n] but the writing system retains the distinction which is also phonemically and phonetically distinct in Mandarin. In Mandarin, "南" is pronounced as [nan], and "蓝" is pronounced as [lan]. The claimed merger in Fuzhou Min has never been supported by quantitative data in either production or perception, and thus the goal of the current study is to provide data about the [l~n] merger in Fuzhou Min.

In the current study, there are three research questions:

> RQ1: To what extent is there evidence of an [l~n] merger in Fuzhou Min production? What are the acoustic cues to [l] and [n] in Fuzhou Min?
>
> RQ2: To what extent is there evidence of an [l~n] merger in perception in Fuzhou Min? Can Fuzhou Min listeners perceptually differentiate between [l] and [n]?

For clarity, henceforth we will use L and N as the prescriptive labels of the phonemes that map to an unmerged system, a distinction that is maintained in the *hanzi* writing system. Moreover, we use [l] and [n] as the phonetic labels that represent the phonetic realizations by Fuzhou Min speakers.

In production, we initially identify the acoustic correlates of the contrast between [l] and [n] in English and Mandarin. These acoustic cues are then examined in Fuzhou Min L and N phonemes to test whether the phonemes are merging. In perception, an AX discrimination task investigates whether Fuzhou Min listeners still perceive the difference between prescriptive L and N, as well as the difference between phonetic [l] and [n]. Finally, a 2AFC identification task investigates how Fuzhou Min listeners identify the forms involved in the [l~n] merger.

If the merger occurs in production, we predict the acoustic correlates of prescriptive L and N to be similar. At the phonetic level, if the merger tends toward [l], the [l] variant will occur more frequently than the [n] variant, and the overall acoustic properties of the productions will be more [l]-like. If the merger tends toward [n], the acoustic properties may tend toward [n]. If the merger results in an intermediate outcome, [l] and [n] will be produced at a similar frequency and the acoustic properties may be ambiguous between [l] and [n].

In perception, if the merger occurs, low discrimination accuracy is expected for prescriptive L and N labels. In addition, Fuzhou Min listeners are also predicted to not perceive the difference between phonetic [l] and [n]. If the merger is not complete or contains a new emerging contrast, Fuzhou Min listeners may still perceive the difference between phonetic [l] and [n], even if they confuse prescriptive L and N. In identification, if the merger tends toward the lateral, we predict more L responses and vice versa for nasal-dominant identifications. If the merger ends up with an intermediate form (two possible scenarios: [l] and [n] occur interchangeably; or the outcome variant is a form acoustically between [l] and [n]), we predict a similar proportion of L and N responses.

RQ3: How do linguistic and social factors modulate the merger?

In both production and perception, two linguistic factors, vowel height and nasal coda, and two social factors, age and gender, are examined. In production, based on previous studies of other Chinese dialects, we expect a higher degree of the merger in the context of low vowels and vowels with a nasal coda because of higher acoustic nasality. In perception, we expect a higher proportion of N responses in identification in the context of low vowels and vowels with a nasal coda, which have more acoustic nasality.

For social factors, older speakers are predicted to show higher degrees of the merger than younger speakers in both production and perception. Given sociolinguistic research that suggests that females are expected to use the innovative [l] forms more frequently than males (Cheng et al., 2019), we predict that females may tend toward the lateral direction more than males.

2 Study 1: production

The aim of Study 1 is to identify the acoustic correlates of [l] and [n] in two languages that have this contrast and to then explore the presence of these correlates in Fuzhou Min to assess the status of the [l~n] merger. Both acoustic analysis and linear discriminant analysis (LDA) were used.

2.1 Acoustic analysis

In the acoustic analysis, we first identified the acoustic correlates that signaled the contrast between [l] and [n] in English and Mandarin and then applied these correlates to test the production of prescriptive L and N in Fuzhou Min. Acoustic cues have been previously observed in at least English for quantifying the nature of nasal and lateral sounds, including F2–F1 spacing (Chiba & Kaijiyama, 1941), F3 frequency (K. Johnson, 2012), relative RMS (root-mean-square) amplitude (Glass, 1984; Miyawaki et al., 1975; Rastatter, 1984), Δ A1 (House & Stevens, 1956; Prahler, 1998), bandwidth of F1 (BW1; House & Stevens, 1956), and A1–P0 (Chen, 1995, 1997) In Fuzhou Min, if the contrast between prescriptive L and N still exists, these acoustic parameters will show a significant difference between prescriptive L and N. If the contrast has collapsed, we expect no significant difference between the acoustic parameters of L and N.

2.1.1 Speakers. Six native English speakers (aged 18–34; 3 females, and 3 males) were recruited from Lawrence, Kansas. Twelve Mandarin speakers (aged 28–62; 6 females and 6 males) were recorded in Lawrence, Kansas, and Beijing, China. Six Mandarin speakers were recruited in Lawrence, and the other six Mandarin speakers were recruited in Beijing. Mandarin speakers speak the standard variety of Mandarin (Hou, 1986) as their native language. Twelve native Fuzhou Min speakers (aged 29–56; 6 females and 6 males) were recorded in a soundproof booth in Fuzhou, China. They were born and brought up in Fuzhou, China. The Fuzhou Min speakers also know Mandarin, the nation-wide official language. All speakers provided written informed consent before the start of the recording.

2.1.2 Materials. The English word list consisted of 9 [l] versus [n] minimal pairs and near-minimal pairs (e.g., [nɛt] vs. [lɛt], [nʌt] vs. [lʌk]) with 7 vowel contexts (/i/, /ɛ/, /æ/, /ɑ/, /ʌ/, /ʊ/, /u/).

For Mandarin, words appearing in everyday life were used. The Mandarin word list contained 7 [l] versus [n] minimal pairs (e.g., [lu] vs. [nu]) in 4 vowel contexts (/i/, /a/, /o/, /u/).

The Fuzhou Min word list included 8 minimal pairs in prescriptive L versus N minimal pairs (e.g., Lu vs. Nu) in 5 vowel contexts (/i/, /ø/, /a/, /ɔ/, /u/).

For Mandarin and Fuzhou Min, the lexical tone was identical within each minimal pair such that the only difference within each pair is the onset consonant. Full word lists (with tone information for Mandarin and Fuzhou Min) are presented in Tables 3 to 5, respectively, in Appendix 1.

2.1.3 Recording procedure. Speakers produced multiple repetitions of the words in isolation in a randomized order, and they took a short break between repetitions. English speakers produced two repetitions in an anechoic chamber and were recorded with an Electro-Voice 767 microphone and a solid-state recorder (Marantz PMD671) at the University of Kansas. Mandarin speakers produced three repetitions. Six Mandarin speakers were recorded in an anechoic chamber at the University of Kansas, and the other six speakers were recorded in a recording booth with a SONY ECM-44B microphone and a Sound Blaster X-Fi Surround 5.1 Pro sound card connected to a PC laptop at Peking University in Beijing. Fuzhou Min speakers produced three repetitions and were recorded in a radio studio with an AKG C544L microphone and an Avid ProTools Mbox Mini sound card connected to a PC laptop in Fuzhou. English and Mandarin were digitally recorded at a sampling rate of 22,050 Hz, and Fuzhou Min was digitally recorded at a sampling rate of 44,100 Hz. All recordings were made with a bit depth of 16. For the two Fuzhou speakers who reported difficulties reading some of the Chinese characters in the Fuzhou Min pronunciation, the first author explained the words to them in Fuzhou Min, without producing any instances of [l] or [n]. The experimenter communicated with each group of speakers only in their native language.

In English, we collected 162 [l] tokens and 162 [n] tokens. In Mandarin, there were 224 [l] tokens and 224 [n] tokens. In Fuzhou Min, we had 288 prescriptive L tokens and 288 prescriptive N tokens.

2.1.4 Acoustic correlates. Audio recordings were manually segmented by the first author in Praat (Boersma & Weenink, 2019). The onset of each target word was marked as the onset of the word-initial consonant, either nasal or lateral. The onset of the F1 transition into the following vowel was considered as the offset of the word-initial consonant. We measured six acoustic correlates based on the acoustic properties of [l] and [n] as documented in previous studies. Figure 1 presents a visual display of the acoustic segmentation.

F2–F1 spacing and F3 frequency. We extracted F1, F2, and F3 at the midpoint of the word-initial consonants ([l] or [n]). F2–F1 spacing is the frequency difference between F2 and F1. For [l], both F2 and F1 are low, while F3 is high (Prahler, 1998; Stevens, 1998). Based on perturbation theory, the complete anterior constriction of [n] causes F1 to slightly decrease and F2 to increase, relative to [ə], resulting in greater F2–F1 spacing (Chiba & Kajiyama, 1941). For F3, comparing [l] and [n], as the length of the primary resonance tube (L) of [l] is shorter than that of [n], the formula $F3 = 3(v/4L)$ predicts that [l] has a higher F3 than [n] when the volume of the resonance tube (v) is constant (K. Johnson, 2012). Indeed, Tabain et al. (2016a, 2016b) reported higher F3 and smaller F2–F1 spacing for [l] ($F3 = 2839$ Hz, $F2–F1 = 1245$ Hz) than [n] ($F3 = 2761$ Hz, $F2–F1 = 1361$ Hz) across three Central Australian languages. Hence, we predict that [l] has a higher F3 frequency than [n] and that large F2–F1 spacing may be a feature of [n] in English and Mandarin. In Fuzhou Min, if the [l]-[n] contrast undergoes a merger, then no difference in either F2–F1 or F3 will be observed.

Relative RMS (root-mean-square) amplitude. We extracted the RMS amplitude across the entire initial consonant and the entire vowel, respectively, with the Praat function "Get root-mean-square"

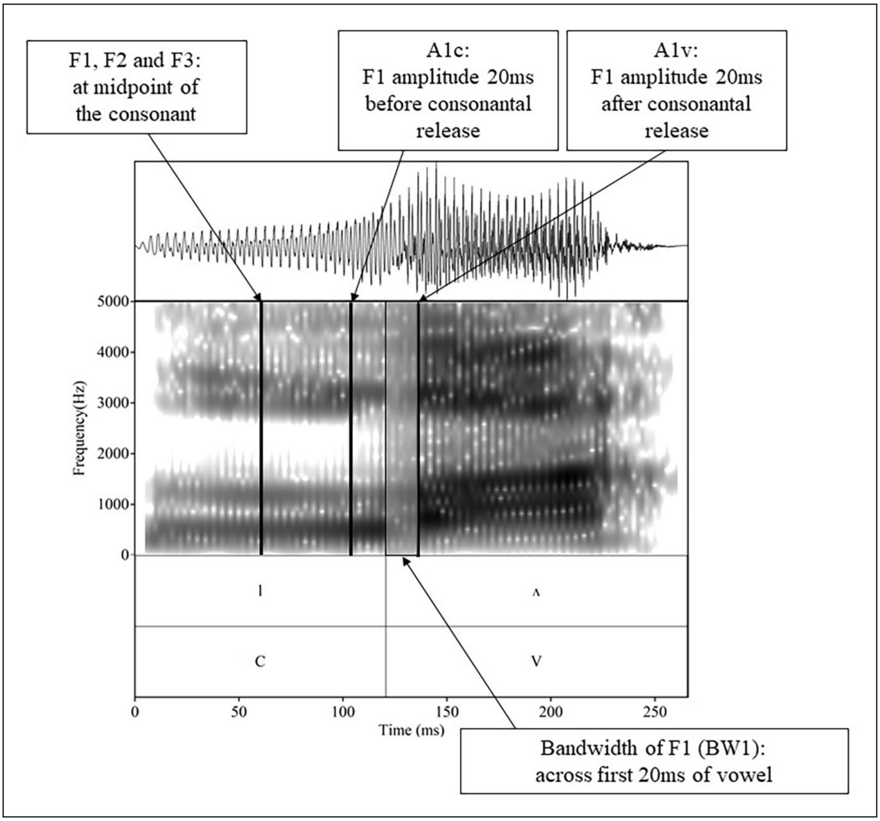

Figure 1. Waveform and spectrogram of [lʌ] in English "luck."

