Open Library of Humanities

Production and perception across three Hong Kong Cantonese consonant mergers: Community- and individual-level perspectives

Lauretta S. P. Cheng, Department of Linguistics, University of Michigan, Ann Arbor, Michigan, United States of America, lspcheng@umich.edu

Molly Babel, Department of Linguistics, University of British Columbia, Vancouver, British Columbia, Canada, molly.babel@ubc.ca

Yao Yao, Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, HKSAR, China, y.yao@polyu.edu.hk

Individual variation is key to understanding phenomena in phonetic variation and change, including the production-perception link. To test the generalizability of this relationship, this study compares community- and individual-level variation across three long-standing consonant mergers in Hong Kong Cantonese speakers: [n]→[l], [ŋ]→[m̩], and [ŋ]↔∅. Concurrently, we document these understudied mergers in a community that has undergone rapid social change in recent decades. Younger (college-aged) and older (middle-aged) Hong Kongers completed a reading production task followed by a forced-choice lexical identification perception task. Group-level results suggest mismatching production and perception: While the community overall distinguished merger pairs in production, younger listeners are more perceptually categorical than older listeners. However, aggregate results obscure the fact that individuals vary substantially in the extent of merging in both perception and production, including many who exhibit complete merger, and that individual-level production-perception correlations were found for [n]→[l] and [ŋ]→[m̩], though not [ŋ]↔∅. Results are discussed in the context of previous research. We find that (i) these mergers have diverged from predicted trajectories of completion, and (ii) overall, prior findings on the production-perception link are generalizable to these consonant mergers.

1. Introduction

While variability within speech communities has long been acknowledged, phonetic variation and change has historically been studied at the level of the community or macro-social demographic groups. Recently, a more concentrated focus has been placed on the characteristics of individual speaker-listeners and their role in innovating and driving change (for overviews, see Coetzee, 2018; Yu & Zellou, 2019), including both social (e.g., Sharma & Sankaran, 2011; Lev-Ari, 2017; Garrett & Johnson, 2013) and cognitive (e.g., Yu, 2010; Yu, Abrego-Collier, & Sonderegger, 2013; Dimov, Katseff, & Johnson, 2012; Kong & Edwards, 2016) dimensions. As Yu and Zellou (2019) state, "the study of individual differences can shed light on the nature of the cognitive representations and mechanisms involved in phonological processing," which in turn can "provide insight into long-standing issues in linguistic variation and change" (p. 131). In other words, bringing together community-level (macro) patterns with individual-level (micro) considerations helps build a fuller, more accurate picture of how speech is produced, perceived, and processed—feeding into our collective understanding of how sound systems vary and change over time (see also Yao & Chang, 2016; Fridland & Kendall, 2012; Voeten, 2020).

The example of focus here is how perception and production systems relate in the context of sound change. As this relationship holds implications for mechanisms behind the initiation and propagation of sound change, a more complete picture is crucial for understanding both the spread of sound change from individuals to entire communities and the progression of community-level change over time. Though many have probed this question, consistent results have been hard to come by both in regards to whether production-perception systems are linked in individuals (e.g., Schultz, Francis, & Llanos, 2012; Schertz, Cho, Lotto, & Warner, 2015; Beddor, Coetzee, Styler, McGowan, & Boland, 2018) and which domain precedes the other during a change-in-progress (e.g., Harrington, Kleber, & Reubold, 2008; Kuang & Cui, 2018; Pinget, Kager, & Van de Velde, 2020). Insofar as the nature of the production-perception relationship has yet to be fully uncovered, attempts to answer this question strongly benefit from an approach comparing patterns at the group and individual levels.

In line with this perspective, the present study examines—at community and individual levels—production, perception, and production-perception alignment in an understudied case of phonetic variation and change. We conduct an apparent-time investigation of three consonant mergers in Hong Kong Cantonese with distinct profiles and trajectories, namely [n]→[l], [ŋ]→[m̩], and [ŋ]↔Ø. While documenting the recent state of these mergers, we test hypotheses of the production-perception relationship, seeking to extend previous findings to a novel set of sound changes of a type—consonant mergers—that has thus far lacked study (with the notable exception of Pinget et al., 2020). In our focus on these particular consonant mergers in Hong Kong Cantonese, we find ourselves in a complex situation involving long-standing changes-in-progress, style-shifting, and sociolinguistic ideologies. Although previous research has largely presented

the variants as a subset of mergers-in-progress within a larger set of merging changes (e.g., Bauer, 1986; Zee, 1999a; To, Mcleod, & Cheung, 2015), scholarship also indicates the presence of formality-based stylistic variation (e.g., Bauer, 1982a; Bourgerie, 1990) and stigmatization (e.g., Chen, 2018). Nevertheless, regardless of whether these mergers have stabilized, completed, or continue to change, they remain situated in the context of diachronic change. Accordingly, we primarily approach these mergers as historical sound change and compare them to the sound change literature while also discussing the socially-structured variation and ideology in individual experiences.

1.1. Production and Perception: Is there a link?

Theories of sound change have long assumed that an individual's perception and production systems must be connected in some way for an individual to perceive a change within their speech community and implement it in their own production repertoire. This is a necessary condition if we posit that changes propagate from individual to individual, though the extent to which this is true (or, under which circumstances) had not been directly explored until recently (see Beddor, 2015). In establishing groundwork for theories about phonetic and sound change, Beddor et al. (2018) examine the time course of speaker-listeners' coarticulatory vowel nasalization in production (from aerodynamic measures) and perceptual patterns (with an eye-tracking paradigm) in American English. They demonstrate within-individual correlations across production and perception for this stable variation not only in the *use* of a cue but in the *time-course of use*: Those who produce early nasal flow in a vowel use temporally early nasalization to identify words.

Although there is a strong theoretical appetite for perceptual and productive repertoires to map neatly onto one another, only some investigations have found individual-level evidence for this connection (e.g., Zellou, 2017; Coetzee, Beddor, Shedden, Styler, & Wissing, 2018; Beddor et al., 2018; Fridland & Kendall, 2012; Kim & Clayards, 2019; Pinget et al., 2020; Voeten, 2020) while others have not (e.g., Kataoka, 2011; Grosvald & Corina, 2012; Schertz et al., 2015; Schultz et al., 2012; Kwon, 2015). Thus, the field has yet to come to a satisfying answer regarding the relevant conditions for matching or mismatching production-perception repertoires (see Yu & Zellou, 2019). Here, we briefly review some possible explanations for these cross-study discrepancies.

One common suggestion put forth to account for inconsistent findings is that the experimental tasks and measures used to assess a link between production and perception were not necessarily comparable, such that production and perception tasks may have in fact been measuring different constructs (see Zellou, 2017; Kim & Clayards, 2019; Grosvald & Corina, 2012). A related concern is that previous measures may not have been sensitive enough to capture similarities across domains (e.g., static versus dynamic measures, Beddor et al., 2018). These concerns point to a

need for careful selection of tasks and variables for comparison. They also link to methodological concerns about validity and reliability in terms of how phonetic variables are assessed (e.g., acoustic measurements versus auditory coding; trained versus naive listeners; speakers versus non-speakers; Evans, Munson, & Edwards, 2018).

From a theoretical perspective, some posit that the context of variation, such as the stability of variation patterns in the community, has consequences for production-perception alignment. Beddor et al. (2018), for example, suggest that an individual's link between perception and production is likely at its weakest during change when variants are in flux and socially-stratified, in comparison to stable, non-socially-indexed variation (see also Harrington, 2012). However, current empirical evidence does not obviously support this position. In terms of stable variation, while Beddor and colleagues do find a link between production and perception at the individual level for vowel nasalization (as does Zellou, 2017), a number of scholars investigating different types of coarticulatory variation failed to do so (e.g., Schultz et al., 2012 on VOT-f0 cue weighting in English; Schertz et al., 2015 on Korean L2 English VOT versus f0; Grosvald & Corina, 2012 on vowel-to-vowel coarticulation), though these direct comparisons may be inconclusive due to the aforementioned methodological differences.

Importantly, many recent studies of sound change do find evidence of individual-level links between domains, despite community-level mismatch (Coetzee et al., 2018; Kuang & Cui, 2018; Pinget et al., 2020; Voeten, 2020). On the surface, these findings run counter to the above prediction as individual-level production-perception relationships can indeed be found during changes-in-progress. Yet, it remains possible that the size of correlation is smaller or that a larger number of individuals demonstrate a lack of relationship, both compatible with an interpretation that the relationship is weaker during change than during stability. Overall, this highlights the need for more systematic, controlled comparison of individual production and perception patterns for variation in different contexts, crucially using comparable measures.

1.2. Production-perception alignment: What matters?

Beyond the simple existence of a production-perception connection, a co-occurring question pertains to the *nature* or *direction* of alignment. In studies of community change, alignment direction has been demonstrated to vary such that, at times, perception appears to lead in sound change while at other times, production appears more advanced. Some have called for attention to the trajectory and stage of change when examining production and perception, arguing that alignment patterns may depend on whether the change is in early or late stages. A slew of recent findings converge to support the notion that direction of misalignment may be yoked to the stage of the change.

Kuang and Cui (2018) documented two vowel changes in Southern Yi where phonation cues are shifting to formant cues. For the incipient /u/ change, all speaker groups were generally

aligned: Phonation was the most important cue in both domains, but F1 was increasingly relied upon in perception. For the ongoing /e/ change, however, older speakers were misaligned, using phonation as the primary cue in production but F1 as the novel primary cue in perception. In contrast, younger speakers were categorized as 'realigned,' using the novel F1 cue in both domains. In sum, misalignment occurred during intermediate stages of change and perception led production for both early- and mid-stage change.

Pinget et al. (2020) show comparable results for regional bilabial stop and labiodental fricative devoicing mergers in Dutch, where perception tends to precede production for change in initial stages. Interestingly, their results indicate a reversal during later stages of change such that perception remained comparatively less merged for individuals whose change is near completion in production. In an investigation of a VOT-f0 cue weighting shift in Afrikaans, Coetzee et al. (2018) similarly found that younger speakers were generally 'realigned' at the individual level, but if they were misaligned, production was more advanced. This late-stage perception lag is consistent with research on vowel mergers showing that, based on expectations of the talker, listeners can utilize cues in perception that they do not distinguish in production (e.g., Hay, Warren, & Drager, 2006; Koops, Gentry, & Pantos, 2008).

Factors relating to the source of sound change may separately influence the manifestation of production-perception misalignment. Voeten (2020), for example, finds that sociolinguistic migrants from Belgium who moved to the Netherlands were variable in their approximation of Netherlandic vowels; in general, although production and perception were linked in this case, production seemed to lead perception. This suggests that differences may arise in production-perception mismatch between cases of internally-driven versus contact-driven changes. In other words, studies of second dialect acquisition (e.g., Voeten, 2020; Evans & Iverson, 2007) find that production appears to lead perception, at least in early stages. In other cases of presumably internal changes occurring across generations, perception leads at first (e.g., Kuang & Cui, 2018; Pinget et al., 2020) until the change reaches late stages.

Finally, the type of change may matter in the coordination of production and perception. Most prior studies investigated cue shifting (e.g., Coetzee et al., 2018; Kuang & Cui, 2018) or boundary shifting (e.g., Harrington et al., 2008; Kleber, Harrington, & Reubold, 2012) changes, with a focus on vowels. To quote Kuang and Cui (2018), does "production and perception have the same relationship across different languages, different types of sound change (e.g., merger, cue shifting, or boundary shifting), and different types of coarticulation?" (p. 196).

In the current study, we highlight mergers as an interesting case of change, unique because they represent a total loss of phonological contrast rather than simply a shifting of cues that maintain contrast. This comparatively drastic phonological change could lead to differences in behaviour surrounding production and perception. To the best of our knowledge, only one study of production and perception has been published on segmental mergers (Pinget et al., 2020), for

which reported results agree with those of other studies. Prior studies on tone mergers in Hong Kong Cantonese also report a link between reduced distinction in production and slower responses in perception (Mok, Zuo, & Wong, 2013; Ou & Law, 2017). Here, we present an investigation of a set of consonant mergers in Hong Kong Cantonese, each with its own socio-historical profiles, to test the robustness of the perception-production link across different types of sound changes.

1.3. Consonant mergers in Hong Kong Cantonese

Hong Kong Cantonese (HKC) provides an interesting opportunity to study mergers, as a large set of consonantal (e.g., Wong, 1941; Zee, 1999a; To et al., 2015) and tonal mergers (e.g., Bauer, Cheung, & Cheung, 2003; Mok et al., 2013; Zhang, 2019) have emerged over the course of the last century. Moreover, the complex and ever-changing socio-political landscape surrounding Hong Kong has introduced a multitude of external influences on HKC, with consequences for the trajectory of these mergers. The last few decades, in particular, have seen rapid changes due to the transition of Hong Kong's political status from a British colony to a Special Administrative Region under Chinese rule in 1997 ('the handover'). While already a highly multilingual environment, one notable change is the increasing presence and influence of Mandarin in the city, alongside Cantonese and English, especially for younger generations (e.g., in the education system; Lee & Leung, 2012; Wang & Kirkpatrick, 2015; Wong, A., 2019). Whether contact with Mandarin has ultimately affected the HKC sound changes is unclear, though both convergence, such as reversing changes that are dissimilar to Mandarin (Bauer & Benedict, 1997), and divergence, for example maintaining HKC-specific changes that are linked to local identity (Whelpton, 1999), are conceivable. This latter possibility is particularly plausible given that many locals continue to identify as 'Hong Kongers' (as opposed to 'Chinese' or 'Hong Kong-Chinese') and specifically view Cantonese as central to this identity (Lai, 2001; Lai, 2011).

Another layer of complexity is that linguistic ideologies (e.g., the public media campaigns and school curriculum changes spearheaded by Professor Richard Ho Man-Wui, a prominent public scholar and professor of Chinese literature) have led to stigmatization of the innovative variants. The use of mergers is termed 懶音 *laan5 jam1*, which translates roughly to 'lazy pronunciation' (also variably denoted as 'lazy accent,' 'lazy articulation,' and 'lazy sounds'). This notion is recognized by Hong Kongers in the present day (Chen, 2018), and is perhaps particularly salient to younger generations due to the introduction of Cantonese pronunciation in a standardized exam in 2007. This prompted official promotion of 'proper Cantonese pronunciation' on TV and radio, as well as in school activities (Lee & Leung, 2012). As a result of the prevailing prescriptivist approach linking conservative variants with standardness, clarity, and formality, a certain level of explicit awareness and attitudes about the various mergers is available to individuals; this may in turn legitimize hypercorrection to the 'proper' variant.

In Hong Kong, these external factors could have plausibly come together to stall or reverse the progression of the mergers, leading to stable variation or preservation of the conservative form in formal registers (e.g., Labov, 1994; Hickey, 2012). Though HKC consonant mergers have been the focus of a sizable number of sociolinguistic and phonetic production studies in the past (e.g., Bauer, 1982a, Zee, 1999a; To et al., 2015), little to no research appears to have revisited these consonant mergers in recent years. The most recent comprehensive study on the status of HKC consonant mergers is To et al. (2015), which reports on data collected in the early 2000s when the earliest post-handover generation were still children. While large in its sample of speakers, this study was limited in scope and too early to assess post-handover outcomes.

Against this background, the present investigation focuses on three consonant mergers in HKC: onset [n]→[l], syllabic [ŋ̩]→[m̩], and onset [ŋ]↔Ø. **Table 1** provides examples. Despite surface similarities, they present a mix of phonological and socio-historical profiles, including differences in phonetic biases, functional load, cross-language correspondences, and metalinguistic awareness (for a detailed review of the historical trajectories, see To et al., 2015). This variation provides an opportunity to extend previous research on the production-perception link to a set of contrasting mergers in a single speech community, testing the extent to which predictions of production and perception hold across different types of sound change in the same individuals while using the same tasks and measures.

Merger	Chinese Character	Jyutping Romanization		English Translation
		Historical	**Innovative**	
[n]→[l]	藍	*laam4*		'blue'
	男	*naam4*	*laam4*	'male'
[ŋ̩]→[m̩]	唔	*m4*		'not'
	五	*ng5*	*m5*	'five'
[ŋ]↔Ø	嘔	*au2*	*ngau2*	'vomit'
	牛	*ngau4*	*au4*	'cow'

Table 1: Examples of Cantonese words involved in each merger, and the direction of change.

1.3.1. [n]→[l] merger

The [n]→[l] onset merger represents a prototypical case of contrast loss. These syllable-initial phonemes, described by Zee (1999b) as an apico-laminal denti-alveolar nasal and apical (denti-) lateral approximant, historically differed in both nasality and tongue blade involvement. There is also an asymmetry in frequency: [l]-onsets occur more often than [n] by both measures of type

frequency (approximately double) and token frequency (approximately six times; Leung, Law, & Fung, 2004).

Since the 1940s or earlier (Wong, 1941 as cited by To et al., 2015; see also Bourgerie, 1990), [n] has been documented to be steadily merging towards [l] in progressively younger generations (Hashimoto, 1972; Yeung, 1980; Bourgerie, 1990; Zee, 1999a). Most recently, To et al. (2015) reported that nearly all of their participants, regardless of age group and gender, produced the innovative form [l] in place of the historical /n/ (94.2% of children versus 94.6% of adults). All available evidence thus points to (near-)completion of the merger by the start of the 21st century.

Because the innovative [l] has become so prevalent, Chen (2018) notes that this merger appears to be comparatively less stigmatized than other pairs such that "most users do not consider them wrong" (p. 7). Nevertheless, hypercorrection from historical [l] to [n] has been described—albeit with some debate—in the literature. In support, Pan (1981) reports [n] realizations of historical /l/ in a word list reading task. Zee (1999a) also notes that there are "hypercorrective individuals" who may use [n]. In contrast, Bourgerie (1990) impressionistically found little evidence of hypercorrection in sociolinguistic interviews, further claiming that no other studies had reported the [l] to [n] direction of change and that "most observers believe that this hypercorrection does not occur" (p. 46). Later studies only investigate historical /n/ items, leaving unknown the degree to which hypercorrection of [l] to [n] is present in HKC.

