



Predicting tonal realizations in one Chinese dialect from another

Junru Wu^{a,b,*}, Yiya Chen^b, Vincent J. van Heuven^{b,c}, Niels O. Schiller^b

^a Dept. Chinese Language and Literature, East China Normal University, 500 Dongchuan Rd., Shanghai 200241, China

^b Leiden University Centre for Linguistics, Leiden Institute for Brain and Cognition, The Netherlands

^c Dept. Applied Linguistics, University of Pannonia, Egyetem utca 10, Veszprém, Hungary

Received 16 February 2015; received in revised form 20 October 2015; accepted 29 October 2015

Available online 5 November 2015

Abstract

Pronunciation dictionaries are usually expensive and time-consuming to prepare for the computational modeling of human languages, especially when the target language is under-resourced. Northern Chinese dialects are often under-resourced but used by a significant number of speakers. They share the basic sound inventories with Standard Chinese (SC). Also, their words usually share the segmental realizations and logographic written forms with the SC translation equivalents. Hence the pronunciation dictionaries of northern Chinese dialects could be easily available if we were able to predict the tonal realizations of the dialect words from the tonal information of their SC counterparts. This paper applies statistical modeling to investigate the tonal aspect of the related words between a northern dialect, i.e. Jinan Mandarin (JM), and Standard Chinese (SC). Multi-linear regression models were built with between-word pitch distance of JM words as the dependent variable and the following were included as the predictors: SC tonal relations, between-dialect tonal identity, and individual backgrounds. The results showed that tonal relations in SC and between-dialect identity, as predictors featuring the relation between the JM and SC tonal systems, are significant and robust predictors of JM tonal realizations. The speakers' sociolinguistic and cognitive backgrounds, together with the tonal merge and neutral tone information within JM, are important for the prediction of JM tonal realizations and affect the way that between-language predictors take effect.

© 2015 Elsevier B.V. All rights reserved.

Keywords: Tone; Translation equivalents; Cognates; Modeling; Individual backgrounds

1. Introduction

1.1. The necessity and sufficiency of modeling under-resourced northern Chinese dialects

Under-resourced languages, featured by the “lack of a unique writing system or stable orthography, limited presence on the web, lack of linguistic expertise, and lack of electronic resources for speech and language processing”

(Besacier et al., 2014: 27), have always been a challenge for both engineers of Human Language Technologies (HLT) and linguists. One of the main reasons behind this challenge is the large amount of phonetic data required, which can be both difficult and expensive to acquire. To tackle this challenge, more and more researchers are transferring information from a related language or dialect to improve the understanding and automatic machine-processing of the under-resourced language. For instance, the automatic speech recognition of Afrikaans was significantly improved using the available Dutch data (Imseng et al., 2014). However, to better incorporate the information from the related language, we need a better understanding of the relations between the two languages or dialects. In this aspect, linguists have carried out studies

* Corresponding author at: Dept. Chinese Language and Literature, East China Normal University, 500 Dongchuan Rd., Shanghai 200241, China. Tel.: +86 (0)2154344874.

E-mail address: jrwu@zhwx.ecnu.edu.cn (J. Wu).

of a wide-range of languages, though linguistic knowledge sometimes needs adaptations to be applied in engineering.

Chinese appears to be anything but under-resourced. For instance, Mandarin Chinese and Shanghai Chinese are already covered by the standardized multilingual text and speech database “GlobalPhone” (Schultz, 2002). Even the (Standard) Mandarin-English bilingual test-to-speech system has seen important breakthroughs (Qian and Soong, 2012). However, compared with the relatively well-investigated Standard Chinese (also referred to as “Mandarin Chinese”, “Standard Mandarin”, or “putonghua”, abbreviated as “SC” in this article), many Chinese dialects are still under-resourced, including most northern dialects.¹ These northern dialects need more attention. First, they are used by a large Chinese population in everyday life (Hamed, 2005; Li, 1988). Second, they are closely related to SC and are often used together with SC. This type of bilingualism comes with frequent code-switching/-mixing and sometimes also results in accented SC speech, which presents challenges for engineers and linguists (Huang et al., 2000; Sproat et al., 2004).

On the other hand, the close relation between the northern dialects and SC is also an attractive resource for the modeling of these dialects. Besides the large overlap in syntactic structure, the northern dialects and SC are very similar in basic sound inventories. For instance, we can find the comparison of the basic sound inventories of major Chinese dialects in a dictionary designed by linguists (Collective_work, 1989). This type of similarity has been proved useful in the sound-to-phoneme modeling in other languages (Imseng et al., 2014; Kamper et al., 2012; Van Heerden et al., 2010). However, there is one additional aspect of the between-dialect relation that may be useful and needs some more exploration. The northern dialects and SC share a high percentage of cognates and frequently borrow from each other² (Norman, 2003). The resulting translation equivalents share the same meaning across dialects and sound similar to each other. These related words are easy to identify because they are written in the same characters across all these dialects using the same logographic writing system. This paper applies statistical modeling to explore the tonal aspects of the related words between a northern dialect and SC. As a preliminary but important step before predicting the dialect pronunciation directly from SC pronunciation, the current study investigated to what extent and in what way a very limited but well available SC resource, the SC tonal categories, can predict the dialectal tonal realizations. We also tried to find out how the SC tonal categories, together with the speaker’s social and cognitive backgrounds can account for the speaker-dependent tonal variability.

1.2. Research background on Jinan Mandarin (JM)

We aim at predicting between-word pitch distances for JM Chinese using the tonal relations of the SC counterparts of the target words. JM is a northern dialect of Chinese. It is used in some local TV shows, but mostly in traditional folk arts, such as in “Shandong Kuaishu”. Most JM speakers also speak SC fluently, and the mutual intelligibility between JM and SC is high (Tang and van Heuven, 2009). Some linguistic descriptions are available for JM. “Jinan Fangyan Cidian” (JM Dialect Dictionary) (Qian, 1997) provides the largest vocabulary but no recording. “Jinanhua Yindang” (The Sound System of JM Dialect) (Qian and Zhu, 1998) provides recordings of 428 monosyllabic characters, 410 words with two or more syllables, and some sentences. Pronunciations of characters are also available in “Hanyu Fangyin Zihui” (Collective_work, 1989). However, these studies are based on the pronunciations by senior speakers many years ago (above 65 years old in 1993, 1998, and 1979).

Our fieldwork in 2012 showed that JM has become more similar to SC and the differences are mainly only retained in the tonal system. First, the usage and knowledge of JM-specific words are largely reduced and JM-specific words are replaced by words with etymologically related SC counterparts. Second, most JM words are now almost identical to their SC counterparts in segmental structure. However, the tonal differences remain between the JM and SC translation equivalents.

As a result, the current JM dialect shares a high percentage of related words with SC, which are almost only different from their SC counterparts in their tonal realizations (pitch contours). Since most non-tonal resources can already be directly transferred from SC, tone is the main potential space for cost reduction when building the pronunciation dictionary. The building cost of a JM pronunciation dictionary could be reduced if we are able to predict the tonal realizations of the JM words from the tonal information of their SC counterparts.

However, many JM words have shown tolerance of different tonal patterns, possibly due to the on-going process of “lexical diffusion”, where new tonal variants have appeared on some words but not on other words originally from the same tonal category (Chen and Wang, 1975; Wang, 1969), and the generalization of JM “neutral tone sandhi” (Qian, 1997), which means some words which were not reported to carry neutral tones are starting to have variants with neutral tone sandhi. As a result, some JM words allow one single tonal pattern (mono-pattern) but the others allow more than one (dual-pattern/multi-pattern). Fig. 1(a) and (b) demonstrate the difference between mono-pattern (i.e. “very”, fei1chang2, /feitsʰaŋ/) and dual-pattern (i.e. “simple”, j1andan, /tɕiantan/). These words were plotted with normalized F0 contours from multiple speakers. Different tonal patterns of the same word can be observed not only in the production of different speakers but also in the production of the same speaker.