(Boersma & Weenink, 2019). The relative amplitude of the initial consonant was defined as the RMS amplitude of the vowel following the consonant minus the RMS amplitude of the initial consonant. For both [l] and [n], an overall decrease in the amplitude of the signal results from the addition of a side branch. The relative amplitude between vowels and consonants was initially proposed for its potential to distinguish different categories of sounds in speech recognition, especially those sounds that share similar acoustic features (Glass, 1984). Although previous research did not directly compare the relative amplitudes of [l] and [n], previous studies suggest that [n] may have a smaller relative amplitude than [l]. Specifically, the relative amplitude of [n] was reported to be 10 dB in real words across syllabic and phonetic contexts (Glass, 1984) and 5 dB in [VnV] nonwords (Rastatter, 1984). The difference between words and nonwords might be due to how they were elicited: a spontaneous speech corpus (Glass, 1984) versus controlled nonwords in a carrier sentence (Rastatter, 1984). While there are no published relative amplitude data for [l], previous speech synthesis studies made the onset of the consonant 15 dB weaker than the onset of the following vowel throughout all steps of a [lV]-[ɹV] continuum (Miyawaki et al., 1975). Thus, we predict that [l] has a larger relative RMS amplitude than [n] in English and Mandarin. In Fuzhou Min, if the [l]-[n] contrast undergoes a merger, then no difference in relative RMS amplitude will be observed.

Δ *A1.* Delta A1 (Δ A1) is defined as the F1 amplitude of the vowel (A1v) minus the F1 amplitude of the consonant (A1c). We extracted A1v 20 ms after consonantal release and A1c 20 ms before consonantal release (Prahler, 1998) with VoiceSauce (Shue et al., 2011). This cue captures the change in F1 amplitude during the consonant-vowel transition for [lV] and [nV] sequences. For [lV] sequences, the resonance of lateral consonants involves a Helmholtz resonator that decreases the energy of A1c (Prahler, 1998). Since the vowel following a lateral consonant does not have much or any nasalization, we predict a larger Δ A1 (=larger A1v—smaller A1c). For [n], a high A1c is expected since the energy of nasal consonants is concentrated in the low-frequency domain. In nasalized vowels, A1v drops due to nasal tract coupling (House & Stevens, 1956). Hence, we expect a smaller Δ A1 (=smaller A1v—larger A1c) in [nV] sequences. Overall, we predict that [lV] sequences have a larger Δ A1 than [nV] sequences in English and Mandarin. In Fuzhou Min, if the [l]-[n] contrast undergoes a merger, then no difference in Δ A1 will be observed.

BW1. We extracted the bandwidth of F1 (BW1) across the first 20 ms of the vowel segment with VoiceSauce (Shue et al., 2011). The bandwidth of F1 (BW1) increases when a vowel is nasalized due to nasal tract coupling (House & Stevens, 1956). Vowels in [nV] sequences are likely to be more nasalized than vowels in [lV] sequences. In English and Mandarin, we hypothesize that vowels have a larger BW1 in [nV] than [lV] sequences, especially at the beginning of the vowel. In Fuzhou Min, if the [l]-[n] contrast undergoes a merger, then no difference in BW1 will be observed.

A1–P0. Following Styler (2017), we used A1–P0 as an index of nasalization. A1 is the amplitude of the first harmonic and P0 is the strongest harmonic below F1. We extracted A1–P0 from a pulse at the first 10% time point of the vowel segment that follows the onset consonant, using a Praat script (Styler, 2017). First, this pulse was copied many times to reach a duration of 500 ms to ensure stable acoustic analysis. Next, the script picked A1 as the harmonic closest to the estimated F1, and identified P0 as the strongest harmonic below A1, according to the definitions of A1 and P0 in Chen (1997). In nasalized vowels, nasal tract coupling causes a decrease in A1 (House & Stevens, 1956). P0 increases due to paranasal sinus cavity coupling (Maeda, 1982). Chen (1995, 1997) found that in English, A1–P0 was significantly lower for nasalized vowels than for oral vowels, and the difference was more consistent for non-high vowels than high vowels. A1–P0 served as an important cue in speech recognition models for vowel nasality (Pruthi, 2007; Styler, 2017). In Mandarin and English, we hypothesize that vowels have a lower A1–P0 (more nasality) in [nV] than [lV] sequences. In Fuzhou Min, if the [l]-[n] contrast undergoes a merger, then no difference in A1–P0 will be observed.

2.1.5 Results. Mixed-effects linear regression models were fitted for each measurement to examine which cues served to distinguish [l] from [n] in English and Mandarin, and how these cues patterned in Fuzhou Min. The dependent variable was the measured value of the acoustic parameter, and the fixed effect was the word-initial consonant ([l] vs. [n]). We used the lme4 package (Bates et al., 2015) for running the model, and the lmerTest package for generating *p* values (Kuznetsova et al., 2017) in R (R Core Team, 2020). The model also accounted for the varying trends across participants, as shown in the formula below:

$$\text{Acoustic measurement} \sim \text{onset} + \left(\text{onset} \,|\, \text{participant}\right)$$

Table 1 summarizes the mean measurements of each acoustic correlate. For all the acoustic measures examined, each of the six acoustic correlates showed some significant difference between [l] and [n] in English and Mandarin. None of the acoustic measures showed any significant difference between [l] and [n] in Fuzhou Min.

F2–F1 spacing and F3 frequency. In English, [l] had significantly smaller F2–F1 spacing ($\beta = 549.43$, $SE = 106.91$, $t = 5.14$, $p = .004$) and significantly higher F3 frequency ($\beta = -0.57$, $SE = 84.32$, $t = -44.86$, $p < .001$) than [n], consistent with our predictions. In Mandarin, [l] and [n]

Table 1. Comparison of the Acoustic Correlates of Nasal and Lateral Consonants in English, Mandarin, and Fuzhou Min.

Acoustic Parameter		English		Mandarin		Fuzhou Min	
		[l]	[n]	[l]	[n]	L	N
F2–F1 spacing (Hz)	*M*	748	1,297	1,061	1,108	1,238	1,259
	SD	262	290	302	403	253	304
F3 frequency (Hz)	*M*	2,886	2,628	2,710	2,623	2,651	2,669
	SD	296	229	431	279	222	201
Relative RMS amplitude (dB)	*M*	7.65	6.82	7.84	6.33	2.57	2.40
	SD	3.05	3.08	3.61	4.17	4.13	4.24
Δ A1 (dB)	*M*	–0.13	–4.89	2.69	–0.86	0.55	0.20
	SD	4.98	6.78	7.42	9.36	11.59	11.47
BW1 (Hz)	*M*	146	225	114	165	177	182
	SD	89	108	114	165	108	109
A1–P0 (dB)	*M*	9.38	3.39	9.89	2.98	2.82	2.25
	SD	4.72	4.01	5.34	5.36	4.82	4.83

Note. SD: standard deviation; RMS: root mean square; BW: bandwidth.

did not significantly differ regarding F2–F1 spacing (β=52.85, *SE*=81.46, *t*=0.65, *p*=.53) or F3 frequency (β=−97.64, *SE*=71.75, *t*=−1.36, *p*=.20). In Fuzhou Min, there was no significant difference between L and N for either F2–F1 spacing (β=20.27, *SE*=39.08, *t*=0.52, *p*=.61) or F3 frequency (β=18.08, *SE*=23.95, *t*=0.76, *p*=.61).

The different pattern of F2–F1 spacing between English and Mandarin may result from the fact that English [l] has coronal and velarized allophones, while Mandarin [l] has only the coronal realization. Velarized [l] has smaller F2–F1 spacing (Sproat & Fujimura, 1993), and can appear in syllable-initial position (De Jong et al., 2019). Slight velarization in English [l] may contribute to the smaller F2–F1 spacing in English (748 Hz) than in Mandarin (1,061 Hz).

The different pattern for F3 frequency between English and Mandarin may be due to the different coarticulatory effects of the following vowel. Rounding serves as a contrastive feature for Mandarin vowels, but not for English vowels. Mandarin [l] had a significantly lower F3 when followed by rounded vowels (2,535 Hz) than unrounded vowels (2,842 Hz), while no such difference was observed between rounded vowels (2,859 Hz) and unrounded vowels (2,940 Hz) in English. F3 frequency in Mandarin [l] may be lower due to stronger coarticulatory effects (e.g., vowel rounding lowered F3 for [l] in Catalan dialects in Recasens & Espinosa, 2005) in Mandarin and more rounded vowels included in Mandarin (3 out of 7 minimal pairs) than in English (2 out of 9 minimal pairs) in the current project. When limiting our comparison to only unrounded vowels, a significant F3 difference between [l] and [n] was observed in both English, $F3_l$=2,861 Hz, F3n=2,636 Hz, *t*(41)=5.29, *p*<.001, and Mandarin, $F3_l$=2,842 Hz, F3n=2,630 Hz, *t*(47)=2.68, *p*=.01. In Fuzhou, there was no significant difference between L and N for either F2–F1 spacing or F3 frequency.

Relative RMS (root-mean-square) amplitude. In Mandarin, [l] had a significantly larger relative RMS amplitude than [n] (β=−1.38, *SE*=0.63, *t*=−2.22, *p*=.05). The pattern in English was in the same direction but not statistically significant (β=−0.73, *SE*=0.81, *t*=−0.91, *p*=.41). The pattern in Mandarin aligned with our prediction.

In Fuzhou, there was no significant difference in relative RMS amplitude between the prescriptive L and *N* (β=−0.17, *SE*=0.28, *t*=−0.61, *p*=.55).

Δ A1. Word-initial [l] showed significantly greater[1] Δ A1 than word-initial [n] in both English ($\beta=-5.14$, $SE=1.87$, $t=-2.75$, $p=.04$) and Mandarin ($\beta=-3.69$, $SE=1.30$, $t=-2.84$, $p=.02$), corresponding to our prediction. The Δ A1 parameter was considered as a distinguishing cue to the [l] versus [n] contrast.