It has further been specifically noted that [n]→[l] varies by speech register such that [n] is associated with formality (Pan, 1981) and careful speech (Zee, 1999a). This is supported by production data, as Bourgerie (1990) found that impromptu, conversational speech included the most [l] realizations (80.5%), trailed by interview speech (58.2%) and public speech (38.6%). Thus, even while casual speech may mostly comprise innovative variants, HKC speakers also have access and exposure to speech registers that include the conservative [n] variant. The presence of stylistic variation complicates the interpretation of merger completion in this speech community across sporadic research reports with varying methodologies.

1.3.2. [ŋ̩]→[m̩] merger

The [ŋ̩]→[m̩] merger involves two historical phonemes, the syllabic velar and bilabial nasal consonants, that were largely non-contrastive. [ŋ̩] historically occurred in relatively few lexical items, limited to the three low tones. Of these, Bauer (1982a) contends that only four occur with any frequency in spoken HKC: two common words (五 *ng5* 'five,' 午 *ng5* 'noon') and two surnames (吳 *ng4* and 伍 *ng5*). On the other hand, [m̩] historically occurred only as a single—albeit highly frequent—morpheme: the negation marker 唔 *m4* 'no, not.' Both categories therefore contained few types but still maintained a phonemic contrast. The single morpheme for /m̩/ combined with the rather limited inventory of /ŋ̩/ words suggests that contrastiveness has been largely uncompromised over the course of the merger.

Though this merger was first documented in the 1980s (Bauer, 1982a; 1982b), evidence suggests that the [ŋ]→[m̩] change was well underway by the late 1970s and likely initiated around the 1940s from labial assimilation for the highly frequent word 'five' (Bauer, 1986). Since then, To et al. (2015) reported that nearly all children (born 1993-1994) and adults (born 1960-1987) produce the innovative [m̩]. Specifically, 98.6% of children and 95.5% of adults used [m̩] for the word 'five.' Given that this study collected data only for the most frequently occurring historical /ŋ/ word, the generalizability of these results to other words of this category is unknown. At face value, however, the [ŋ]→[m̩] merger appears to be near completion at the community level, and possibly complete for a great many individual speakers.

Some early sources demonstrate general or public awareness of the [ŋ]→[m̩] merger: Newspapers mentioned homophony of the negative morpheme *m4* and the surname *ng4* in 1980 and 1981, while a Chinese dictionary listed 'five' with both velar and bilabial pronunciations (Lau, 1977, cited by Bauer, 1982a). As one of the mergers identified by proper pronunciation campaigns, it can be assumed that awareness has been maintained, though the degree of salience or stigmatization of this particular merger is uncertain. Finally, like [n]→[l], the conservative [ŋ] is associated with formality and can occur in style-shifting (e.g., [m̩] is produced more often during spontaneous speech than story or word list reading; Bauer, 1982a).

1.3.3. [ŋ]↔Ø merger

Unlike the former two mergers, the [ŋ]↔Ø merger—involving the syllable-initial velar nasal and its null-initial (Ø) counterpart (also referred to as zero-initial, phonetically either a vowel or glottal stop onset; Bourgerie, 1990; Chen, 2018)—is described as historically allophonic, such that [ŋ] onset occurred with low tones while null-initial occurred with high tones. While this merger is not technically a merger of phonemic categories, it is a merger of lexical classes and is discussed as a merger within the Cantonese literature (e.g., To et al., 2015). To complicate matters, this merger has also undergone a reversal in direction over time, leading to a loss of the historical pattern of complementary distribution.

Since the start of the 20th century, both historical [ŋ] and Ø have shown evidence of merging towards the other (note some exceptions of [ŋ] onset occurring with mid to high tones; Ball, 1907 as cited in To et al., 2015). Specifically, reports suggest that Ø→[ŋ] took place in the first half of the century (Wong, 1941 as cited in To et al., 2015), which was subsequently supplanted by a [ŋ]→Ø change in the second half (Yeung, 1980; Bourgerie, 1990). Along with the change in merger direction, younger speakers (roughly born since the 1960s–70s) used null-initial more often than not for *both* historical /ŋ/ and Ø (Young, 1980; Bourgerie, 1990; Zee, 1999a). Likewise, the children in To et al. (2015) produced more null-initial variants for both historical /ŋ/ (65.0%) and historical Ø (94.2%), in contrast to the adult proportions of Ø production for historical /ŋ/ (37.5%) and historical Ø (68.7%). However, both groups differentiated historical

/ŋ/ and Ø historical categories, using null-initial more for historically Ø items. Thus, in spite of the bidirectional merger and variation in both historical classes, the two categories remain distinct to a degree from a community standpoint, although the original tone-based complementary distribution does appear to be lost.

A further complexity arises when we compare the results of To et al.'s (2015) survey to Bourgerie (1990). In To et al. (2015), children, as compared to adults, produced significantly more Ø for /ŋ/ (65.0% versus 37.5%) and significantly less [ŋ] for Ø (5.8% versus 31.3%). These group values are conspicuously similar to Bourgerie's (1990) children and young adult group values of Ø for [ŋ] (66.2% versus 36.7%) and [ŋ] for Ø (0% versus 25.4%), despite the fact that Bourgerie's child speakers (born 1970s–80s) correspond roughly to the adult cohort of To et al.'s study (born 1960–87). Since the methodologies of these two studies were quite different, direct comparisons may not be appropriate; however, taken at face value, there appears to have been no progression of the sound changes at the community level between the 1980s to the 2000s.[1]

One potential explanation is that stigmatization of the null-initial as 'lazy' led to age-based patterning across the community, such that adult populations tend to produce the more 'proper' and 'prestigious' [ŋ] variant for both historical classes. Indeed, according to Chen (2018), this merger is more socially stigmatized compared to mergers like [n]→[l], perhaps due to its relative rarity. As with the other mergers, style-shifting is reported to occur for [ŋ]→Ø (Bourgerie, 1990) such that impromptu speech featured the highest rate of null-initials for historical [ŋ] words (35.6%), followed by interview (21.8%) and public speech (8.7%). Further investigation, particularly of speakers born in the 1990s in more recent years, is necessary before coming to a conclusion about the status of the [ŋ]↔Ø variation.

1.4. The present study

The present study investigates the production and perception of the [n]→[l], [ŋ̍]→[m̩], and [ŋ]↔Ø mergers across two generations in Hong Kong, examining both group-level and individual-level patterns. By studying phonetic variation and change in this context, we hope to contribute new insights to unresolved questions in the sound change literature, including those on the phonetic basis of sound change. Our research objectives are two-fold.

The first goal is to test how production-perception patterns generalize across sound changes with varying attributes. In doing so, we seek to clarify the factors that modulate the existence and nature of the production-perception link, for which evidence has been inconsistent.

---

[1] In a study on the frequencies of all Cantonese word initial sounds, Ng and Kwok (2004) do report that, comparing to an earlier survey from the 1970s (Fok, 1979), the overall occurrence of [ŋ-] dropped from 2.9% to 1.91%. This suggests that there has been a general decrease of [ŋ] in favor of Ø.

Characteristics that may be relevant include source of change (internal versus contact-induced), stage of change (e.g., early- versus late-stage), and type of change (shifting versus merging). The present study explores the comparison between three internally-driven consonant mergers within the same speaker-listeners using a consistent methodology. While [n]→[l] and [n̩]→[m̩] appear to similarly be late-stage changes (compared to the mid-to-late stage [ŋ]→Ø) according to recent report, these mergers are further differentiated by a host of other factors. We interpret these case studies in the context of previous results that have focused on cue-shifting and boundary-shifting vowel changes, with particular attention to the results of Pinget et al. (2020) who studied a pair of consonant voicing mergers.

The second objective is to add to documentation on the progression of these mergers specifically in the post-handover era of Hong Kong, revisiting several open threads in the HKC literature. Hong Kongers who grew up in the years since 1997 have had vastly different formative experiences from preceding generations, including increased exposure to Mandarin and reinforcement of 'proper pronunciation' ideologies that brand the mergers as 'lazy' (Chen, 2018). What can we glean about the status of the mergers across generations, taking into account these shifts in the sociolinguistic landscape? Because this aspect of the study is descriptive in nature, we seek to generally explore the patterning of variant use across demographic groups. Nonetheless, because of the historical context of these mergers as sound change, we place special interest in the extent to which age co-varies with usage of each phonetic variant and/or degree of lexical contrast between each pair. Specifically, based on their documentation of children's production patterns in the early-2000s, To et al. (2015) determined that the [n]→[l] and [ŋ]→[m̩] mergers were near completion while Ø-initial was gaining ground over [ŋ]. Do these patterns and their predicted trajectories hold out? Alongside providing a more recent picture of production norms, we contribute to this body of literature by adding a perceptual component through a word categorization task.

2. Methods

2.1. Participants

As part of a larger project, data was collected from Cantonese speakers in both Hong Kong and Vancouver, Canada but only the Hong Kong data are discussed in this paper. This study received approval from the Human Subjects Ethics Sub-committee of the Hong Kong Polytechnic University (HSEARS20160829002), and participants gave written consent before taking part in the experiment. Data collection took place between September 2016 and February 2017. To examine the merger trajectories in apparent time, participants were recruited across two age groups: older middle-aged individuals and younger college-aged individuals.

Fifty-one participants were recruited from the Hong Kong Polytechnic University. The older generation consisted of 23 speakers (M = 54.09 years, *SD* = 5.20, range = 44–60), 12 of whom identified as female and 11 male. The younger generation consisted of 28 speakers (M = 19.42 years, *SD* = 2.28, range = 17–27), 13 of whom identified as female and 15 male. All reported Cantonese as a first language and all had lived in Hong Kong since birth, except one young man who had moved to Hong Kong before age three. Due to recording errors, production data from two participants were excluded; as such, only 49 participants were included in the production and production-perception analyses. Along with Cantonese, all participants reported some level of ability to speak and/or understand English and Mandarin, with variation in reported proficiency across domains. Summary demographics for the participants are provided in **Tables 2** and **3**.

2.2. Materials

2.2.1. Production stimuli

Production stimuli consisted of 64 Cantonese words presented visually in Chinese orthography[2] accompanied by an English translation (full word list in Appendix). Each target historical onset (/l/, /n/, /ŋ/, Ø) appeared in four words while each target historical syllabic nasal (/ŋ̩/, /m̩/) appeared in three, totalling 22 target words. The remaining 42 consisted of filler word pairs with non-target Cantonese contrasts, including onset stop aspiration (/ts/-/tsʰ/, /t/-/tʰ/, /k/-/kʰ/) and other merging consonants (onset /k/-/kʷ/; coda /n/-/ŋ/, /t/-/k/).

Age Group	Gender	n	Age	Cantonese AoA	English AoA	Dominance Score
Older	F	12	55.25 (4.90)	0	6	–81 (39)
	M	11	52.82 (5.46)	0	4	–95 (39)
Younger	F	13	18.77 (1.74)	0	3	–91 (25)
	M	15	20.00 (2.59)	0	3	–84 (26)

Table 2: Summary demographic information, including mean age (standard deviation), median Cantonese and English age of acquisition (AoA), and mean Cantonese-English language dominance score (standard deviation) as calculated by the BLP.

---

[2] While including orthography in our tasks departs from methodology in previous research (e.g., picture naming) and likely elicits particular styles of speech that are more formal than spontaneous speech, because target word selection was constrained by the limited options for the [ŋ]→[m̩] and [ŋ̩-]↔Ø mergers (especially with the requirement of minimal pairs for the perception task), the available target words were not easily picturable. In order to maintain comparability across domains, we chose to include orthography for both production (reading) and perception (word selection) tasks; thus, production and perception tasks were well-matched methodologically. For more discussion, see Section 4.1.1.

Age Group	Gender	n	Cantonese		English		Mandarin	
			Speak	Under-stand	Speak	Under-stand	Speak	Under-stand
Older	F	12	5.50	5.50	3.92	4.25	2.25	2.75
	M	11	5.82	5.91	3.82	4.27	3.18	3.73
Younger	F	13	5.46	5.38	3.77	3.69	3.62	4.08
	M	15	5.07	5.13	3.43	3.73	2.93	3.87

Table 3: Mean ratings (on a scale from 0–6) of self-reported speaking and comprehension abilities for Cantonese, English, and Mandarin per demographic group.

2.2.2. Perception stimuli

Perception stimuli consisted of three 13-step continua generated between Cantonese real-word minimal pairs, one per merger as listed in **Table 4**. To create these continua, the target minimal pairs were produced by a 33-year-old Cantonese speaker and trained linguist who grew up in Hong Kong. Stimuli were digitally recorded at a sample frequency of 44.1 kHz in a sound-attenuated booth using an AKG C520 headset mic with a USBPre 2 Pre-Amp. The speaker read from transcriptions of each target word, accompanied by the Chinese orthography, to elicit maximally contrastive pronunciations. She was asked to produce each pair naturally but as similarly as she could (aside from the target contrast) to facilitate more natural-sounding synthesis. Each target word was produced at least three times. The clearest and most similar recorded pairs were selected for synthesis. Decisions were based on duration, pitch, and intonation similarity assessed by the first author via auditory and spectral analysis in Praat (Version 5.4.08, Boersma & Weenink, 2015).

These six natural productions were then used as endpoints to create three word-pair continua using TANDEM-STRAIGHT in Matlab (Kawahara et al., 2008). The entire word forms were used as the endpoints, which, given TANDEM-STRAIGHT's global synthesis methods involving acoustic decomposition and generation, allows for natural-sounding resynthesis that retains redundant co-variation. Twenty-five acoustically-equidistant steps were synthesized between the duration of the target words, with each pair time-aligned at acoustic landmarks (e.g., phone boundaries) to facilitate morphing. From these, every odd-numbered step was selected to result in thirteen equidistant steps. The resulting continuum steps gradually morph from one endpoint token to the other, thus including in the synthesis an array of phonetic differences across the tokens. The natural endpoints were carefully selected to be as similar as possible in terms of non-contrastive elements. As such, the main perceptible change across continua is between the target sounds in each pair. The continua are available at https://osf.io/mk2v4/?view_only=51cfc7a974f345678e226944023dcb39. **Figure 1** illustrates the transition between the

initial sound in the aak1-ngaak1 continuum (i.e., the beginning of the vowel [aː] and the entire duration of [ŋ]) with waveforms and spectrograms of the midpoint and endpoint recordings. Judgments from the first author and two other Cantonese-speaking colleagues were used to select the most natural-sounding continuum per merger. In all, the 13 variants from each continuum totalled 39 tokens.

Target Merger	IPA	Jyutping Romanization	Chinese orthography	English translation
[n]→[l]	[lou] - [nou]	*lou5-nou5*	老-腦	'old'- 'brain'
[ŋ]→[m̩]	[m̩] - [ŋ]	*m4-ng4*	唔-吳	'no, not'- 'Ng (surname)'
[ŋ]↔Ø	[aːk˺] - [ŋaːk˺]	*aak1-ngaak1*	握-呃	'shake [hands]' - 'deceive'

Table 4: Minimal word pairs for each synthesized continuum.
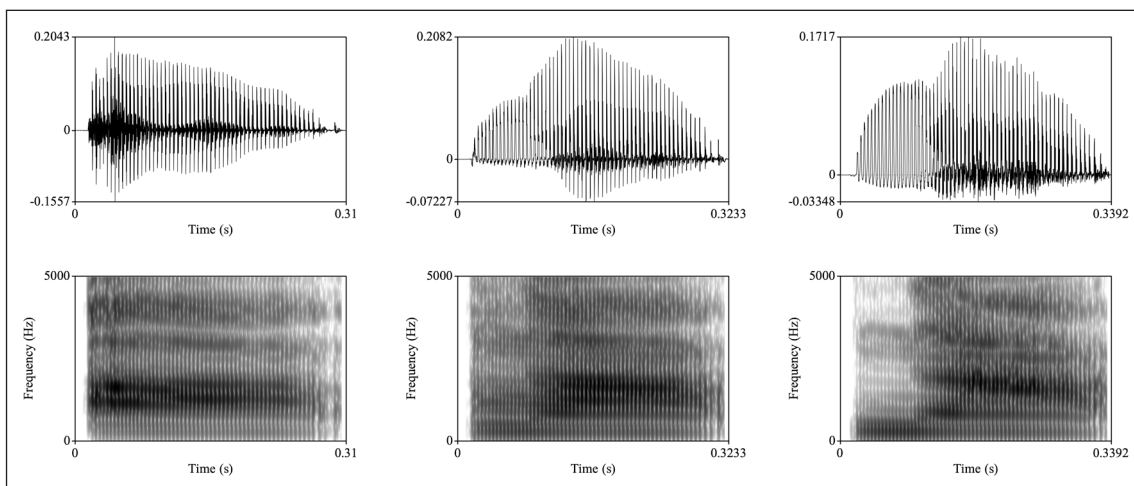
Figure 1: Waveforms (top) and spectrograms (bottom) of the endpoints and midpoint of the final 13-step *aak1-ngaak1* continuum. From left to right: null-initial [aːk˺] (step 1), maximally ambiguous initial (step 7), velar nasal-initial [ŋaːk˺] (step 13).