¹ The term “northern dialects” is sometimes distinguished from “Mandarin dialects”, which are even more similar to Standard Chinese (Hamed, 2005). Here we use it in a more general way, following Li (1988).

² However, the cognates and loan words are difficult to distinguish for closely related dialects.

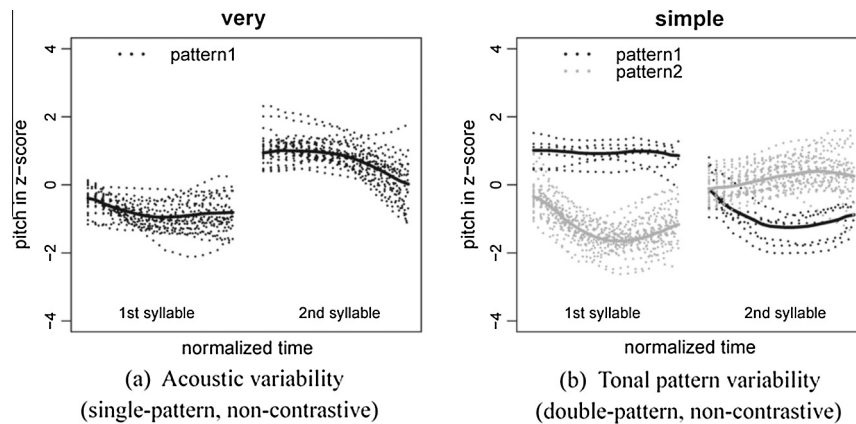


Fig. 1. Examples of mono-pattern (a) and dual-pattern (b) JM words, adapted from (Wu et al., 2014).

Table 1

Percentage of characters, which follow systematic correspondence in JM and Beijing Mandarin (BM, which the phonological system of SC is based on).

| BM | Tone 1 high-level | Tone 2 high-rising | Tone 3 low-rising or dipping | Tone 4 high-falling | Total |
|----|-------------------|--------------------|------------------------------|---------------------|-------|
| JM | 81% (low-)rising | 76% high-falling | 70% high-level | 75% low-falling | 76% |

This adds difficulty to the modeling of JM tonal realization.

1.3. Systematic correspondence and phonological similarity

In the present study, we try to implement two mechanisms which linguists have long been aware of in the modeling. The first mechanism is called “systematic correspondence” (Dyen, 1963; Meillet and Ford, 1967). For two related dialects, “there is a significant number of words with similar meanings whose phonemes correspond systematically” (Dyen, 1963: 634). Systematic correspondence can be measured via cross-dialectal comparison (Chen, 1973). Since the phonology of SC is based on the Beijing Mandarin (BM) pronunciation, we calculated the correspondence of the 4 JM tones and the 4 BM tones from the phonological transcriptions of 2722 monosyllabic Chinese characters collected in 1979 (Collective work, 1989), with polyphones counted multiple times,³ resulting in 3679 pairs. Table 1 shows the percentage of pairs, which follow the systematic correspondence rule. For instance, within the items carrying Tone 1 (high-level) in BM, 81% carry the low-rising tone in JM. The strength of systematic correspondence is likely to be even higher in current JM.⁴

The second mechanism is “phonological similarity”. Phonological similarity across dialects or languages comes from two sources. (1) Cognates are inherently similar to

each other. For instance, Dutch “donder” sounds like its English cognate “thunder” because they were derived from the same proto-Germanic word *thunraz (Harper, 2001). (2) Loan words and their sources are similar to each other too. For instance, Dutch “computer” sounds identical to its English translation equivalent “computer” because Dutch borrowed this word from English. Cognates and loanwords are difficult to differentiate when two dialects are involved, because most of the time the borrowed form is also a cognate. Compared with the differentiation based on word origin, the degree of between-language similarity is a more practical standard. In the relation between JM and SC, since the JM-SC related words are almost always with the same segmental structure, the most significant differences are in tone. If we keep minor pitch variation on the acoustic level out of consideration, the related words are either identical or different in their tonal realizations. For instance, the disyllabic Chinese word “thanks” (xie4xie5) /çiɛçiɛ/ carries the falling + (low) neutral tones in SC and can have an almost identical tonal realization in JM, while “very” (fei1chang2) /feitʂʰaŋ/ carries high-level + rising tones in SC but has a totally different low + high-falling tone in JM. In the present study we distinguish only whether the JM word sounds identical to its SC counterpart. The between-dialect identity is taken as another main predictor in the current study.

Although both mechanisms are potentially useful for predicting the tonal realization of a JM word from its SC counterpart, the tonal phonological similarity can disrupt the effect of tonal systematic correspondence. For instance, as shown in Fig. 1(b), the disyllabic word “need” in JM can follow the systematic correspondence rule with SC; then from the known tonal category of its SC counterpart (Tone 1 + Tone 4) we can predict that the JM “need” (xu1yao4) /çyiau/ carries low-rising + low-falling tones,

³ For instance, according to the corpus the character for the Chinese quantifier “ge” allows both Tone1 and Tone2 in Beijing and both low-rising and low-falling tone in Jinan. The resulting pairs would be the Cartesian product of the BM and JM sets, with an amount of four.

⁴ However, how to handle polyphones and individual variability in choosing tonal patterns needs to be considered more carefully in implementation.

different from the high-level + falling tones in SC. On the other hand, the JM word “need”, can also carry tones almost identical to the tonal realization of its SC counterpart (with high-level + falling tone); in this case it disrupts the systematic correspondence rule between JM and SC. How the two mechanisms interact with each other in detail remains an open question.

Although the effects of systematic correspondence and phonological similarity have received full attention from linguists since the time of “Grimm’s Law” (early 19th century), more efforts are still needed to bridge the gap between the linguistic theories and the technical applications. In the present research, we try to incorporate both mechanisms in the same statistical model and investigate their potential applications in predicting JM tonal realization using SC data.

1.4. Disyllabic tonal combinations and sandhi

We target disyllabic words in the present study. The majority of modern Chinese words are disyllabic, taking up 56% in the Microsoft Chinese dictionary (Wu and Jiang, 2000). Although in some Chinese dialects trisyllabic tonal realizations cannot be directly predicted from the corresponding disyllabic tonal sandhi patterns, e.g. in Tianjin Mandarin (Li and Chen, 2016), disyllabic foot is the most frequent, the least constrained, and the standard foot in Chinese speech prosody (Feng, 2001; Li, 2002). This means the results from disyllabic words can be applied on multisyllabic words because they are usually realized with combinations of disyllabic and monosyllabic feet.

In SC, tonal realizations of disyllabic words are largely predictable from the citation tones of their monosyllabic components, either via tonal co-articulation or via sandhi rules (Xu, 1994). Tone sandhi means the morpheme in combination carries a tonal variant different from the variant it carries in isolation. The Tone 3 sandhi in SC is well-known. Tone 3 before another Tone 3 sounds like Tone 2 but retains its minor acoustic difference from Tone 2 (Peng, 2000; Yuan and Chen, 2014). Tone 2 and the allophonic variants of Tone 3 have also been shown to be processed differently during speech preparation (Chen et al., 2011). In speech production and visual lexical access both the sandhi form of Tone 3 can activate and be activated by both Tone 2, which overlaps with its pitch contour, and Tone 3, which overlaps with its phonemic representation (Nixon et al., 2014). We have marked tonal categories of SC words according to the tonal citation forms and include whether the SC tonal categories of a pair of words are the same as one of the predictors.