In Fuzhou, there was no significant difference in Δ A1 between prescriptive L and N ($\beta=-0.31$, $SE=1.02$, $t=-0.31$, $p=.77$).

BW1. BW1 of vowels following [n] was significantly larger than that of vowels after [l] in English ($\beta=89.69$, $SE=16.64$, $t=5.09$, $p=.004$) and Mandarin ($\beta=51.09$, $SE=9.15$, $t=5.58$, $p<.001$), as predicted. Thus, BW1 was a distinguishing cue as well.

In Fuzhou, there was no significant difference in BW1 between prescriptive L and N ($\beta=4.67$, $SE=0.44$, $t=-2.01$, $p=.66$).

A1–P0. A1–P0 was significantly lower for vowels following [n] than for vowels after [l] in both English ($\beta=-6.57$, $SE=0.56$, $t=-11.83$, $p<.001$) and Mandarin ($\beta=-6.25$, $SE=0.57$, $t=-11.00$, $p<.001$), aligning with our prediction and serving as a distinguishing cue.

In Fuzhou, there was no significant difference in A1–P0 between prescriptive L and N ($\beta=-0.89$, $SE=0.44$, $t=-2.01$, $p=.07$).

In conclusion, the acoustic analyses identified three distinguishing cues that worked for both English and Mandarin (Δ A1, BW1, and A1–P0), and three cues that worked for one of the two languages, including F2–F1 spacing (English), F3 frequency (English), and relative RMS amplitude (Mandarin). Crucially, in Fuzhou, none of the six acoustic correlates showed a significant difference between prescriptive L and N, suggesting the collapse of the contrast between the originally distinct phonemes.

2.2 Linear discriminant analysis

Having established the collapse of the phonemic contrast between prescriptive L and N in Fuzhou, we conducted a linear discriminant analysis (LDA) to test whether Fuzhou Min speakers were producing acoustically distinct [l] and [n] tokens. Classifiers trained with English and Mandarin acoustic measurements were applied to categorize the Fuzhou Min word-initial consonants.

2.2.1 Analysis procedure. LDA was performed with the six acoustic correlates described in Section 2.1.4. The values of each measurement were transformed into z-scores within each speaker. Discriminant analysis was performed via SPSS version 26 (IBM Corp, 2019), with the stepwise method and leave-one-out cross-validation. Since we found that English and Mandarin rely on slightly different acoustic parameters to signal the [l]-[n] contrast, we trained two classifiers, one with English data and one with Mandarin data, respectively. The accuracies for the English and Mandarin classifiers were similar (see Results below).

2.2.2 Results. The classifiers showed good to very good accuracy in classifying English (95.8%) and Mandarin (73.3%) [l] and [n]. When categorizing Fuzhou Min tokens, the English classifier showed an accuracy of 54.7% and the Mandarin classifier reached an accuracy of 53.7%. As accuracy for Fuzhou Min was based on prescriptive labels, the low accuracy in Fuzhou Min again suggests the collapse of the contrast between prescriptive L and N.

The LDA predictions from the Mandarin classifier were modeled to test the effects of linguistic and social factors. The Mandarin classifier was selected over the English classifier, since Mandarin shares a similar syllable structure and syllable inventory with Fuzhou Min. LDA predictions included two important outputs, the binary label of whether the token is acoustically [l] or [n], and the probability that reflects the degree to which the token is a typical [l] or [n]. When the

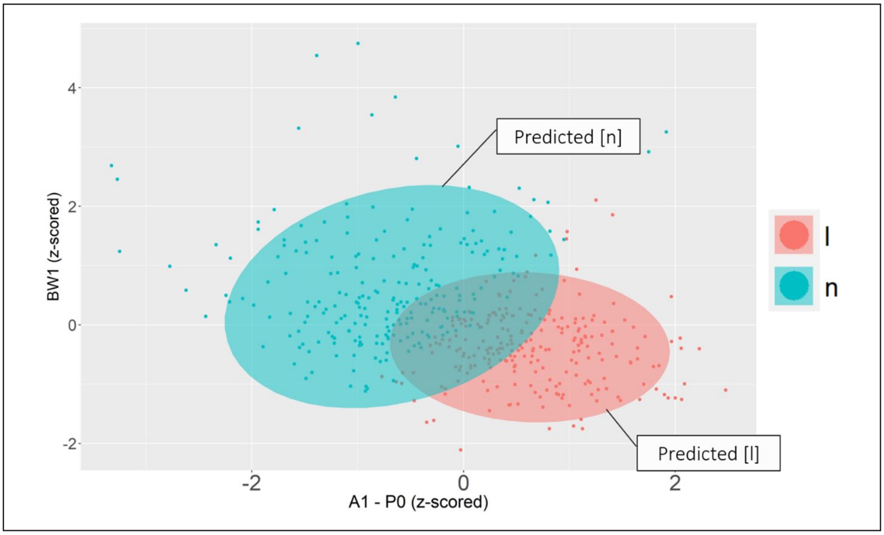

Figure 2. Clustered distributions of Fuzhou Min production based on LDA labels.
Note. The y-axis represents the normalized BW1 values. The x-axis shows the normalized A1–P0 values. The red dots indicate the predicted acoustic [l], and the green dots show the predicted acoustic [n].

probability is between 0 and 0.5, the LDA prediction labeled the token as [n]. A token with a probability between 0.5 and 1 was labeled as [l].

We modeled the probability of being categorized as an acoustic [l] in the LDA for Fuzhou Min segments with mixed-effects linear regression, using the *lme4* package (Bates et al., 2015) for running the model, and the *lmerTest* package for generating p values (Kuznetsova et al., 2017) in R v4.0.3.

LDA onset predictions were included to examine whether the probability of being categorized as an acoustic [l] or [n] significantly differed among Fuzhou Min segments. The model with the lowest Akaike information criterion (AIC)[2] values was selected with a forward step-wise approach. Predicted onset, Vowel height, Nasal coda, and Gender were coded as categorical variables with a mean-centered coding (−0.5 vs. 0.5), while Age was coded as a mean-centered continuous variable. (The same practice of model selection and coding scheme was used in Sections 3.1.4 and 3.2.4 as well). The best model included Predicted onset ([l] vs. [n]) and Vowel height (high vs. low) as fixed effects, and Participant as the random intercept. Nasal coda, and Age and Gender of the speaker did not significantly improve the model.

Overall, the probability of being classified as [l] or [n] was significantly different ($\beta = -0.57$, $SE = 0.01$, $t = -44.86$, $p < .001$). The distribution of the Fuzhou Min [l] and [n] tokens was plotted in the acoustic space[3] in Figure 2. Figure 2 shows the two acoustic cues (A1–P0 and BW1) that contributed the most to the discriminant function. The clusters based on the LDA predictions suggest that in Fuzhou Min, the [l] and [n] tokens were acoustically distinct. The complete model output is presented in Table 6 in Appendix 2.

However, Fuzhou Min speakers did not map the [l] and [n] onto distinct words that were labeled by prescriptive L and N. As shown in Figure 3, the prescriptive L and N (phonemes in the unmerged system) overlapped substantially in the acoustic space, aligning with the specific acoustic parameter finding. Both prescriptive L and N were variably realized as acoustic [l] or [n] in production.

As for directionality, the overall probability of any token being categorized as [l] (0.54) was significantly greater than chance, $t(575) = 3.22$, $p = .001$, and the number of [l] categorizations

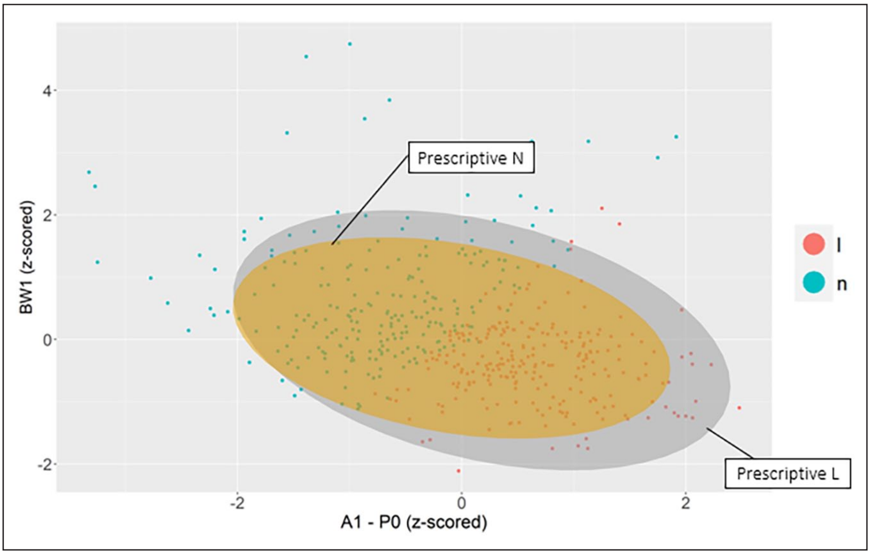

Figure 3. Clustered distributions of Fuzhou Min production based on prescriptive labels.
Note. The y-axis represents the normalized BW1 values. The x-axis shows the normalized A1–P0 values. The yellow ellipse indicates the phonetic space of prescriptive N, and the gray ellipse corresponds to the phonetic space of prescriptive L. The red dots indicate the predicted acoustic [l], and the green dots show the predicted acoustic [n].

(n=329) was greater than that of [n] categorizations (n=247). These results suggest that the merger tends toward the [l] direction.

There was also an effect of vowel height. Segments were significantly less likely to be classified as [l] and more likely as [n] when followed by low vowels ([a], [an], [on]) than high vowels ([i], [u], [un], [øn]) ($\beta=-0.04$, $SE=0.01$, $t=-3.15$, $p=.002$), as shown in Figure 4. The vowel height effect may be explained by the acoustic properties of low vowels. A low vowel has a higher F1, a larger F1 bandwidth, and a lower F1 amplitude than a high vowel, all acoustic properties that are more similar to those of a nasal sound (Stevens, 1998, pp. 260–271).

In conclusion, the LDA established that Fuzhou Min speakers do produce acoustically distinct [l] and [n] tokens. However, these acoustically distinct tokens do not map onto the prescriptive L and N categories, with both [l] and [n] mapping onto both prescriptive L and N categories. As for the linguistic and social factors, low vowels contributed to more nasality and there was no impact of social factors. Overall, the data show the merging of [l] and [n], along with the directionality toward [l], suggesting an initial pattern similar to the findings of the merger in Cantonese, with 90% of Cantonese speakers pronouncing prescriptive N with a phonetic [l] based on transcription data (To et al., 2015).

3 Study 2: perception

Study 2 examined how Fuzhou Min listeners discriminated and identified the merging sounds. The participants first completed an AX discrimination task and then a 2AFC identification task.

3.1 AX discrimination

The AX discrimination task tested the perception of the contrast between prescriptive L and N in Fuzhou Min and whether Fuzhou Min listeners still perceive the acoustic difference between [l] and [n].