2.2.3. Language questionnaire

Participants completed the Bilingual Language Profile (BLP; Birdsong, Gertken & Amengual, 2012), which was used to quantify Cantonese-English language dominance scores. A small number of modifications were made to better reflect the language experiences of the population (e.g., accounting for the education system in Hong Kong and the role of Mandarin in Hong Kong, alongside Cantonese and English). Additional questions about biographical information, language background, and language experiences were included to adjudicate participant inclusion criteria. Scores were calculated from the BLP portion of the questionnaire. The BLP is bound between

–218 to +218. Zero indicates balanced dominance between the two languages while negative and positive scores represent relative Cantonese and English dominance, respectively. The range was negative (–165, –2), indicating Cantonese dominance for all speakers. Mean values and standard deviation for the four demographic groups are reported in **Table 2**. An analysis of variance with the BLP dominance score as the dependent measure and Age and Gender as independent variables revealed no significant effects. This confirms the uniform Cantonese dominance in the sample, and BLP scores are not considered further.

2.2.4. Post-task awareness interview

An exit interview was administered to examine metalinguistic awareness about the mergers in the participant's own speech and experiences. The interview was conducted in English or Mandarin, supplemented by Cantonese when necessary to ensure comprehension.[3] Participants were presented with a list of minimal or near-minimal word pairs designed to historically contrast the target sounds. For each pair, they were asked to say the words as they would in casual speech, explain any differences between their productions, and name the prescribed pronunciation if they were aware of any. At the end, participants were asked if they were familiar with the term 懶音 *laan5 jam1* 'lazy pronunciation,' what they thought it meant, and whether they believe (or have been told by others) that they use it in their own speech.

2.3. Procedure

2.3.1. Experiment

Participants were seated alone at a computer in a sound-attenuated booth where they were presented with the production task followed by the perception task. Instructions were provided visually in English on the screen. Upon completion, the experimenter returned to the room to answer any questions while participants filled out the language questionnaire, then conducted the exit interview. To reduce the chance that participants would change their speech behaviour due to realizing the purpose of the experiment (i.e., our interest in the target sounds), the production task, which included fillers, was ordered prior to the perception task, which involved only the target sounds.

In the self-paced production task, Chinese characters and the English translation were visually presented using E-Prime 2.0 software (Schneider, Eschman, & Zuccolotto, 2007) in three randomized blocks, where each word appeared once per block (192 trials in total). Participants

---

[3] Interviews were not conducted fully in Cantonese mainly due to limitations of available personnel, but an added advantage is that we minimize the potential for an interviewers' pronunciation of Cantonese (e.g., if they used merged variants) to have influenced participants' response to minimal pair comparisons and 'lazy pronunciation'-related questions. In addition, we note that because participants only interacted with the experimenter outside the main task and that the post-task interview was conducted only for additional context, there is no reason to expect this choice to impact the study results in any meaningful way.

were asked to read the Chinese characters out loud. Productions were digitally recorded using an AKG C520 headset microphone connected to a UR22MKII interface. Participants pressed the zero key on the keyboard to begin a two-second recording for each trial. They were allowed to re-record as many times as they wished before moving on to the next word by pressing the spacebar. Participants spent approximately 20–30 minutes on this task.

Perception stimuli were auditorily presented in a two-alternative forced choice lexical identification task. To reduce influences of explicit knowledge about 'proper' pronunciation, participants were instructed to respond as quickly as possible and not overthink their response. Using E-Prime 2.0, the synthesized Cantonese words were played over AKG K77 Perception headphones at a comfortable listening level, accompanied by a visual display of the appropriate minimal pair words in Cantonese orthography and the English translation. For each trial, participants heard the audio stimulus once and saw two word choices labelled with '1' or '5' (on the left or right side of the screen). They were asked to press the button on the button box (i.e., '1' or '5') that corresponded to the word they heard. If no response was registered within three seconds, the next trial began automatically. Each token was repeated three times throughout the experiment, randomly presented over three blocks (117 trials in total).

2.3.2. Production coding

To assess production, two phonetically-trained Cantonese speakers who grew up in Hong Kong coded the onset of each item, or in the case of the syllabic nasals, the sole segment comprising the word. Items were blocked by talker, randomized, and presented to the coders blind, without knowledge of the intended lexical items. Coders categorized the onset from a closed set, presented orthographically as six options: 'l,' 'm,' 'n,' 'ng,' 'a vowel,' or 'other.' The 204 items identified by both transcribers as 'other' were coded by the first author. These items either did not include a usable production (e.g., recording errors where the word was cut off or missing) or included a mispronounced production (e.g., *saam1* 'clothing' instead of *lau1* 'coat').

The coders agreed on 4933 trials. Inter-rater reliability for the two coders was calculated in R (R Core Team, 2020) using the Kappa statistic for two raters unweighted using the kappa2() method from irr() (Gamer, Lemon, Fellows, & Singh, 2019) from the perspective of whether there was agreement on the transcription of each item's coded onset. These data from the two coders are reported in **Table 5**. Kappa scores ranged from values that are in the range described as near perfect agreement for historical /l/ and /n/-initial words to moderate agreement scores for the historical syllabic nasals. To resolve disagreement, all items (n = 725) on which the two coders disagreed were presented in a similar format to the first author: blinded as to their lexical identity and organized by talker. Sixteen of these disagreement items were removed due to recording errors or mispronunciations. The ultimate coding for each item was then determined as whichever category two out of the three coders agreed.

Historical Phoneme	Number of Items Coded	Kappa Statistic	*p*-value	Percentage Agreement
[l]	1067	0.82	<0.001	95.9%
[n]	1327	0.85	<0.001	90.4%
[m̩]	801	0.48	<0.001	92.5%
[ŋ̩]	801	0.46	<0.001	79.3%
[ŋ]	798	0.60	<0.001	84.5%
Ø	1068	0.59	<0.001	81.0%

Table 5: Inter-rater reliability measures for the auditory coding of critical items from the perspective of the historical phoneme.

3. Results

3.1. Production

3.1.1. [n]→[l] merger

Data for the [n]→[l] merger (n = 1158) were analyzed in a logistic mixed effects regression model with the probability of [l] productions ([l] = 1, [n] = 0) as the dependent measure. All categorical variables were sum coded (Historical Pattern: /l/ = 1, /n/ = –1; Talker Age: Older = 1, Younger = –1; Talker Gender: Male = 1, Female = –1) and entered as possible main effects and interactions. There were random intercepts for Subject and Item, with Historical Pattern as a by-subject random slope.[4]

	B	Standard Error	*z*-value	*p*-value
Intercept	**2.4026**	**0.52**	**4.62**	**3.83E-06**
Historical Pattern	**1.1601**	**0.3818**	**3.038**	**0.00238**
Age Group	0.2192	0.4088	0.536	0.59182
Gender	0.2537	0.4084	0.621	0.5344
Historical Pattern : Age Group	–0.297	0.2121	–1.4	0.16156
Historical Pattern : Gender	0.1822	0.2112	0.863	0.38839
Age Group : Gender	–0.0898	0.4077	–0.22	0.82567
Historical Pattern : Age Group : Gender	0.1713	0.2109	0.812	0.41668

Table 6: Model output for [n] and [l] merger in production. Significant factors (α level = 0.05) are bolded.

---

[4] The model syntax used was: glmer(ProportionL ~ HistoricalPattern * AgeGroup * Gender + (1+HistoricalPattern|Subject) + (1|Item), family="binomial").

The model output is summarized in **Table 6**. There was a significant intercept, and a significant effect of Historical Pattern. These results indicate that [l] productions are overall more likely, and that [l] productions are even more likely on words that are historically pronounced with /l/. These results are visualized in **Figure 2**, separated by the non-significant factors of Talker Age and Gender in order to transparently present the individual variation that is present. The figures present a barplot to indicate the group patterns, along with individual data points for each individual; these individual data points are somewhat transparent so that individual overlap is observable. Lines connect an individual's data points for the two word classes. While the group averages indicate that both historically /l/ and /n/ words are usually pronounced with [l], **Figure 2** illustrates how there is both group-level separation between historical categories (confirming the model result) and considerable variation on the individual level.
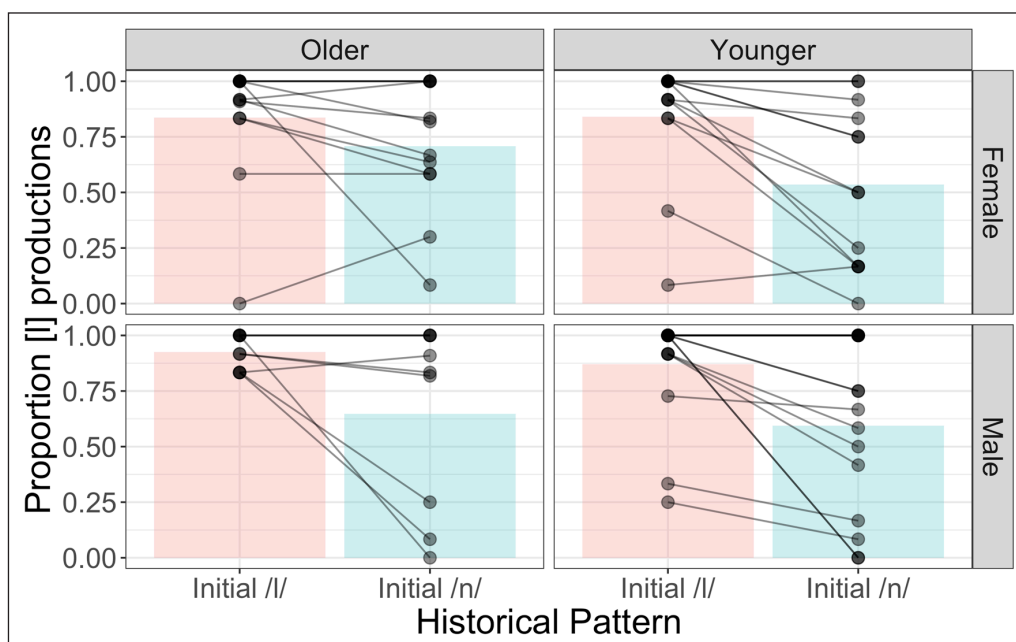
Figure 2: Group means and individual data points for proportion [l] productions for historically /l/- and /n/-onset words, faceted by age and gender. The lines connecting data points connect values for an individual. The shading of the individual points is to allow the visualization of individual overlap.

To better visualize the individual-level data, we also present each individual's mean proportion of the innovative variant for the two historical categories as a scatter plot in **Figure 3**. Here, the bottom right quadrant represents conservative contrastive productions ([n] for /n/ and [l] for /l/), the top right quadrant represents innovative merged productions ([l] for /n/, [l] for /l/), the bottom left represents hypercorrective merged productions ([n] for

/n/, [n] for /l/), and the top left represents a fully flipped contrastive production ([l] for /n/, [n] for /l/). Most participants are clustered in the top right quadrant, realizing /n/ as [l] the majority (or all) of the time, though there is considerable variation ranging to fully historically contrastive for a small number of participants in the bottom right corner. Participants who lean conservative include a mix of older and younger individuals. Five participants who are mainly younger (two younger men, two younger women, and one older woman) appear relatively merged in the hypercorrective direction, realizing a majority of both /n/ and /l/ words as [n]. As expected, none are contrastive in the opposite-to-historical direction. This visual inspection of the individual variation further supports the non-significance of age and gender as a predictor of innovative variant productions.
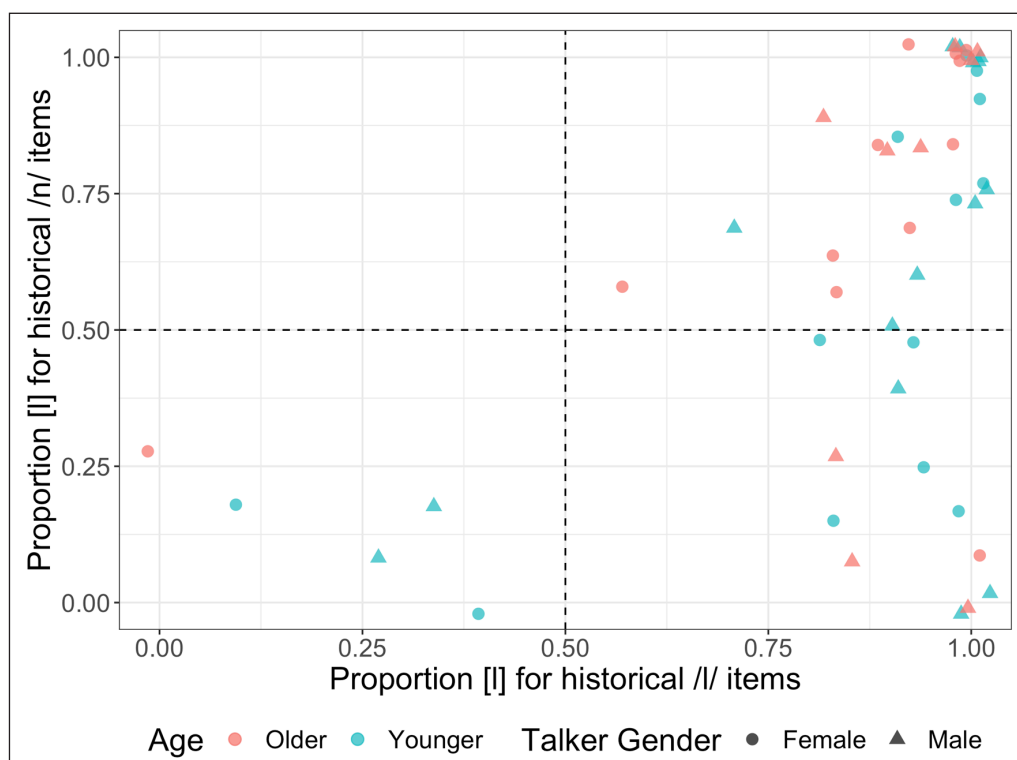
Figure 3: Proportion [l] for historically /n/ items (y-axis) and historically /l/ items (x-axis) for each individual. Women are plotted with circles and men with triangles. Older speakers are in red and younger speakers in blue. The values have been jittered to minimize overlap.

3.1.2. [ŋ̩]→[m̩] merger

Data for the [ŋ̩]→[m̩] merger (n = 872) were analyzed with a logistic mixed effects regression model in an identical fashion to those for the [n]→[l] merger, with the variant-specific variables

coded analogously (i.e., dependent variable as the probability of [m̩] productions: [m̩] = 1, [ŋ̍] = 0; Historical Pattern: /m̩/ = 1, /ŋ̍/ = –1).

	B	Standard Error	z-value	p-value
Intercept	**3.21905**	**0.57616**	**5.587**	**2.31E-08**
Historical Pattern	**1.08923**	**0.49451**	**2.203**	**0.02762**
Age Group	**0.71872**	**0.36322**	**1.979**	**0.04784**
Gender	**1.14734**	**0.3618**	**3.171**	**0.00152**
Historical Pattern : Age Group	–0.2848	0.21884	–1.301	0.19312
Historical Pattern : Gender	0.05902	0.21788	0.271	0.78649
Age Group : Gender	–0.03475	0.35812	–0.097	0.92271
Historical Pattern : Age Group : Gender	–0.04729	0.21276	–0.222	0.8241

Table 7: Model output for [ŋ̍] and [m̩] merger in production. Significant factors (α level = 0.05) are bolded.

The model output is summarized in **Table 7**. There was a significant intercept, along with significant effects of Historical Pattern, Age Group, and Gender. There were no significant interactions. The results are visualized in **Figure 4** separated by Historical Pattern, Age Group, and Gender with the bars showing the group pattern, overlaid with individual results. These visualizations corroborate the statistical output. Overall, talkers are more likely to use [m̩] for both historical word classes, but the effect of Historical Pattern confirms that [m̩] is more likely for historically /m̩/ words. In addition, older speakers are more likely to show the innovative pattern, using higher rates of [m̩], while male participants are significantly more likely to use [m̩] than female participants.

