Journal of Phonetics 87 (2021) 101060

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Research Article

Phonetic accommodation of tone: Reversing a tone merger-in-progress via imitation

Yuhan Lin [a,b], Yao Yao [a,c,*], Jin Luo [a]

[a] Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, HKSAR, China
[b] School of Foreign Languages, Shenzhen University, China
[c] PolyU-PekingU Research Centre on Chinese Linguistics, HKSAR, China

ARTICLE INFO

Article history:
Received 4 February 2020
Received in revised form 29 March 2021
Accepted 7 April 2021
Available online 11 May 2021

Keywords:
Phonetic imitation
Merger reversal
Sound change
Cantonese tone

ABSTRACT

Previous literature has documented phonetic accommodation for various segmental and suprasegmental features, but the accommodation of tone remains under-explored. The current study contributes to the literature by investigating the accommodation of two merging tones in Hong Kong Cantonese, mid-level Tone 3 (T3) and low-level Tone 6 (T6), in a speech shadowing experiment. Specifically, we ask whether shadowers will reverse the merging trend after exposure to a model talker with distinct T3–T6 productions and if so, what factors will modulate the accommodative behaviors. Evidence for phonetic imitation is observed, but the effect varies by shadower's baseline production of T3–T6 distinction and across different model talkers. Shadowers with less baseline tonal distinction exhibit greater degrees of imitation, suggesting that greater linguistic distance facilitates imitation. More robust imitation is observed in the young model talker condition, but the effect is most likely driven by the talker's idiosyncratic production. Shadowers' impression of the model talker and attitudes towards ongoing changes in the language did not show substantial impact on imitation. We discuss the implications of these findings for theories of speech perception and production.

1. Introduction

Phonetic accommodation refers to the phenomenon of spontaneously adjusting one's speech when conversing with another talker.[1] In theory, the result of adjustment could be either sounding more similar to the interlocutor (i.e. convergence) or more different (i.e. divergence); but in reality, convergence is more widely documented than divergence, probably because imitation is a fundamental human behavior. Indeed, spontaneous phonetic imitation plays a central role in almost all core aspects of language learning and language use. Infants acquire spoken words by imitating the speech sounds they hear (Kuhl & Meltzoff, 1996); adults who migrate to a new country or region gradually pick up the ambient language/accent by interacting with the locals (Chang, 2012; Sancier & Fowler, 1997). Imitation is especially important for sound change. When an innovative pronunciation is spreading in a linguistic community, speakers come into contact with the new sound when talking to other members of the community; if a speaker imitates and adopts the new sound in her future production, not only will she exhibit the sound change herself, she will also advance the sound change by passing it on to the next interlocutor (Labov, 2001; Siegel, 2010). In other words, phonetic imitation on the individual level is the main vehicle for the propagation of sound change in the community.

1.1. Phonetic imitation: empirical evidence

Most of the empirical evidence for phonetic imitation comes from speech shadowing studies in the lab. In a typical shadowing task (Goldinger, 1998; Namy, Nygaard, & Sauerteig, 2002; Shockley, Sabadini, & Fowler, 2004; etc.), the participant (i.e. shadower) first produces an item as they usually do (i.e. baseline production), then hears and repeats after a model talker's production of the same item (i.e. shadowing production), and finally produces the item again (i.e. post-exposure production). Imitation is examined by comparing the baseline and

[*] Corresponding author at: Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hung Hom, Hong Kong.
E-mail address: ctyaoyao@polyu.edu.hk (Y. Yao).

[1] Throughout the paper, we use the terms "accommodation" and "(spontaneous) imitation" interchangeably, both referring to phonetic accommodation (imitation) specifically.

https://doi.org/10.1016/j.wocn.2021.101060
0095-4470/© 2021 The Authors. Published by Elsevier Ltd.

shadowing/post-exposure productions, in terms of either perceptual similarity ratings (from the AXB task) or acoustic–phonetic properties (see Pardo, Urmanche, Wilman, & Wiener (2017) for a review).

The ubiquity of phonetic imitation attests to the malleability of an individual's sound system, which allows for constant evolution and restructuring across the lifespan (Harrington, Palethorpe, & Watson, 2000; Sankoff, 2019). Nevertheless, imitation does not *always* occur where it could, and even when imitation is observed, it does not *always* occur to the fullest possible extent. In some sense, this is not surprising. We know that language acquisition is often imperfect: it takes children quite a few years after producing the first word to achieve adult-like pronunciation norms (Dodd, Holm, Hua, & Crosbie, 2003; Menn, 1983; Smit, Hand, Freilinger, Bernthal, & Bird, 1990); adult second language/dialect learners rarely pass as native-sounding, even after years of immersion in the second language/dialect environment (Flege, Munro, & MacKay, 1995; Nycz, 2011; Piske, MacKay, & Flege, 2001; Siegel, 2010). Similarly, a sound change may never be complete, resulting in an extended period of synchronic variation. A long line of literature has discussed how oscillations in the phonological system are shaped by a multitude of linguistic and social factors (Bybee, 2006; Flege, 1995; Labov, 1994, 2001; Ohala, 1993; Pierrehumbert, 2002; etc.). It stands to reason that phonetic imitation—a critical process in sound change and variation—is likewise bounded by various constraints.

The most obvious constraints are related to biophysiological differences across talkers, which determine the upper bounds of how much we can sound like someone else. Previous literature reveals at least three additional groups of factors that could modulate the propensity and extent of imitation. The first group regards how the imitator parses relevant phonetic information, both perceptually and linguistically. For example, speakers who are better able to use acoustic–phonetic information when perceiving vowels are more effective at imitating vowels (Kim & Clayards, 2019); speakers with stronger attention focus in general produce greater imitation in voice onset time (VOT) (Yu, Abrego-Collier, & Sonderegger, 2013). In a similar vein, Garnier et al. (2013) reported fMRI data showing that the levels of neural activations in multiple brain regions during the listening stage were predictive of the degree of imitation in fundamental frequency (F0) in the subsequent production stage. Furthermore, imitation is influenced by the speaker's phonological knowledge. Among others, Nielsen (2011) reported VOT lengthening by English-speaking shadowers after the exposure to hyper-aspirated /p/-initial words, but no corresponding VOT shortening after the exposure to under-aspirated stimuli. The asymmetry is commonly attributed to contrast preservation (Flemming, 1996, 2004; Liljencrants & Lindblom, 1972; Wedel, Kaplan, & Jackson, 2013), as voiceless stops with short-lag VOTs may jeopardize the contrast between voiced and voiceless stops. Mitterer and Ernestus (2008) further proposed that the target of phonetic imitation was strictly phonological—not phonetic—based on results from a speeded shadowing experiment in Dutch, which found imitation only for phonemically important acoustic cues (e.g. prevoicing) but not for subphonemic detail (e.g. length of prevoicing and /r/ variant).

The second group of factors concerns the linguistic distance between the imitator and the model talker, but the effect is not straightforward. On one hand, speakers seem to have a greater tendency to imitate less familiar productions (Nye & Fowler, 2003) and atypical voices (Babel, McGuire, Walters, & Nicholls, 2014), which is in line with Goldinger's (1998) finding of low-frequency (hence less familiar) words showing greater imitation than high-frequency words. In the case of cross-dialectal imitation, greater imitation is found in speakers whose dialectal background is further away from the model talker's (Walker & Campbell-Kibler, 2015), and in sounds that exhibit greater cross-dialectal differences (for example, Babel (2012) reported the greatest imitation in low vowels /æ ɑ/, which vary substantially across American English dialects). On the other hand, several studies reported a negative correlation between linguistic distance and imitation, especially when the model production is perceived as non-native or foreign. Native English speakers are more likely to imitate other native English speakers than they would imitate English-L2 speakers (Kim, Horton, & Bradlow, 2011). English and Spanish native speakers tend to only imitate the VOT regions that are used in their native languages but not the unfamiliar, "non-native" VOT regions (Olmstead, Viswanathan, Aivar, & Manuel, 2013). Taken together, linguistic distance constrains imitation in a non-linear manner: when distance is either too small (not perceptible) or too large (not native), imitation is impeded.

The third group of modulating factors are social factors (e.g. gender, attitude, and social awareness), and the associated effects are highly variable. Studies on gender effects (see Coles-Harris 2017 for a recent review) have documented opposite patterns of talker gender effects (e.g. Namy, Nygaard, & Sauerteig, 2002; Pardo, 2006) as well as null effects (Pardo, Jordan, Mallari, Scanlon, & Lewandowski, 2013; Pardo et al., 2017). Attitude and awareness are both complex psychological constructs that are hard to characterize or measure**.** Overall a more positive attitude toward the model talker and their speech variety predicts greater convergence (Babel, 2010; Babel et al., 2013; Pardo, Gibbons, Suppes, & Krauss, 2012; Yu et al., 2013), but the effects are quite nuanced (Babel, 2012). As for awareness, previous literature presents mixed findings regarding whether and how imitation is influenced by the speaker's knowledge of the linguistic variable: While some claimed that greater awareness of the phonetic variants should lead to more imitation (Drager & Kirtley, 2016; Trudgill, 1981), a counterexample was presented in Babel (2010), with New Zealand English speakers converging the most toward Australian English model speech in the DRESS vowel, a cross-dialectal distinction that is *not* salient for the imitators. This suggests that awareness by itself is probably neutral, and therefore high-awareness items may be either more or less likely to be imitated, depending on whether other factors (such as attitude) are in favor.

As reviewed above, much of the acoustic–phonetic evidence for imitation comes from convergence in VOT (Fowler, Brown, Sabadini, & Weihing, 2003; Nielsen, 2011; Olmstead, Viswanathan, Aivar, & Manuel, 2013; Sanchez, Miller, & Rosenblum, 2010; Shockley, Sabadini, & Fowler, 2004; Yu, Abrego-Collier, & Sonderegger, 2013) and vowel spectra

(Babel, 2010, 2012; Pardo et al., 2012, 2017; Walker & Campbell-Kibler, 2015; etc.), but an increasing number of shadowing studies are attending to the imitation of F0. Convergence in both raw F0 (Aubanel & Nguyen, 2020; Babel & Bulatov, 2012; Garnier, Lamalle, & Sato, 2013; Pardo et al., 2017; Postma-Nilsenovà & Postma, 2013; Sato et al., 2013) and more complicated intonational patterns (D'Imperio, Cavone, & Petrone, 2014; German, 2012; Kim, 2016; Mantell & Pfordresher, 2013; Michelas & Nguyen, 2011; Postma-Nilsenovà & Postma, 2013; Wisniewski, Mantell, & Pfordresher, 2013) have been reported in previous studies.

An interesting finding arising from this growing literature is the disparity between linguistic and non-linguistic processing of pitch. Mantell and Pfordresher (2013) asked participants to imitate both spoken utterances and wordless melodies, and found that the imitation of absolute F0 was more accurate for wordless melodies, but the imitation of relative pitch contour was more accurate for spoken utterances. One way to interpret this finding is that relative pitch is more important than absolute pitch for linguistic processing, which often requires F0 normalization. In this regard, lexical tone is probably the most prominent example of utilizing relative pitch variation to encode linguistic information (i.e. lexical contrast). However, to the best of our knowledge, there has not been any formal investigation of tonal imitation. Previous studies of F0 imitation, most of which focused on absolute F0, examined non-tonal languages such as English and French, where F0 does not provide contrastive cues for lexical identification.