It is important to note that the same combination of monosyllabic morphemes in JM can yield different disyllabic tonal realizations and the resulting variation is not totally predictable from the known predictors. According to Qian’s (1997) description, JM has two types of sandhi. One type, the so-called “normal” tonal sandhi maintains the distinctions across monosyllabic citation tones, and

the tonal realizations of the disyllabic word need to be predicted from the tones of the citation forms of both syllables (Qian, 1997). The second type concerns the neutral tone. The neutral tonal sandhi merges the tonal realizations of different tones on the second syllable of a disyllabic word, so that the pitch contour of the disyllabic word can be predicted from (but is not necessarily identical to) the citation tone of the first syllable. For instance, “hen” (mu3ji1) /mutɕi/ (with high-level + low-rising tone) and “morning” (zao3shang4) /tsauʂaŋ/ (with high-level + high-falling tone) are different in the second citation tone, but both can be realized with a “low + high-level” tonal contour following the neutral tone sandhi rule. The JM neutral tone has different variants depending on the different citation tones of the previous syllables. For instance, the JM neutral tone can be realized as a low-falling tone following the low-rising citation tone, as a high-level tone following the high-falling or the high-level citation tone, and as a high-falling tone following the low-falling citation tone. Moreover, different from SC, the tonal realization of the syllable before the neutral tone is also different from its citation form. For instance, before the neutral tone, the rising citation tone is realized as a falling tone, the high-falling citation tone is realized as a rising tone, the high-level citation tone is realized as a low tone, and the low-falling citation tone is realized as a high-level tone (Qian, 1997).

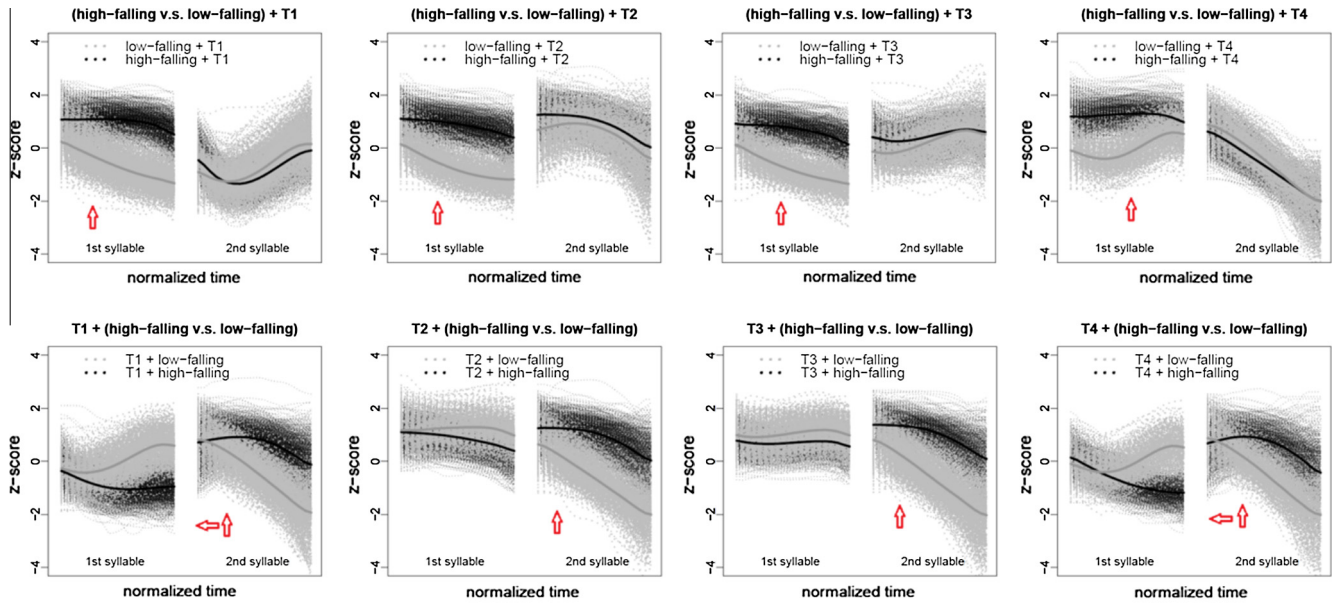
Most JM words carrying neutral tone sandhi have counterparts in SC, which also carry neutral tone. However, in our corpus, we have observed some JM words which allow tonal patterns following both neutral tonal sandhi and “normal” tonal sandhi rules. For instance, the two patterns of “simple” (jian3 dan4) /tɕiantan/ in Fig. 1(b) follow the two types of sandhi rules, respectively. Considering this phenomenon, whether the JM word in the specific rendition carries a neutral tonal sandhi or “normal” sandhi is included as a predictor in the present study.

Note that the JM neutral tone sandhi with “high-level” citation tone on the first syllable usually results in sandhi forms almost identical to its SC counterpart. For instance, JM “snack” (dian3xin5) /tʂienɕin/ (with high-level + neutral → low + high-level) sounds almost identical to “snack” (dian3xin1) /tʂienɕin/ (with dip + high-level → low + high-level) in SC. Thus some JM words identical to their SC counterparts result from JM neutral tone sandhi, not necessarily from borrowing.

1.5. Potential merging of JM tonal categories

According to earlier descriptions, the high-falling and low-falling tones are similar in JM. We analyzed the monosyllabic words recorded in 1998 (Qian and Zhu, 1998) and found that the distributions of pitch contours of the two tones were very similar but still distinguishable in monosyllabic words. In our more recent corpus, the two tones are clearly distinguishable in disyllabic words, as shown in Fig. 2.

Moreover, we found that some middle-aged speakers weaken the falling part of JM high-falling tone. As a result,



All 42 speakers

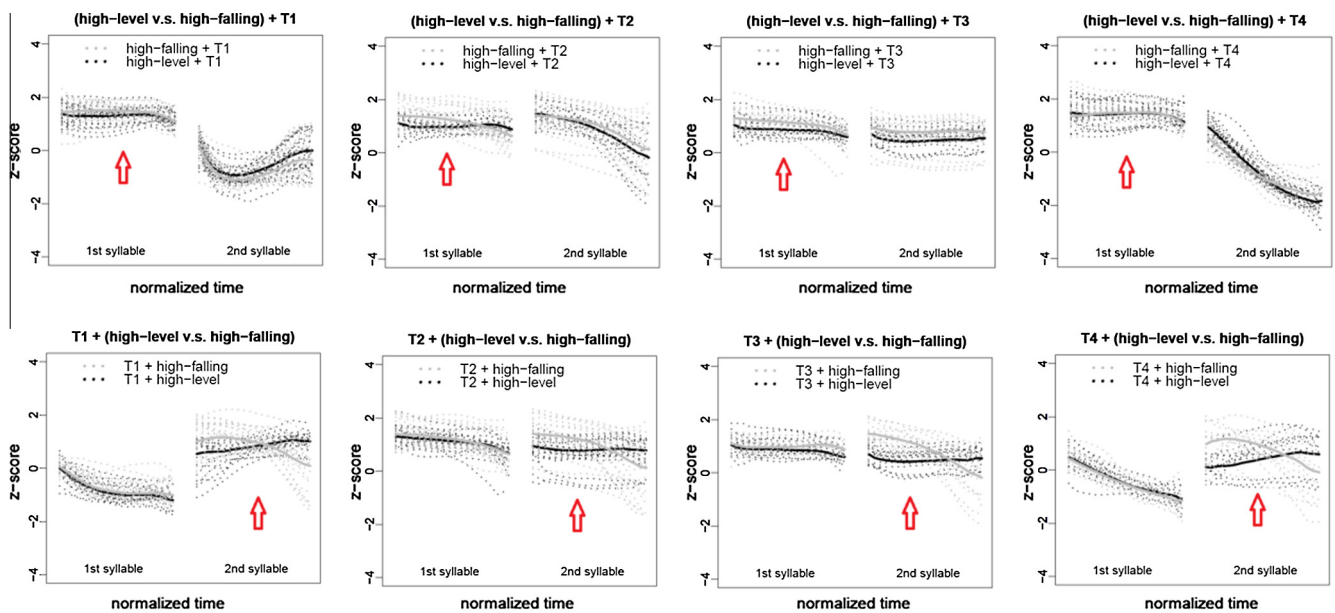
Fig. 2. The comparison of the JM high-falling and low-falling tones in the first syllable (first row) and the second syllable (second row) in disyllabic words.