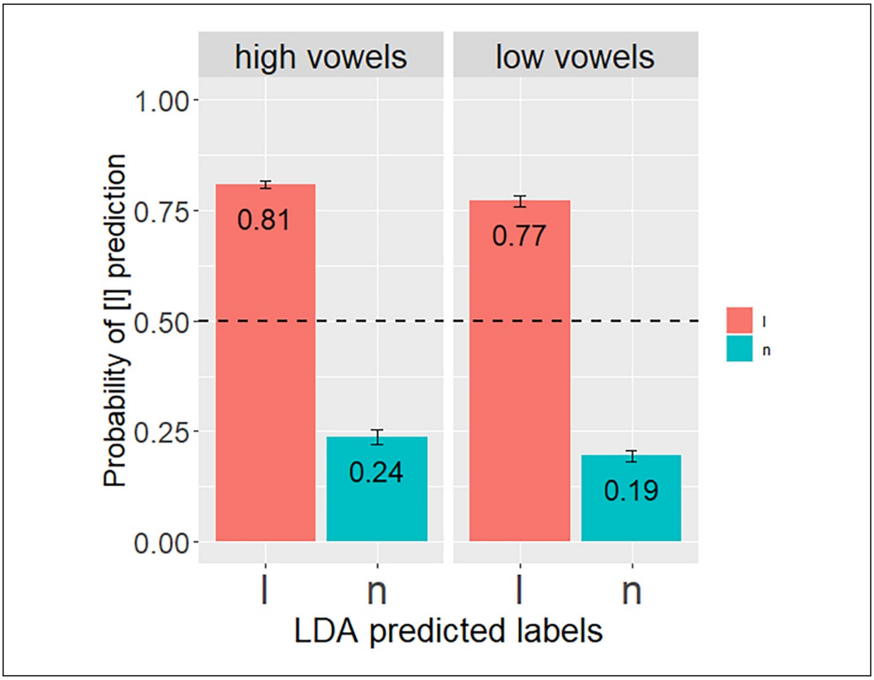

Figure 4. Mean probability of being classified as [l] across onsets and vowels.
Note. The x-axis lists the LDA predicted labels (i.e., [l] vs. [n]). Red bars represent [l], and green bars indicate [n]. The y-axis represents the average probability of being categorized as [l]. If the probability is larger than 0.5, a token is labeled as [l]. If not, the token is labeled as [n]. Error bars indicate standard error of the mean.

3.1.1 Participants. Thirty native Fuzhou Min listeners (aged 26–58; 13 females and 17 males) participated in the experiment in Fuzhou, China, and all participants provided written informed consent before the experiment. These participants did not overlap with the Fuzhou participants in Study 1. We recruited 15 younger (⩽45 years) and 15 older participants (>45 years). The age ranges were selected based on K. Johnson and Song (2016). Each participant received 70 Chinese Yuan (around 10 U.S. dollars) after completing the experiment.

3.1.2 Stimuli. The stimuli were selected from the audio recordings in Study 1. To ensure a reasonable experiment duration, we randomly chose eight speakers (two younger females, two older females, two younger males, and two older males) from the 12 Fuzhou speakers. All stimuli used in the perception experiment were produced by the selected eight Fuzhou Min talkers. To best contrast vowel height and the presence versus absence of the nasal coda, four rimes /a, an, u, un/ and all three repetitions of each stimulus were included. For each trial, the two stimuli always came from the same speaker, and two physically identical sounds never occurred in the same trial. Table 2 illustrates all four types of trials for the [li]-[ni] pair produced by the speaker Female #01. In total, there were 128 trials (4 rimes × 4 types of trials × 8 speakers). Durations varied across different rimes and were normalized to the average of each rime over the whole syllable within Praat (Boersma & Weenink, 2019): /a/ (261 ms), /an/ (453 ms), /u/ (395 ms), and /un/ (430 ms). Analyses showed no differences in consonant duration between prescriptive L and N in Fuzhou Min. The intensity was normalized to 70 dB and naturalness was confirmed by a native Fuzhou Min listener.

Table 2. The Four Types of Trials Used in the AX Discrimination Task.

Trial type	AA	AB	BA	BB
Stimulus 1	*Li*-F01-1st	*Li*-F01-3rd	*Ni*-F01-3rd	*Ni*-F01-2nd
Stimulus 2	*Li*-F01-2nd	*Ni*-F01-3rd	*Li*-F01-3rd	*Ni*-F01-1st

Note. F01 represents female speaker 1; 1st, 2nd, 3rd refer to the repetition of the stimulus.

3.1.3 Procedure. The inter-stimulus interval (ISI) was 750 ms to encourage linguistic processing (Werker & Logan, 1985). After the auditory presentation via head-mounted headphones (SONY WH-1000XM2) of both stimuli on each trial, the listener had up to 3,000 ms to respond either "same" or "different" by pressing "F" or "J" on the keyboard with their left and right hand, respectively. Immediately after the response or 3,000 ms after the offset of the stimulus, the next trial was presented. Response times were measured from the offset of the second stimulus sound in the AX task. The trials were repeated three times in three separate blocks, with each block consisting of one repetition of all the trials. In total, each listener made 384 responses in the AX discrimination task, which took around 30 minutes. There was a practice session before the formal experiment, with 8 trials that were not included in the actual experiment. The task was implemented via Paradigm (Tagliaferri, 2005) on a PC laptop.

3.1.4 Results. A total of 11,520 responses were collected and 10,251 responses (89%) remained in the analysis, after excluding the 1,269 responses with a response time shorter than 200 ms.

Discrimination accuracy. Two labeling systems were used to calculate accuracy. To look for converging evidence of the nasal-lateral merger in perception, we coded the stimuli with prescriptive L and N, which mark the nasal and lateral onsets in an unmerged system. To investigate the status of the merger, we also labeled the auditory stimuli with the LDA classification membership as [l] or [n] to reflect the acoustic properties of the stimuli. Discrimination accuracy for each of these two labeling systems was calculated accordingly. For prescriptive conditions, if a trial such as "*Lan* versus *Lan*" elicited a "same" response, the response would be considered as correct. If this trial elicited a "different" response, it would be considered incorrect. In prescriptive conditions, 5,100 trails were "same," and 5,151 trials were different. For LDA-based conditions, a "same" response to "[lan] versus [lan]" was treated as correct, while a "different" response to such a trial was treated as incorrect. In LDA-based conditions, there were 6,609 "same" trials, and 3,643 "different" trials.

When coded with prescriptive labels, the proportion of correct responses across all participants was 48%. In the different conditions, there was a correct proportion of 22% across the 5,100 responses and in the same condition, there was a correct proportion of 75% across the 5,151 responses. To examine listeners' sensitivity to the phonemic contrast, we calculated the A′ score based on the participants' hit and false alarm rates (e.g., Polka et al., 2001). As a nonparametric index between 0 and 1, higher A′ scores reflect greater sensitivity to the signal difference, and 0.5 represents chance level (Grier, 1971; Stanislaw & Todorov, 1999). In a one-sample t test, Fuzhou Min listeners' A′ scores coded with prescriptive labels were significantly lower than 0.5, $M=0.46$, $t(29)=-2.71$, $p=.007$, suggesting that they were not sensitive to the difference between prescriptive L and N and providing perceptual evidence of the [l~n] merger in Fuzhou Min.

When the same discrimination data were analyzed with LDA labels, the proportion of correct responses across all participants was 57%, with a correct proportion of 22% across the 3,642 responses in the different conditions, and a correct proportion of 76% across the 6,609 responses in the same conditions. The A′ scores with the LDA-based analysis were also significantly lower than 0.5, $M=0.47$, $t(29)=-2.13$, $p=.02$, indicating that Fuzhou Min listeners did not perceive the

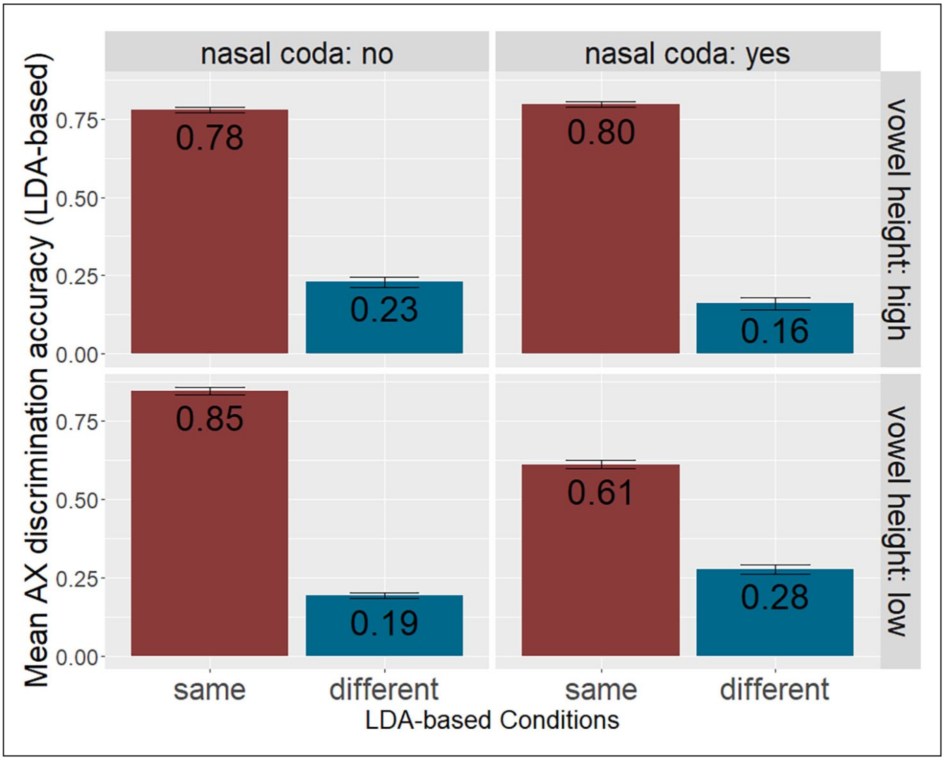

Figure 5. Mean accuracy for discrimination of same and different stimuli (LDA labels) across vowel height and nasal coda contexts in Fuzhou Min.

Note. The horizontal axis shows the experimental conditions (same vs. different). The vertical axis presents the mean accuracy. Different panels indicate vowel height and nasal coda contexts. Error bars represent standard error of the mean.

acoustic difference between [l] and [n]. A′ scores calculated in both labeling systems provide perceptual evidence of the Fuzhou Min [l~n] merger.

Impact of linguistic and social factors on discrimination accuracy. Having established the collapse of the contrast between prescriptive L and N, we used mixed-effects logistic regression (Bates et al., 2015; Kuznetsova et al., 2017) to test whether linguistic and social factors contributed to different degrees of sensitivity to the acoustic difference between [l] and [n]. The dependent variable was the accuracy coded with the LDA labels. The best model included Experimental condition (same vs. different, using LDA labels), Vowel height (high vs. low), Nasal coda (present vs. absent), Age of listener (mean-centered continuous variable), and their interactions as fixed effects, and Participants as random intercepts. Gender did not significantly improve the model.