The main goal of the current study is to fill this gap in the imitation literature by investigating how speakers imitate lexical tones. Specifically, we report a shadowing study regarding two Hong Kong Cantonese (HKC) tones—mid-level Tone 3 and low-level Tone 6—that are undergoing a merger. In doing so, we also contribute to the research of phonetic imitation in a less well-studied context, i.e. sound change and variation.

1.2. Phonetic imitation: theoretical accounts

Since imitation involves both perceptual input and production output, multiple theories of speech perception and production have been applied to account for this phenomenon, including motor theory (Liberman & Mattingly, 1985), direct realist theory (Fowler, 1986), exemplar theory (Goldinger, 1998; Johnson, 1997; Pierrehumbert, 2002), the interactive alignment account (Gambi & Pickering, 2013; Pickering & Garrod, 2007, 2013), and the Communication Accommodation Theory (Giles et al., 1991; Giles, 1973). The theoretical discussion of this study is mainly situated in the frameworks of the exemplar theory and the interactive alignment account, both of which can accommodate linguistic as well as social constraints on imitation.

In an exemplar model, each category is represented by a "cloud" of tokens (i.e. exemplars) that the speaker has previously encountered and remembered, where acoustic-phonetically similar tokens are located close to each other and dissimilar tokens are farther away. Pierrehumbert (2001) lays out several key principles regarding how production proceeds in an exemplar model: first, a certain location in the cloud is selected as the production goal, then the *n*-nearest exemplars—in terms of activation-weighted distance from the chosen location—will be averaged in phonetic values to compute the final target of production. Following these principles, imitation occurs when the target of production moves toward a newly encountered exemplar as the result of the averaging calculation.

The interactive alignment account, on the other hand, assumes that listeners are constantly making predictions at different linguistic levels (semantics, syntax, phonetics) about upcoming speech from their interlocutors, and that prediction can happen along two routes, prediction-by-association and prediction-by-simulation. The association route is based on the listener's past experience of language comprehension—analogous to how we can predict the sound of thunders after seeing lightnings. The simulation route calls on both perception and production processes, and is thus more relevant for imitation. Using the simulation route, the listener covertly imitates the interlocutor's unfolding speech, derives the production commands for the heard speech and the upcoming speech, and then uses a forward production model to predict what the upcoming speech should sound like. Thus, if a listener is exposed to an accent that is different from her own, her simulation models may undergo gradual revision to achieve more accurate predictions. When doing so, updates to the forward production model will in turn impact the listener's own production system due to shared representations, resulting in a phonetic drift toward the ambient accent.

While both theories can explain imitation, they resort to different mechanisms to account for the constraints on imitation. The exemplar theory allows various linguistic and social factors (e.g. frequency, recency, salience, attitude, awareness) to adjust the activation levels of relevant exemplars, and hence influence the calculation of the production target. The interactive alignment account strictly predicts that the goodness of alignment (i.e. degree of convergence) relies on the degree of similarity between the two talkers' language systems. The more similar the two talkers are, the more likely they will predict each other's speech behaviors through the simulation route (as opposed to the alternative route by association), and the more likely their predictions will be accurate—thus perpetuating alignment. Social factors such as language attitudes come into play through the mediation of similarity (Gambi & Pickering, 2013), as positive attitudes toward an accent tend to correlate with increased experience, familiarity, and linguistic similarity.

1.3. Current study

Despite the inherent link between imitation and sound change, there has not been much investigation of speech accommodation in a sound change context (a notable exception is Babel et al., 2013; see the discussion below). Compared to other imitation scenarios, ongoing sound changes present an interesting and unique environment. For example, linguistic distance between the imitator and the model talker will be defined as the relative degree of advancedness in the change, as opposed to cross-linguistic/dialectal distance used in previous studies. In addition, changes-in-progress are known to be influenced by asymmetries between perception and production (e.g. Labov, Karen, & Miller, 1991; Yu, 2007) and strong effects of age, gender, attitude, and other social factors (Labov, 2001), all of which create a distinctive context for potential imitation to occur between two interlocutors who are at different stages of the sound change. Examining phonetic imitation in a sound change environment will not only further our understanding of the imitation phenomenon, but also illuminate the underlying mechanisms of the propagation of sound change.

In this study, our central inquiry is whether speakers would tend to reverse an ongoing sound change—specifically, a merger-in-progress—via imitation. While mergers have been commonly observed, merger reversals are deemed rare, if not impossible (Garde, 1961). It is certainly not realistic to expect speakers who have completely merged the two categories to fully uncover the contrast in one experimental session. What we are interested in is whether speakers who are en route to a merger—who may or may not have completed the merger—would show a reversing trend by decreasing the degree of mergedness in their production. Previous works by Babel et al. (2013) and Yao and Chang (2016) demonstrate that speakers of a language/dialect X, where two phonological categories are merging, do show trends of reversing the merger during or after shadowing the speech of another (closely related) language/dialect Y, where the two categories are clearly distinguished. Specifically, Babel et al. (2013) examined New Zealand English speakers shadowing Australian English, while Yao and Chang (2016) examined Shanghainese speakers exposed to Mandarin.

Apart from the extension to tonal imitation (both Babel et al., 2013; Yao & Chang, 2016 examined vowel merger reversals), the current study furthers this line of research in two other key aspects. First, both the shadowers and the model talkers in the current study are from the *same* linguistic community (regional and dialectal) but differing in how advanced they are in the sound change, with the model talkers being the least advanced (i.e. most conservative) and producing the canonical, unmerged pronunciations. The same-language/dialect design is more generalizable to sound changes with internal origins. In addition, while the previous studies focused on the effects of attitude and amount of L2 exposure on imitation, we add to the discussion the factors of age (generation) and the shadower's initial sound change status. We manipulate the age difference between the shadower and the model talker, as age is a strong correlate of pronunciation variation amidst a sound change; furthermore, we ask whether shadowers, who are at varying stages of the merger in their baseline production, would shift their perception/production of the two categories to the same degree toward the unmerged model productions.

Before we delve into the detail of the current study, we will first provide a brief description of the tones in Hong Kong Cantonese (HKC), the majority language of Hong Kong and a member of the Chinese language family. Among the six phonological tones in HKC (see Fig. 1), of interest here are the mid-level Tone 3 (T3) and the low-level Tone 6 (T6), which are mainly contrasted by tone height while being nondistinguishable in tonal contour or duration (Bauer & Benedict, 1997; Peng, 2006; Zee, 1998). A trend to merge the two tones, mostly through T6 raising (Zhang, Zhang, & Xu, 2019), is part of a group of (ongoing) sound changes that constitute a novel HKC accent known as the "lazy pronunciation", which is characterized by the elimination of multiple phonological contrasts (Bauer, 1986; Bauer, Cheung, & Cheung, 2003; Mok, Zuo, Wong, 2013; To, Mcleod, & Cheung, 2015). As the "lazy" accent is widely associated with the younger generations by the general public (which is not necessarily true, see Fung & Lee, 2019; To et al., 2015), various efforts have been put forward by schools and the mass media to "correct" the young people's speech (e.g. Ho, 2001), which in turn raises the awareness about the accent in the society.

Compared to other sound changes associated with the "lazy" accent, the T3–T6 merger is less advanced in progress and overall less noticed by speakers. Previous studies (Fung & Lee, 2019; Mok et al., 2013) reported incomplete T3–T6 merger in production and almost intact distinction in perception even among those who have merged in production (>97% accuracy in the AX discrimination task, although reaction times are slower (Mok et al., 2013)). Our own survey data indicate that young speakers are generally not aware of the T3–T6 merger being part of the "lazy" accent, as none listed it as an example of "lazy" pronunciation.

1.4. Hypotheses

Our main hypothesis is that shadowers will reverse the trend of merging T3 and T6 by increasing the tonal distinction after exposure to unmerged, canonical productions from the model talker. Furthermore, we predict that the likelihood (and degree) of imitation is influenced by age difference (between

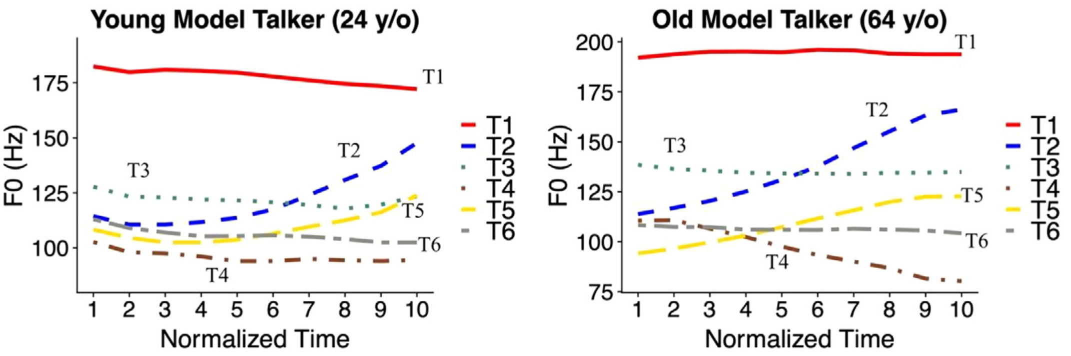

Fig. 1. F0 of six Cantonese lexical tones on the syllable [si] by the two model talkers. In the 5-point system proposed in Chao (1930), the six tones are transcribed as T1 [55], T2 [25], T3 [33], T4 [21], T5 [23], and T6 [22], respectively.

the shadower and the model talker) and the shadower's baseline merger status. To facilitate the investigation, we include two model talkers, one young and one old, and a group of young shadowers. Both model talkers produce clear distinctions between T3 and T6 (see Fig. 1 and *Section 2.4.1* for more detail). Two competing predictions are formed regarding the effect of age difference. On one hand, a young model talker, who is close in age to the shadowers, may induce greater imitation, due to peer influence (Eckert, 2003). On the other hand, unmerged T3 and T6 productions from an older model talker may better match the stereotype of older talkers sounding more standard, and thus produce a greater imitation effect (Drager, 2010; Walker & Hay, 2011).

Since the T3–T6 merger is very much in progress and that the strongest merging trend is observed in production, we expect the young shadowers to be distributed over a range of T3–T6 mergedness—from less merged (i.e. greater distinction) to more merged—in their baseline production. In this paper, we use the terms "non-merging shadowers" and "merging shadowers" to refer to two broadly defined groups, showing a weak or strong trend of merging the two tones, respectively. How should a shadower's baseline merger status affect their imitation of unmerged model productions? Two alternative predictions can be formed. Non-merging shadowers may be more likely to imitate, because they are perceptually more sensitive to the acoustic–phonetic distinction of T3 and T6, evidenced by faster reaction times in AX discrimination in previous studies. Conversely, it is also likely for merging shadowers to imitate more, due to a larger phonetic distance from the model speech, which, importantly, is not large enough to be considered as cross-linguistic.