the realization of the JM high-falling tone is instead more similar to the JM high-level tone and maintains its contrast with the JM low-falling tone. This change holds for both monosyllabic and disyllabic words. As shown in Fig. 3, the high-falling and high-level tones are very similar and the merge is more salient in non-final position.

As shown in Fig. 4, the difference between the JM low-rising and low-falling tones is also largely reduced when they appear in non-final position, except when the following tone is a low-rising one.

Due to the sandhi rules, the disyllabic combination of JM high-level + neutral tone is realized as low + high, very similar to the tonal realizations of JM low-rising + high-level, low-falling + high-level, low-rising + high-falling and low-falling + high-falling. The disyllabic combination of JM low-falling + neutral tone is realized as high-level + falling, very similar to the tonal realization of JM high-level + low-falling. These are depicted in Fig. 5.

The above-mentioned types of potential merging are marked and taken into consideration in our modeling.



Speaker 18

Fig. 3. This speaker (Speaker 18) merges JM high-falling and high-level tones in the first syllable (first row) and almost merges them in the second syllable (second row) in disyllabic words.

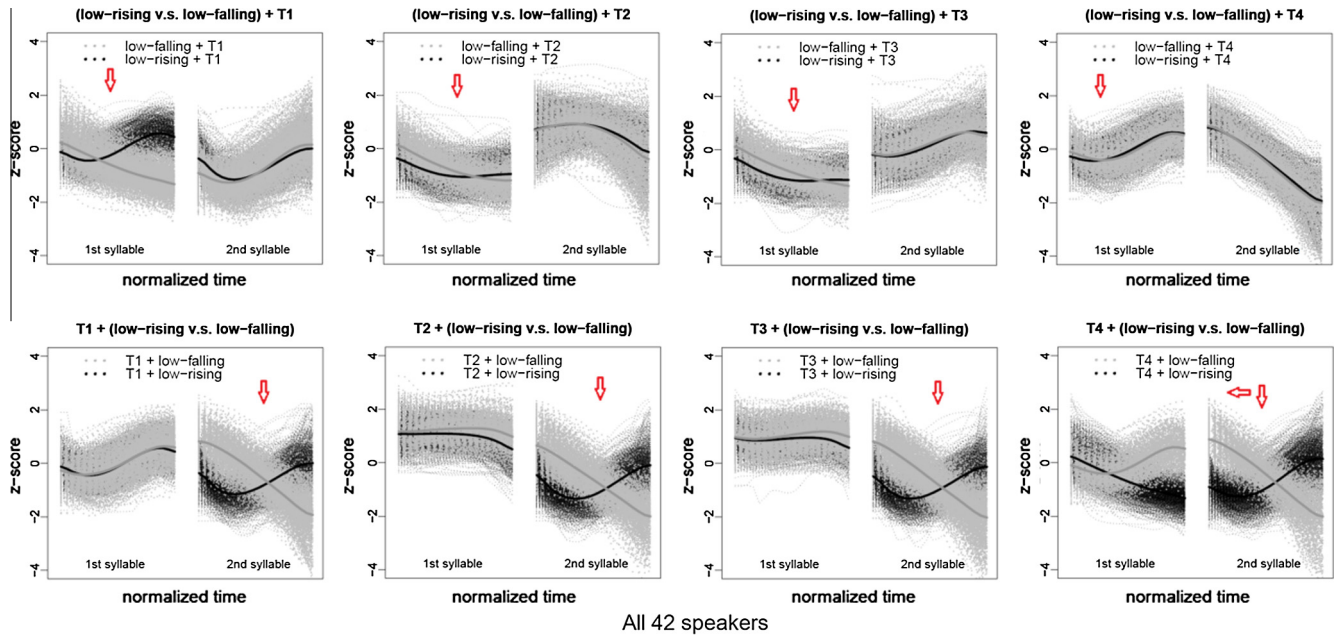


Fig. 4. The difference between JM low-rising and low-falling tones is largely reduced in the first syllable (first row) but the difference is maintained in the second syllable.

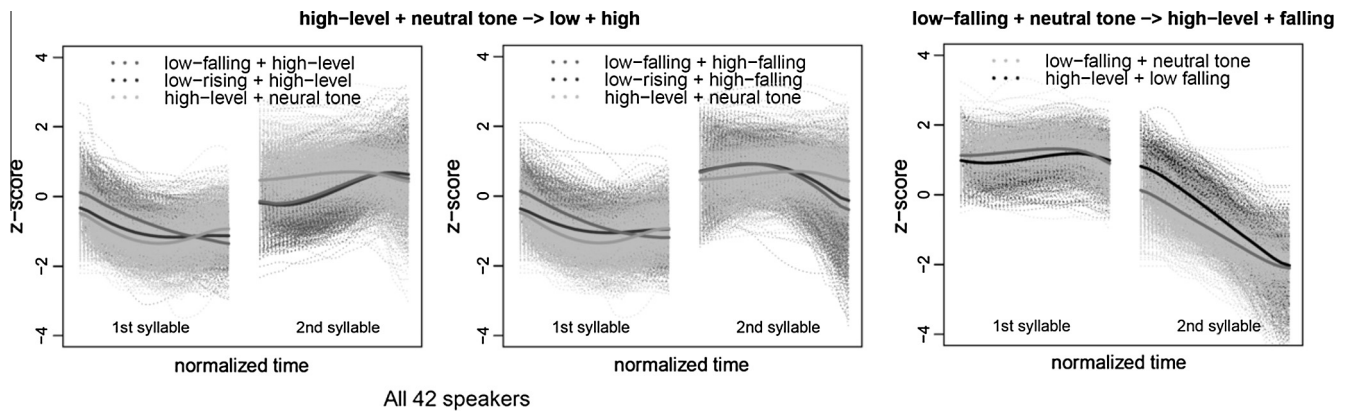


Fig. 5. JM high-level + neutral tone is realized as low + high, very similar to the tonal realizations of JM low-rising/low-falling + high-level and low-rising/low-falling + high-falling.

1.6. Word frequency

The frequency effect has long been known and discussed by psycholinguists in research on lexical access (Dell, 1990; Grainger, 1990; Levelt, 1999; Oldfield and Wingfield, 1965). The general finding is that frequent words and forms are accessed more quickly.

We would like to know whether word frequency affects the tonal relation across JM words and whether word frequency modulates the tonal effects of the systematic correspondence between SC and JM. Moreover, we would like to see whether including word frequency information will improve the modeling of JM tonal realizations.

Limited by the sharing of logographic writing system in Chinese, we have no access to any dialect-specific word frequency data. We use the Chinese word frequency based on film subtitles (Cai and Brysbaert, 2010). This resource

can be taken as mainly in SC, but we cannot exclude the contribution of JM speakers.