Fuzhou Min listeners showed a complete merger in perception. As shown in Figure 5, although the accuracy in acoustically different conditions was consistently below 50%, the sensitivity to the acoustic difference between [l] and [n] was slightly greater in certain linguistic contexts. . Fuzhou Min listeners were slightly more sensitive to the acoustic difference between [l] and [n] when there was no nasal coda ($\beta = -0.28$, $SE = 0.06$, $z = -4.77$, $p < .001$), and when a low vowel and a nasal coda co-occurred ($\beta = -0.23$, $SE = 0.12$, $z = -1.93$, $p = .05$), showing an accuracy of 28%. An age effect was also observed within the context of the complete merger. The interaction between the

Experiment condition and Age showed ($\beta = 0.09$, $SE = 0.01$, $z = 13.54$, $p < .001$) that the younger listeners were more accurate (26%) in the different conditions than the older listeners (17%). Table 7 in Appendix 2 presents the statistical details of the model.

Acoustic properties of the stimuli influenced discrimination. Two additional analyses were conducted to understand the nature of the vowel height and nasal coda impact. We tested the relationship between acoustics and discrimination accuracy with mixed-effects logistic regression and explored the relationship between linguistic factors and the acoustic correlates with mixed-effects linear regression. The acoustic properties here refer to the acoustic distance between the two stimuli in a trial.

To explore the relationship between acoustic cues and discrimination responses, mixed-effects logistic regression models were fitted with discrimination accuracy as the dependent variable. The fixed effects in the best model included Experimental condition (same vs. different, using LDA labels), acoustic cues (Δ A1, A1–P0, BW1, relative RMS amplitude), and interactions between Experimental condition and Δ A1, BW1, and relative RMS amplitude, respectively. Participants were included as random intercepts. The increased acoustic distance between two stimuli correlated with higher accuracies in the different conditions ($\beta = 1.52$, $SE = 0.14$, $z = 11.27$, $p < .001$). The values of coefficients indicate different weights of the acoustic cues: A1–P0 ($\beta = 0.53$, $SE = 0.03$, $z = 15.19$, $p < .001$) > RMS amplitude ($\beta = 0.39$, $SE = 0.04$, $z = 9.70$, $p < .001$) > Δ A1 ($\beta = 0.20$, $SE = 0.04$, $z = 5.60$, $p < .001$) > BW1 ($\beta = -0.03$, $SE = 0.04$, $z = -0.80$, $p = .42$). The outputs of the models are reported in Table 9 in Appendix 2.

To explore the relationship between acoustic cues and discrimination responses, the fixed effects of the mixed-effects linear regression model included Experimental condition (coded in LDA labels, [l] and [n]), Vowel height, Nasal coda, and interactions between Condition and Vowel height, and between Condition and Nasal coda. Speakers of the stimuli were entered into the model as random intercepts. Models were fitted for each of the six cues, in which the dependent variable is the normalized acoustic distance between stimulus A and stimulus X in a trial (e.g.,$|F3_A – F3_X|$). Within the different trials, low vowels had a larger acoustic distance for BW1 ($\beta = 0.36$, $SE = 0.01$, $t = 28.51$, $p < .001$), Δ A1 ($\beta = 0.37$, $SE = 0.01$, $t = -17.74$, $p < .001$), F2–F1 ($\beta = 0.08$, $SE = 0.02$, $z = 4.15$, $p < .001$), and relative RMS amplitude ($\beta = 0.43$, $SE = 0.01$, $t = 30.59$, $p < .001$). The presence of a nasal coda increased the acoustic distance for BW1 ($\beta = 0.31$, $SE = 0.01$, $t = 25.95$, $p < .001$), Δ A1 ($\beta = 0.19$, $SE = 0.01$, $t = 14.2$, $p < .001$), and relative RMS amplitude ($\beta = 0.19$, $SE = 0.01$, $t = 14.20$, $p < .001$) as well. For a comprehensive report of the models, see Table 8 in Appendix 2.

In conclusion, A′ scores based on prescriptive labels suggested that Fuzhou Min listeners did not perceive the difference between prescriptive L and N. Moreover, the A′ scores coded with LDA labels indicated that Fuzhou Min listeners also did not perceive the acoustic difference between the [l] and [n] tokens. Further analyses showed that linguistic factors modulated Fuzhou Min listeners' sensitivity to the acoustic difference between [l] and [n]. Fuzhou Min listeners were more sensitive to the [l]-[n] acoustic difference when there is no nasal coda, and they were more sensitive when a low vowel and a nasal coda co-occurred. Low vowels and nasal codas increased the acoustic distance between [l] and [n] and Fuzhou Min listeners showed a corresponding increased sensitivity in these contexts. As for social factors, an age effect was observed, with younger listeners being more sensitive to the acoustic difference between [l] and [n] than older listeners. It should be noted that despite the varying degrees of sensitivity in different contexts, overall, Fuzhou Min listeners do not distinguish prescriptive L and N or acoustic [l] and [n]. The AX discrimination data illustrate a complete [l~n] merger in perception in Fuzhou Min. Given this evidence of a complete merger, we next investigate the direction of the merger in perception, which will be tested with a 2AFC identification experiment.

3.2 Two-alternative forced choice identification

3.2.1 Participants. The 30 Fuzhou listeners in the 2AFC identification task were the same as those in the AX discrimination task.

3.2.2 Stimuli. The auditory stimuli were the same as those in the AX discrimination task, but each token was presented individually. In total, there were 192 trials (2 onsets × 4 rimes × 8 speakers × 3 repetitions by the speaker).

3.2.3 Procedure. The listeners heard a stimulus over headphones and had 3,000 ms to select either "L" or "N" via the "F" or "J" keys, respectively, on the keyboard. The options of lexical "L" and "N" were visually displayed on the screen as Chinese characters together with the prescriptive spelling (e.g., *lan*蓝 vs. *nan*南). The experiment took around 12 minutes via Paradigm (Tagliaferri, 2005) on a PC laptop.

3.2.4 Results. A total of 5,760 responses were collected from the 2AFC identification task. After 81 responses were excluded due to a response time less than 200 ms, 5,679 (99%) responses were included in the analysis. In this section, we report the proportion of L and N responses rather than accuracy. The term accuracy implies that there is a correct answer. However, in the context of a complete merger, there is no correct answer for the listeners. The 2AFC identification experiment was designed to examine the directionality of the complete merger in perception.

Identification responses. After aggregating within the same participant and within the same item, the mean proportion of L responses was 53%, which was significantly greater than chance, $t(239)=1.98$, $p=.02$. The mean proportion of N responses was 47%. Fuzhou Min listeners' identification of the merging sounds tended toward L.

The experimental condition was coded in LDA labels to examine how Fuzhou Min listeners identified the acoustic [l] and [n], regardless of whether the sound originally belonged to the prescriptive L or N category. For the stimuli with an acoustic [l] as the word-initial consonant ($n=3{,}224$), the mean L response proportion was 53%, and the mean N response proportion was 47%. The stimuli with an acoustic [n] as the onset ($n=2{,}455$) elicited a mean proportion of 52% L responses and a mean proportion of 48% N responses. Fuzhou Min listeners showed a tendency to identify the stimulus as L, regardless of whether the sound is acoustically more [l]-like or [n]-like.

Impact of linguistic factors on the identification responses. Mixed-effect logistic regression was fitted to the binary response data, with Experimental condition, Vowel height, and Nasal coda as the fixed effects and Participants as the random intercepts in the best model. Age and gender of participants did not significantly improve the fit of the model. In general, the proportions of L and N responses were both close to 50%, while in certain contexts more L responses were elicited. The acoustic [l] condition elicited more L responses (53%) than the acoustic [n] condition (52%; $\beta=-0.13$, $SE=0.06$, $z=-2.09$, $p=.04$), but the difference is very small. As illustrated by Figure 6, low vowel contexts received more L responses (63%) than high vowel contexts (48%; $\beta=0.55$, $SE=0.06$, $z=8.97$, $p<.001$) in the [l] conditions. And the same pattern appeared in the acoustic [n] conditions, with 56% L responses in low vowel contexts, and 44% L responses in high vowel contexts. Vowels without a nasal coda elicited more L responses in both the acoustic [l] (62%) and the acoustic [n] conditions (64%) than vowels with a nasal coda (acoustic [l]: 42%; acoustic [n]: 44%; $\beta=-0.88$, $SE=0.06$, $z=-15.48$, $p<.001$). Table 10 in Appendix 2 lists the statistical output of the model.

Impact of acoustic properties on identification. To test whether the effect of vowel height and nasal coda in identification is influenced by the acoustic properties of the stimuli, regression analyses were conducted between the acoustics of the stimuli and the participants' identification responses as well as between the acoustics of the stimuli and the linguistic factors.

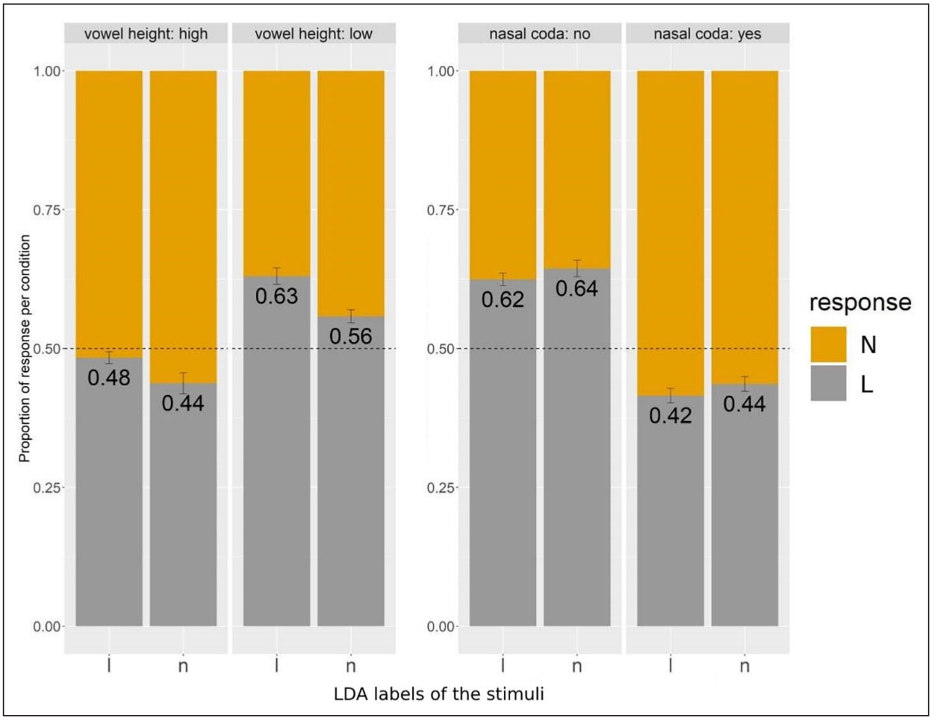

Figure 6. Mean response proportions in LDA-based conditions.
Note. The left panel includes high and low vowels. The right panel includes vowels with and without a nasal coda. The horizontal axis represents the experimental conditions (acoustic [l] vs. [n]). The vertical axis shows the proportion of identification responses. Error bars indicate standard error of the mean.