Figure 5 presents the individual proportion scatterplot for [ŋ̍]→[m̩], as described for [n]→[l]. The vast majority of participants are located in the top right quadrant, representing innovative productions where both /m̩/ and /ŋ̍/ are realized as [m̩] the majority of the time. Many individuals are fully merged, clustering in the top right corner. There is some variation in proportions of [m̩] productions, though no participants produce flipped categories, and two younger women produce a hypercorrective pattern (majority [ŋ̍] for both /ŋ̍/ and /m̩/). Overall, this visualization makes clear that the individuals who produce variable, conservative, or hypercorrective patterns are overwhelmingly younger women (along with some older women and younger men), confirming the model Age and Gender effects where women and younger participants skew towards producing [ŋ̍].
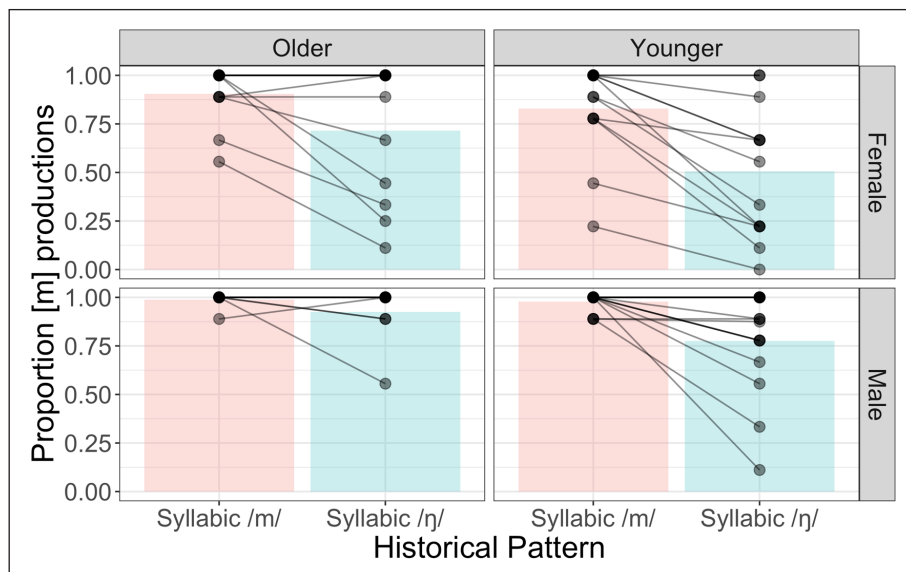
Figure 4: Group means and individual data points for proportion [m̩] productions for historically syllabic /m̩/ and /ŋ̍/ words, faceted by age and gender. The lines connecting data points connect values for an individual. The shading of the individual points is to allow the visualization of individual overlap.
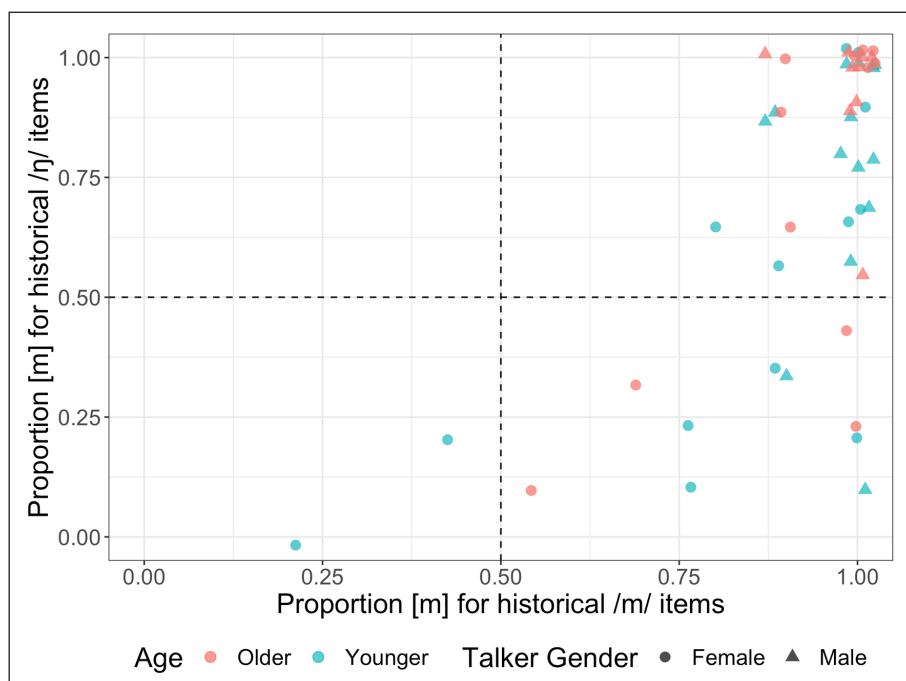

Figure 5: Proportion [m̩] for historically /ŋ̍/ items (y-axis) and historically /m̩/ items (x-axis) for each individual. Women are plotted with circles and men with triangles. Older speakers are in red and younger speakers in blue. The values have been jittered to minimize overlap.

3.1.3. [ŋ]↔∅ merger

Data for the [ŋ]↔∅ merger (n = 1023) were analyzed similarly to the two previous mergers with a logistic mixed effects regression model. The variant-specific variables were coded with null-initial assumed as the innovative variant (i.e., dependent variable as the probability of vowel-initial productions: ∅ = 1, [ŋ] = 0; Historical Pattern: ∅ = 1, [ŋ] = –1).

	B	Standard Error	z-value	p-value
Intercept	**–2.83844**	**0.64677**	**–4.389**	**1.14E-05**
Historical Pattern	**0.58628**	**0.28279**	**2.073**	**0.0382**
Age Group	0.43279	0.59628	0.726	0.468
Gender	0.60362	0.58385	1.034	0.3012
Historical Pattern : Age Group	0.10844	0.16359	0.663	0.5074
Historical Pattern : Gender	–0.09481	0.16497	–0.575	0.5655
Age Group : Gender	0.62266	0.58525	1.064	0.2874
Historical Pattern : Age Group : Gender	–0.22533	0.18756	–1.201	0.2296

Table 8: Model output for vowel initial and initial [ŋ] merger in production. Significant factors (α level = 0.05) are bolded.

The model output is summarized in **Table 8**. There was a significant intercept, and a significant effect of Historical Pattern. The results are visualized in **Figure 6** separated by Historical Pattern, along with the nonsignificant effects of Age Group and Gender. Again, the bars demonstrate the group pattern and individuals' results are presented as a top layer. Participants are more likely to use [ŋ] for all of these lexical items, with the effect of Historical Pattern indicating that vowel initial productions are more likely on historically vowel initial words, but these are still, overall, unlikely. Some clearly prefer vowel-initial onsets, however; individuals from three of the four demographic groups use vowel-initial onsets 100% of the time for both historically vowel-initial and [ŋ]-initial word classes.

These results can be seen more clearly in the individual scatterplot (**Figure 7**). Unlike the other two mergers, the vast majority of participants are located in the bottom left quadrant, representing merged productions where both historical null and [ŋ] are overwhelmingly realized as [ŋ]-initial, including many individuals who are fully merged (clustering in the bottom left corner). Although no age effect was detected in the model, visual inspection of the individual data provides some insight into trends of variation. Specifically, the few participants who show a comparatively conservative pattern (lower right quadrant) are all older while all but one of the participants who show merging towards null (top right quadrant) are younger.
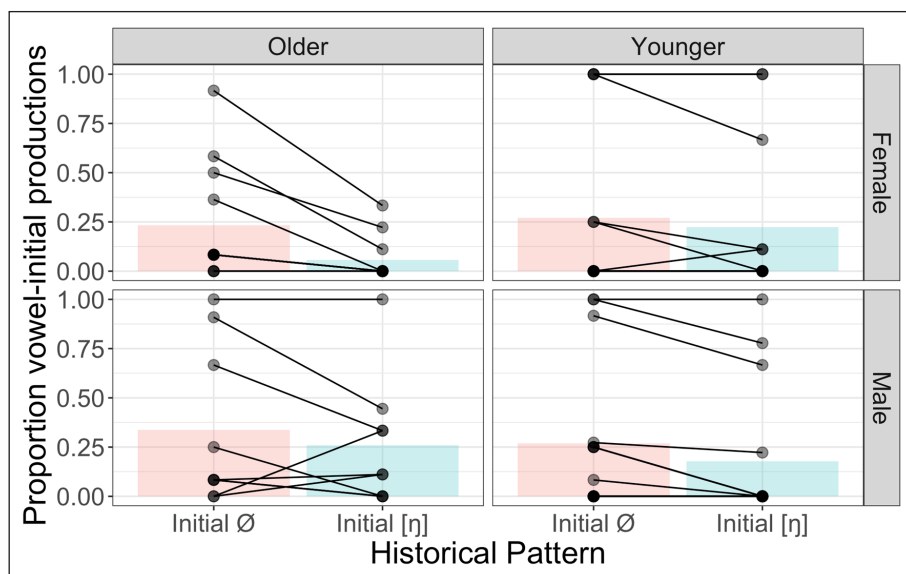
Figure 6: Group means and individual data points for proportion vowel-initial productions for historically [ŋ]- and null-initial words, faceted by age and gender. The lines connecting data points connect values for an individual. The shading of the individual points is to allow the visualization of individual overlap.
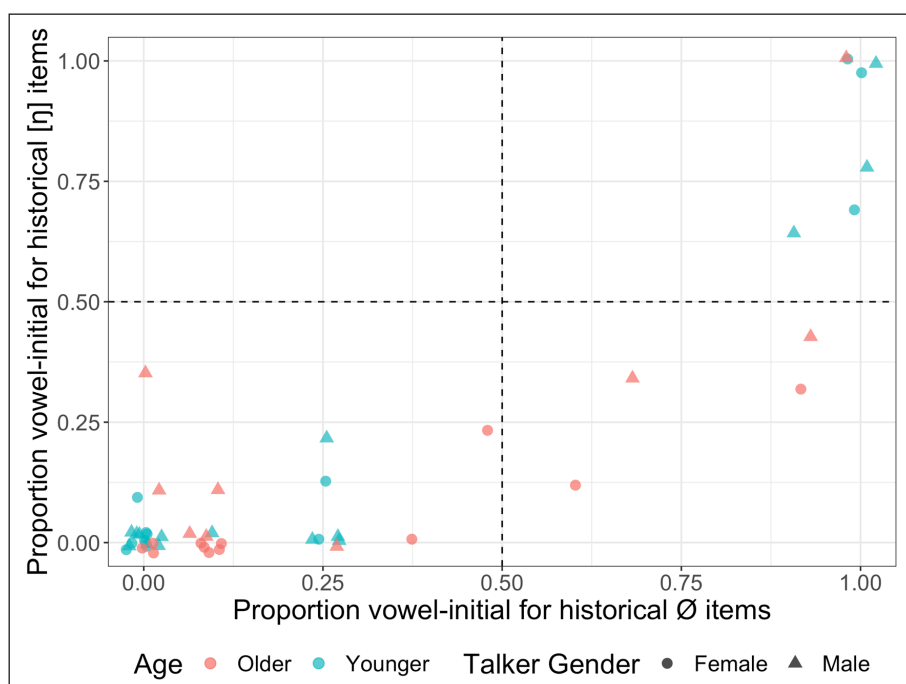

Figure 7: Proportion vowel-initial productions for historically [ŋ]-initial items (y-axis) and historically null-initial items (x-axis) for each individual. Women are plotted with circles and men with triangles. Older speakers are in red and younger speakers in blue. The values have been jittered to minimize overlap.

3.1.4. Summary of production

Group-level production results indicate that all three Hong Kong Cantonese merger pairs maintain some form of community-wide contrast, as indicated by a statistically significant effect of historical category. However, underlying these community averages was a wide variety of merging patterns at the individual level. For [n]→[l], speakers ranged from maximally merged to maximally contrastive, while for [ŋ]→[m̩] and [ŋ]↔Ø, the majority of speakers were either fully merged or demonstrated an intermediate contrast. There was also a subset of individuals who were 'hypercorrective' such that they produced mainly conservative variants for both historical categories (e.g., [n] for both /n/ and /l/). Although degree of contrastiveness varied substantially across individuals, in none of the three mergers is it predicted by demographic factors like age or gender (i.e., no significant interactions were found), which suggests that these mergers should not be characterized as changes-in-progress.

3.2. Perception
3.2.1. [n]→[l] merger

Null responses were removed, accounting for just over 1.5% of the data. The remaining data (n = 3265) were fit to a logistic mixed effects regression model predicting the likelihood of a historical /l/ word response (/l/ = 1, /n/ = 0). Continuum step was centered and scaled, and Talker Age and Gender were sum coded (Age: Older = 1, Younger = –1; Talker Gender: Male = 1, Female = –1). Subject was a random effect with Step as a by-subject random slope.[5]

	B	Standard Error	*z*-value	*p*-value
Intercept	**–0.44226**	**0.11441**	**–3.866**	**0.000111**
Step (centered)	**–1.85118**	**0.23359**	**–7.925**	**2.28E-15**
Age Group	**0.28777**	**0.11327**	**2.54**	**0.01107**
Gender	0.10879	0.11218	0.97	0.332137
Step : Age Group	**0.69128**	**0.23094**	**2.993**	**0.002759**
Step : Gender	0.1023	0.22859	0.448	0.654494
Age Group : Gender	0.05584	0.11223	0.498	0.618813
Step : Age Group : Gender	–0.1741	0.22868	–0.761	0.446471

Table 9: Model output for [n]→[l] merger in perception. Significant factors (*α* level = 0.05) are bolded.

[5] Model syntax: glmer(PropNovel ~ Step_centered * AgeGroup * Gender + (1 + Step_centered|Subject), family = "binomial")

The model output is reported in **Table 9**, and the results are visualized in **Figure 8**. There was a significant intercept, indicating that /l/ word responses were more likely, and main effects of Step and Age. The interaction between Step and Age was also significant. While, overall, participant response functions varied predictably by continuum step with more [n]-like realizations receiving more /n/ word responses, younger listeners present comparatively more extreme responses at the continuum endpoints, resulting in a steeper response function.
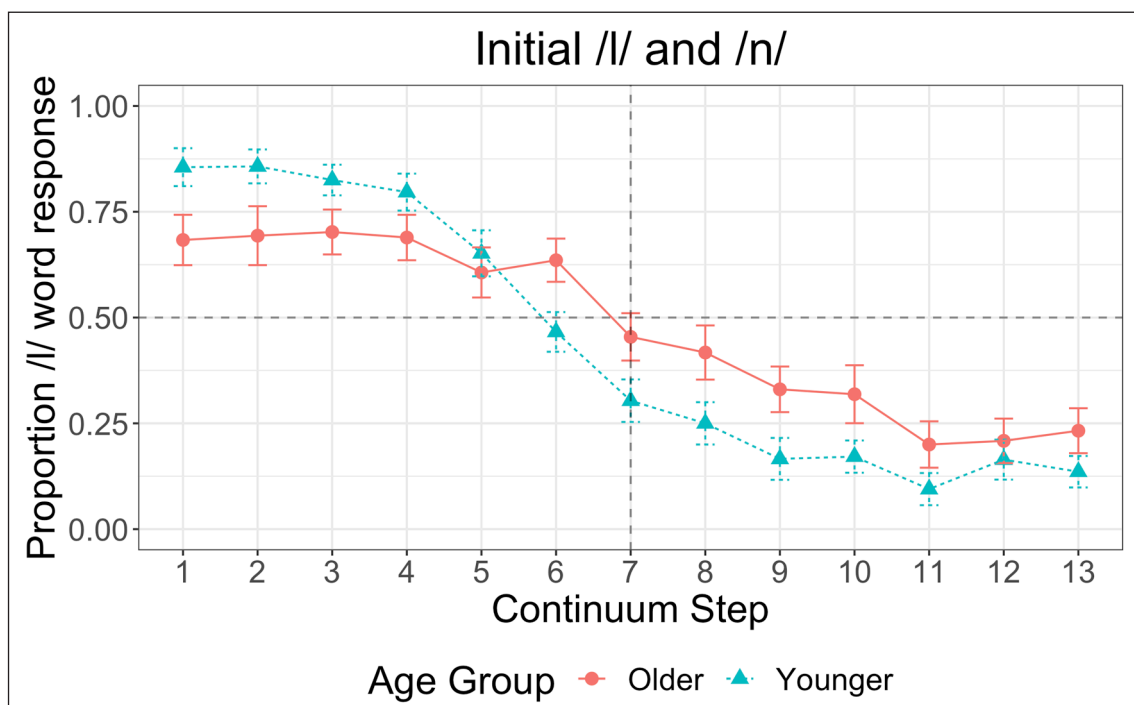
Figure 8: Proportion /l/ word responses for the [n]→[l] merger by continuum step. Step 1 of the continuum is a canonical [l] pronunciation and Step 13 is a canonical [n] pronunciation. The dashed blue line and triangles represent the data of the younger listeners, while older listeners are shown in solid red lines and circles. The error bars show standard error.

To explore the individual differences in the discreteness of these lexical items, individual values were quantified using the by-subject contrast coefficient slope (CCS), following the methods provided by Casillas (2021). This value, which can be taken as a measure of category 'crispness' (Morrison, 2007; Casillas & Simonet, 2016), is an estimate of the slope of each individual's sigmoid at its steepest point. More extreme values indicate more crisp or discrete perceptual categories, while values closer to zero are taken to indicate a less discrete contrast,

which we interpret as evidence of a merger in this case. As a concept, this crispness value acknowledges that phonological categories can exist on a gradient in terms of their categorical contrasts (see, for example, Hall, 2013). These category crispness values were estimated by fitting new logistic regression models with continuum Step (centered and scaled) as the only fixed effect. The left panel in **Figure 9** shows the individual values for the [n] and [l] contrast. There are two extreme outliers with crispness values below –5; both of these individuals had nearly discrete response functions such that the first six steps of the continuum were categorized 100% of the time as /l/ and the last six steps of the continuum was categorized 100% of the time as /n/. This contrasts with the majority of participants who had more gradient sigmoid functions. The central tendency of the value was negative (Mean = –0.35, Median = –0.10) and the range was substantial (–5.33, 0.09).[6] These data, visualized in more fine-grained detail in **Figure 10**, suggest that at the individual level, the degree of category crispness varied considerably. Younger women represent the most variable demographic group, demonstrating a very flat distribution and including comparatively many individuals with crisper categorization; younger men show a similar pattern which is notably flatter than both older groups.
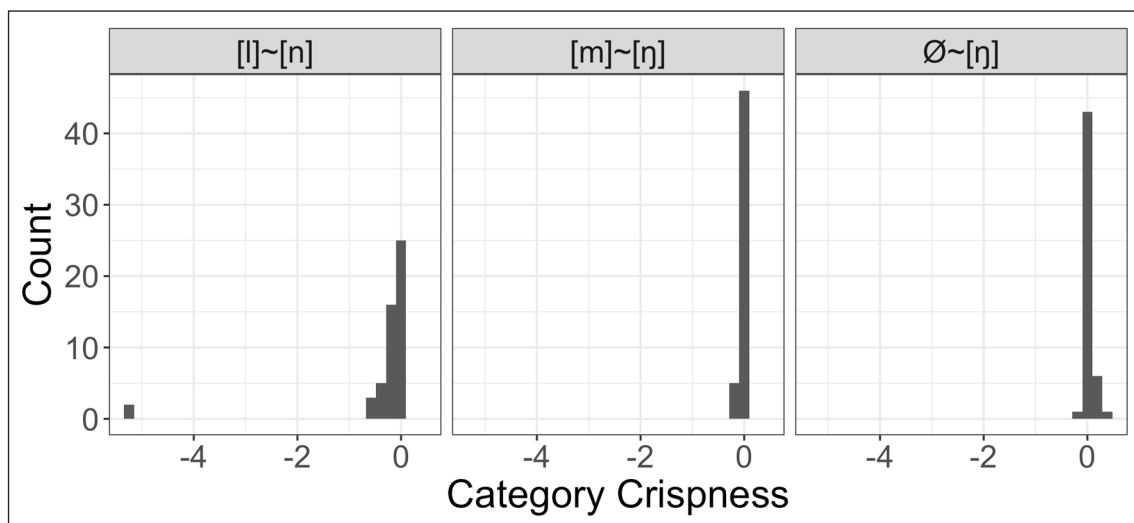

Figure 9: Histograms illustrating the distributions of category crispness values for the three continua.