1.7. Individual backgrounds

Individual variation is interesting for both linguists and technology experts. System developers working on speaker adaptation have been rather successful in dealing with pure acoustic deviations via speech normalization and changing Hidden-Markov-Model (HMM) parameters (Leggetter and Woodland, 1995; Woodland, 2001). However, when separate models are built for different speaker types and/or when the pronunciation dictionary also needs to be adapted for different accents, the cost increases and the speaker type is difficult to decide (Huang et al., 2000; Woodland, 2001). On the other hand, sociolinguists have proven that these phonological variations can largely be

predicted from socio-backgrounds (Labov, 2006; Weinreich et al., 1968). Moreover, socio-backgrounds can not only group the speakers but can also index the speakers along a continuum. It would be beneficial if measurable individual backgrounds are introduced into the model.

Earlier studies reported segmental variation across and within JM individuals, regarding age and speech style (Cao, 1991; Qian, 1997; Qian and Zhu, 1998). They distinguished “old” JM from “new” JM and “Wendu” (literal style) from “Baidu” (colloquial style). They also did some quantitative analyses. However, these studies did not pay attention to the relation between tonal variation and individual backgrounds and after 20 years the JM words’ segmental structures are mostly identical to those of their SC counterparts.

We intend to incorporate individual backgrounds into the predicting model and investigate their statistical effects on JM tonal realization. Our corpus was collected in 2012. It covers a greater age range of urban native JM speakers and includes individual backgrounds on both social and cognitive aspects. We expect to quantify the age effect observed earlier (Qian, 1997) and we are also interested in which of the other aspects take effects.

Beside the previously investigated factors, such as gender, age, and education backgrounds, we also take the speakers’ experience with both languages into consideration. Previous studies on bilingualism have shown that language of education and language proficiency affect the pattern of code-switching in bilingual speech production (Carter et al., 2011). Whether these factors influence the systematic correspondence in general needs further investigation. Language proficiency is influenced by language exposure. Hence frequencies of language usage were taken into consideration. Beside sociolinguistic backgrounds, the cognitive aspects may also affect individual variation of speech production. For instance, tonal awareness, as a subset of phonological awareness, reveals listeners’ aptitudes for discriminating and identifying tones (Chen, 2004; Shu et al., 2008). It also affects the processing of tonal variants in JM lexical access (Wu and Chen, 2014). Additionally, the speaker’s digit-naming speed and auditory working memory are also taken into consideration. Few studies have shown the relevance of auditory working memory in speech production. Even bilinguals seem to be similar to monolinguals in auditory working memory (Bialystok et al., 2008) and simultaneous interpreters seem to have no advantage in retaining auditory information (Signorelli et al., 2011). However, would auditory working memory influence the systematic correspondence shown in the bilinguals’ production? All these factors are taken into consideration, together with their interaction with age.

It is reasonable to assume that the effects of cognitive and socio-linguistic backgrounds are at least partly mediated by age in the present study. Age is easy to measure. However, age is related to both the aging of the individuals and the change of the society. On the one hand, cognitive aging affects the speakers’ cognitive performances. Older

speakers have declined auditory working memories and slower reaction times. On the other hand, socio-linguistic backgrounds change across generations and affect the change of the tonal system. With the promotion of SC and social progress in China, younger speakers use more SC and less JM, receive higher education, and are more likely to receive their literacy educations in SC. The change of socio-linguistic backgrounds can also influence some cognitive backgrounds. With the introduction of the alphabetic system “pinyin”, the younger generations received more training in benefit of their acuity to tones.

Nevertheless, it is also reasonable to hypothesize that the cognitive and socio-linguistic backgrounds could affect JM tonal realization beyond the effect of age. Individuals of the same age have different cognitive aptitudes and individuals from the same generation have different socio-linguistic experiences. Which cognitive and socio-linguistic backgrounds have unmediated effects on the JM tonal system?

1.8. Research predictions

The present study focuses on the influences of the following factors on JM tonal realization: systematic correspondence, phonological similarity, and individual backgrounds. Other covariates are also taken into consideration, including the neutral tone sandhi and potential tonal merging in JM and word frequency.

As mentioned above, the systematic correspondence is related to the between-word tonal relations in both dialects. The between-word tonal relation, whether measured on a scale or dichotomously, is comparable across different tonal categories. Thus, answers about the between-word tonal relation can be technically applied before identifying the specific tonal categories. Also, the answers to the theoretical questions are category-independent and more general. We choose between-word pitch distance as the dependent variable to quantify the effects and interactions of the above-mentioned factors. Based on the linguistic knowledge of systematic correspondence (described in Section 1.3), we expect them to affect the between-word pitch distance in JM in the following way.

The systematic correspondence mechanism predicts that, if two words share their tonal categories in one dialect, their counterparts are also more likely to share tonal categories in the other dialect. Sharing tonal categories means smaller pitch distance. Thus, considering the pitch distance between two JM words, the distance is more likely to be smaller if their counterparts share tonal categories in SC. JM disyllabic words whose SC counterparts share the tonal categories on both syllables should show smaller pitch distances compared to JM disyllabic words whose counterparts only share the tonal category on just one syllable.

The effect of systematic correspondence should be robust if neither of the two JM words is realized identically to its SC counterpart. However, the effect of systematic correspondence should be disrupted when one of the JM words borrows its tonal realization from SC, especially when the

two words share tonal categories in SC. For instance, “direct” (zhi2jie1) /tʃɿtɕiɛ/ and “leave” (li2kai1) /likʰai/ share the same “Tone 2 + Tone 1” (high-rising + high-level) tones in SC. Systematic correspondence predicts that they are very likely to share the same “high-level + low-rising” tones in JM and see a relatively small between-word pitch distance. However, if the speaker borrows the SC form of “leave” (li2kai1) /likʰai/ (with high-rising + high-level) into JM and keeps the JM native high-level + low-rising tone for “direct” (zhi2jie1) /tʃɿtɕiɛ/, the between-word pitch distance should be larger than expected.

On the other hand, the predicting power of the two words’ tonal relation in SC should be rebuilt when both of the JM words borrow their tonal realizations from SC. Moreover, the predicting power should be stronger, because the effect is no longer mediated by systematic correspondence, which does not control all of the JM vocabulary (see Table 1). In the present case, the pitch distance between two JM words directly reflects the pitch distance of their SC counterparts. For instance, if both JM “leave” (li2kai1) /likʰai/ and “direct” (zhi2jie1) /tʃɿtɕiɛ/ borrow the SC form, it is sure that they share the high-level + low-rising tones just like in SC and have a small between-word pitch distance.

Note that, when a JM word is realized identically to its SC counterpart, it is difficult to decide from the surface realization whether it is due to borrowing or the coincidence of JM neutral tone sandhi (see Section 1.5). Nevertheless, whether the between-dialect identity is due to borrowing or the specific JM neutral tone sandhi, it should work similarly in most cases.

Also, the JM neutral tone sandhi itself should work in a similar way as the between-dialect identity. For instance, “body” (shen1ti3) /ʃəntʰi/ and “clear” (qing1chu3) /tɕʰiŋtɕʰu/ share the same high-level + low-rising (“Tone1 + Tone3”) tones in SC. Systematic correspondence predicts that they are very likely to share the same low + high-level tones in JM and see a relatively small pitch distance. However, “clear” is usually produced with neutral tone sandhi in JM and realizes with a low-falling + low pitch contour. As a result, the between-word pitch distance should be larger than expected.

We are also interested in the effects of individual backgrounds. The present study does not focus on the pure physical gender differences which can be normalized, such as the gender effect on the pitch range (Chen, 2011; Peng et al., 2012). The present study works on the variants which simple normalization cannot handle, namely the JM tonal variants related to the speakers’ social backgrounds.