Not every acoustic cue influenced the identification responses. Mixed-effect logistic regression models were fitted to the binary identification responses, and the best model only contained BW1 and F2–F1 as fixed effects and Participants as the random effects. An increase in BW1 (acoustically more nasal-like) correlated with more L responses ($\beta=0.07$, $SE=0.03$, $z=2.34$, $p=.02$). An increase in F2–F1 (acoustically more nasal-like) was accompanied by a decrease in L responses ($\beta=–0.10$, $SE=0.03$, $z=-3.14$, $p=.002$). A detailed statistical report is provided in Table 12 in Appendix 2.

In the mixed-effect linear regression model with the acoustic cues (normalized within each speaker) as the dependent variable, the fixed effects included Experimental condition ([l] vs. [n]), Vowel height, Nasal coda, the interaction between Experimental condition and Vowel height, and the interaction between Experimental condition and Nasal coda, and Participants as random intercepts. Vowel height and nasal coda influenced all six acoustic correlates. Specifically, BW1 increased in low vowels ($\beta=1.00$, $SE=0.02$, $t=44.28$, $p<.001$). When a nasal coda was present, BW1 ($\beta=0.35$, $SE=0.02$, $t=16.81$, $p<.001$) and F2–F1 spacing ($\beta=12$, $SE=0.03$, $t=4.13$, $p<.001$) increased. A detailed statistical report is provided in Table 11 in Appendix 2.

In conclusion, Fuzhou Min listeners were slightly more likely to identify the merger as L than as N, suggesting that the merger tended toward the lateral direction. Low vowels and vowels with a nasal coda elicited more L responses than their counterparts. Acoustic properties explained the effects of the linguistic factors. F2–F1 and BW1 were the two acoustic cues that correlated with the identification responses but showed different patterns. The pattern of F2–F1 indicated that more acoustic nasality led to more nasal responses. However, the effect of BW1 suggested that more acoustic nasality contributed to more lateral responses. Neither age nor gender influenced Fuzhou Min listeners' identification.

4 Discussion

The present study investigated the status of the phonemic contrast between word-initial nasal and lateral consonants in Fuzhou Min. Acoustic analyses, as well as discrimination and identification results all, provide clear evidence for a merging of the lateral [l]—nasal [n] contrast in Fuzhou Min.

4.1 Production and perception evidence for the [l~n] merger

In production, we first identified 6 acoustic parameters (F2–F1 spacing, F3 frequency, relative RMS amplitude, Δ A1, BW1 and A1–P0) that signaled the contrast between nasal and lateral consonants in Mandarin and English (both languages with the nasal-lateral contrast). Next, we established that none of these acoustic parameters significantly differed across prescriptive L and N in Fuzhou Min. Acoustically, then, there is evidence of a merger in Fuzhou Min. Also, the phonetic space of prescriptive L and N substantially overlapped. The linear discriminant analysis did identify acoustically distinct [l] and [n] tokens, but Fuzhou Min speakers did not map acoustic [l] and [n] onto distinct words labeled by prescriptive L and N. The onsets of both L and N words were variably realized as acoustic [l] and [n] in production.

In perception, Fuzhou Min listeners' low discrimination accuracy and A′ scores indicated that they were not sensitive to the difference between prescriptive L and N or between acoustic [l] and [n]. Listeners could not reliably tell the two forms apart. In addition, results from the identification experiment showed that listeners' responses slightly favored L over N responses and that these responses are not based on the acoustic categorization of the stimuli.

Based on both production and perception data, we were able to provide evidence of a complete phonological merger in Fuzhou. The acoustic distinction between [l] and [n] was not perceived by Fuzhou Min listeners. This pattern is different from a *near-merger* (Labov, 1994). In a near-merger, there is still a phonetic distance between the prescriptive categories, but in Fuzhou, the acoustic distinction did not map onto prescriptive L and N. In Fuzhou Min, among the L tokens, 61% were realized as [l] and 39% as [n]. For the N tokens, 53% were realized as [l] and 47% as [n]. When Fuzhou Min listeners identified the merging sounds, 53% of [l] sounds were perceived as L and 47% as N; 52% of [n] sounds elicited L responses, and 48% elicited N responses. Fuzhou Min speakers were producing a sound contrast that they did not perceive. A similar misalignment between production and perception has also been identified in other sound changes in progress, such as Dutch obstruent devoicing (Pinget et al., 2020). Hence, we consider the [l~n] merger in Fuzhou Min as a complete merger in both production and perception. It should be noted that we collected production and perception data from different Fuzhou Min participants. Further exploration regarding the link between production and perception should elicit production and perception from the same group of participants.

Our data also illustrated the directionality of the [l~n] merger in Fuzhou Min. The merger tended toward the lateral. The acoustic properties of all the Fuzhou Min speakers' productions are tilted toward [l]. In perception, Fuzhou Min listeners were slightly more likely to identify the merging sounds as prescriptive L than N. A preference for lateral sounds has also been observed, to a greater extent, in Cantonese. When producing prescriptive N words, 95% of adult Cantonese speakers used the [l] realization, based on trained speech pathologists' transcription (To et al., 2015). Previous studies of the [l~n] merger in Chinese languages utilized nasalance (Shi, 2015; Shi & Liang, 2017; Shi & Xiang, 2010), perceptual experiments (Cheng, 2017; K. Johnson & Song, 2016), and transcription (To et al., 2015) to quantify the degree of merger. In the current study, we were able to integrate evidence from both production and perception to quantify the merger in Fuzhou Min. In addition, we identified six acoustic correlates that signal the [l]-[n] contrast, which could be applied to other languages with the [l~n] merger as a quantitative tool.

4.2 Merger can surface in a gradient manner across linguistic and social contexts

In production, the overall acoustic properties of the merger are slightly more [l]-like, since the mean probability of the Fuzhou Min tokens being categorized as [l] (0.54) was greater than chance. The probability of being categorized as [l] increased in high vowels, showing that Fuzhou Min speakers' production of the merger varied across linguistic contexts. The merger surfaced in a gradient manner.

In AX discrimination, Fuzhou Min listeners' sensitivity to the difference between acoustic [l] and [n] also slightly increased when the low vowel and the nasal coda co-occurred, which was attributed to the enlarged acoustic distance of several cues (i.e., Δ A1, BW1, relative RMS amplitude). In identification, low vowels and the absence of nasal codas contributed to more L responses, and the increase in L responses correlated with larger BW1 (leaning toward [n]) and smaller F2–F1 (leaning toward [l]), which may involve two distinct mechanisms. The effect of F2–F1 may be explained based on the acoustics. F2–F1 was more lateral-like in low vowels and vowels without a nasal coda. When F2–F1 was more lateral-like, listeners reported more L responses. The effect of BW1, however, may be compatible with a perceptual compensation account of coarticulation (e.g., Fowler, 2006; Lotto & Holt, 2006; Lotto & Kluender, 1998; Mann & Repp, 1980). BW1 signaled more nasality in low vowels and vowels with a nasal coda, and Fuzhou Min listeners may have attributed the nasality signaled by BW1 to the nasal context, instead of to the target onset consonant. Thus, Fuzhou Min listeners were more likely to identify an ambiguous onset consonant as the lateral consonant when they paid attention to BW1. Perceptual compensation for coarticulation has also been claimed to be the source of innovative variants that later propagate throughout the whole speech community, resulting in a sound change (Ohala, 1993; Yu, 2010). The difference between F2–F1 and BW1 may also be due to the temporal location at which the cue occurs in the signal: F2–F1 was measured at the temporal midpoint of the onset consonants, and bandwidth of F1 was averaged over the first 20 ms of the vowel following the onset consonant. The time-course of the signal may influence the perception of an ongoing sound change (Beddor et al., 2018). In addition to offering a phonetic explanation for the effect of linguistic contexts (Warren et al., 2007), we were able to use acoustic data to support the effects of vowel height and nasal coda.

While previous studies examined lexical frequency as a linguistic factor that influences sound mergers (e.g., Shen, 1990; Todd et al., 2019), the lexical frequency could not be controlled in the present study because frequency data for Fuzhou Min are not available. Instead, to ensure natural productions, we selected highly common words in Fuzhou Min. Future research could include familiarity data as an index of lexical frequency and examine whether extent of the merger is modulated by lexical frequency.

As for the effect of social context, younger Fuzhou Min listeners were more likely to perceive the difference between [l] and [n] than older listeners. The participants provided a subjective rating of their Fuzhou accent in Mandarin via a language background questionnaire. We coded "no accent" as 0, "slight accent" as 1, and "moderate accent" as 2. A lower accent rating indicates greater Mandarin proficiency. Younger Fuzhou speakers had a lower average accent rating (0.73) than older Fuzhou speakers (0.93) in Mandarin. Hence, it is reasonable to postulate that younger Fuzhou speakers were exposed to Mandarin more than older Fuzhou speakers. Exposure to Mandarin seems to serve as a strong explanation of the age effect, as the Mandarin influence was also found in the Nanjing [l~n] merger as well as the reversals of the Shanghainese vowel merger (Yao & Chang, 2016). Contact with a language having the contrast had similar age effects in bilingual communities. In Vancouver Cantonese, younger listeners' showed more categorical identification of the [l~n] merger (Cheng, 2017), and in Toronto Korean, younger speakers' showed greater VOT distance between lenis and aspirated stops (Kang & Nagy, 2016), relative to the older members in each of the language communities. The [l~n] merger in Fuzhou Min may be complete, but the influence of Mandarin has re-introduced the two phonetic categories ([l] and [n]) to the younger Fuzhou Min community.

The lack of a gender effect indicated that female and male Fuzhou Min speakers did not differ regarding their production and perception of the [l~n] merger. This result is consistent with Labov's (1994) sound change model regarding generational gender difference, which predicts that when a sound change is close to completion, female and male speakers converge in their speech behavior (p. 309). The current absence of an effect of gender differs from the gender effect found in Vancouver Cantonese (Cheng et al., 2019), in which older females used the more innovative form of the merger. Our older group (aged 45–56 in production; aged 45–58 in perception) was slightly younger than the older group (aged 52–64) in Cheng et al. (2019), and it is possible that the Fuzhou Min older participants were in a more advanced state of the merger than the Vancouver Cantonese older participants.