Lastly, we also include in the study an exploratory investigation of attitudinal effects, regarding the attitudes towards both the model talker and the "lazy" accent. Following the existing literature, we predict that a more positive impression of the model talker should predict greater imitation, but a more positive attitude toward the "lazy" accent should predict less imitation of the model talker's conservative speech.

2. Methods

2.1. Shadowers

The shadower group consisted of 63 young native HKC speakers[2] (aged 18–25, born between 1993 and 2001; 20M, 43F), randomly assigned to the Young (*N* = 31) or Old (*N* = 32) Model Talker condition.[3] All the shadowers were students in a local university at the time of testing, and none of them was enrolled in a music major.

2.2. Stimuli

All the items used in the experiment are monosyllabic words in HKC, each represented by a traditional Chinese character. To minimize the chance of the shadower not recognizing the characters, we only used high-frequency characters (>3500 occurrences in the *Chinese Character Database* (Research Centre for Humanities Computing Chinese University of Hong Kong, 2003)) or characters often used in colloquial HKC (e.g. *gwai3* 'expensive', etc.), and avoided characters with multiple readings (i.e. homographs). The production task used 27 critical items (14 T3 and 13 T6, including three minimal pairs) in the baseline and post-shadowing production blocks and a subset of 15 items (8 T3 and 7 T6, including two minimal pairs) in the shadowing blocks (see Appendix for a full list). The perception task (AX discrimination) used 12 critical minimal pairs (including the three pairs used in the production tasks), which formed 12 AX trials (where the two stimuli were different, roughly balanced in the order of tones) and 12 AA trials (where the two stimuli were the same).[4] A matching number of filler trials with T1 (high-level [55]) or T2 (high-rising [25]) items were included in each block, and there was no repeated syllable (including tone) in the items within a block. The use of filler items also allows the shadowers to have a general idea of the model talker's tonal range. To prevent the participant from developing task-specific strategies and focusing only on tonal contrasts in the AX task, filler AX trials all featured segmental contrasts.

The auditory stimuli were recorded by two male model talkers from substantially different age groups (Young: 24 yr male; Old: 64 yr male). Both model talkers had extensive training in Cantonese linguistics, and clearly distinguished T3 and T6 in production with an average difference of 11–24 Hz (see Table 1; more details on model talker's tonal distinctions in *Section 2.4.1*). The recording took place in a sound-attenuated booth at a sampling rate of 44.1 kHz; the model talkers read each test item in isolation, critical items appearing before the controls, both groups ordered alphabetically based on jyutping romanization. All auditory stimuli were scaled to 70 dB in average intensity in Praat (Boersma & Weenink, 2019).

2.3. Procedure

Fig. 2 shows the order of events in the shadowing experiment. In the baseline and post-shadowing production blocks, shadowers produced the characters in isolation, with each character displayed in the middle of the screen for 2500 ms after a fixation period of 500 ms. In the shadowing blocks, the auditory stimulus started 200 ms before the display of the character, and the character was displayed for 2500 ms. Shadowers were instructed to read out the character after the model talker. The two shadowing blocks cycled through the same set of stimuli, with no break between the two. The baseline and post-shadowing perception blocks used the

[2] Given the high degree of societal multilingualism, all the shadowers also speak English and Mandarin Chinese as a second or third language. Since English is non-tonal and Mandarin only has one level tone (i.e. high-level Tone 1), it is unlikely for either language to influence the HKC T3–T6 merger. However, seven shadowers also reported being exposed to Hakka (*N* = 3) or Chaoshan Min (*N* = 4) at home, both of which are Chinese languages with more than one level tone. To examine possible influence of Hakka/Min on the imitation results, the four best-fit models reported in Section 4.1 were re-run without these shadowers, and all the significant results sustained.

[3] Besides the 63 participants reported here, seven additional HKC speakers also took part in the study, but their data were excluded from analysis, because of technical errors during recording (*N* = 3) or high proportions (>50%) of creaky voice in the production (*N* = 4).

[4] Two T3 items and 1 T6 item were accidentally included the practice trials for the production task and 1 T3–T6 minimal pair was in the perception practice trials. They were excluded from the analysis and were thus not counted here.

Table 1
Mean and standard deviation (Hz) of the model talkers' T3 and T6 productions.

	T3	T6
Young	126.0 Hz (SD = 2.9)	102.8 Hz (SD = 2.6)
Old	103.2 Hz (SD = 4.1)	92.1 Hz (SD = 2.9)

same AX discrimination task, where shadowers pressed designated keys to indicate their responses (same/different) to the question "Are the two Cantonese pronunciations the same?". Reaction time was calculated from the onset of the second syllable and only responses within the first 5000 ms were collected. Each block started with some practice trials: two in a production task and eight (with feedback) in a perception task. Order of trials was randomized in each block, and shadowers were allowed to take self-paced breaks. On average, a complete session lasted around 20 minutes. The experiment was conducted using OpenSesame 3.2.5 (Mathôt, Schreij, & Theeuwes, 2012) in a soundproof booth, with a desktop computer connected to a headset and a head-mounted AKG C520 microphone.

After the shadowing experiment, the shadowers also completed a sociolinguistic questionnaire consisting of three sections: impression of model talker, language attitudes, and demographic information (See the Supplementary File for a full list of questions). In the first section, shadowers provided evaluations/estimation of the model talker's age and ten personal traits ("native-sounding", "standard-sounding", "well-educated", "authoritative", "friendly", "likable", "lively", "introverted", "cool", and "unnatural"; each item rated on a 6-point Likert scale from "1 – strongly disagree" to "6 – strongly agree"). The items were intended to reflect three well-established dimensions of personality traits in the literature on speech evaluation (Zahn & Hopper, 1985): status, solidarity, and dynamism. The second section elicits shadowers' attitudes towards the younger generation's HKC pronunciation, given the widespread belief that they speak with a "lazy accent". Specifically, the shaodwers rated six attitudinal statements and answered an open-ended question about what they think are the differences between young and old speakers' HKC accents. In the last section, shadowers provided their own gender, age, and linguistic backgrounds.

2.4. Analysis

2.4.1. Measures of T3–T6 distinction

A total of 5292 critical production tokens (84 syllables * 63 shadowers) were collected. All the acoustic analysis was conducted in Praat. Each token was hand marked for syllable boundaries and voiced portions (from the first to the last regular vibration cycle), with seriously creaky portions (with sudden drops in F0) manually checked and excluded from subsequent calculation of F0. Mean F0 was first averaged from automatically extracted raw F0 at eight equidistant points in the middle two thirds of the voiced portion, and then normalized by speaker and block so that tone heights can be compared across speakers and blocks. Following previous studies of Chinese tones (Fung & Wong, 2011; Shi & Wang, 2006), we used the T scale (see the formula in (1)) for normalization, where $F0_{min}$ and $F0_{max}$ come from the lowest- and highest-F0 points of the filler tokens (T1 and T2) by the same shadower in the same block. One might notice that while T1 is the highest tone in HKC, T2 is *not* the lowest tone even though it has quite a low start (see Fig. 1). We did not use lower tones like T4 or T5 because either the tone is involved in another merger with one of the critical tones (e.g. T4 and T6) or their lowest point is often creaky and hard to measure. Since the same procedure was applied to all the shadowers and blocks, the purpose of normalization should still be effectively achieved.

$$T = 5 * ((log(F0_x) - log(F0_{min}))/\ (log(F0_{max}) - log(F0_{min}))) \qquad (1)$$

where $F0_x$ is a raw F0 value in Hz, $F0_{max}$ and $F0_{min}$ represent the maximum and minimum raw F0 value of a given shadower in a given block. The resulting T value ranges from 0 to 5.

Unless otherwise stated, all the F0 measures in subsequent reporting refer to the normalized F0 (in T). If shadowers indeed imitate the model talkers, we would expect greater differences between the F0 values of T3 and T6 in shadowing and post-shadow blocks than in baseline. An additional measure of T3–T6 distinction is the Kolmogorov–Smirnov score ("K–S score" hereafter), which is generated by the namesake test and measures the distributional overlap between two sets of unidimensional observations (Egbert & Laflair, 2018). Similar to the Pillai score used to measure the distributional overlap in two-dimensional vowel formants (Babel et al., 2013; Hall-Lew, 2010), the K–S score ranges from 0 to 1, with a greater value indicating greater distinction. We calculated the K–S score for the F0 values of T3 and T6 by block by shadower. Thus, imitation of unmerged T3–T6 model productions will manifest as increasing K–S scores during/after shadowing compared to the baseline.

We also calculated the normalized F0 and K-S scores for the two model talkers, based on their recordings of the stimuli, in order to confirm that they produce sufficient T3–T6 distinction. Both model talkers' T3 and T6 are significantly different in terms of normalized F0 (*p*s < 0.001 in the *t*-tests), yielding an average tonal difference of 1.27 T (Young) and 0.8 T (Old), respectively. Both talkers' K–S scores are close to 1 (Young: 1; Old: 0.95), indicating a (near-)complete separation of the two tones' F0 distributions. Between the two talkers, the Young Model Talker exhibits a greater tonal distinction than the Old Model Talker in terms of both mean F0 and K–S score.

For the perception tasks, accuracy and (log-transformed) reaction time in AX trials were analyzed, excluding trials with

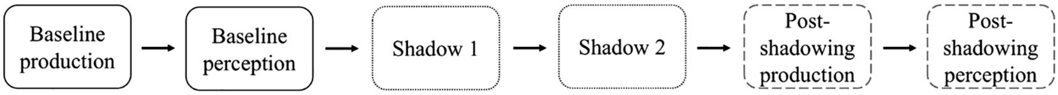

Fig. 2. Procedure of the shadowing experiment. Each box represents an experimental block.

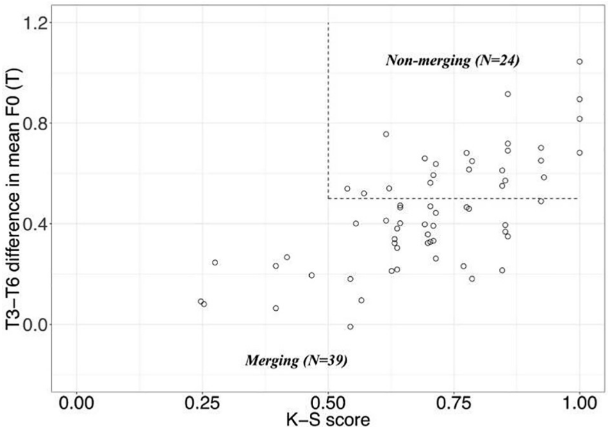

Fig. 3. Division of merging (F0: *M* = 0.30 T, *SD* = 0.13; K–S score: *M* = 0.64, *SD* = 0.17) and non-merging (F0: *M* = 0.67 T, *SD* = 0.13; K–S score: *M* = 0.81, *SD* = 0.14) shadowers based on T3–T6 difference in mean F0 and K–S score. The dashed lines indicate the boundary (mean F0 difference = 0.5 T and K–S score = 0.5) between the two groups.

reaction time two standard deviations away from the mean (24 trials * 63 shadowers = 1512 trials in total; 3.51% excluded).