Except gender, most of the other aspects of individual backgrounds are related to age. Older speakers are more proficient in JM, use JM more frequently, and mostly received literacy education in JM. It is natural to predict that their JM pronunciation should be less aligned with SC. However, we know that JM is related to SC via both systematic correspondence and phonological similarity. Does it mean that older JM speakers’ between-word pitch

distance should be less sensitive to the words’ tonal relation in SC? Or does it mean that older JM speakers produce JM words acoustically less similar to SC? Or are both true? On the other hand, older speakers usually have poor auditory working memories due to aging, and lower tonal awareness because they did not receive proper education of pinyin (the Chinese alphabetic writing system). Do these cognitive predictors have independent effects besides age? We will apply statistical analyses to answer these questions.

2. Material and methods

2.1. Corpus preparation

The speech data used in the present study were collected from 42 JM native speakers in 2012 (see Section 2.2 for details). Each speaker read 400 disyllabic Chinese words in JM. The written words were selected from a corpus of Chinese film subtitles (Cai and Brysbaert, 2010). One list of 200 high-frequency words was selected from the 10% disyllabic Chinese words with the highest word frequency. In a similar way, we selected the other list of 200 low-frequency words. In each list, there are 10 words for each of the 20 disyllabic tonal combinations. The high and low frequency lists were presented to the speakers in two blocks with a self-paced rest break in between. The words in each list were presented in a different random order for each speaker. After the speakers finished producing a word, they pressed a key to see the next word.

We used Praat (Boersma and Weenink, 2001) to extract pitch contours. Only pitch contours on the rhymes were extracted. A trained phonetician listened to each recording, looked at the spectrogram, and manually marked the rhyme of each syllable. Also, in this process, recordings with speech and recording errors were excluded from the corpus. Afterwards, the pitch contours were converted from hertz to semitones with 100 Hz as the base and then transformed into *z*-scores based on the speakers’ means and standard deviations (Chen, 2011; Lobanov, 1971). This normalization removed the pitch range difference across speakers, which is not the main focus of the present study. The normalized pitch contours were then interpolated to 20 points per-syllable to remove the difference in duration. Since each speaker produced a list with many different tones, the multidimensional distribution of the dataset involves inherent clusters. Thus, we chose a density-based local approach to eliminate possible outliers (Breunig et al., 2000). We calculated Local Outlier Factors (LOF) for each speaker’s pitch contours. Any pitch contour with an LOF greater than 1.5 (Breunig et al., 2000) and belonging to the 2.5% with the highest integral density was eliminated from the corpus.

2.2. Individual backgrounds

We collected both sociolinguistic and cognitive backgrounds from the speakers. The sociolinguistic backgrounds

included the speaker's age, gender, education level, reported proficiencies and frequencies of JM and SC, language of literacy education, and the dialects they use with their primary social relations. All speakers except one received formal education, of which 57% reached college level and the rest reached middle school level. As for the literacy education, 26% of the speakers received it in JM, 56% received it in SC, and 18% received it in a combination of JM and SC. The cognitive backgrounds were found by earlier studies to be related to vocabulary, reading, and comprehensive skills, including the speaker's digital naming speed (Torgesen and Davis, 1996) in JM and SC, auditory working memory in JM (Gathercole et al., 1994), and tonal awareness of JM and SC (Shu et al., 2008). The distributions of the scale variables are plotted in Fig. 6.

2.3. Model fitting

A “between-word pitch distance” was used as the crucial dependent variable in the modeling. This was practically defined as the Euclidean distance (Deza and Deza, 2009) of the pitch contours between each pair of JM words produced by the same speaker. For each speaker, each normalized pitch contour was taken as a Euclidean vector. For each pitch contour, each of the 20 time points was taken as one dimension of the vector. Then Euclidean distances were calculated for each two vectors of the same speaker, yielding 53,628–79,800 between-word pitch distances for each speaker. Similar acoustic distance matrices have been used in studies investigating the correlation between speech perception and production (Iverson et al., 2003).

The following two sets of predictors were included in the modeling in line with the research predictions. The first set includes linguistic predictors, within which pairs of words are nested. Table 2 shows the structures and explanations of these predictors. The SC tonal relations (on the first and second syllables) were calculated from the standard phonological transcriptions (in pinyin) of SC words. A phonetician with Putonghua Proficiency Test Certificates – Level 1B judged whether the JM word was produced (almost) identically to its SC counterpart, taking native monolinguals' and bilinguals' similarity rating of a subset of the corpus in another study into consideration (Wu et al., in prep.). Whether the two JM words are undergoing tonal merging and on which syllable(s) they are merging were predicted from the tonal categories of their SC counterparts first and then manually verified. Whether the JM word carries a neutral tone was predicted from the tonal category of its SC counterpart first and then manually verified. Word frequency was imported from the recording list (Cai and Brysbaert, 2010) and converted into two categories. The second set includes predictors based on individual backgrounds, within which speakers are nested and they were collected together with the recordings. Table 3 shows the structures and explanations of these predictors.

We performed exploratory linear-mixed-effects (LME) analyses on the between-word pitch distance data, using

R (R_Core_Team, 2013), lme4 (Bates et al., 2013), and lmerTest (Kuznetsova et al., 2013). Considering the size of the dataset, building all the predictors into one model would exceed the limitation of computing power. Moreover, a model with too many predictors would suffer from multi-collinearity across predictors and yield uninterpretable results. Thus, we did not build a model including all the predictors. Instead we built smaller models using different subsets of the data and subsets of the predictors to investigate the importance and robustness of different predictors. Then we built the important predictors and their interactions together into a more general model.

We fitted two sets of LME models. The first set of LME analyses focused on the effects of the SC tonal relation and the other linguistic predictors. We built one separate model for each speaker. Each speaker-wise model included all the fixed effects of the six nominal linguistic predictors in Table 2 (tonal relation on the first SC syllables, tonal relation on the second SC syllable, between-dialect identity, merge, neutral tone, word frequency), and their two-way and three-way interactions, as well as the random intercept of the SC tonal combinations of the two words. These models were trimmed and the results of the final models are reported here.

The second set of LME analyses focused on exploring the effects of individual backgrounds (see Table 3) and their interactions with the SC tonal relations predictor (a combination of the tonal relations on the first SC syllables and second syllables). To avoid unnecessary rank-deficiency and accommodate for the limit of computing power, these analyses were performed on the averaged between-word pitch distance, which was collapsed across pairs and aggregated by the combination of speaker, SC tonal relation, and between-dialect identity. A separate model was built for each level of between-dialect identity because the speaker-wise models showed that SC tonal relations function differently with different levels of between-dialect identity. We did not build all the predictors of individual backgrounds at once into one model because of two main reasons: first, the collapsed data could not support so many predictors; second, the multi-collinearity between these predictors would blur the interpretation of mediated effects and result in a model with unclear causality and unreliable direction of main effects. Instead, we first included these predictors separately in smaller models to investigate their independent effects and then, after statistically removing the collinearity, built a selected subset of the predictors together into a full model to investigate their interactions and non-mediated effects.