5 Conclusion

In conclusion, in this experimental study of the [l~n] merger in Fuzhou Min, evidence from both speech production and perception indicates the suspension of the phonemic contrast between prescriptive L and N. While two distinct acoustic variants are sometimes produced, these two variants do not map onto prescriptive L and N. In perception, although Fuzhou Min listeners overall did not perceive the acoustic distinction, younger listeners demonstrated increased sensitivity to the acoustic difference, most likely due to contact with Mandarin. These data are most likely a consequence of younger listeners' greater contact with and use of Mandarin and may suggest a re-emerging contrast in Fuzhou Min. The current study also found that the merger was modulated by linguistic (vowel height and nasal coda) factors. We suggested two mechanisms that may be involved in the effects of vowel height and nasal coda, including acoustics and perceptual compensation for coarticulation. The current research contributes to our understanding of the widespread [l~n] merger in Chinese languages with quantitative data from both production and perception.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The first author received financial support for recruiting participants from the University of Kansas Linguistics Graduate Student Scholarship

ORCID iD

Ruoqian Cheng https://orcid.org/0000-0001-5779-7848

Notes

1. Δ A1 values are negative, so a negative coefficient indicates a greater difference between the $A1_c$ and $A1_v$.
2. Akaike information criterion (AIC) reflects the goodness of a statistical model in model comparison. Smaller AIC values indicate a better fit for the model (Akaike, 1981).
3. A1–P0 and BW1 were selected for their largest weights in the discriminant function. The inclusion of A1–P0 sacrificed 102 tokens out of the 576 tokens. The 102 tokens with missing A1–P0 consisted of 2 tokens from [a], 5 tokens from [an], 26 tokens from [i], 10 tokens from [on], 35 tokens from [u], 20 tokens from [un], and 4 tokens from [yn]. Across speakers, they included 24 tokens in the older female

group, 24 tokens in the older male group, 31 tokens in the younger female group, and 23 tokens in the younger male group.

References

Akaike, H. (1981). This Week's Citation Classic. *Current Contents Engineering, Technology, and Applied Sciences*, *12*(51), Article 42.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Beaman, K. V., & Tomaschek, F. (2019, July 4). Loss of historical phonetic contrast across the lifespan: Articulatory, lexical, and social effects on sound change in Swabian. *PsyArXiv*. https://doi.org/10.31234/osf.io/f8kd5

Beddor, P. S., Coetzee, A. W., Styler, W., McGowan, K. B., & Boland, J. E. (2018). The time course of individuals' perception of coarticulatory information is linked to their production: Implications for sound change. *Language*, *94*(4), 931–968.

Boersma, P., & Weenink, D. (2019). *Praat: Doing phonetics by computer* [Computer program] (Version 6.0.52). http://www.praat.org/

Chae, S. Y. (1995). *External constraints on sound change: The raising of /o/ in Seoul Korean* [Dissertation]. University of Pennsylvania.

Chambers, J. K. (1995). *Sociolinguistic theory: Linguistic variation and its social significance*. Basil Blackwell.

Chen. (1998). *福州方言研究* [The study of Fuzhou Eastern Min]. 福建人民出版社 [Fujian People's Publishing House].

Chen, M. Y. (1995). Acoustic parameters of nasalized vowels in hearing-impaired and normal-hearing speakers. *The Journal of the Acoustical Society of America*, *985*(5), 2443–2453.

Chen, M. Y. (1997). Acoustic correlates of English and French nasalized vowels. *The Journal of the Acoustical Society of America*, *1024*(4), 2360–2370.

Cheng, L. (2017). *Mergers-in-progress in Hong Kong and Vancouver Cantonese-English bilinguals* [BA thesis]. University of British Columbia.

Cheng, L., Chung, S., & Babel, M. (2019, June 21–23). *Cantonese consonant mergers-in-progress in Hong Kong and Vancouver, BC* [Conference session]. 5th Workshop on Sound Change, University of California, Davis, Davis, CA, United States.

Chiba, T., & Kajiyama, M. (1941). *The vowel: Its nature structure*. Tokyo-Kaiseikan.

Clark, L., Watson, K., & Maguire, W. (2013). Introduction: What are mergers and can they be reversed? *English Language & Linguistics*, *17*(2), 229–239.

De Jong, K., Berkson, K., Lulich, S. M., Myers, S., & Bohnert, A. (2019). The lingual topography of American English laterals in onsets and codas. *The Journal of the Acoustical Society of America*, *145*(3), Article 1928.

Fowler, C. A. (2006). Compensation for coarticulation reflects gesture perception, not spectral contrast. *Perception & Psychophysics*, *68*(2), 161–177.

Glass, J. R. (1984). *Nasal consonants and nasalized vowels: An acoustic study and recognition experiment* [Master's thesis]. Massachusetts Institute of Technology.

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin*, *75*(6), 424.

Hay, J., Drager, K., & Thomas, B. (2013). Using nonsense words to investigate vowel merger. *English Language & Linguistics*, *17*(2), 241–269.

Hou. (1986). *北京话音档* [Beijing Phonemic Inventory and vocabulary]. 上海教育出版社 [Shanghai Education Publishing House].

House, A. S., & Stevens, K. N. (1956). Analog studies of the nasalization of vowels. *Journal of Speech and Hearing Disorders*, *21*(2), 218–232.

IBM Corp. (2019). *IBM SPSS statistics for Windows* (Version 26.0).

Johnson, D. E. (2007). Stability and change along a dialect boundary: The low vowel mergers of Southeastern New England. *University of Pennsylvania Working Papers in Linguistics*, 13(1), Article 7.

Johnson, K. (2012). *Acoustic and auditory phonetics*. John Wiley & Sons.

Johnson, K., & Song, Y. (2016). *Gradient phonemic contrast in Nanjing Mandarin* (UC Berkeley phonetics and phonology lab annual report 12(1)). https://escholarship.org/content/qt6fk2m1sw/qt6fk2m1sw_noSplash_4d44f18cc772d390a80ce83a5bfacd8e.pdf?t=pe69w4

Kang, Y., & Nagy, N. (2016). VOT merger in heritage Korean in Toronto. *Language Variation and Change*, *28*(2), 249–272.

Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., & Banno, H. (2008, March). Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. In 2008 IEEE International Conference on Acoustics, Speech and Signal Processing (pp. 3933–3936). IEEE.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26.

Labov, W. (2006). *The social stratification of English in New York City*. Cambridge University Press. (Original work published 1966)

Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania Press.

Labov, W. (1994). *Principles of linguistic change internal factors*. Blackwell.

Labov, W. (2001). *Principles of linguistic change social factors*. Blackwell.

Lotto, A. J., & Holt, L. L. (2006). Putting phonetic context effects into context: A commentary on Fowler (2006). *Perception & Psychophysics*, *68*(2), 178–183.

Lotto, A. J., & Kluender, K. R. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics*, *60*(4), 602–619.

Maeda, S. (1982). The role of the sinus cavities in the production of nasal vowels. In *IEEE international conference on acoustics, speech, and signal processing* (ICASSP'82, Vol. 7, pp. 911–914). Institute of Electrical and Electronics Engineers.

Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics*, *28*(3), 213–228.

Mather, P.-A. (2012). The Social Stratification of /r/ in New York City: Labov's Department Store Study Revisited. *Journal of English Linguistics*, 40(4), 338–356.

Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, *185*(5), 331–340.

Nycz, J. (2013). New contrast acquisition: Methodological issues and theoretical implications. *English Language & Linguistics*, *17*(2), 325–357.

Ohala, J. J. (1993). The phonetics of sound change. *Historical Linguistics: Problems and Perspectives*, *237*, Article 278.

Pinget, A. F., Kager, R., & Van de Velde, H. (2020). Linking variation in perception and production in sound change: Evidence from Dutch obstruent devoicing. *Language and Speech*, 63(3), 660–685.

Polka, L., Colantonio, C., & Sundara, M. (2001). A cross-language comparison of/d/–/ð/ perception: Evidence for a new developmental pattern. *The Journal of the Acoustical Society of America*, *109*(5), 2190–2201.

Prahler, A. A. M. (1998). *Analysis and synthesis of the American English lateral constant* [Master's thesis]. Massachusetts Institute of Technology.

Pruthi, T. (2007). *Analysis, vocal-tract modeling, and automatic detection of vowel nasalization* [Doctoral dissertation]. University of Maryland.

Rastatter, M. P. (1984). Relative amplitude and time characteristics of denasalized/m/, /n/, and/ŋ/ consonants. *Language, Speech, and Hearing Services in Schools*, *154*(4), 253–261.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org

Recasens, D., & Espinosa, A. (2005). Articulatory, positional and coarticulatory characteristics for clear/l/and dark/l/: Evidence from two Catalan dialects. *Journal of the International Phonetic Association*, *35*, 1–25.

Sebastian, S., Sreedevi, N., Lepcha, A., & Mathew, J. (2015). Nasalance in cochlear implantees. *Clinical and Experimental Otorhinolaryngology*, *83*(3), 202.

Shen, Z. (1990). Lexical diffusion: A population perspective and a mathematical model. *Journal of Chinese Linguistics*, *18*(1), 159–201.

Shi, X. (2015). 成都话响音的鼻化度——兼论其/n, l/不分的实质及类型 [The nasality degree of sonorants in Chengdu dialect]. *中国语音学报* [Chinese Journal of Phonetics], 10, 92–100.

Shi, X., & Liang, L. (2017). 南京话响音的鼻化度——兼论/n, l/不分的实质 [An analysis of sound nasality on sonorants in Nanjing dialect: The essence of the confusion between /n/ and /l/]. *南京师大学报 (社会科学版)* [Journal of Nanjing Normal University, Social Science edition], 21, 153–160.

Shi, X., Ran, Q., & Shi, F. (2010). 北京话响音鼻化度的初步分析 [Preliminary analysis of nasalance in Mandarin]. *当代语言学* [Contemporary Linguistics], *4*, 348–355.

Shi, X. & Xiang, N. (2010). 武汉话语音的鼻化度考察 [An investigation on sound nasality in Wuhan dialect]. *语言研究*. [Studies in Language and Linguistics], *30*, 49–54.

Shi, Z., & Zhang, J. (2017). 普通话单字音通音声母的鼻化分析 [An analysis of the nasality of approximant initials in the monosyllabic Mandarin characters]. *语言文字应用* [Applied, Linguistics], 4, 39–48.

Shue, Y.-L., Keating, P., Vicenik, C., & Yu, K. (2011). VoiceSauce: A program for voice analysis. *Proceedings of the ICPHS XVII* (pp. 1846–1849). http://www.phonetics.ucla.edu/voicesauce/

Soo, R., Johnson, K. A., & Babel, M. (2021). Sound change in spontaneous bilingual speech: A corpus study on the Cantonese n-l merger in Cantonese-English bilinguals. *Proceedings of Interspeech, 2021*. https://doi.org/10.31219/osf.io/9c3w6

Sproat, R., & Fujimura, O. (1993). Allophonic variation in English /l/and its implications for phonetic implementation. *Journal of Phonetics*, *213*(3), 291–311.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149.

Stevens, K. N. (1998). *Acoustic phonetics*. MIT Press.

Styler, W. (2017). On the acoustical features of vowel nasality in English and French. *The Journal of the Acoustical Society of America*, *1424*(4), 2469–2482.

Tabain, M., Butcher, A., Breen, G., & Beare, R. (2016a). An acoustic study of nasal consonants in three Central Australian languages. *The Journal of the Acoustical Society of America*, *139*(2), 890–903.

Tabain, M., Butcher, A., Breen, G., & Beare, R. (2016b). An acoustic study of multiple lateral consonants in three Central Australian languages. *The Journal of the Acoustical Society of America*, *139*(1), 361–372.