Figure 10: Histograms illustrating the distribution of individual category crispness scores for the [n]→[l] merger continuum; only crispness scores between –0.6 to 0.3 are presented to show the variation amongst individuals excluding the two extreme outliers. Density plots are underlaid to represent the distribution per demographic group. Women (top) and men (bottom) are plotted separately while older speakers are colored red and younger speakers colored blue.

3.2.2. [ŋ]→[m̩] merger

To analyze the [ŋ]→[m̩] merger, null responses were removed, accounting for just over 2% of the data. The remaining data (n = 3248) were fit to a logistic mixed effects regression model predicting the likelihood of a historical /m̩/ word response (/m̩/ = 1, /ŋ/ = 0). All other model specifications were the same as [n]→[l].

The model output is reported in **Table 10**, and results are visualized in **Figure 11**. There was a significant intercept, indicating that /m̩/ responses were more likely, and there were main effects of Step, a two-way interaction between Age and Gender, and a three-way interaction between Step, Age, and Gender. Older men (M = 0.64) and younger women (M = 0.62) were more likely than younger men (M = 0.54) and older women (M = 0.49) to categorize the continuum as /m̩/. These demographic variables interact with step, and **Figure 11** makes clear that the perceptual changes by step are much smaller than those for the [n]→[l] merger. The word categorization functions suggest that all listener groups are quite advanced in the merger, with younger women presenting the biggest distinction between /m̩/ and /ŋ/.

	B	Standard Error	*z*-value	*p*-value
Intercept	**0.41768**	**0.13979**	**2.988**	**0.002809**
Step (centered)	**–0.40383**	**0.10457**	**–3.862**	**0.000113**
Age Group	–0.05142	0.13962	–0.368	0.712672
Gender	0.0954	0.13971	0.683	0.494698
Step : Age Group	0.15525	0.10419	1.49	0.136216
Step : Gender	0.13829	0.10425	1.327	0.184662
Age Group : Gender	**0.36936**	**0.13979**	**2.642**	**0.008239**
Step : Age Group : Gender	**–0.2807**	**0.10455**	**–2.685**	**0.007257**

Table 10: Model output for the [ŋ]→[m̩] merger in perception. Significant factors (*α* level = 0.05) are bolded.
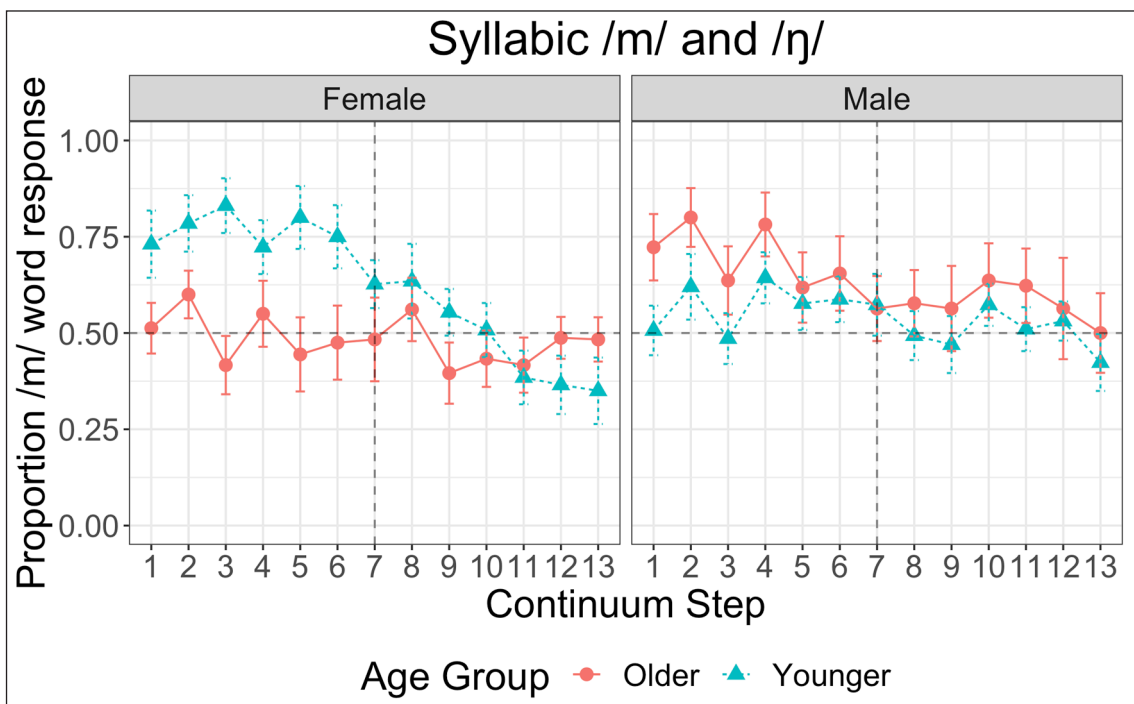
Figure 11: Proportion /m̩/ word responses for the [ŋ]→[m̩] merger by continuum step with different panels for male and female listeners. Step 1 of the continuum is a canonical [m̩] pronunciation and Step 13 is a canonical [ŋ] pronunciation. The dashed blue line and triangles represent the data of the younger listeners, while older listeners are shown in solid red lines and circles. The error bars show standard error.

Individual differences in recognition performance for [ŋ̩]→[m̩] were also analyzed in terms of category crispness. The middle panel in **Figure 9** presents the distribution of these values. The range of values is attenuated compared to [n]→[l] (–0.23, 0.08), and the measures of central tendency are lower (Mean = –0.03, Median = –0.014), indicating that the contrast is overall less crisp for /ŋ̩/ and /m̩/. **Figure 12** additionally reveals that while most demographic groups accordingly present distributions sharply peaking near zero, younger women as a group present a flatter distribution with increased variability towards higher crispness values. The majority of values are negative, however, indicating consistency in the direction of the category distinction that does exist.
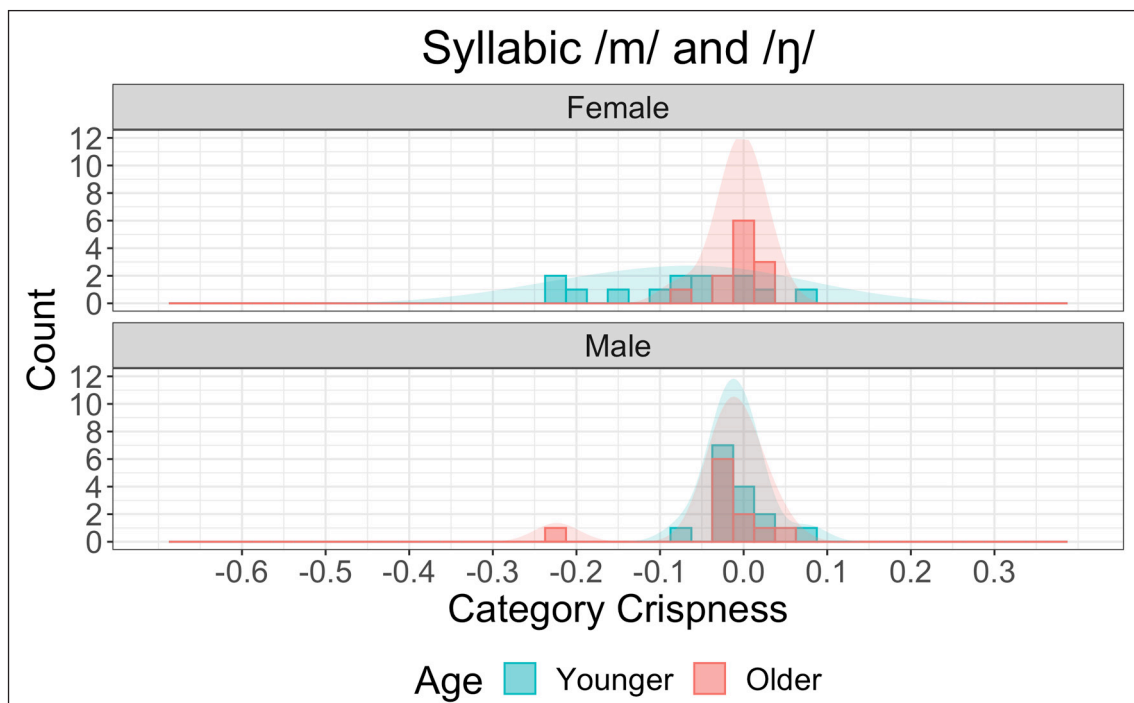

Figure 12: Histograms illustrating the distribution of individual category crispness scores for the [ŋ̩]→[m̩] merger continuum. Density plots are underlaid to represent the distribution per demographic group. Women (top) and men (bottom) are plotted separately while older listeners are colored red and younger listeners colored blue.

3.2.3. [ŋ-]↔Ø merger

Lastly, the responses to the [ŋ]↔Ø continua were analyzed in a similar manner. Null responses were removed, accounting for just over 1% of the data. The remaining data (n = 3277) were fit to a logistic mixed effects regression model predicting the likelihood of a historical

vowel-initial word response (Ø = 1, /ŋ/ = 0). All other model specifications were the same as above.

The model output is reported in **Table 11**, and the results are visualized in **Figure 13**. There was a significant intercept, indicating that [ŋ] responses were more likely, and there was a main effect of Step. As the two-way interaction between Step and Age (p = 0.05912) approached the threshold for statistical significance, we include it in **Figure 13** for consideration. Younger listeners trended towards more categoricity in their response patterns.

	B	Standard Error	z-value	p-value
Intercept	**0.21985**	**0.09718**	**2.262**	**0.02368**
Step (centered)	**0.35784**	**0.13257**	**2.699**	**0.00695**
Age Group	–0.06238	0.0971	–0.642	0.52062
Gender	–0.04629	0.09708	–0.477	0.63344
Step : Age Group	–0.25023	0.13259	–1.887	0.05912
Step : Gender	–0.16481	0.13245	–1.244	0.21341
Age Group : Gender	0.07286	0.09704	0.751	0.45271
Step : Age Group : Gender	0.12243	0.13234	0.925	0.3549

Table 11: Model output for the [ŋ]↔Ø-initial model in perception. Significant factors (α level = 0.05) are bolded.

Individual differences for [ŋ]↔Ø were also quantified in terms of category crispness. While the mean (Mean = 0.029) and median (Median = 0.009) both are positive, the range of values (–0.22, 0.31) spans 0 to a greater degree than those for the previous two continua. Individuals' data for these continua are shown in the right panel of **Figure 9**. Not only do the values for [ŋ]↔Ø indicate that listeners are not at all crisp in their category boundary, the varying signs of the crispness value indicates that individuals differ in the direction of their categorization of these items. That is, individuals have different mappings of which lexical item is vowel-initial and which is [ŋ]-initial. **Figure 14** indicates that younger women again stand out with more distributed scores, particularly including a slight skew towards crisper positive values. While younger men show a strongly peaked distribution centered on zero, there were also a few individuals who look rather different in showing relatively crisp categorization.
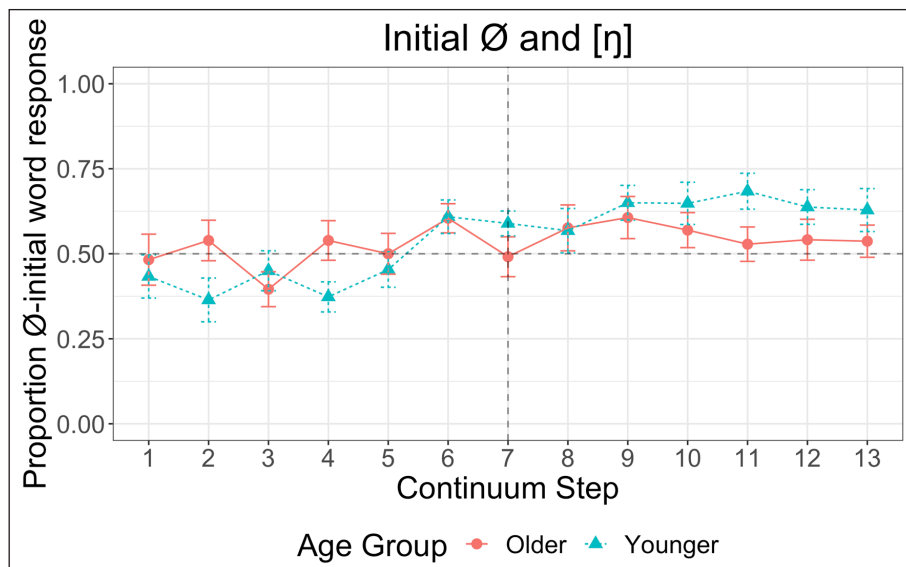
Figure 13: Proportion vowel-initial word responses for the [ŋ]↔Ø-initial merger by continuum step. Step 1 of the continuum is a canonical vowel-initial pronunciation and Step 13 is a canonical [ŋ] onset pronunciation. The dashed blue line and triangles represent the data of the younger listeners, while older listeners are shown in solid red lines and circles. The error bars show standard error.

Figure 14: Histograms illustrating the distribution of individual category crispness scores for the [ŋ]↔Ø-initial merger continuum. Density plots are underlaid to represent the distribution per demographic group. Women (top) and men (bottom) are plotted separately while older listeners are colored red and younger listeners colored blue.

3.2.4. Summary of perception

Group-level perceptual results indicate that each merger is characterized by different patterns, but unlike production, age appears to be a relevant factor. Simplifying slightly, younger listeners were generally more categorical: significantly so for [n]→[l], a trending effect for [ŋ]↔∅, and a case where younger women alone appear more categorical for [ŋ]→[m̩]. In other words, younger listeners were often less merged in recognition than older participants. Individual category crispness values further demonstrate that individuals varied from no differentiation between lexical items (present for all mergers) to fully categorical (for /n/ and /l/). Across mergers, the /n/ and /l/ categories were the most discrete, followed by /m̩/ and /ŋ̍/. The [ŋ] and null-initial lexical items elicited the least discrete patterns, and the + /– signs of the crispness values further indicate that listeners categorize these items in opposing directions.

3.3. Production-Perception

To understand the relationship between perception and production at the individual level for the three mergers under investigation, we quantified the degree of mergedness for perception and production. Mergers in perception were quantified using the absolute value of the by-subject crispness values described above. Higher values indicate more discrete perceptual categories, while lower values are taken to indicate a merger of perceptual categories.

Mergers in production were quantified as the absolute value of the difference between proportions of the more novel pronunciation (i.e., [l], [m̩], null) for historical /l/, /m̩/, and null-onsets words and proportions of the more novel pronunciation for the historical /n/, /ŋ̍/, and [ŋ]-onset words. This quantification means that individuals who produce, for example, a full merger 100% of the time (e.g., 100% [l] for historical /l/ and /n/ words) will have an equivalent merger score as those who exclusively show hypercorrection (e.g., 0% [l] for historical /n/ and /l/ words; thus 100% [n] for both lexical classes) and those who exhibit a variable mix of pronunciations for both lexical sets (e.g., 50% [l] for historical /n/ and /l/ words). Crucially, however, participants of these types are demonstrating a lack of a reliable difference in pronunciation variants for the lexical sets, which we take as an indication of a merger of these categories at the lexical level.

Importantly, for the production-perception analysis, we removed the two perceptual crispness outliers (see **Figure 9**), leaving 47 individuals for consideration. We did so because, first, their values were so exceptionally different (values less than –5 compared to an overall CCS median of –0.021 and IQR of 0.081) that it was neither possible to conduct any reasonable comparison between them and the remaining data, nor within the remaining data due to the expanded range. Second, based on visual inspection of individual response curves, the extreme jump in numerical value appears not to be linearly reflective of actual change in categoricity given

that the next highest values within a reasonable range (approximately 0.6) looked similar and highly categorical;[7] as a result, although treating the next highest values as maximum crispness is a simplification, using them to represent strongly categorical perceptual responses is a fair representation of the data and exclusion of the outliers is not expected to alter conclusions.

Figure 15 visualizes the individual patterns for merger in perception and production for the three mergers. Merger in production is on the x-axis. For ease of interpretation, the x-axis is reversed in these visualizations to represent the time course of merger, which proceeds from 100% contrast (no merger) to 0% contrast (full merger). Merger in perception is on the y-axis, represented by absolute category crispness scaled to a range of zero to one.[8] Low values along either dimension indicate a stronger merger, with values of zero indicating a complete merger of linguistic categories. A dashed line is included to represent the hypothetical fitted line if production and perception were perfectly correlated, assuming, as a simplification, the maximum value across all three mergers as maximum categoricity.

These measures of merger in perception and production are moderately correlated for the [n]→[l] merger [$t(47) = 4.10, r = 0.51, p = 0.0002$][9] and [ŋ]→[m̩] merger [$t(47) = 4.13, r = 0.52, p = 0.001$]. These positive correlations are represented as negative in the plots due to x-axis reversal, but indicate the degree to which perception and production move in tandem. There was no relationship between merger in perception and production for the [ŋ]↔Ø merger [$t(47) = -0.06, r = -0.009, p = 0.95$]. The regression lines in the panels present these relationships. Note that the y-axes on these figures are uniform to illustrate the different distributions of crispness values, and this makes the strength of the correlation for [ŋ]→[m̩] look substantially less strong than for [n]→[l], which is not the case numerically.