2.4.2. Coding variables

Shadower's initial merger status was coded in a binary variable using their productions in the baseline block. A shadower was coded as Non-merging if they distinguished T3 and T6 by more than 0.5 T and had a K-S score of over 0.5 (midpoint of the range); otherwise the shadower was coded as Merging. The threshold of 0.5 T was chosen because it was roughly the midpoint between 0 (i.e. no difference between T3 and T6) and the average of the two model talkers' T3–T6 difference ((1.27 + 0.8)/2 = 1.035 T). As shown in Fig. 3, almost all the shadowers are distributed in the range of 0 and 1 T on the y-axis. Using this criterion, the Young Model Talker condition has 15 non-merging (F = 10) and 16 merging (F = 11) shadowers, whereas the Old Model Talker condition has 9 non-merging (F = 6) and 23 merging (F = 16) shadowers.

Since the current study hypothesizes that the degree of imitation is mediated by the model talker's age, we asked the shadowers to guess the age of the model talker in the sociolinguistic questionnaire. The young model talker was indeed perceived to be younger (M = 25.7 yr) than the older talker (M = 38.2 yr) by at least one generation, although the perceived age difference was smaller than the actual difference. For simplicity, we coded perceived talker age as a binary variable, indicating whether the shadower recognized the talker as Young (i.e. a peer, <30 yr) or Old (≥30 yr). Overall shadowers are less accurate in predicting the old model talker's age, with nine (out of 32) shadowers misidentifying the old talker as Young, whereas only four (out of 31) shadowers misidentified the young talker as Old.

Lastly, we also examined how imitation is affected by the shadowers' impression of the model talker and their general attitudes towards young HKC speakers' accent. Exploratory factor analyses were conducted on the ratings for the 10 model talker personality traits (i.e. talker impression) and the six language attitude statements, separately. As a dimension reduction process, factor analysis was used to "[analyze] patterns of correlations to uncover empirically distinct latent constructs" (Sakaluk & Short, 2017, p. 1) and to generate values for the new latent variables for further statistical testing. The optimal number of factors was determined by parallel analysis and the scree test (Sakaluk & Short, 2017). The two factor analyses yielded three factors of talker impression (which we name as TALKER LIKABLE, TALKER STANDARD, and TALKER INTROVERTED) and two factors of language attitudes (which we name as YOUNG ACCENT NEGATIVE and YOUNG ACCENT POSITIVE). Tables 2 and 3 show the loading of each survey item for each factor, with greater values indicating larger portions of variance explained by the item. As shown in Table 2, the main contributors to the TALKER LIKABLE factor are traits typically associated with solidarity (e.g. "friendly", "likeable", "lively"), while the TALKER STANDARD factor is mainly informed by "standard-sounding" and "native-sounding", both centered around correctness. The TALKER INTROVERTED factor is a bit difficult to interpret, but the items with higher loadings ("introverted", "cool" and "authoritative") seem to suggest a somewhat distanced stance. Similarly, Table 3 shows that the YOUNG ACCENT NEGATIVE factor is mainly informed by negatively valenced attitudes toward young HKC accent (or negatively informed by positive-valenced attitudes), whereas the YOUNG ACCENT POSITIVE factor is informed by solidarity-related or positively valanced evaluations. For each factor, a score was calculated for each shadower based on their responses to the sociolinguistic evaluations. The ranges for the five factors are as follows: TALKER LIKABLE (−2.84 ~ 1.69),

Table 2
Factor analysis on ratings of talker traits, with loadings of each survey item for each factor.

Survey items	Factor 1 (TALKER LIKABLE)	Factor 2 (TALKER STANDARD)	Factor 3 (TALKER INTROVERTED)
The speaker sounds friendly	0.88		
The speaker sounds likable	0.95		
The speaker sounds lively	0.57		0.29
The speaker is native-sounding	0.35	0.64	
The speaker is standard-sounding		0.99	
The speaker sounds introverted			0.55
The speaker sounds educated	0.48	0.21	0.27
The speaker sounds authoritative	0.25	0.32	0.43
The speaker sounds cool	0.33		0.49
The speaker sounds unnatural	−0.43		

Table 3
Factor analysis on language attitudes, with loadings of each survey item for each factor.

Survey items	Factor 1 (YOUNG ACCENT NEGATIVE)	Factor 2 (YOUNG ACCENT POSITIVE)
The young generation speaks HKC with a lazy accent	0.67	0.23
The young generation's HKC accent needs improvement	0.99	
The young generation's HKC accent is likable		0.51
I like the young generation's HKC accent		1.00
The young generation's HKC pronunciation differs from that of their parents	0.29	
The young generation's HKC accent is standard-sounding	−0.49	

TALKER STANDARD ($-3.13 \sim 0.93$), TALKER INTROVERTED ($-2.78 \sim 3.22$), YOUNG ACCENT NEGATIVE ($-3.25 \sim 1.69$) and YOUNG ACCENT POSITIVE ($-3.41 \sim 2.34$). Given the abovementioned meanings of the five factors, we argue that greater values for the three talker factors indicate more *positive* talker impressions, and greater values for the YOUNG ACCENT POSITIVE factor or lesser values for the YOUNG ACCENT NEGATIVE factor indicate more *positive* attitudes towards the younger generation's HKC.

2.4.3. Statistical analysis

For the production data, linear mixed-effects models on mean F0 and K-S score were constructed for Young and Old Model Talker conditions separately. The full set of independent variables consists of critical predictors including Tone (T3, T6)[5], Block (Baseline, Shadow1, Shadow2, Post-shadow), shadower's Merger Status (Non-merging, Merging), and scores from sociolinguistic evaluations (TALKER LIKABLE, TALKER STANDARD, TALKER INTROVERTED, and YOUNG ACCENT POSITIVE, YOUNG ACCENT NEGATIVE), and control predictors such as shadower Gender (F, M), Perceived Talker Generation (Old, Young), and presence of Minimal Pair counterpart (Y, N). The models on mean F0 started with main effects of all the independent variables, a two-way interaction between Tone and Block, and three-way interactions between Tone, Block and each of the remaining critical predictors. Thus, an overall imitation effect will manifest as a Tone × Block interaction in the predicted direction, and effects of Merger Status or sociolinguistic variables on imitation should manifest as three-way interactions with Tone × Block. The initial models had a minimal random structure with only by-speaker and by-item intercepts. The models then went through a backward elimination procedure, which iteratively eliminated non-significant predictor terms (determined by the change in log-likelihood of the model) until all the remaining predictors were significant. Then, the resulting models, while retaining all the fixed-effects terms, underwent a similar process of selecting random predictors, starting with a maximal random structure (with by-speaker and by-item random intercepts and all possible slopes) and reducing the random structure until the models converged. Models of K-S score were constructed in a similar manner, except that Tone was not included as a fixed-effect predictor and Item was not included in the random structure. An overall imitation effect would manifest as a main effect of Block in the predicted direction, whereas effects of Merger Status and sociolinguistic variables would manifest as two-way interactions with Block.

For the perception data, given the ceiling accuracy rates (Young Model Talker: >99%; Old Model Talker: >95%), which is consistent with previous reporting, we did not model the accuracy rates. Mixed-effects models were built on reaction time only, for the two talker conditions separately, with a similar procedure as the models on mean F0.

All the models were built and tested with the lme4 (Bates, Baechler, Bolker, & Walker, 2015) and lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017) packages in R (R Core Team, 2018). In *Section 3*, we only report results from the final models, using a significance level of 0.05.

Table 4
Summary of fixed-effects terms in the model on mean F0 for the Young Model Talker condition; random effects = (1 + Tone + Block | Shadower) + (1 + Merger Status | Syllable).

	β	*t*	*p*
(Intercept)	1.641	6.110	<0.001
Tone = T6	−0.715	−11.467	<0.001
Block = Shadow 1	0.409	1.775	n.s.
Block = Shadow 2	0.288	1.598	n.s.
Block = Post-shadow	0.199	1.260	n.s.
Merger Status = Merging	0.860	3.904	<0.001
Perceived Talker Generation = Young	0.540	2.176	0.037
Tone = T6 : Block = Shadow1	0.011	0.217	n.s.
Tone = T6 : Block = Shadow2	−0.118	−2.358	0.018
Tone = T6 : Block = Post-shadow	0.079	1.924	n.s.
Tone = T6 : Merger Status = Merging	0.442	6.865	<0.001
Block = Shadow1 : Merger Status = Merging	−0.704	−2.195	0.036
Block = Shadow2 : Merger Status = Merging	−0.581	−2.315	0.027
Block = Post-shadow : Merger Status = Merging	−0.393	−1.789	n.s.
Tone = T6: Block = Shadow1 : Merger Status = Merging	−0.468	−6.894	<0.001
Tone = T6: Block = Shadow2 : Merger Status = Merging	−0.309	−4.557	<0.001
Tone = T6: Block = Post-shadow : Merger Status = Merging	−0.233	−4.102	<0.001

3. Results

3.1. Production

In the Young Model Talker condition, the model on mean F0 reveals a significant interaction of Tone × Block × Merger Status, indicating the influence of merger status on the degree of imitation (see Table 4 for model detail). Specifically, as shown in Fig. 4, merging shadowers exhibit clear patterns of hypothesized phonetic imitation: the T3–T6 difference increases significantly during and after shadowing (Shadow1: $M = 0.75$ T; Shadow2: $M = 0.72$ T; Post-shadow: $M = 0.43$ T) compared to the baseline production ($M = 0.28$ T; all at $p < 0.001$), with the greatest increase during the two shadowing blocks. On the other hand, the non-merging shadowers only enlarge the T3–T6 distinction in the second shadowing block ($M = 0.85$ T) compared to the baseline ($M = 0.71$ T; $\beta = -0.118$, $t = -2.358$, $p = 0.018$), and the imitation effect diminishes in the post-shadowing block (M = 0.63 T; p = n.s.). For confirmation, the merging and non-merging shadowers were modelled separately and the same imitation effects sustained. The model also shows an effect of perceived talker generation, with shadowers who believe the talker to be young producing overall higher F0 values, but the factor does not interact with Tone or Block, suggesting no effect on imitation.

Similar effects of Merger Status on imitation are shown in the model on K-S score (see Table 5). Merging shadowers show a significant increase in K-S score (i.e. greater separation between T3 and T6) during the shadowing blocks (Shadow1: $M = 0.89$; Shadow2: $M = 0.92$) compared to the baseline ($M = 0.62$; both ps < 0.001); a similar trend is observed in the post-shadowing block ($M = 0.7$) albeit non-significant. As for non-merging shadowers, no by-block differences are observed (Baseline: $M = 0.84$; Shadow1: $M = 0.9$; Shadow2: $M = 0.91$; Post-shadow: $M = 0.82$; all at p = n.s.). Additionally, the model shows significant interactions of Block × Gender and Block × YOUNG ACCENT NEGATIVE. Other things being equal, male shadowers exhibit greater degree of

[5] The first listed category in each categorical variable is set to be the reference level.