Tagliaferri, B. (2005). *Paradigm* (Version 1). Perception Research Systems.

Tao, Y. (1956). *闽音研究* [A study of sound systems of Min]. 科学出版社 [China Science Publishing & Media].

Tian, H. (2009). 汉语方言 “泥”“来” 二母相混类型研究 [Types of the nasal and lateral mergers in Chinese dialects]. 河北师范大学学报: 哲学社会科学版 [*Journal of Hebeinormal University, Philosophy and Social Science Edition*], *32*(1), 108–113.

To, C. K., McLeod, S., & Cheung, P. S. (2015). Phonetic variations and sound changes in Hong Kong Cantonese: Diachronic review, synchronic study and implications for speech sound assessment. *Clinical Linguistics & Phonetics*, *29*(5), 333–353.

Todd, S., Pierrehumbert, J. B., & Hay, J. (2019). Word frequency effects in sound change as a consequence of perceptual asymmetries: An exemplar-based model. *Cognition*, *185*, 1–20.

Wang, W. S.-Y. (1969). Competing changes as a cause of residue. *Language*, *45*(1), 9–25.

Warren, P., Hay, J., & Thomas, B. (2007). The loci of sound change effects in recognition and perception. *Laboratory Phonology*, *9*, 87–112.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, *37*(1), 35–44.

Wolf, C., & Jiménez, E. (1979). A sound change in progress: Devoicing of Buenos Aires /ž/. *Manuscript*. Unpublished Manuscript, University of Pennsylvania.

Wolfram, W., & Schilling-Estes, N. (2006). *American English: Dialects and variation* (2nd ed.). Blackwell.

Yao, Y., & Chang, C. B. (2016). On the cognitive basis of contact-induced sound change: Vowel merger reversal in Shanghainese. *Language*, *92*(2), 433–467.

Yu, A. C. L. (2010). Perceptual compensation is correlated with individuals' "autistic" traits: Implications for models of sound change. *PLOS ONE*, *58*(8), Article e11950.

Zhang, W., & Levis, J. M. (2021). The Southwestern Mandarin/n/-/l/ merger: Effects on production in Standard Mandarin and English. *Frontiers in Communication*, *6*, Article 639390.

Appendix I

Word Lists Used in the Production Experiment (Study 1).

Table 3. English Word List.

Vowel	[n]	[l]	[b] (Filler)	[d]/[t] (Filler)	[g]/[k] (Filler)	[s] (Filler)
i	**neat**	**leat**	beat	deep	keep	sip ([ɪ])
ɛ	**net**	**let**	bet	debt	get	set
æ	**nap**	**lap**	bat	Dad	gap	-
æ	**nan**	**land**	ban	-	-	sand
ʊ	**nook**	**look**	book	-	cook	shoot ([u])
u	**noon**	**loon**	boom	tomb	-	soon
ʌ	**nut**	**luck**	buck	duck	-	suck
ʌ	**none**	**lung**	bun	done	gun	sun
ɑ	**knock**	**lock**	ball	dog	call	saw

Table 4. Mandarin Word List.

[n]			[l]		
IPA	orthography	meaning	IPA	orthography	meaning
ni51	逆	“backward”	li51	力	“power”
nu35	奴	“slave”	lu35	炉	“stove”
noŋ35	农	“agriculture”	loŋ35	龙	“dragon”
noŋ35	浓	“concentrated”	loŋ15	聋	“deaf”
na51	捺	“press down”	la51	辣	“spicy”
na51	纳	“take in”	la51	腊	“smoked”
nan35	南	“south”	lan35	蓝	“blue”

Note. IPA: International Phonetic Alphabet.

Table 5. Fuzhou Min Word List.

[n] origin			[l] origin		
IPA	orthography	meaning	IPA	orthography	meaning
niʔ4	日	“sun”	liʔ4	力	“power”
nu51	奴	“slave”	lu51	炉	“stove”
nun51	农	“agriculture”	lun51	轮	“tyre”
nøŋ51	侬	“human”	løŋ51	聋	“deaf”
nɔŋ242	嫩	“small”	lɔŋ242	卵	“egg”
naʔ4	捺	“press down”	laʔ4	辣	“spicy”
naʔ4	纳	“take in”	laʔ4	粒	“particle”
nan51	南	“south”	lan51	蓝	“blue”

Appendix 2

Results of statistical tests

Table 6. Summary of the Mixed-Effects Linear Regression Model for the LDA Predicted Probability.

Predictor	Estimate (β)	Standard error	Degree of freedom	*t*-value	*p*-value
(Intercept)	0.79	0.01	573.00	96.61	< 0.001
LDA predictions[a]	–0.57	0.01	573.00	–44.86	< 0.001
Vowel height[b]	–0.04	0.01	573.00	–3.14	0.002

Note. LDA: linear discriminant analysis.
[a] [l] = −0.5, [n] = 0.5. [b] high = −0.5, low = 0.5.

Table 7. Summary of the Mixed-Effects Logistic Regression Model for the LDA-Labeled Discrimination Accuracy with the Best Fit.

	Estimate (β)	Standard error	*z*-value	*p*-value
(Intercept)	–0.06	0.05	–1.09	0.28
Condition (LDA)[a]	2.60	0.06	43.55	< 0.001
Vowel height[b]	0.00	0.06	0.01	0.99
Nasal coda[c]	–0.28	0.06	–4.77	< 0.001
Age	0.00	0.01	–0.03	0.98
Condition: Vowel height	–0.49	0.12	–4.14	< 0.001
Conditionsame: Nasal coda	–0.61	0.12	–5.19	< 0.001
Conditionsame: Age	0.09	0.01	13.54	< 0.001
Vowel height: Nasal coda	–0.23	0.12	–1.93	0.05
Condition: Vowel height: Nasal coda	–2.31	0.23	–9.83	< 0.001

[a] [l] = −0.5, [n] = 0.5. [b] high = −0.5, low = 0.5. [c] no = −0.5, yes = 0.5. [d] female = −0.5, male = 0.5.

Table 8. Summary of the Mixed-Effects Linear Regression Models for the Acoustic Distance between Stimuli (LDA Labels).

Predictor	A1–P0	Bandwidth of F1	Δ A1	F2–F1	F3	Relative RMS amplitude
(Intercept)	1.02*** (0.01)	0.63** (0.10)	0.95*** (0.09)	1.07*** (0.10)	0.87*** (0.08)	0.95*** (0.09)
Condition (LDA)[a]	–0.71*** (0.02)	–0.24*** (0.01)	0.15*** (0.01)	–0.12*** (0.02)	0.17*** (0.02)	0.15*** (0.41)
Vowel height[b]	–0.02 (0.331)	0.36*** (0.01)	0.43*** (0.01)	0.08*** (0.02)	–0.17*** (0.02)	0.43*** (0.01)
Nasal coda[c]	–0.09*** (0.02)	0.31*** (0.01)	0.19*** (0.01)	–0.02 (0.02)	–0.11*** (0.02)	0.19*** (0.01)
Condition: Vowel height	0.14** (0.04)	0.04 (0.03)	0.12*** (0.03)	–0.55 (0.04)	–0.44*** (0.04)	0.12*** (0.03)
Condition: Nasal coda	0.41*** (0.04)	–0.02 (0.02)	0.02 (0.03)	–1.20*** (0.04)	–0.07 (0.04)	0.02 (0.03)

Note. LDA: linear discriminant analysis; RMS: root mean square. Standard errors are reported in parentheses.
[a] [l] = −0.5, [n] = 0.5. [b] high = −0.5, low = 0.5. [c] no = −0.5, yes = 0.5.
*$p < .05$, ** $p < .01$, *** $p < .001$.

Table 9. Summary of the Mixed-Effects Logistic Regression Models for Accuracy of Discrimination (Coded in Prescriptive and LDA Labels).

Predictor	Accuracy coded in LDA labels
(Intercept)	–1.33*** (0.07)
Condition[a]	1.52*** (0.14)
Distance of A1–P0	0.53*** (0.03)
Distance of Δ A1	0.20*** (0.04)
Distance of BW1	–0.03 (0.04)
Distance of relative RMS amplitude	0.39*** (0.04)
Condition: Distance of Δ A1	–0.96*** (0.07)
Condition: Distance of BW1	–0.49*** (0.08)
Condition: Distance of relative RMS amplitude	0.37*** (0.08)

Note. LDA: linear discriminant analysis; BW: bandwidth; RMS: root mean square. Standard errors are reported in parentheses.
[a] Different = −0.5, same = 0.5.
*$p < .05$, ** $p < .01$, *** $p < .001$.

Table 10. Summary of the Mixed-Effects Logistic Regression Model for the Identification Responses to LDA-Based Conditions with the Best Fit.

	Estimate (β)	Standard error	*z*-value	*p*-value
(Intercept)	0.15	0.13	1.21	0.22
Condition[a]	–0.13	0.06	–2.09	0.04
Vowel.height[b]	0.55	0.06	8.97	< 0.001
Nasal coda[c]	–0.88	0.06	–15.48	< 0.001

Note. LDA: linear discriminant analysis; BW: bandwidth; RMS: root mean square.
[a] Acoustic [l] = −0.5, acoustic [n] = 0.5. [b] High = −0.5, low = 0.5. [c] No = −0.5, yes = 0.5.

Table 11. Summary of the Mixed-Effects Linear Regression Models for the Acoustic Measurements of Identification Stimuli (LDA Labels).

Predictor	A1–P0	Bandwidth of F1	∆ A1	F2–F1	F3	Relative RMS amplitude
(Intercept)	–0.03	–0.03	0.08*	–0.08	–0.08	0.02***
	(0.06)	(0.05)	(0.02)	(0.11)	(0.08)	(0.12)
Condition (LDA)[a]	–1.57***	0.67***	–0.52***	0.05	–0.26***	–0.30***
	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Vowel height[b]	0.15***	1.00***	0.13***	0.03	0.20***	0.01
	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Nasal coda[c]	0.13***	0.35***	–0.05	0.12***	0.04	0.45***
	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.03)
Condition: Vowel height	0.13**	0.05	0.46***	0.28***	0.19***	0.31***
	(0.04)	(0.04)	(0.06)	(0.06)	(0.05)	(0.06)
Condition: Nasal coda	0.26***	–0.02	0.20***	0.02	0.33***	–0.70***
	(0.04)	(0.04)	(0.06)	(0.06)	(0.05)	(0.06)

Note. RMS: root mean square. Standard errors are reported in parentheses.
[a]Acoustic [l] = −0.5, acoustic [n] = 0.5. [b]High = −0.5, low = 0.5. [c]No = −0.5, yes = 0.5.
* $p < .05$, ** $p < .01$, *** $p < .001$.

Table 12. Summary of the Mixed-Effects Logistic Regression Models for Identification Responses.

Predictor	Probability of responding L
(Intercept)	0.17
	(0.12)
Bandwidth of F1	0.07*
	(0.03)
F2–F1	–0.10**
	(0.03)

Note. Standard errors are reported in parentheses.
* $p < .05$, ** $p < .01$, *** $p < .001$.