To assess the direction of misalignment within an individual, we calculate the difference between scaled production and perception values per merger (DiffPP, following Pinget et al., 2020). Zero indicates completely aligned production and perception, regardless of whether individuals are fully merged, fully contrastive, or at an intermediate value. Scores above zero represent individuals whose production is more contrastive (less merged) than perception while scores below zero represent individuals whose perception is more contrastive (less merged) than production. Visualized in **Figure 16**, this provides another way to look at the data, in essence representing the direction and distance of each individual from the hypothetical perfect correlation line in **Figure 15**.

---

[7] Because the distance was too great, attempts to apply transformations for skewed data were not successful in creating a reasonable distribution; thus, we ultimately chose to remove the outliers.

[8] Due to the removal of the two outliers prior to scaling and visualizations, the maximum scaled value in Figures 12 and 13 represents the next highest CCS value, which was 0.632.

[9] To be conservative, the correlation provided in text includes the two extreme values. Removing those two data points from the correlation does not affect the direction or interpretation of the results [$t(45) = 4.54, r = 0.56, p < 0.001$].
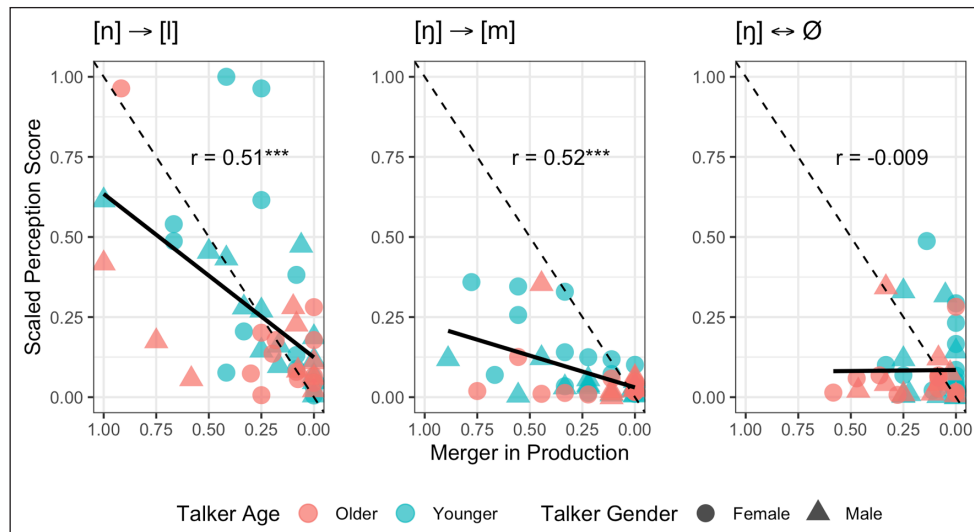
Figure 15: Degree of merger in perception (y-axis) versus production (x-axis) for the three mergers. Merger in perception is the absolute value of category crispness and the merger in production is quantified as the absolute value of the difference in the proportions of the novel pronunciation for the two lexical sets for each merger. The range of the x-axis is from 1 (fully contrastive) to 0 (fully merged) to reflect the time course of change. The solid lines represent fitted lines to the data while the dashed lines represent a hypothetical fitted line if production and perception were perfectly correlated.
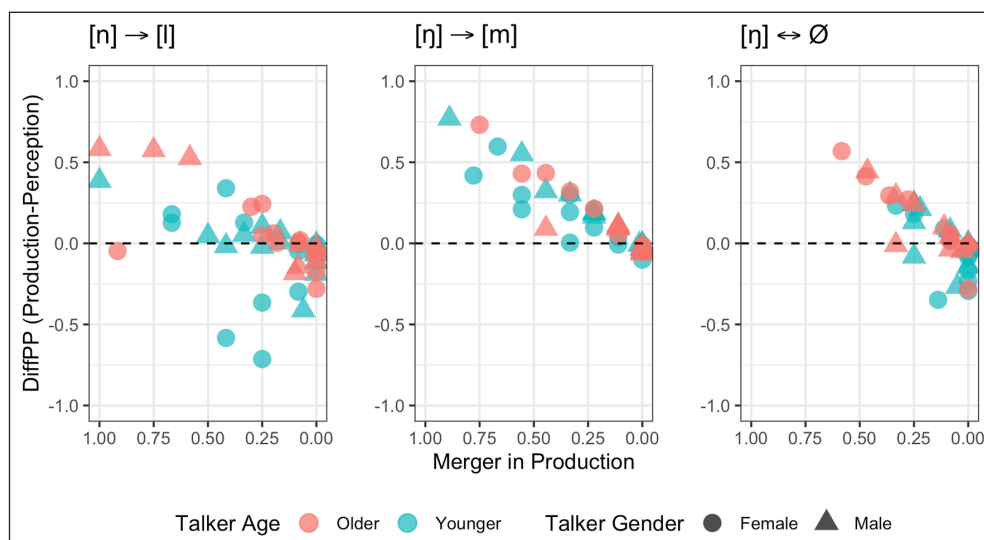

Figure 16: DiffPP scores, calculated as the difference of production and perception measures (y-axis) versus production (x-axis, reversed) for the three mergers. DiffPP scores above zero represent individuals whose production is more contrastive (less merged) than perception while scores below zero represent individuals whose perception is more contrastive (less merged) than production.

The left panel in **Figure 15** illustrates the reliable correlation between merger in /n/ and /l/ production and perceptual category crispness in a forced choice task of word selection. Individuals who make smaller differences in their rates of [l] production for historical /l/ and /n/ categories are more likely to be merged in perception; this is seen as the clustering in the bottom right corner. At the same time, the subset of participants who show a difference in [l] pronunciation rates for historical /n/ and /l/ words are more likely to exhibit a crisper perceptual contrast. Nevertheless, the relationship between production and perception is far from perfect. As seen in the left panel of **Figure 16**, many individuals show production-perception misalignment by falling above or below the zero line. A number of observations are notable. First, multiple individuals who produce little to no contrast (on the right side of the plot) fall below zero, demonstrating contrast in perception despite none in production. Conversely, misaligned individuals who produce some contrast (roughly greater than 0.2) tend to fall above zero, demonstrating more contrast in production than perception. In other words, perception appears to lead production in merger, but when merger is (nearly) complete in production, perception appears to lag (i.e., retain contrast). Finally, age appears to play a role such that younger individuals appear more likely to show lower DiffPP scores relative to older individuals with similar production patterns. In particular, three younger women defy the general trend, showing intermediate production contrast but with comparatively greater perceptual categoricity than production.

The middle panel in **Figure 15** illustrates the reliable correlation between merger in /m̩/ and /ŋ̍/ production and perceptual category crispness in a forced choice task of word selection. Specifically, individuals who exhibit a smaller difference in their rates of production of the more innovative [m̩] for historically /m̩/ and /ŋ̍/ words are more likely to be merged in perception, while those producing larger contrasts were more likely to show crisper perceptual distinction. The distribution of perceptual crispness values is contracted compared to those values for the [n]→[l] merger, indicating that, overall, this contrast is less perceptually robust.[10] Again, while there is a significant correlation, the relationship is not isomorphic: There is substantial variability in the alignment of production of these variants and listeners' ability to systematically distinguish the lexical items by historical category in perception. Taking the maximum perceptual crispness value for [n]→[l] as the upper limit,[11] the overall pattern in the middle panel of **Figure 16** is that individuals who are not fully merged show more contrastiveness in production than perception. Nearly all participants fall roughly at or above zero; a few individuals do show a small degree of

---

[10] It's worth noting that regardless of its participation in a merger, the auditory-acoustic contrast between syllabic nasals compared to /n/ and /l/ is less robust.

[11] While this may not be a foolproof approach due to the less perceptually robust nature of nasal place of articulation contrasts, the use of the upper boundary of [n]→[l] crispness values serves as an estimate, given that 'maximum' perceptual crispness values are unknown for this particular contrast.

misalignment towards perception, but otherwise, if not aligned, perception leads production. In addition, younger participants tend to show lower DiffPP scores, indicating less alignment skew towards production contrast.

Lastly, the third panel of **Figure 15** illustrates the lack of correlation between the merger of historically [ŋ] and vowel-initial items in production and perceptual category crispness in a forced choice task of word selection (of *aak1* 'shake' and *ngaak1* 'deceive,' specifically). The range of perceptual crispness values are similar to those for [ŋ]→[m̩], but the production range is more limited. This reaffirms the production results where individuals either mainly produce [ŋ] or null-initial. In **Figure 16**, the vast majority of participants merge [ŋ] and null-initial in both production and perception but those who do not follow a similar pattern to [n]→[l]: Those who produce little to no contrast fall below zero (perception lags behind production) while misaligned individuals who produce some contrast (roughly greater than 0.2) fall above zero (perception leads production). Further, there appears to be an age-based pattern whereby younger individuals make up the majority of production-merged individuals who fall below zero, showing a contrast in perception despite none in production.

3.3.1. Summary of production-perception

Production-perception analyses, which correlate an individual's degree of merger in production to their degree of merger in perception, reveal a mixed bag of results. We find moderate production-perception correlations for [n]→[l] and [ŋ]→[m̩] despite different degrees of overall mergedness, but no correlation for [ŋ]↔Ø. At the same time, while all mergers reveal a similar overall trend of misalignment such that if not aligned, production remains more contrastive than perception, [n]→[l] and [ŋ]↔Ø show an additional pattern where individuals with merged production show contrast in perception despite little to none in production. Finally, age-based trends suggest that younger individuals are less misaligned than older individuals in the direction of production contrast but more misaligned in the direction of perception contrast.

4. Discussion

The two aims of the study were to (1) test generalizations of the production-perception link and (2) describe apparent-time patterning of the merger variants in production and perception. We first discuss the descriptive data relative to previous documentation of these mergers (Sections 4.1 and 4.2). As this study was not designed to provide a definitive or comprehensive overview of the completion status of the mergers, we discuss various possible interpretations consistent with the data but do not diagnose the situation further. Then, we discuss the production-perception results situated in the sound change literature (Sections 4.3 and 4.4). **Tables 12–14** summarize the full set of findings for each merger.

Table 12: Summary of results for *[nl]→[l]*. Individual productions patterns (n = 49) are described based on their proportion of [l] for /n/ as either *innovative* (>0.75), *variable* (0.75–0.25), or *conservative* (<0.25); if proportion of [l] for /l/ was under 0.5, individuals were instead described as *hypercorrective*. Individual perception patterns (n = 51) are described based on scaled category crispness scores as *merged* (<0.25), *intermediate* (0.25–0.75), or *categorical* (>0.75).

		Production (PROD)	Perception (PERC)	Production-Perception
Community	Overall	[l]-variant more likely than [n]	/l/-word more likely than /n/	• PROD and PERC are overall *matched* • But *mismatched* by age: Across-the-board contrast (PROD) versus age-related contrast (PERC)
	Lexical classes	**Contrast** between /l/ and /n/	**Contrast** between /l/ and /n/	
	Age	None	*Younger* more contrastive than older	
	Gender	None	None	
Individual	Overall	• **Most (n = 25, 51%) use mainly [l] for both (innovative)** • Some (n = 12, 25%) use [l] for both somewhat variably • Few (n = 7, 14%) use mainly [n] for /n/ and [l] for /l/ (conservative) • Few (n = 5, 10%) use mainly [n] for both (hypercorrective)	• **Most (n = 33, 65%) had little to no contrast (merged)** • Some (n = 13, 25%) had intermediate contrast • Few (n = 5, 10%) had close to discrete contrast (categorical)	• PROD and PERC are overall *correlated* ($r = 0.52$***) When misaligned: • **Perception leads early:** Those who contrast *more* in PROD tend to contrast relatively *less* in PERC • Younger contrast least • **Perception lags late:** Those who contrast *less* in PROD tend to contrast relatively *more* in PERC • Younger contrast most
	Age	None	Those most contrastive are nearly all *younger*	
	Gender	None	None	

Table 13: Summary of results for [ŋ̍]→[m̩]. Individual productions patterns (n = 49) are described based on their proportion of [m̩] for /ŋ̍/ as either *innovative* (>0.75), *variable* (0.75–0.25), or *conservative* (<0.25); if proportion of [m̩] for /m̩/ was under 0.5, individuals were instead described as *hypercorrective*. Individual perception patterns (n = 51) are described based on scaled category crispness scores as *merged* (<0.25), *intermediate* (0.25–0.75), or *categorical* (>0.75).

		Production (PROD)	Perception (PERC)	Production-Perception
Community	Overall	[m̩]-variant more likely than [ŋ̍]	[m̩]-variant more likely than [ŋ̍]	• PROD and PERC are overall *matched* • But *mismatched* by age and gender: Across-the-board contrast (PROD) versus age/gender-related contrast (PERC)
	Lexical classes	**Contrast** between /m̩/ and /ŋ̍/	Contrast between /m̩/ and /ŋ̍/	
	Age	*Older* use more [m̩] than younger	*Younger women* more contrastive than older	
	Gender	*Men use more [m̩] than women*		
Individual	Overall	• **Most (n = 29, 59%) use mainly [m̩] for both (innovative)** • Some (n = 13, 27%) use [m̩] for both somewhat variably • Few (n = 5, 10%) use mainly [m̩] for /m̩/ and [ŋ̍] for /ŋ̍/ (conservative) • Few (n = 2, 4%) use mainly [ŋ̍] for both (hypercorrective)	• **Majority (n = 46, 90%) had little to no contrast (merged)** • Few (n = 5, 10%) had intermediate contrast • None had close to discrete contrast (categorical)	• PROD and PERC are overall *correlated (r = 0.51***)* When misaligned: • **Perception leads early:** Those who contrast *more* in PROD tend to contrast relatively *less* in PERC • Younger contrast least • **No evidence of perception lagging**
	Age	*Younger* more likely to use [ŋ̍]	Those most contrastive are nearly all *younger women*	
	Gender	*Women more likely to use [ŋ̍]*		

Table 14: Summary of results for [ŋj]↔Ø. Individual productions patterns (n = 49) are described based on their proportion of Ø for /ŋ/ as either *innovative* (>0.75), *variable* (0.75–0.25), or *conservative* (<0.25); if proportion of Ø was under 0.5, individuals were instead described as *hypercorrective*. Individual perception patterns (n = 51) are described based on scaled category crispness scores as *merged* (<0.25), *intermediate* (0.25–0.75), or *categorical* (>0.75).

		Production (PROD)	Perception (PERC)	Production-Perception
Community	Overall	[ŋj]-variant more likely than Ø	[ŋj]-variant more likely than Ø	• PROD and PERC are overall *matched* • But (trend towards being) *mismatched* by age: Across-the-board contrast (PROD) versus age-related contrast (PERC)
	Lexical classes	**Contrast between Ø and /ŋ/**	**Contrast between Ø and /ŋ/**	
	Age	None	Trend towards younger more contrastive than older	
	Gender	None	None	
Individual	Overall	• Few (n = 5, 10%) use mainly Ø for both (innovative) • Few (n = 5, 10%) use Ø for both somewhat variably • Few (n = 2, 4%) use mainly Ø for Ø and [ŋj] for /ŋ/ (conservative) • **Most (n = 37, 76%) use mainly [ŋj] for both (hypercorrective)**	• **Majority (n = 50, 98%) had little to no contrast (merged)** • Few (n = 1, 2%) had intermediate contrast • None had close to discrete contrast (categorical)	• PROD and PERC are overall *not correlated* (r = –0.009) When misaligned: • **Perception leads early:** Those who contrast *more* in PROD tend to contrast relatively *less* in PERC • **Perception lags late:** Those who contrast *less* in PROD tend to contrast relatively *more* in PERC • Younger contrast most
	Age	None	Those most contrastive are generally *younger*	
	Gender	None	None	

4.1. Past versus present in production

In To et al. (2015), the [n]→[l] merger appeared near-complete in both adults and children (mean rates of [l] for /n/ around 94%), but in the current data, /n/ is realized as [l] between 54% (younger women) to 71% (older women) of the time. Within each group, speakers also varied considerably as to whether they realized /n/ as [l], ranging from 0% to 100%—in fact, 13 of the 49 speakers were fully merged. Further, no age or gender pattern was evident for speakers who were mainly contrastive. This suggests that the linguistic situation is more complicated than a simple story of change-in-progress where the contrast is maintained by the older generation. In addition to variation in /n/, there was variation in the amount of [l] produced for /l/, which has been generally ignored in previous literature due to the assumption that /l/ was not changing. That some speakers produce a significant amount of [n] for /l/ (up to 100% in one speaker's case) confirms that [l]→[n] 'hypercorrection' does exist, contrary to Bourgerie's (1990) assertion. For at least these individuals, it seems that the social meaning of 'properness' indexed to [n] (Pan, 1981) can be applied to both historical /n/ and /l/. Altogether, this suggests that the two variants [l] and [n] are dually-mapped to the same lexical items regardless of historical lexical class (Samuel & Larraza, 2015); this is indicative of a merger at the lexical level (Soo & Babel, 2020). Listeners have implicit knowledge—and possibly explicit awareness, given the pronunciation campaigns and social meaning attributed to the variants—that /n/ and /l/ are merged, but they lack the knowledge of which items were historically /l/, allowing for both [n] and [l] pronunciation variants for both lexical classes.