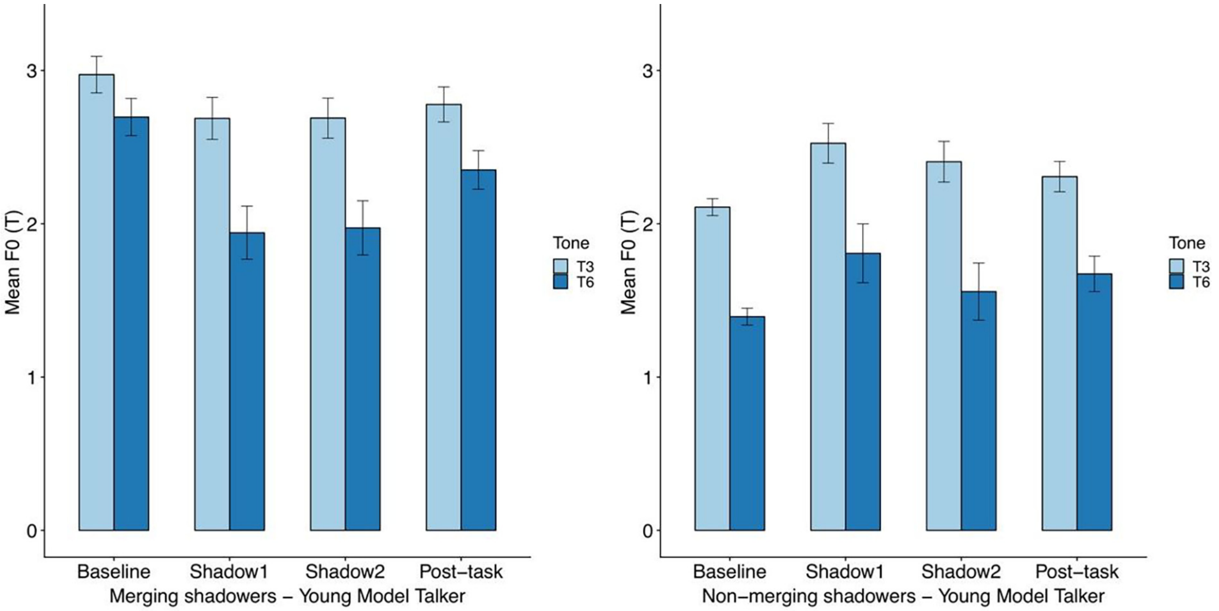

Fig. 4. Mean F0 by Tone and Block in Young Model Talker condition, produced by merging (left) and non-merging (right) shadowers. Error bars indicate 95% confidence intervals.

Table 5
Summary of fixed-effects terms in the model on K–S score for the Young Model Talker condition; random effects = (1 | Shadower).

	β	t	p
(Intercept)	0.859	22.226	<0.001
Block = Shadow1	0.036	0.832	n.s.
Block = Shadow2	0.011	0.250	n.s.
Block = Post-shadow	−0.036	−0.844	n.s.
Merger Status = Merging	−0.214	−4.424	<0.001
Gender = Male	−0.059	−1.152	n.s.
YOUNG ACCENT NEGATIVE	−0.033	−1.621	n.s.
Block = Shadow1 : Merger Status = Merging	0.189	3.532	<0.001
Block = Shadow2 : Merger Status = Merging	0.227	4.243	<0.001
Block = Post-shadow : Merger Status = Merging	0.096	1.797	n.s.
Block = Shadow1 : Gender = Male	0.091	1.599	n.s.
Block = Shadow2 : Gender = Male	0.170	2.976	0.004
Block = Post-shadow : Gender = Male	0.054	0.949	n.s.
Block = Shadow1 : YOUNG ACCENT NEGATIVE	0.086	3.879	<0.001
Block = Shadow2 : YOUNG ACCENT NEGATIVE	0.018	0.789	n.s.
Block = Post-shadow : YOUNG ACCENT NEGATIVE	0.010	0.444	n.s.

Table 6
Summary of fixed-effects terms in the model on mean F0 for the Old Model Talker condition; random effects = (1 + Tone + Block | Shadower) + (1 + Merger Status | Syllable).

	β	t	p
(Intercept)	2.249	10.788	<0.001
Merger Status = Merging	0.415	2.155	0.039
Tone = T6	−0.552	−9.287	<0.001
Gender = Male	−0.489	−2.617	0.013
Block = Shadow1	−0.237	−1.660	n.s.
Block = Shadow2	−0.228	−1.946	n.s.
Block = Post-shadow	0.204	1.170	n.s.
Merger Status = Merging : Tone = T6	0.201	4.805	<0.001
Tone = T6 : Gender = Male	−0.105	−2.644	0.013

imitation than female shadowers in Shadow2; those with a more negative attitude toward young generation's accent imitate more, but only in Shadow1.

In the Old Model Talker condition, the model on mean F0 reports no significant imitation effects (i.e. no Tone × Block interactions) or further effect of Merger Status on imitation (see Table 6). There is a main effect of Merger Status (p = 0.039) and a significant interaction of Merger Status and Tone (p < 0.001), indicating that merging shadowers overall produced higher F0 for T3 than non-merging shadowers and a closer distance between T3 and T6 throughout the experiment. Despite the lack of significance of Tone × Block × Merger Status, when the mean F0 values of merging and non-merging shadowers' T3 and T6 are plotted separately (see Fig. 5), we see a trend for merging shadowers to increase the T3–T6 distance during and after shadowing compared to the baseline (Baseline: M = 0.32 T; Shadow1: M = 0.49 T; Shadow2: M = 0.49 T; Post-shadow: M = 0.41 T; all at p = n.s.), while the T3–T6 distance in non-merging shadowers is stable across blocks (Baseline: M = 0.61 T; Shadow1: M = 0.57 T; Shadow2: M = 0.61 T; Post-shadow: M = 0.5 T; all at p = n.s.). The model also reveals effects regarding shadower gender, suggesting that male shadowers have lower F0 for T3 and even lower for T6.

The K–S model for the Old Model Talker condition shows a significant main effect of Block, providing evidence for imitation (see Table 7). Overall the distribution of T3 and T6 becomes more distinct during shadowing (Shadow 1: M = 0.78, Shadow 2: M = 0.81) and after shadowing (M = 0.76), compared to the baseline (M = 0.67); all at p < 0.05), but there is no further modulation from Merger Status. Although the model doesn't show any significant interaction of Block × Merger Status (all at p = n.s.), the mean K–S scores indicate a trend of increasing distinction in (post-)shadowing blocks (compared to baseline) for merging shadowers (Baseline: M = 0.64; Shadow1: M = 0.79; Shadow2: M = 0.78; Post-shadow: M = 0.75), but no similar trend for non-merging shadowers (Baseline: M = 0.75; Shadow1: M = 0.77; Shadow2: M = 0.9; Post-shadow: M = 0.77), except for maybe in the Shadow2 block.

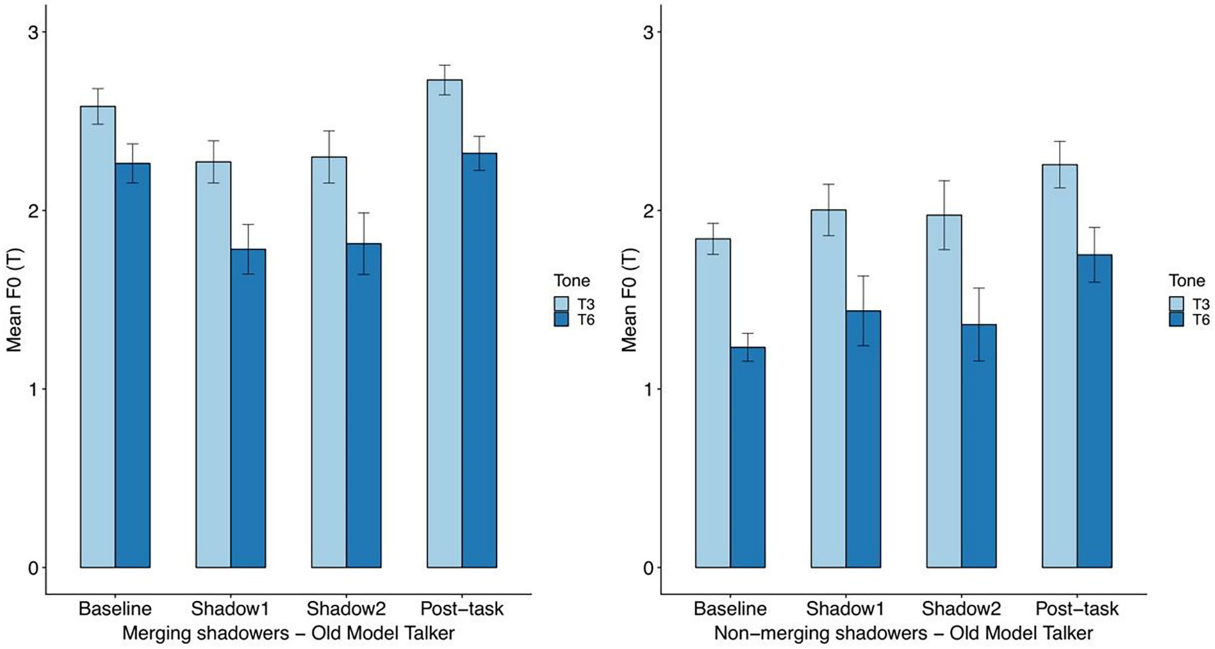

Fig. 5. Mean F0 by Tone and Block in Old Model Talker condition, produced by merging (left) and non-merging (right) shadowers. Error bars indicate confidence intervals.

Table 7
Summary of fixed-effects terms in the model on K–S score for the Old Model Talker condition; random effects = (1 | Shadower).

	β	*t*	*p*
(Intercept)	0.674	23.832	<0.001
Block = Shadow1	0.110	2.980	0.004
Block = Shadow2	0.139	3.775	<0.001
Block = Post-shadow	0.082	2.221	0.029
YOUNG ACCENT NEGATIVE	0.057	2.458	0.020

Table 8
Summary of fixed-effects terms in the model on the log-transformed reaction time for both talker conditions; random effects = (1 + Block | Shadower) + (1 + Block | Syllable).

	β	*t*	*p*
YOUNG MODEL TALKER CONDITION			
(Intercept)	2.965	221.933	<0.001
Block = Post-shadow	−0.018	−1.904	n.s.
YOUNG ACCENT NEGATIVE	0.025	2.562	0.016
Block = Post-shadow : YOUNG ACCENT NEGATIVE	−0.014	−2.116	0.043
OLD MODEL TALKER CONDITION			
(Intercept)	3.005	257.572	<0.001
Block = Post-shadow	−0.022	−2.465	0.021

In addition, the model shows a significant effect of YOUNG ACCENT NEGATIVE (β = 0.057; p = 0.020), indicating that those who have a more negative view of young generation's accent tend to show greater T3–T6 distinction overall.

3.2. Perception

As discussed in *Section 2*, AX discrimination accuracy is over 95% for both model talker conditions and thus not analyzed statistically. As for reaction time, as shown in Table 8, shadowers in both talker conditions became faster in the post-shadowing block (the pattern is trending for the Young Model Talker Condition, p = 0.07), suggesting a practice effect, with a further interaction between Block and YOUNG ACCENT NEGATIVE in the Young Model Talker Condition that indicates greater acceleration in shadowers with more negative views.