In the separate analyses of individual backgrounds, each model included SC tonal relations, one aspect of the individual backgrounds (interval predictors centralized by subtracting the mean), and their interaction as the fixed predictors, as well as Speaker as the random intercept. However, in the combined analysis of individual backgrounds, in order to investigate the non-mediated effects of the individual backgrounds, all the interval predictors

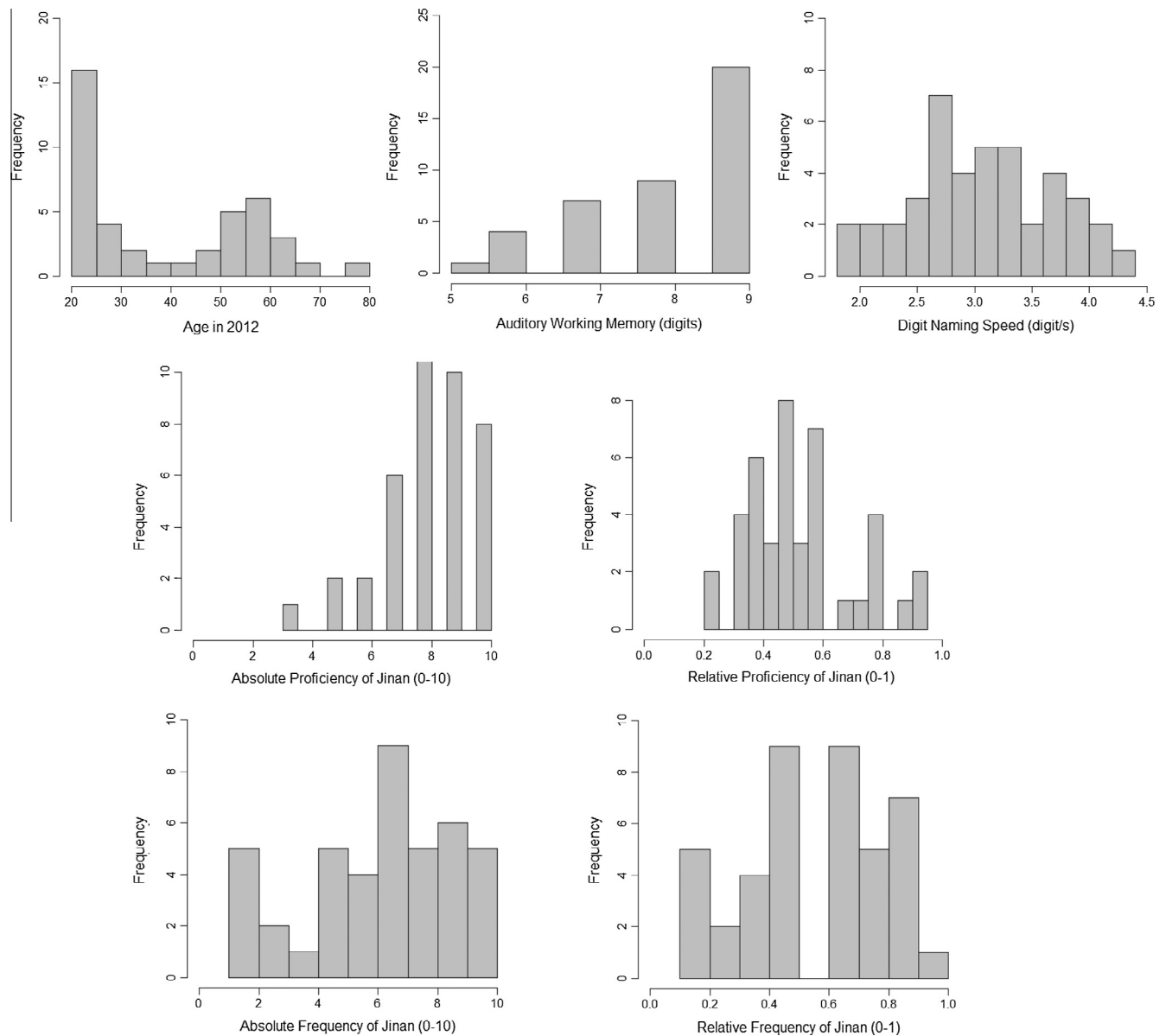


Fig. 6. The distribution of individual backgrounds: age (top row left), auditory working memory (top row middle), digital naming speed (top row right), absolute proficiency of JM (middle row left), relative proficiency of JM (middle row right), absolute frequency of JM (bottom row left), and relative frequency of JM (bottom row right).

Table 2

Linguistic predictors.

| Predictor | Structure | Explanation |
|---|-----------|---|
| Tonal relation on the first SC syllables | 2 Levels | Whether or not the counterparts of the first syllables are from the same tonal category in SC; an indicator of systematic correspondence |
| Tonal relation on the second SC syllables | 2 Levels | Whether or not the counterparts of the second syllables are from the same tonal category in SC; and indicator of systematic correspondence |
| Between-dialect identity | 3 Levels | Whether neither, one, or both of the two words is/are identical to its/their counterpart(s) in SC |
| Merge | 5 Levels | Whether or not the tones of the two words are undergoing merging on neither, the second, the first, both, or the combination of the two syllables in JM |
| Neutral tone | 2 Levels | Whether or not this pair involves neutral tones |
| Word frequency | 3 Levels | Whether or not the two words are from different word frequency groups, both from the high frequency group, or both from the low frequency group |

Table 3
Predictors of individual backgrounds.

| Predictor | Structure | Explanation |
|--------------------------------|-----------|--|
| Gender | 2 Levels | Male or female |
| Age | Interval | The speaker's age in 2012 |
| JM absolute proficiency | Interval | Self-rated JM proficiency on a 1–10 scale |
| JM relative proficiency | Interval | The proportion of self-rated JM proficiency in the speaker's total language proficiency |
| JM absolute frequency | Interval | Self-rated JM frequency on a 1 to 10 interval |
| JM relative frequency | Interval | The proportion of self-rated JM frequency in the speaker's total language frequency |
| Education | 2 Levels | The highest education the speaker has received (middle school, college) |
| Language of literacy education | 3 Levels | Whether the speaker received literacy education in JM, SC, or a mixing of both dialects |
| Tonal awareness in JM | Interval | The correct rate in JM tonal oddity test |
| Tonal awareness in SC | Interval | The correct rate in SC tonal oddity test |
| Digit-naming speed | Interval | Digit-naming speed in JM (words/per second) |
| Auditory working memory | Interval | How many digits the speaker can recall in correct order immediately after the digits are presented in JM |

of individual backgrounds (Age, JM absolute proficiency, JM absolute frequency, digit-naming speed, tonal awareness in JM, and tonal awareness in SC) were first centralized and standardized. Then Kappa (Baayen, 2011; Belsley et al., 2005) and pair-wise Pearson correlations were calculated for these interval predictors to quantify the problem of collinearity after which the multicollinearity was reduced via residualization (Baayen et al., 2006; Jaeger, 2010). The nominal predictors of individual backgrounds (Gender, Education, and Language of literacy education) were also considered for their multicollinearity and separate models were fitted accordingly on subsets of data to see whether the effects persisted.

All these models were fitted in an exploratory way. We first built full models, including all the predictors of interest and their two-way and three-way interactions as the fixed predictors, as well as the random intercept of the SC tonal combinations of the two words. When there were unrealized combinations of predictors, which revealed multi-collinearity and would cause rank deficiency in the modeling, the corresponding interaction terms were removed. A backward elimination was then performed to remove non-significant effects, using p -values calculated from F tests based on Sattethwaite's method (Kuznetsova et al., 2013). Finally, we carried out post-hoc tests and calculated the least squares means and confidence intervals for the nominal predictors (Kuznetsova et al., 2013). As for the interval predictors and their interactions with the nominal predictors, we calculated and plotted the slopes of estimated mean distances.