The [ŋ]→[m̩] merger was likewise reported by To et al. (2015) to be near-complete. While older adults (particularly men) show comparable or slightly decreased production of [m̩] for /ŋ̍/ in the current data (70–100% now versus 91–100% before), the younger cohort shows clearly decreased [m̩] usage roughly 15 years later (50–75% now versus 97–100% before). While interactions with age were not statistically significant, this result seems linked to trending differences in historical category merger across demographic groups: Nearly all older men are fully merged (92.6% [m̩] for /ŋ̍/, 98.8% [m̩] for /m̩/), many older women (71.6% [m̩] for /ŋ̍/, 90.5% [m̩] for /m̩/) and younger men (77.6% [m̩] for /ŋ̍/, 97.8% [m̩] for /m̩/) are relatively merged, but a sizable number of younger women make moderate-to-large contrasts (50.4% [m̩] for /ŋ̍/, 82.9% [m̩] for /m̩/). We additionally see gender differences in the preferred variant influenced by some women using high rates of both hypercorrective and conservative [ŋ̍]. Again, this behaviour indicates dual-mapping of [ŋ̍] to both historical /ŋ̍/ and /m̩/ lexical items, implying a certain degree of lexical merger.

For the [ŋ]↔Ø merger, children in To et al. (2015) produced more null onsets for both historical [ŋ] and null-initial (65.0% and 94.2%) while adults used the historical variants more often (37.5% and 68.7%). Yet, rather than a continuing trend of increased null, the current findings indicate that all groups produce *less* null-initial for both historical categories, at mean

rates below 33.6% null (i.e., above 66.4% [ŋ]). Individual patterns show that there is large variation in behaviour of merging and choice of variant that, to a certain degree, pattern relative to age. Descriptively, those who show a higher degree of merger or are merged to null-initial are more likely younger, while those who maintain distinction between lexical classes are more likely older. Thus, while merger of the [ŋ]- and Ø-initial lexical classes may indeed have progressed, variant preference seems to have reversed dramatically.

4.1.1. What has changed?

Several factors could plausibly lead to the discrepancies between current results and prior predictions about the trajectory of these consonant mergers. We consider here four possible sources: methodological differences, style-shifting, age-grading, and community-wide change.

In terms of methodology, we used several words per merger pair as opposed to the single word used in To et al. (2015). Rather than representing a true difference in norms over time, the larger set of lexical items could have simply provided a more accurate depiction of variant usage across the lexicon. However, by-word examination of the current results rules out this explanation: Production of the same words used in To et al. (2015) demonstrates consistently lower rates of [l], [m̩], and null-initial.[12] Specifically, the group rates of [l] for *naam4* 'male' are similar to the aggregate values, ranging from 35.9% for young women to 74.1% for older men (compared to 94%). Similarly, rates of [m̩] for *ng5* 'five' range from 51.2% to 88.9% (compared to 95–98%) while rates of null-initial for *ngau4* 'cow' range from 2.8% to 22.2% (compared to 37–65%).

Another methodological difference between our study and the previous offers the possibility that speakers may have been style-shifting to a formal register in the current data given our lab-based reading task, but would have produced more innovative variants in a more casual setting without the influence of orthography (e.g., the picture naming task in To et al., 2015). This interpretation could be congruent with the across-the-board increases in [n] and [ŋ]. Given the prescriptivist norms in the local context, task-based style-shifting is plausible and cannot be ruled out as an explanation for our diverging results (a point we return to below).

A third possible scenario is that of age-grading[13] due to linguistic marketplace pressure, wherein individuals change their production patterns over different life stages, linked to the social value of 'standard' variants (for reviews of age-grading and lifespan changes, see Wagner,

---

[12] A direct comparison could not be made for the historical null-initial word, as the word *uk1* 'house' used by To et al. (2015) was not used in the current study.

[13] See Wagner (2012) for discussion on the multiple definitions or criteria of the term 'age-grading' (community stability, repeating patterns across generations, marketplace pressure) and why there isn't a clear, consensus differentiation between 'age-grading' and 'lifespan change.' We reference a particular approach to age-grading described by Wagner (2012), but we recognize it is not unanimously accepted.

2012; Cheshire, 2006). If age-preferential language use were the explanation in the current data, the generation studied in To et al. (2015) as children may have begun to use more of the 'formal' prestige variants, namely [n] and [ŋ], now that they are young adults and likely in more professional settings. Like style-shifting, this could explain the increase in conservative variants in the younger generation across all three mergers. On the other hand, interpretation of the results as age-grading is less clearly motivated for older adults based on these data. For one, the adult group in To et al. (2015) cannot be directly matched to our older adult sample, and two, it is not fully clear what an age-grading explanation would predict for this generation, which spans middle-aged to retirement-aged individuals. Future data collection specifically targeting age-grading will be necessary to adjudicate such an interpretation.

Lastly, production norms may have changed due to reversal of the phonological merger in younger generations. Given that older men appear to be merging [ŋ]→[m̩] at a similar rate as previously reported, the current results could be consistent with an incipient reversal. For [l]→[n] and [ŋ]↔Ø, however, younger speakers did not make a larger contrast relative to the older generation. Perhaps these age-undifferentiated results reflect a community-wide change of language ideologies due to increase in awareness of the 'proper' social meaning of [n] and [ŋ] since the advent of the 'proper pronunciation' campaigns in the late 2000s. Notably these events occurred after To et al. (2015) conducted their study. This social shift may have led to more prestige-related style-shifting in the present day as compared to the past.[14]

If changes in production were indeed driven by an ideological shift, the age-related results for [ŋ]→[m̩] could be seen as representing an effect of register that is particularly pronounced in younger speakers; that is, the social association of [ŋ] with 'properness' may not exist as strongly for older speakers (or, alternatively, older speakers may be less invested in abiding to newer social norms). Moreover, women were more likely to use high rates of [ŋ] for both categories, including historical /m̩/. Given that women have been suggested to (a) be more aware of social stigma and (b) use more prestige variants (e.g., Labov, 1990), this is compatible with the interpretation that prestige motivates the maintenance of contrast or overall increased use of the conservative variant. The sharp shift in [ŋ]↔Ø norms from majority null-initial in To et al. (2015) to majority [ŋ] in the current study could also stem from drastically increased salience of the social meaning of [ŋ] as 'proper,' aligning with Chen's (2018) assertion that null-initial is particularly stigmatized as 'lazy.'

---

[14] Merger reversal across contexts and context-based style-shifting cannot be distinguished in the current data, as we do not compare contexts. As well, they may not be mutually exclusive: More prestige-related style-shifting could co-occur with and/or precede incipient reversals in production outside of formal contexts, both motivated by increased awareness and attitudes towards the social meanings of the merger variants. As we propose these outcomes to have the same source, we discuss them together.

It is clear that, to fully understand the patterns found in these data, speech style and attitudes must be accounted for. Thus, limitations of the current study include a lack of consideration for multiple registers in the production task and attitudes about 'lazy pronunciation.' Accounting for variable styles would have allowed for more conclusive interpretation of descriptive results, especially in comparison to previous literature. At the same time, while we were limited in our methodological choices due to constraints of the stimuli, the formal style elicited in our task does not invalidate our findings, including those of the production-perception analyses. Speakers have a repertoire at their disposal, and we tapped into a particular style of production, which may well have been influenced by our methodological choices; regardless, we need to account for the linguistic knowledge presented to us. Moreover, our choice of an isolated, context-free, single word production task accompanied by orthography is well matched with the perception task, which is characterized by all those features as well. This well-matched formality allows for interpretable (mis)alignment patterns across production and perception.

With respect to 'lazy pronunciation,' although the exit interviews confirm that all participants knew what 'lazy pronunciation' was, we did not assess awareness or attitudes for each particular merger. Attitudes can vary substantially, potentially influencing patterns of both production and perception. For example, although many participants characterized the mergers as 'incorrect' and 'lazy,' one young man expressed more neutral or positive sentiments, describing the pronunciations as 'more convenient' (see also Lin, Yao, & Luo, 2021 on both positive and negative attitudes about the younger generation's accent). Knowledge of the degree of stigmatization for each merger, along with individual attitudes towards the stigmatization, could help contextualize variation across individuals as well as any potential community-level reversals that are motivated by group orientation towards or away from the existing social meanings of the merger variants (see, e.g., Labov, Rosenfelder, & Fruehwald, 2013; Tamminga, 2019). Continued research is needed to document whether merger reversals eventually emerge in production across the Hong Kong community; this work would benefit from comparison of multiple speech styles, including more naturalistic speech, as well as inclusion of awareness and attitude assessments to form a clearer picture of production norms across both age and context.

4.2. Perception versus production in the community

Unlike production, a general theme in perception is that younger participants, particularly women, responded to historical lexical classes more categorically than older participants, though the degree of categoricity was objectively low for mergers involving [ŋ].[15] Considering that

---

[15] We acknowledge the possibility that an age-based effect of hearing sensitivity (or cognitive processing abilities) may have influenced perceptual results to some extent and note that we did not specifically control for this confounding factor. At the same time, because younger men often patterned with older individuals (most noticeably for [ŋ]→[m]), we take the position that the age trend cannot simply be reduced to a difference in auditory-perceptual abilities.

age effects are not robustly supported by evidence in production, how can the perceptual age effects be interpreted? Younger listeners have had more exposure to the historical contrasts due to growing up with 'proper pronunciation' prescriptivism and language contact with Mandarin. One possibility is that, due to these experiences, they may have adapted to flexibly recognize the contrast in perception, regardless of whether they make a contrast between the two in their own productions (see Samuel & Larraza, 2015 on adaptation to variable pronunciations or 'errors'; Sumner & Samuel, 2009 on covert experience with pronunciation variants on word recognition timing). Cognitively, this would be a dual-mapping of two pronunciation variants to a single phonological or lexical category for a style or register difference. This accommodation to phonetic variation is akin to interpretations of late-stage sound changes where perception 'lags' behind production due to the more advanced (e.g., younger) listeners' need to maintain these representations to cope with speech produced by less advanced (e.g., older) speakers (see Pinget et al., 2020; Coetzee et al., 2018). In other words, they are flexible in perception while showing stability in production.

Another possibility is that, rather than simply demonstrating perceptual flexibility, younger speaker-listeners are in fact beginning to reverse the mergers in their recognition of lexical items, prior to showing clear signs of reversal in production. That is, a phonological change across generations may truly be occurring, where perception leads production in the opposition direction to the historical change. As suggested earlier, this explanation could align with qualitative age-based trends found in [ŋ]→[m̩] production. Only future investigations on the continuing trajectory of these variants can shed light on these open questions.

An alternative explanation does not rely on listeners' decontextualized phonological representations but involves social representations construed from the model speaker's voice. That is, differences may have arisen due to younger listeners perceiving the model speaker as a peer while older listeners perceived the same speaker as younger (Hay et al., 2006; Koops et al., 2008). Specifically, if there is an association between the younger generation and increased 'lazy pronunciation' held by older speakers (see discussion in Lin et al., 2021), we might expect that younger listeners (without such an association) perceive the stimuli more veridically while older listeners perceive the stimuli with less distinction. This possibility cannot be teased apart from other possible interpretations and must be left to future research.

4.3. Individual evidence for the production-perception link

Moving on to the production-perception link, our individual-level production-perception results reveal a link between production and perception for two consonantal mergers, corroborating recent positive findings from the sound change literature. In particular, it allows us to more confidently generalize the findings from Pinget et al. (2020) that a production-perception link

can be found for *consonantal mergers*, just like more frequently-studied types of vowel-related sound change. Given the rather different properties of mergers to shifting-type changes (e.g., loss of contrast), the consistency of detecting an individual-level production-perception relationship is interesting and strengthens the claim of an underlying mechanistic link between systems that drives sound change.

On the contrary, no production-perception correlation was found for the [ŋ]↔Ø merger. Since the other two merger pairs demonstrated a correlation, the type of change (i.e., merger rather than cue shifting, for example) cannot be the constraining factor, contributing evidence against the hypothesis that production-perception relationships vary by type of change. Still, it is possible that the specific nature of this change is relevant. That is, the other two mergers are 'standard' contrast-loss mergers, unlike [ŋ]↔Ø which was originally allophonic and exhibited bidirectional merging. The nature of this change (including the changes in direction) may have contributed to the apparently unlinked representations. Given this, it may be beneficial for future research to consider more fine-grained variability among types of change and how it may influence the existence of a production-perception link. We additionally note that these findings could have resulted from the choice of perceptual stimuli and task, which involved lexical identification between items in a purported minimal pair that did not follow historical tone allophony patterns (i.e., involved an exception). With the current data, it is not possible to tease apart whether this null effect is due to the nature of this particular merger or aspects of the experiment. Future work would benefit from examining the [ŋ]↔Ø merger further in perception.

4.4. Direction of production-perception misalignment

Despite correlation differences, similar patterns of misalignment were uncovered across all three merger pairs. In general, individuals making a contrast in production showed comparatively less perceptual contrast, indicating that perceptual merger is more advanced than production merger. On the other hand, many individuals with nearly or fully merged production for [n]→[l] and [ŋ]↔Ø—if not 'realigned' through full merger in production and perception—in fact show evidence of categoricity in perception; this suggests that at these late stages of production merger, perception is less advanced, or lagging. Given the data at hand, these results fall in line with previous reports of production-perception misalignment with regards to stage of change (Pinget et al., 2020; Coetzee et al., 2018). However, the switch from perception-led to production-led change was reported in Pinget et al. (2020) to be roughly at the halfway point of merger while here, only those who were nearly or fully merged exhibited production leading perception. Whether this difference relates to the fact that the present mergers appear to have reached a state of socially conditioned variation rather than continued merger or whether such a pattern is dependent on the particular situation requires further investigation.

In addition, younger participants on the whole tend towards less misalignment in the direction of production contrast and/or more in the direction of perceptual contrast. This confirms that community-level perceptual and production results, where younger individuals showed more contrast in perception, can be sourced to the individual level. A final observation is noteworthy. Unlike the other two mergers, few individuals showed misalignment in the direction of perceptual contrast for the [ŋ̩]→[m̩] merger, even for those merged in production. Why does this difference arise? While speculative, it seems conceivable that [ŋ̩]→[m̩] operates on a reduced perceptual scale due to the lack of acoustic-perceptual salience for nasals (Harnsberger, 2000; Narayan, 2008); if so, a different approach to assessing misalignment may be needed in future in work that attempts to compare across multiple cases of variation and change.

5. Conclusion

This study investigates production and perception of the [n]→[l], [ŋ̩]→[m̩], and [ŋ]↔Ø mergers in Hong Kong Cantonese, on one hand documenting the present-day status of these long-standing sound changes and on the other hand extending hypotheses about the production-perception relationship in the context of sound change. Production results demonstrate substantial amounts of individual variability but a lack of clear age- or gender-related variation, indicating that the picture for each merger is less straightforward than continued or completed merger. Perceptual results find that younger speaker-listeners are less merged than the older generation, potentially due to adaptation to high phonetic variability. While this project did not collect the data necessary to fully diagnose the current stage of the mergers, the results do suggest that [n]→[l] and [ŋ]↔Ø no longer appear to be changes-in-progress, though incipient reversal is a possibility for [ŋ̩]→[m̩]; instead, usage of the formal variants appear inflated compared to previous records, suggesting that 'proper pronunciation' ideologies are at play. Another possibility is that the mergers have stabilized into context-conditioned style-shifting, a scenario that Labov (2001) proposes may historically be more common than changes going to completion (p. 75). Future research can disentangle these possibilities.

A birds-eye view of our results on the production-perception link reveals that, on the whole, it parallels the literature: Some level of linkage evidently exists between production and perception systems, though it is not always strong and not always found. We show here that, in the same population using the same methodology, a production-perception correlation can be found moderately for two mergers yet not at all for another. Our finding of production-perception coupling for two additional consonant mergers specifically bolsters the conclusion that convergent results in the literature are generalizable beyond certain types of sound change. However, while we aimed to use comparable measures for production and perception, we also acknowledge that our experimental choices may still have impacted results. Ultimately, we do

not solve the problem of methodological inconsistency, but we urge future research to take these issues into careful consideration.

Interestingly, when a correlation was detected, they were of very similar magnitude, even though these mergers are characterized by rather different profiles (e.g., functional load), trajectories, and perceptual salience. This direct comparison of magnitudes indicates that there is a modicum of consistency about the production-perception relationship across cases; notably, such a comparison is difficult to make across studies due to the variability in context, methodology, participants, and more (see Pinget et al., 2020, who nevertheless report that previous studies often find moderate correlations).

That production and perception systems constrain but do not determine each other is sensible as we expect flexibility in perception beyond variability exhibited in production (e.g., we understand and accept a much larger range of speech than we tend to produce; see Samuel & Larraza, 2015). It cannot simply be that all exposure in perception affects production, as effects of exposure in stored representations are most likely mediated by encoding (see attention-weighting approaches to exemplar theoretic models of speech perception, Sumner et al., 2014; Drager & Kirtley, 2016; Johnson, 1997). In the same way, production is mediated by other factors, such as speaker agency, and does not reflect purely linguistic aspects of experience. If this is the case, however, what is the expected extent of transfer between perception and production, and what cognitive factors influence the degree of coupling?

One route forward is to consider moving beyond a search for evidence on the existence of a production-perception link to a more nuanced examination of when and to what degree we are able to detect a relationship. In this vein, we encourage researchers to apply a controlled comparative approach with a larger range of case studies in order to paint a broader typology of sound change and the extent of coupling between production-perception systems. By focusing attention on theorizing why there would not be a link for any particular case, what mechanisms underlie such connections, and what regulates the strength of relationship across controlled groups, variables, or tasks, we can move closer to understanding the persistently mixed results of production-perception studies and the underlying cognitive processes.

Additional File

The additional file for this article can be found as follows:

- **Appendix.** Word list for production task. DOI: https://doi.org/10.16995/labphon.6461.s1

Competing Interests

The authors have no competing interests to declare.

References

Ball, J. D. (1907). *Cantonese Made Easy: A Book of Simple Sentences in the Cantonese Dialect, with Free and Literal Translations, and Directions for the Rendering of English Grammatical Forms in Chinese* (3rd ed.). Kelly & Walsh, Limited.