3.3. Post-hoc analysis

As reported in *Section 3.1*, the group of shadowers that exhibited the greatest imitation was the merging shadowers in the Young Model Talker group. We conducted two post-hoc analyses on this group to confirm whether the observed imitation is targeting the phonological contrast or merely the surface phonetics. The first analysis examines the tonal productions in both normalized and absolute F0 values. As shown in Fig. 6(a and b), the reversal of the T3–T6 merger—measured by normalized F0—is mainly driven by the lowering of T6, which in effect undoes the reported raising of T6 that caused the T3–T6 merger (Zhang et al., 2019). Importantly, the merger reversal is not likely to have resulted from simply imitating absolute F0. The strongest evidence for this comes from female shadowers, whose absolute F0 goes up in the shadowing and post-task blocks compared to the baseline (see Fig. 6(c)), probably as a result of hyperarticulation, moving further away from the male model talker's low absolute F0 (cf. Table 1). In other words, the imitating shadowers converged with the model talker in terms of enlarged distance between T3 and T6 in the normalized tonal space but not in terms of absolute F0. This argument receives further support from the fact that the increase in (normalized) T3–T6 distinction is also found in T3/T6 items that were **not** presented in the shadowing stage. To this end, we conducted the second analysis by refitting the model on mean F0 for the Young Model

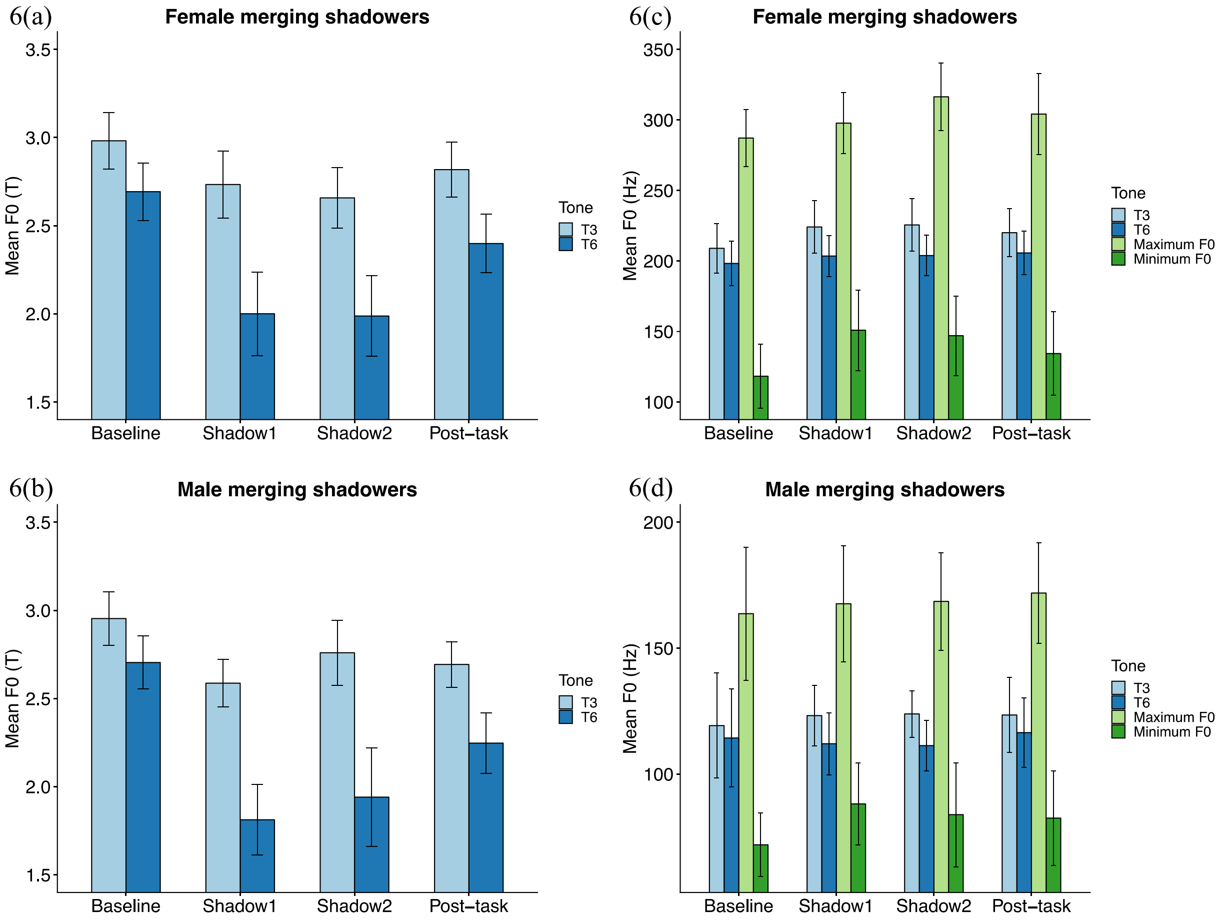

Fig. 6. Mean F0 in normalized ((a) and (b)) and absolute ((c) and (d)) F0 by Tone by Block produced by merging shadowers in the Young Model Talker condition (F = 11, M = 5). The plots of absolute F0 also show the group mean for maximum F0 and minimum F0 for reference. Error bars indicate 95% confidence intervals.

Talker condition with two modifications: (1) only production data from the baseline and post-shadowing blocks (15 shadowed and 12 unshadowed items in each block; see *Section 2.2* for detail and the Appendix for the full list) were included, and (2) a three-way interaction term among Tone, Block and Shadow Status (Y, N) was added. The newly added interaction produced no significant effect, while other model results remained unchanged, suggesting that degree of imitation is **not** modulated by whether the item has been shadowed. That is to say, both shadowed and unshadowed items demonstrate imitation effects to the same degree, consistent with previous findings of VOT imitation spreading to untrained items (Nielsen, 2011) and providing further evidence that the imitation occurs at the level of lexical tone contrast.

4. Discussion

4.1. Summary

In this study, we demonstrate that speakers can enlarge the distinction of two merging tones, T3 and T6 in HKC, by shadowing a model talker who clearly makes the distinction. It is worth noting that the current study uses high-frequency/familiarity words, which are not the most conducive to imitation, according to previous research (Goldinger, 1998). Our results show that more imitation happened in the Young Model Talker condition than in the Old Model Talker condition, and that the amount of imitation was modulated by the shadower's baseline merger status, with merging shadowers imitating more than non-merging shadowers. The enlarged T3–T6 distinction is mainly driven by the lowering of T6, which reverses the historical process of T6 raising. Social factors overall produced limited effects, except for YOUNG ACCENT NEGATIVE. Whether or not a word has a minimal pair counterpart had no effect, either. Post-hoc analyses further revealed that imitation did not result from mimicking absolute F0, and that both trained and untrained items showed the same degree of tonal adjustments. Both these findings suggest that imitation took place at the lexical tone level. Additionally, although more than half of the shadowers belonged to the merging group, all the shadowers achieved high accuracy in the AX discrimination, echoing previous findings of merging productions but distinct perception of T3 and T6 (Fung & Lee, 2019; Mok et al., 2013).

4.2. Selective imitation

How to account for the effects of model talker age and shadower's baseline merger status? We discuss the latter first. The observed effect of merger status is compatible with the prediction that a larger linguistic distance—either phonetically or phonologically—between the shadower and the model talker leads to greater imitation (Babel, 2012; Walker & Campbell-Kibler, 2015). The imitation of the phonological contrast may be aided by the incompleteness of the merger in perception, allowing even merging shadowers to be attentive to the perceptual contrast.

Now let's turn to the effect of model talker age. We originally formed two predictions, based on peer influence (predicting more imitation of the young talker) or age–accent coherence (predicting more imitation of the old talker). However, the two model talkers happened to also differ in the magnitude of the T3–T6 distinction in their production, with the young talker showing greater distinction than the old talker. Thus, the planned comparison of Young versus Old Model Talker conditions is confounded by the talkers' production idiosyncrasy. Indeed, we argue that production idiosyncrasy is a better explanation for the current findings. Compared to the old model talker's production, the young model talker's T3–T6 distinction was more pronounced and probably more perceptible, thus easier to imitate. Furthermore, having a more salient T3–T6 distinction made the young model talker's production more different from the shadowers', which could also facilitate imitation, given the finding of the positive link between linguistic distance and amount of imitation.

The hypothesized talker age effect, on the other hand, is largely undermined by the fact that the young shadowers were in general not good at recognizing the old talker's age (mean estimated age = 38.2 yr vs. actual age of 64 yr), probably due to their lack of experience interacting with senior speakers (the old talker's voice sounds normal for his age to the authors' ears). The shadowers' inaccuracy in age perception effectively cancelled out our manipulation of model talker age, which in turn invalidates the talker age-based account. Moreover, when we tested the effect of perceived talker generation in the statistical models, this predictor failed to produce any significant effect on imitation, indicating that accommodation was not affected by whether or not the shadower recognized the model talker as a peer.

4.3. Limited social effects

We only observed limited effects of social factors on imitation from the YOUNG ACCENT NEGATIVE factor, which encodes an overall agreement with the view that the young generation's accent is nonstandard and needs improvement. On the surface, these effects suggest that the strength of such agreement positively predicts the degree of imitation of the standard accent, but a closer look reveals that the effects are in fact driven by a small number of outlier shadowers (5 out of 63) who scored exceptionally low (>1.5 SD from the mean) on YOUNG ACCENT NEGATIVE. In other words, it was the shadowers who strongly *disagreed* with negative evaluations of the young accent that resisted the imitation, yielding the observed effects of YOUNG ACCENT NEGATIVE. When this subset of shadowers' data were removed from the models, the effects of YOUNG ACCENT NEGATIVE diminished.

We impute the general lack of social effects in the current study to the shadowers' complex and multidimensional attitudes toward the lazy accent. The shadowers overall acknowledge the societal consensus that "the young generation speaks with a lazy accent" (mean rating = 4.87 out of 6), but simultaneously ascribe to the views that "the young generation's HKC accent is likable" (mean rating = 4.08) and "the young generation's HKC accent needs improvement" (mean rating = 4.38). In other words, the "lazy" accent is characterized by both low status and high solidarity (Garrett, 2001; Giles & Billings, 2004).

While previous studies reported effects of shadowers' subjective impression of the model talker (Babel et al., 2014; Yu et al., 2013), our results fail to support such accounts. The discrepancy is probably due to the overall positive impression projected toward the model talkers in the current experiment (mean rating on "friendly" = 4.35, SD = 1.03; mean rating on "likable" = 4.25, SD = 1.03).

4.4. Theoretical implications

The current finding can be accounted for by the exemplar theory (Johnson, 1997; Pierrehumbert, 2002). A possible scenario in a merging shadower's system is as follows: the shadower starts with a production target of T6 that is close to the target of T3, but repeated exposure to the model speech adds new exemplars of T6, which have lower tone heights, to the cloud. Because the new exemplars are both recent and salient (i.e. standing out relative to the other exemplars, see Drager & Kirtley, 2016), they have high activation levels and thus weigh more in the calculation of production targets, causing future T6 productions to lower in F0 and effectively differentiating T3 and T6. The same new exemplars will be added to a non-merging shadower's T6 cloud during shadowing; however, the potential perturbation they can cause—assuming they also participate in the calculation of production target—will be smaller compared to the scenario with a merging shadower, because the original target of T6 in this situation is already quite low in F0. For the same reason, exposure to the young model talker's speech, which features greater T3–T6 distinction, may create a more pronounced shift in production than exposure to the old model talker's speech. Lastly, as discussed in *Section 4.3*, shadowers who have an exceptionally strong positive bias for the merged "lazy" accent may resist the trend to imitate by inhibiting the new exemplars.

An alternative route of imitation is via updating the abstract representation of tonal categories. While exemplar theories assume that phonetic detail is stored in exemplars without speaker normalization (Johnson, 1997), phonological categories can be abstracted in a bottom-up manner through statistical regularities (Pierrehumbert, 2002, 2016). Thus, with increasing exposure to the model talker, the shadower may update the distributional statistics of the normalized F0 of T6 with information from the new exemplars, which will in turn feed into the computation of the production target.

While the interactive alignment account (Gambi & Pickering, 2013; Pickering & Garrod, 2007, 2013) correctly predicts the possibility of shadowers approaching the model talker in tonal targets, its underlying mechanisms of alignment—following from the principles of the prediction-by-simulation route—predicts greater convergence in more similar (i.e. less distant) talker-listener pairs, which runs against the current finding of greater convergence between more dissimilar model talker-shadower pairs. One possible way to reconcile the theory with current results is to impute to the processing differences between non-merging and merging shadowers. Gambi and Pickering (2013) propose that when resources are limited, predictions and simulations of higher-level (e.g. phonological) information will be prioritized over predictions and simulations of lower-level (e.g. phonetic) information. Therefore, if we assume that non-merging shadowers attend **more** to the phonological contrast between T3 and T6 in the simulation process than merging shadowers did, they would have **less** resources for simulating phonetic detail. Nevertheless, this view should be subject to further scrutiny. For one thing, we already know that merging shadowers were likely processing the phonological tonal contrast as well, as their imitation was not simply at the phonetic level. Besides, it is unclear how this account would explain why more convergence was observed toward the young model talker than the old model talker. Another possible account for why non-merging shadowers did not imitate more is that they might be simultaneously subject to pressure from contrast preservation, as further lowering T6—when it is already clearly lower than T3—may risk eliminating the contrast with the low-falling T4 (Mok et al., 2013). However, since the current data do not include productions of T4, the validity of this account will need to be confirmed in future research.

5. Conclusion and future work

In this paper, we report a shadowing experiment with two merging tones in Cantonese. Our results demonstrate that a tonal merger-in-progress can be reversed—to a certain degree—via phonetic imitation, but the likelihood and extent of the imitated reversal depends on the shadower's baseline merger status and the magnitude of distinction in the model speech. The co-presence of mechanisms that promote and restrain imitation allows a sound change to resonate within the community in a gradual manner (Labov, 1994; Trudgill, 1986). In future research, the paradigm of the current study can be applied to several alternative settings, e.g. with older shadowers and with model talkers who have already merged T3 and T6 in production, in order to achieve a fuller understanding of the multifaceted role of phonetic imitation in the context of an ongoing sound change in the community.

Funding source

This research is supported by a research grant from the Hong Kong Polytechnic University (project ID: P0001916).

CRediT authorship contribution statement

Yuhan Lin: Methodology, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Yao Yao:** Conceptualization, Methodology, Software, Formal analysis, Writing - original draft, Writing - review & editing, Supervision, Project administration, Funding acquisition. **Jin Luo:** Conceptualization, Methodology.

Acknowledgements

We are grateful to David C. S. Li and Eugene Wong for helping with creating the stimuli, Chang Liu and Xuan Wang for assistance with the experiments, and the anonymous reviewers and the editor Bernd Möbius for their critical and constructive comments and suggestions on previous versions of the manuscript. We would also like to thank the audiences of the 21st International Conference on Yue Dialects and the 19th International Congress of Phonetic Sciences for helpful comments and suggestions.

Appendix A. Critical items used in the shadowing experiment

Jyutping is the romanization system for Cantonese, developed by the Linguistic Society of Hong Kong. The number at the end of each syllable represents its tone.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.wocn.2021.101060.

Character	Jyutping	IPA	Gloss	Baseline & Post-shadowing	Shadowing	AX discrimination
報	bou3	/pou/	‘report’	–	–	+
部	bou6	/pou/	‘part’	–	–	+
帝	dai3	/tɐi/	‘emperor’	–	–	+
第	dai6	/tɐi/	‘sequence’	–	–	+
到	dou3	/tou/	‘arrive’	+	+	+
導	dou6	/tou/	‘direct’	+	+	+
凍	dung3	/tʊŋ/	‘freeze’	–	–	+
動	dung6	/tʊŋ/	‘move’	–	–	+
富	fu3	/fu/	‘rich’	–	–	+
負	fu6	/fu/	‘load’	–	–	+
記	gei3	/kei/	‘record’	+	–	+
技	gei6	/kei/	‘skill’	+	–	+
據	geoi3	/køy/	‘occupy’	–	–	+
具	geoi6	/køy/	‘tool’	–	–	+
貴	gwai3	/kwɐi/	‘expensive’	–	–	+
櫃	gwai6	/kwɐi/	‘cupboard’	–	–	+
意	ji3	/ji/	‘idea’	–	–	+
義	ji6	/ji/	‘righteousness’	–	–	+
最	zeoi3	/tʃøy/	‘most’	–	–	+
罪	zeoi6	/tʃøy/	‘crime’	–	–	+
至	zi3	/tʃi/	‘reach’	+	+	+
自	zi6	/tʃi/	‘self’	+	+	+
政	zing3	/tʃɪŋ/	‘government’	–	–	+
靜	zing6	/tʃɪŋ/	‘quiet’	–	–	+
代	doi6	/tɔi/	‘replace’	+	+	–
快	faai3	/fai/	‘rapid’	+	+	–
價	gaa3	/ka/	‘price’	+	–	–
故	gu3	/ku/	‘ancient’	+	+	–
限	haan6	/han/	‘boundary’	+	+	–
氣	hei3	/hei/	‘air’	+	+	–
抗	kong3	/kʰɔŋ/	‘resist’	+	+	–
另	ling6	/lɪŋ/	‘another’	+	–	–
未	mei6	/mei/	‘not yet’	+	–	–
面	min6	/min/	‘face’	+	+	–
務	mou6	/mou/	‘affairs’	+	–	–
派	paai3	/pʰai/	‘branch’	+	–	–
破	po3	/pʰɔ/	‘break’	+	+	–
配	pui3	/pʰui/	‘match’	+	–	–
素	sou3	/sou/	‘plain’	+	–	–
送	sung3	/sʊŋ/	‘give’	+	+	–
太	taai3	/tʰai/	‘too’	+	–	–
話	waa6	/wa/	‘speech’	+	+	–
壞	waai6	/wai/	‘spoiled’	+	–	–
換	wun6	/wun/	‘change’	+	+	–
問	man6	/mɐn/	‘ask’	+	–	–

References

Aubanel, V., & Nguyen, N. (2020). Speaking to a common tune: Between-speaker convergence in voice fundamental frequency in a joint speech production task. *PLoS ONE, 15*(5), 1–16. https://doi.org/10.1371/journal.pone.0232209.

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society, 39*(4), 437–456. https://doi.org/10.1017/s0047404510000400.

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics, 40*(1), 177–189. https://doi.org/10.1016/j.wocn.2011.09.001.

Babel, M., & Bulatov, D. (2012). The role of fundamental frequency in phonetic accommodation. *Language and Speech, 55*(2), 231–248. https://doi.org/10.1177/0023830911417695.

Babel, M., McAuliffe, M., & Haber, G. (2013). Can mergers-in-progress be unmerged in speech accommodation?. *Frontiers in Psychology, 4*(SEP), 1–14. https://doi.org/10.3389/fpsyg.2013.00653.

Babel, M., McGuire, G., Walters, S., & Nicholls, A. (2014). Novelty and social preference in phonetic accommodation. *Laboratory Phonology, 5*(1), 123–150. https://doi.org/10.1515/lp-2014-0006.

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

Bauer, R. S. (1986). The microhistory of a sound change in progress in Hong Kong Cantonese. *Journal of Chinese Linguistics, 14*(1), 1–42.

Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese phonology*. Berlin: Mouton de Gruyter.

Bauer, R. S., Cheung, K.-H., & Cheung, P.-M. (2003). Variation and merger of the rising tones in Hong Kong Cantonese. *Language Variation and Change, 15*(2), 211–225. https://doi.org/10.1017/S0954394503152039.

Boersma, P., & Weenink, D. (2019). Praat: Doing phonetics by computer [Computer program] Retrieved from. *Version, 6*(1), 05 http://www.praat.org/.

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language, 82*(4), 711–733.

Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics, 40*(2), 249–268. https://doi.org/10.1016/j.wocn.2011.10.007.

Chao, Y. R. (1930). A system of "tone-letters". *Le Maître Phonétique, 45*, 24–27.

Coles-Harris, E. H. (2017). Perspectives on the motivations for phonetic convergence. *Language and Linguistics Compass, 11*(12), 1–21. https://doi.org/10.1111/lnc3.12268.

D'Imperio, M., Cavone, R., & Petrone, C. (2014). Phonetic and phonological imitation of intonation in two varieties of Italian. *Frontiers in Psychology, 5*(OCT), 1–10. https://doi.org/10.3389/fpsyg.2014.01226.

Dodd, B., Holm, A., Hua, Z., & Crosbie, S. (2003). Phonological development: A normative study of British English-speaking children. *Clinical Linguistics and Phonetics, 17*(8), 617–643. https://doi.org/10.1080/0269920031000111348.

Drager, K. (2010). Speaker age and vowel perception. *Language and Speech, 54*(1), 99–121. https://doi.org/10.1177/0023830910388017.

Drager, K., & Kirtley, 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: Cambridge University Press.

Eckert, P. (2003). Language and adolescent peer groups. *Journal of Language and Social Psychology, 22*(1), 112–118. https://doi.org/10.1177/0261927X02250063.

Egbert, J., & Laflair, G. T. (2018). Statistics for categorical, nonparametric, and distribution-free data. In A. Phakiti, P. De Costa, L. Plonsky, & S. Starfield (Eds.), The Palgrave handbook of applied linguistics research methodology (pp. 523–539). Palgrave MacMillan. https://doi.org/10.1057/978-1-137-59900-1

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–272). Baltimore: York.

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America, 97*(5), 3125–3134. https://doi.org/10.1121/1.413041.

Flemming, E. (1996). Evidence for constraints on contrast: The dispersion theory of contrast. In C.-S.K. Hus (Ed.), UCLA working papers in phonology 1 (pp. 86–106).

Flemming, E. (2004). Contrast and perceptual distinctiveness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *The phonetic bases of phonological markedness* (pp. 232–276). Cambridge: Cambridge University Press.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics, 14*, 3–28.

Fowler, C. A., Brown, J. M., Sabadini, L., & Weihing, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language, 49*, 396–413. https://doi.org/10.1016/S0749-596X(03)00072-X.

Fung, R. S. Y., & Lee, C. K. C. (2019). Tone mergers in Hong Kong Cantonese: An asymmetry of production and perception. *The Journal the Acoustical Society of America, 146*(5), EL424. https://doi.org/10.1121/1.5133661.