Bauer, R. S. (1982a). *Cantonese Sociolinguistic Patterns: Correlating Social Characteristics of Speakers with Phonological Variables in Hong Kong Cantonese* (Doctoral dissertation, University of California, Berkeley). https://escholarship.org/uc/item/1sn2f75h

Bauer, R. S. (1982b). Lexical Diffusion in Hong Kong Cantonese: "Five" Leads the Way. *Proceedings of the 8th Annual Meeting of the Berkeley Linguistics Society*, 550–561. DOI: https://doi.org/10.3765/bls.v8i0.2037

Bauer, R. S. (1986). The Microhistory of a sound change in progress in Hong Kong Cantonese / 香港粤语中一个正在进行的音变的微观历史. *Journal of Chinese Linguistics*, *14*(1), 1–42. http://www.jstor.org/stable/23754216

Bauer, R. S., & Benedict, P. K. (1997). Modern Cantonese Phonology. In *Modern Cantonese Phonology*. De Gruyter Mouton. http://www.degruyter.com/document/doi/10.1515/9783110823707/html. DOI: https://doi.org/10.1515/9783110823707

Bauer, R. S., Cheung, K., & Cheung, P. (2003). Variation and merger of the rising tones in Hong Kong Cantonese. *Language Variation and Change*, *15*(02). DOI: https://doi.org/10.1017/S0954394503152039

Beddor, P. S. (2015). The Relation between Language Users' Perception and Production Repertoires. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences* (pp. 1–9). the University of Glasgow.

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. DOI: https://doi.org/10.1353/lan.2018.0051

Birdsong, D., Gertken, L. M., & Amengual, M. (2012). *Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism*. Centre for Open Educational Resources and Language Learning, University of Texas at Austin. https://sites.la.utexas.edu/bilingual/

Boersma, P., & Weenink, D. (2015). *Praat: Doing phonetics by computer* (5.4.08) [Computer software]. www.praat.org

Bourgerie, D. S. (1990). *A quantitative study of sociolinguistic variation in Cantonese* (Doctoral dissertation, The Ohio State University). http://search.proquest.com/docview/303872748/abstract/856EDC18ADFD4B8DPQ/1

Casillas, J. V. (2021). *Exploring phonemic boundaries using logistic regression*. https://www.jvcasillas.com/posts/2021-05-15_logistic_regression_and_phonemic_boundaries/

Casillas, J. V., & Simonet, M. (2016). Production and perception of the English /æ/–/ɑ/ contrast in switched-dominance speakers. *Second Language Research*, *32*(2), 171–195. DOI: https://doi.org/10.1177/0267658315608912

Chen, K. H. Y. (2018). Ideologies of Language Standardization: The Case of Cantonese in Hong Kong. In J. W. Tollefson & M. Pérez-Milans (Eds.), *The Oxford Handbook of Language Policy and Planning* (Vol. 1). DOI: https://doi.org/10.1093/oxfordhb/9780190458898.013.22

Cheshire, Jenny. 2006. Age- and generation-specific use of language. Sociolinguistics: An international handbook of the science of language and society, ed. by U. Ammond, N. Dittmar and K. Mattheier, 1552–63. Berlin, Germany: Walter de Gruyter. DOI: https://doi.org/10.1515/9783110171488.2.8.1552

Coetzee, A. W. (2018). Individual and community level variation in phonetics and phonology. In R. Mesthrie & D. Bradley (Eds.), *The Dynamics of Language: Plenary and focus lectures from the 20th International Congress of Linguists* (p. 17).

Coetzee, A. W., Beddor, P. S., Shedden, K., Styler, W., & Wissing, D. (2018). Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis. *Journal of Phonetics*, *66*, 185–216. DOI: https://doi.org/10.1016/j.wocn.2017.09.009

Dimov, S., Katseff, S., & Johnson, K. (2012). Social and personality variables in compensation for altered auditory feedback. In M.-J. Solé & D. Recasens (Eds.), *The initiation of sound change: Perception, production, and social factors* (Vol. 323, pp. 185–210). John Benjamins Publishing Company. DOI: https://doi.org/10.1075/cilt.323.15dim

Drager, K., & Kirtley, M. J. (2016). Awareness, Salience, and Stereotypes in Exemplar-Based Models of Speech Production and Perception. In A. M. Babel (Ed.), *Awareness and Control in Sociolinguistic Research* (pp. 1–24). Cambridge University Press. DOI: https://doi.org/10.1017/CBO9781139680448.003

Evans, B. G., & Iverson, P. (2007). Plasticity in vowel perception and production: A study of accent change in young adults. *The Journal of the Acoustical Society of America, 121*(6), 3814–3826. DOI: https://doi.org/10.1121/1.2722209

Evans, K. E., Munson, B., & Edwards, J. (2018). Does Speaker Race Affect the Assessment of Children's Speech Accuracy? A Comparison of Speech-Language Pathologists and Clinically Untrained Listeners. *Language, Speech, and Hearing Services in Schools, 49*(4), 906–921. DOI: https://doi.org/10.1044/2018_LSHSS-17-0120

Fok, C. Y. Y. (1979). The frequency of occurrence of speech sounds and tones in Cantonese. In L. Robert (Ed.), *Hong Kong Language Paper* (pp. 150–157). Hong Kong University Press.

Fridland, V., & Kendall, T. (2012). Exploring the relationship between production and perception in the mid front vowels of U.S. English. *Lingua, 122*(7), 779–793. DOI: https://doi.org/10.1016/j.lingua.2011.12.007

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2019). *irr: Various Coefficients of Interrater Reliability and Agreement* (0.84.1) [Computer software]. https://rdrr.io/cran/irr/

Garrett, A., & Johnson, K. (2013). Phonetic bias in sound change. In A. C. L. Yu (Ed.), *Origins of Sound Change* (pp. 51–97). Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199573745.003.0003

Grosvald, M., & Corina, D. P. (2012). The production and perception of sub-phonemic vowel contrasts and the role of the listener in sound change. In M.-J. Solé & D. Recasens (Eds.), *The initiation of sound change: Perception, production, and social factors* (Vol. 323, pp. 77–100). John Benjamins Publishing Company. DOI: https://doi.org/10.1075/cilt.323.08gro

Hall, K. C. (2013). A typology of intermediate phonological relationships. *The Linguistic Review, 30*(2), 215-275. DOI: https://doi.org/10.1515/tlr-2013-0008

Harnsberger, J. (2000). On the relationship between identification and discrimination of non-native nasal consonants. *Journal of the Acoustical Society of America, 110,* 489–503. DOI: https://doi.org/10.1121/1.1371758

Harrington, J. (2012). The coarticulatory basis of diachronic high back vowel fronting. *The Initiation of Sound Change,* 103–122. http://www.jbe.platform.com/content/books/9789027273666-cilt.323.10har. DOI: https://doi.org/10.1075/cilt.323.10har

Harrington, J., Kleber, F., & Reubold, U. (2008). Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. *The Journal of the Acoustical Society of America, 123*(5), 2825–2835. DOI: https://doi.org/10.1121/1.2897042

Hashimoto, O. Y. (1972). *Phonology of Cantonese* (Vol. 1). Cambridge University Press.

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics, 34*(4), 458–484. DOI: https://doi.org/10.1016/j.wocn.2005.10.001

Hickey, R. (2012). Internally- and externally-motivated language change. In J. M. Hernández-Campoy & J. C. Conde-Silvestre (Eds.), *The Handbook of Historical Sociolinguistics* (pp. 387–407). Wiley-Blackwell. DOI: https://doi.org/10.1002/9781118257227.ch21

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In J. Mullennix (Ed.), *Talker Variability in Speech Processing.* (pp. 145–165.). Academic Press.

Kataoka, R. (2011). *Phonetic and Cognitive Bases of Sound Change* (Doctoral dissertation, University of California, Berkeley). https://escholarship.org/content/qt3m20112c/qt3m20112c.pdf

Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., & Banno, H. (2008). Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing,* 3933–3936. DOI: https://doi.org/10.1109/ICASSP.2008.4518514

Kim, D., & Clayards, M. (2019). Individual differences in the link between perception and production and the mechanisms of phonetic imitation. *Language, Cognition and Neuroscience, 34*(6), 769–786. DOI: https://doi.org/10.1080/23273798.2019.1582787

Kleber, F., Harrington, J., & Reubold, U. (2012). The Relationship between the Perception and Production of Coarticulation during a Sound Change in Progress. *Language and Speech, 55*(3), 383–405. DOI: https://doi.org/10.1177/0023830911422194

Kong, E. J., & Edwards, J. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics, 59*, 40–57. DOI: https://doi.org/10.1016/j.wocn.2016.08.006

Koops, C., Gentry, E., & Pantos, A. (2008). The effect of perceived speaker age on the perception of PIN and PEN vowels in Houston, Texas. *University of Pennsylvania Working Papers in Linguistics, 14*(2). https://repository.upenn.edu/pwpl/vol14/iss2/12

Kuang, J., & Cui, A. (2018). Relative cue weighting in production and perception of an ongoing sound change in Southern Yi. *Journal of Phonetics, 71*, 194–214. DOI: https://doi.org/10.1016/j.wocn.2018.09.002

Kwon, H. (2015). *Cue Primacy and Spontaneous Imitation: Is Imitation Phonetic or Phonological?* (Doctoral dissertation, University of Michigan).

Labov, W. (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change, 2*(2), 205–254. DOI: https://doi.org/10.1017/S0954394500000338

Labov, W. (1994). *Principles of linguistic change. Vol. 1: Internal factors.* Blackwell.

Labov, W. (2001). *Principles of linguistic change. Vol. 2: Social factors* (Digital print). Blackwell.

Labov, W., Rosenfelder, I., & Fruehwald, J. (2013). One hundred years of sound change in Philadelphia: Linear incrementation, reversal, and reanalysis. *Language, 89*(1), 30–65. http://www.jstor.org/stable/23357721. DOI: https://doi.org/10.1353/lan.2013.0015

Lai, M.-L. (2001). Hong Kong Students' Attitudes Towards Cantonese, Putonghua and English After the Change of Sovereignty. Journal of Multilingual and Multicultural Development, 22(2), 112–133. DOI: https://doi.org/10.1080/01434630108666428

Lai, M. L. (2011). Cultural identity and language attitudes – into the second decade of postcolonial Hong Kong. Journal of Multilingual and Multicultural Development, 32(3), 249–264. DOI: https://doi.org/10.1080/01434632.2010.539692

Lau, S. (1977). A Practical Cantonese-English Dictionary. The Government Printer.

Lee, K. S., & Leung, W. M. (2012). The status of Cantonese in the education policy of Hong Kong. Multilingual Education, 2(1), 2. DOI: https://doi.org/10.1186/10.1186/2191-5059-2-2

Leung, M.-T., Law, S.-P., & Fung, S.-Y. (2004). Type and token frequencies of phonological units in Hong Kong Cantonese. Behavior Research Methods, Instruments, & Computers, 36(3), 500–505. DOI: https://doi.org/10.3758/BF03195596

Lev-Ari, S. (2017). Talking to fewer people leads to having more malleable linguistic representations. PLOS ONE, 12(8), e0183593. DOI: https://doi.org/10.1371/journal.pone.0183593

Lin, Y., Yao, Y., & Luo, J. (2021). Phonetic accommodation of tone: Reversing a tone merger-in-progress via imitation. Journal of Phonetics, 87, 101060. DOI: https://doi.org/10.1016/j.wocn.2021.101060

Mok, P. P. K., Zuo, D., & Wong, P. W. Y. (2013). Production and perception of a sound change in progress: Tone merging in Hong Kong Cantonese. Language Variation and Change, 25(03), 341–370. DOI: https://doi.org/10.1017/S0954394513000161

Morrison, G. (2007). Logistic regression modelling for first and second language perception data. In M.-J. Solé, P. Prieto, & J. Mascaro (Eds.), Segmental and prosodic issues in Romance phonology (pp. 219–236). John Benjamins. https://research.aston.ac.uk/en/publications/logistic-regression-modelling-for-first-and-second-language-perce. DOI: https://doi.org/10.1075/cilt.282.15mor

Narayan, C. (2008). The acoustic-perceptual salience of nasal place contrasts. Journal of Phonetics, 36, 191–217. DOI: https://doi.org/10.1016/j.wocn.2007.10.001

Ng, M. L., & Kwok, I. C. L. (2004). Frequency of occurrence of Cantonese word initials, finals and tones. Asia Pacific Journal of Speech, Language and Hearing, 9(3), 189–199. DOI: https://doi.org/10.1179/136132804805575868

Ou, J., & Law, S.-P. (2017). Cognitive basis of individual differences in speech perception, production and representations: The role of domain general attentional switching. Attention, Perception, & Psychophysics, 79(3), 945–963. DOI: https://doi.org/10.3758/s13414-017-1283-z

Pan, P. G. (1981). Prestige forms and phonological variation in Hong Kong Cantonese speech (Doctoral dissertation, The University of Hong Kong). DOI: https://doi.org/10.5353/th_b4389299

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. DOI: https://doi.org/10.1177/0023830919880206

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

Samuel, A. G., & Larraza, S. (2015). Does listening to non-native speech impair speech perception? Journal of Memory and Language, 81, 51–71. DOI: https://doi.org/10.1016/j.jml.2015.01.003

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics*, *52*, 183–204. DOI: https://doi.org/10.1016/j.wocn.2015.07.003

Schneider, W., Eschman, A., & Zuccolotto, A. (2007). *E-Prime 2.0 professional.* Psychology Software Tools.

Schultz, A. A., Francis, A. L., & Llanos, F. (2012). Differential cue weighting in perception and production of consonant voicing. *The Journal of the Acoustical Society of America, 132*(2), EL95–EL101. DOI: https://doi.org/10.1121/1.4736711

Sharma, D., & Sankaran, L. (2011). Cognitive and social forces in dialect shift: Gradual change in London Asian speech. *Language Variation and Change*, *23*(3), 399–428. DOI: https://doi.org/10.1017/S0954394511000159

Soo, R., & Babel, M. (2020, November 19). *Recognition & representation of Cantonese sound change variants in the lexicon.* 61st Annual Meeting of the Psychonomics Society.

Sumner, M., Kim, S. K., King, E., & McGowan, K. (2014). The socially weighted encoding of spoken words: A dual-route approach to speech perception. *Frontiers in Psychology*, *4*, 1015. DOI: https://doi.org/10.3389/fpsyg.2013.01015

Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language*, *60*(4), 487–501. DOI: https://doi.org/10.1016/j.jml.2009.01.001

Tamminga, M. (2019). Interspeaker covariation in Philadelphia vowel changes. *Language Variation and Change, 31*(2), 119–133. DOI: https://doi.org/10.1017/S0954394519000139

To, C. K. S., Mcleod, S., & Cheung, P. S. P. (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. DOI: https://doi.org/10.3109/02699206.2014.1003329

Voeten, C. C. (2020). Individual differences in the adoption of sound change. *Language and Speech*, 1–37. DOI: https://doi.org/10.1177/0023830920959753

Wagner, S. E. (2012). Age grading in sociolinguistic theory. *Language and Linguistics Compass*, *6*(6), 371–382. DOI: https://doi.org/10.1002/lnc3.343

Wang, L., & Kirkpatrick, A. (2015). Trilingual education in Hong Kong primary schools: An overview. *Multilingual Education, 5*(1), 3. DOI: https://doi.org/10.1186/s13616-015-0023-8

Whelpton, J. (1999). The future of Cantonese: Current trends. *Hong Kong Journal of Applied Linguistics*, *4*(1), 43–60.

Wong, A. D. (2019). Authenticity, belonging, and charter myths of Cantonese. *Language & Communication*, *68*, 37–45. DOI: https://doi.org/10.1016/j.langcom.2018.10.011

Wong, S. L. (1941). *A Chinese syllabary pronounced according to the dialect of Canton.* Chung Hwa Book Co.

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. DOI: https://doi.org/10.1353/lan.2016.0031

Yeung, S. (1980). *Some aspects of phonological variations in the Cantonese spoken in Hong Kong* [Master's thesis]. Hong Kong University.

Yu, A. C. L. (2010). Perceptual compensation is correlated with individuals' "autistic" traits: Implications for models of sound change. *PLoS ONE, 5*(8), e11950. DOI: https://doi.org/10.1371/journal.pone.0011950

Yu, A. C. L., Abrego-Collier, C., & Sonderegger, M. (2013). Phonetic Imitation from an Individual-Difference Perspective: Subjective Attitude, Personality and "Autistic" Traits. *PLoS ONE, 8*(9), e74746. DOI: https://doi.org/10.1371/journal.pone.0074746

Yu, A. C. L., & Zellou, G. (2019). Individual differences in language processing: Phonology. *Annual Review of Linguistics, 5*(1), 131–150. DOI: https://doi.org/10.1146/annurev-linguistics-011516-033815

Zee, E. (1999a). Change and variation in the syllable-initial and syllable-final consonants in Hong Kong Cantonese / 香港粤语中声母及韵尾辅音之变化与变异. *Journal of Chinese Linguistics, 27*(1), 120–167. JSTOR. http://www.jstor.org/stable/23756746

Zee, E. (1999b). Chinese (Hong Kong Cantonese). In I. P. Association (Ed.), *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet* (pp. 58–60). Cambridge University Press.

Zellou, G. (2017). Individual differences in the production of nasal coarticulation and perceptual compensation. *Journal of Phonetics, 61*, 13–29. DOI: https://doi.org/10.1016/j.wocn.2016.12.002

Zhang, J. (2019). Tone mergers in Cantonese: Evidence from Hong Kong, Macao, and Zhuhai. *Asia-Pacific Language Variation, 5*(1), 28–49. DOI: https://doi.org/10.1075/aplv.18007.zha