Fung, R. S. Y., & Wong, C. S. P. (2011). Acoustic analysis of the new rising tone in Hong Kong Cantonese. In Proceedings of ICPhS XVII (pp. 715–718). Hong Kong. Aug 17-21, 2011.

Gambi, C., & Pickering, M. J. (2013). Prediction and imitation in speech. *Frontiers in Psychology, 4*(June), 1–9. https://doi.org/10.3389/fpsyg.2013.00340.

Garde, P. (1961). Réflexions sur les différences phonétiques entre les langues slaves. *Word, 17*, 34–62.

Garnier, M., Lamalle, L., & Sato, M. (2013). Neural correlates of phonetic convergence and speech imitation. *Frontiers in Psychology, 4*, 1–15. https://doi.org/10.3389/fpsyg.2013.00600.

Garrett, P. (2001). Language attitudes and sociolinguistics. *Journal of Sociolinguistics, 5*, 626–631.

German, J. S. (2012). Dialect adaptation and two dimensions of tune. In Proceedings of the 6th International Conference on Speech Prosody, SP 2012 (pp. 430–433). Shanghai, China.

Giles, H. (1973). Accent mobility: A model and some data. *Anthropological Linguistics, 15*, 87–105.

Giles, H., & Billings, A. C. (2004). Assessing language attitudes: Speaker evaluation studies. In D. Alan & E. Catherine (Eds.), *The handbook of applied linguistics* (pp. 187–209). Malden, MA: Blackwell Publishing.

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Commuincation, context, and consequence. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of accommodation: Developments in applied sociolinguistics* (pp. 1–68). Cambridge: Cambridge University Press.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*(2), 251–279.

Hall-Lew, L. (2010). Improved representation of variance in measures of vowel merger. In *Proceedings of meetings on acoustics* (pp. 1–10). Baltimore, Maryland: Cambridge University Press. https://doi.org/10.1121/1.3385271.

Harrington, J., Palethorpe, S., & Watson, C. (2000). Monophthongal vowel changes in Received Pronunciation: An acoustic analysis of the Queen's Christmas broadcasts. *Journal of the International Phonetic Association, 30*(1–2), 63–78. https://doi.org/10.1017/S0025100300006666.

Ho, W. H. (2001). *Yueyin zi xue tigang [An outline for the self-study of Cantonese Chinese pronunciation]*. Hong Kong: Hong Kong Education Publishing Company.

Johnson, K. (1997). Speech perception without speaker normalization. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145–166). San Diego, CA: Academic Press.

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. https://doi.org/10.1080/23273798.2019.1582787.

Kim, J. (2016). Prosodic accommodation in Seoul Korean accentual phrases. In Proceedings of the 2016 international conference on speech prosody (pp. 395–399). Boston, MA. https://doi.org/10.21437/speechprosody.2016-81.

Kim, M., Horton, W. S., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology, 2*(1), 125–156. https://doi.org/10.1515/labphon.2011.004.

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America, 100*, 2425–2438. https://doi.org/10.1007/978-1-4419-1698-3_1426.

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

Labov, W. (1994). *Principles of linguistic change: Volume I: Internal factors*. Oxford: Wiley-Blackwell Publishing.

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

Labov, W., Karen, M., & Miller, C. (1991). Near-mergers and the suspension of phonemic contrast. *Language Variation and Change, 3*, 33–74.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory ofspeech perception revised. *Cognition, 21*, 1–36.

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language, 48*(4), 839–862.

Mantell, J. T., & Pfordresher, P. Q. (2013). Vocal imitation of song and speech. *Cognition, 127*(2), 177–202. https://doi.org/10.1016/j.cognition.2012.12.008.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314–324.

Menn, L. (1983). Development of articulatory, phonetic, and phonological capabilities. In B. Butterworth (Ed.), *Language production* (pp. 3–50). London: Academic Press.

Michelas, A., & Nguyen, N. (2011). Uncovering the effect of imitation on tonal patterns of French accentual phrases. In Proceedings of the 12th Annual Conference of the International Speech Communication Association (Interspeech 2011) (pp. 973–976). Florence, Italy.

Mitterer, H., & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition, 109*(1), 168–173. https://doi.org/10.1016/j.cognition.2008.08.002.

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. https://doi.org/10.1017/S0954394513000161.

Namy, L. L., Nygaard, L. C., & Sauerteig, D. (2002). Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology, 21*(4), 422–432. https://doi.org/10.1177/026192702237958.

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics, 39*(2), 132–142. https://doi.org/10.1016/j.wocn.2010.12.007.

Nycz, J. R. (2011). *Second dialect acquisition: Implications for theories of phonological representation* PhD thesis. New York University.

Nye, P. W., & Fowler, C. A. (2003). Shadowing latency and imitation: The effect of familiarity with the phonetic patterning of English. *Journal of Phonetics, 31*, 63–79.

Ohala, J. J. (1993). The phonetics of sound change. In C. Jones (Ed.), *Historical linguistics: Problems and perspectives* (1st Ed., pp. 237–278). Routledge.

Olmstead, A. J., Viswanathan, N., Aivar, M. P., & Manuel, S. (2013). Comparison of native and non-native phone imitation by English and Spanish speakers. *Frontiers in Psychology, 4*(July), 1–7. https://doi.org/10.3389/fpsyg.2013.00475.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America, 119*(4), 2382–2393. https://doi.org/10.1121/1.2178720.

Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. (2012). Phonetic convergence in college roommates. *Journal of Phonetics, 40*(1), 190–197. https://doi.org/10.1016/j.wocn.2011.10.001.

Pardo, J. S., Jordan, K., Mallari, R., Scanlon, C., & Lewandowski, E. (2013). Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures. *Journal of Memory and Language, 69*(3), 183–195. https://doi.org/10.1016/j.jml.2013.06.002.

Pardo, J. S., Urmanche, A., Wilman, S., & Wiener, J. (2017). Phonetic convergence across multiple measures and model talkers. *Attention, Perception, and Psychophysics, 79*(2), 637–659. https://doi.org/10.3758/s13414-016-1226-0.

Peng, G. (2006). Temporal and tonal aspects of Chinese syllables: A corpus-based comparative study of Mandarin and Cantonese. *Journal of Chinese Linguistics, 34*(1), 134–154.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences, 11*(3), 105–110. https://doi.org/10.1016/j.tics.2006.12.002.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *The Behavioral and Brain Sciences, 36*(4), 329–347. https://doi.org/10.1017/S0140525X12001495.

Pierrehumbert, J. B. (2001). Exemplar dynamics. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 137). Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.45.08pie.

Pierrehumbert, J. B. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology VII* (pp. 101–140). Berlin, New York: Mouton de Gruyter.

Pierrehumbert, J. B. (2016). Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics, 2*(1), 33–52. https://doi.org/10.1146/annurev-linguist-030514-125050.

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics, 29*(2), 191–215.

Postma-Nilsenovà, M., & Postma, E. (2013). Auditory perception bias in speech imitation. *Frontiers in Psychology, 4*(NOV), 1–8. https://doi.org/10.3389/fpsyg.2013.00826.

R Core Team (2018). *R: A language and environment for statistical computing*. Vienna: Austria. Retrieved from http://www.r-project.org/.

Research Centre for Humanities Computing Chinese University of Hong Kong (2003). Chinese character database: With word-formations Retrieved October 1, 2018, from http://humanum.arts.cuhk.edu.hk/Lexis/lexi-can/.

Sakaluk, J. K., & Short, S. D. (2017). A methodological review of exploratory factor analysis in sexuality research: Used practices, best practices, and data analysis resources. *Journal of Sex Research, 54*(1), 1–9. https://doi.org/10.1080/00224499.2015.1137538.

Sanchez, K., Miller, R. M., & Rosenblum, L. D. (2010). Visual influences on alignment. *Journal of Speech, Language, and Hearing Research, 53*, 262–273.

Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics, 25*, 421–436.

Sankoff, G. (2019). Language change across the lifespan: Three trajectory types. *Language, 95*(2), 1–36.

Sato, M., Grabski, K., Garnier, M., Granjon, L., Schwartz, J. L., & Nguyen, N. (2013). Converging toward a common speech code: Imitative and perceptuo-motor recalibration processes in speech production. *Frontiers in Psychology, 4*(JUL), 1–14. https://doi.org/10.3389/fpsyg.2013.00422.

Shi, F., & Wang, P. (2006). Beijinghua danziyin shengdiao de tongji fenxi [A statistical analysis of the tones in Beijing Mandarin]. *Dangdai Yuyanxue [Modern Linguistics], 1*, 33–40.

Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics, 66*(3), 422–429. https://doi.org/10.3758/BF03194890.

Siegel, J. (2010). *Second dialect acquisition*. Cambridge University Press.

Smit, A., Hand, L., Freilinger, J., Bernthal, J., & Bird, A. (1990). The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders, 55*, 779–798.

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 and Phonetics, 29*(5), 333–353. https://doi.org/10.3109/02699206.2014.1003329.

Trudgill, P. (1981). Linguistic accommodation: Sociolinguistic observations on a sociopsychological theory. In R. Hendrick, C. Masek, & M. F. Miller (Eds.), *Papers from the Parasession on language and behavior* (pp. 218–337). Chicago, IL: Chicago Linguistics Society.

Trudgill, P. (1986). *Dialects in contact*. Oxford: Blackwell Publishing.

Walker, A., & Campbell-Kibler, K. (2015). Repeat what after whom? Exploring variable selectivity in a cross-dialectal shadowing task. *Frontiers in Psychology, 6*(May), 1–18. https://doi.org/10.3389/fpsyg.2015.00546.

Walker, A., & Hay, J. (2011). Congruence between 'word age' and 'voice age' facilitates lexical access. *Laboratory Phonology, 2*(1), 219–237. https://doi.org/10.1515/labphon.2011.007.

Wedel, A., Kaplan, A., & Jackson, S. (2013). High functional load inhibits phonological contrast loss: A corpus study. *Cognition, 128*(2), 179–186. https://doi.org/10.1016/j.cognition.2013.03.002.

Wisniewski, M. G., Mantell, J. T., & Pfordresher, P. Q. (2013). Transfer effects in the vocal imitation of speech and song. *Psychomusicology: Music, Mind, and Brain, 23*(2), 82–99. https://doi.org/10.1037/a0033299.

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. https://doi.org/10.1017/CBO9781107415324.004.

Yu, A. C. L. (2007). Understanding near mergers: The case of morphological tone in Cantonese. *Phonology, 24*, 187. https://doi.org/10.1017/S0952675707001157.

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). https://doi.org/10.1371/journal.pone.0074746.

Zahn, C. J., & Hopper, R. (1985). Measuring language attitudes: The speech evaluation instrument. *Journal of Language and Social Psychology, 4*(2), 113–123. https://doi.org/10.1177/0261927X8500400203.

Zee, E. (1998). Chinese (Hong Kong Cantonese). In *Illustrations of the IPA* (pp. 58–60).

Zhang, J., Zhang, Y., & Xu, D. (2019). A variationsit approach to tone categorization in Cantonese. *Chinese Language and Discourse, 10*(1), 1–16.