3. Results and discussion

The speaker-wise models showed that the words' SC tonal relations, the between-dialect identity, their two-way and three-way interaction, and the SC tonal combination all affected JM between-word pitch distance. The proportion of variance accounted for by the final models (R^2) ranges between 0.25 and 0.60 (mean = 0.43, median = 0.44), indicating that the predictive power of these models ranges between medium and large. The analysis of individual backgrounds showed that the effects of SC

tonal relations were modulated by the speakers' sociolinguistic and cognitive backgrounds. In every final model, the fixed predictor of the SC tonal combinations of the two words was kept, indicating that SC tonal combination was robust in predicting the between-word pitch distance in JM. In the following sections, results are reported and interpreted based on F statistics and post-hoc estimates of the models (Kuznetsova et al., 2013). Estimates yielded by the model summaries are averaged across all the individual models and reported in Appendix A.

3.1. Systematic correspondence works: effects of SC tonal relation on JM between-word pitch distance

The effects of SC tonal relations reveal the effect of the systematic correspondence mechanism, which predicts that two words that share tonal categories in SC would have a smaller between-word pitch distance in JM.

In the speaker-wise models, the main effect of the tonal relation on the first SC syllables was significant for most speakers, $F \in [7.45, 522.19]$, $p < 0.05$, except for Speaker08, $F = 2.96$, $p = 0.08$ and Speaker20, $F = 3.84$, $p = 0.05$, while the main effect of the tonal relation on the second SC syllable was significant in 24 of the models, $F \in [4.55, 77.92]$, $p < 0.05$ but insignificant in 18 of the models, $F \in [0.00, 3.43]$, $p > 0.05$, indicating that the tonal relation on the first SC syllables was more robust than that on the second syllable as a predictor. The two-way interaction of the tonal relations on the first and second SC syllables was significant in 7 of the models, $F \in [4.31, 9.75]$, $p < 0.05$ but insignificant in 35 of the models, $F \in [0.00, 3.60]$, $p > 0.05$, indicating that the tonal relation in SC was relatively independent of the first and the second syllables in predicting JM between-word pitch distance. In Fig. 7, the estimated means and confidence intervals from the speaker-wise models were plotted in clusters according to the conditions. The combination of the tonal relations on the first and second SC syllables [whether neither (nn), only the first (yn), only the 2nd (ny), both (yy) syllable(s) of the SC counterparts are from the same tonal category] is represented with color-coded clusters and labels on the horizontal axis and the different levels of between-dialect

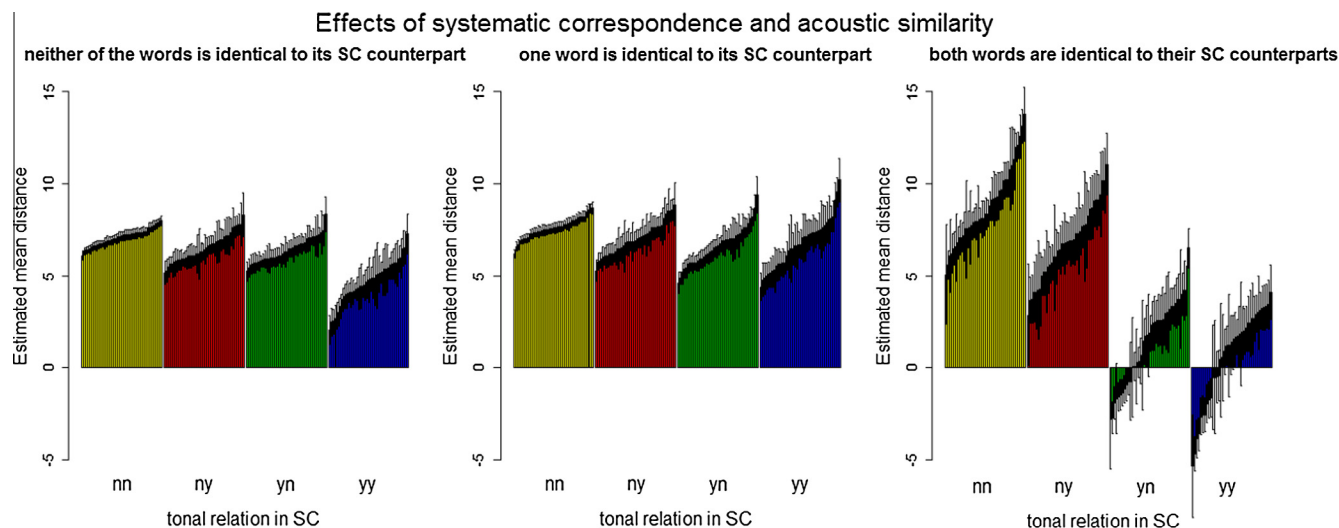


Fig. 7. The interaction of the tonal relation in SC and between-dialect identity on the estimated between-word pitch distance of JM words. Individual estimates under the same condition were clustered according to tonal relation in SC [neither (nn), only the first (yn), only the 2nd (ny), both (yy) syllable(s) are from the same tonal category]. The estimates were split into three plots according to between-dialect identity [neither (left), one (middle), or both (right) of the two words is/are identical to its/their counterpart(s) in SC].

identity were plotted in separate planes. As shown in Fig. 7 (left and right planes), when two words shared the tonal category of their first or second syllables in SC, the between-word pitch distance of their counterparts in JM was also reduced. Sharing more tonal categories in SC reduced the between-word pitch distance in JM. Also sharing the tonal category on the first syllable reduces the distance more than sharing it on the second syllable. Although the effects of SC tonal relations were largely removed when only one of the two JM words was identical to its SC counterpart, in the next session we will discuss the causes.

3.2. Phonological similarity interrupts and reinstalls systematic correspondence: effects of between-dialect identity

The phonological similarity mechanism predicts that, when one of the two JM words was realized almost identically to its SC counterpart, the predicting power of systematic correspondence would be disrupted on this word. Under this condition, whether or not this word shares tonal categories with another word in SC would no longer predict the pitch distance between these two words, unless the other word is also realized almost identically to its SC counterpart. The results fully support the theoretical predictions.

In the final speaker-wise models, the main effect of between-dialect identity was significant in all the models, $F \in [5.36, 689.90]$, $p < 0.05$. However, the direction was inconsistent. More importantly, between-dialect identity interacts with SC tonal relations. The two-way interaction of the tonal relation on the first SC syllables and between-dialect identity was significant for all the speakers, $F \in [7.10, 589.70]$, $p < 0.05$. The two-way interaction of

the tonal relation on the second SC syllable and between-dialect identity was significant in 40 of the models, $F \in [3.32, 127.10]$, $p < 0.05$ but insignificant in 2 of the models, $F \in [1.34, 2.68]$, $p > 0.05$, indicating that this interaction was robust across speakers but less reliable than its counterpart involving the first syllable. The 3-way interaction of the tonal relation on the first SC syllables and second syllable and the between-dialect identity was significant in all of the models, $F \in [3.19, 131.00]$, $p < 0.05$, except that the term was removed for Speaker17 and Speaker26 due to missing combinations. Looking back to Fig. 7, when one of the JM words was identical to its SC counterpart (left plane), sharing tonal categories in SC reduced the between-word pitch distance in JM, revealing the power of the systematic correspondence mechanism. A similar pattern was also found when both of the words were identical to their SC counterparts (right plane). In this situation the JM between-word pitch distance represents the between-word pitch distance of their SC counterparts.⁵ However, when one JM word was identical to its SC counterpart and the other was not, the effects of SC tonal relations were largely removed. This pattern is consistent with the prediction from the phonological similarity mechanism.

Additionally, the difference in between-word pitch distance induced by different SC tonal relations was greater when both of the words were identical to their SC counterparts than when neither of the words was identical to its SC counterpart. This is reasonable. After all, the SC tonal relation is indirectly related to JM between-word pitch distance via a JM tonal relation in the former case but is directly

⁵ Here the model estimated unrealistic negative values for some speakers when the SC counterparts are from the same tonal category in the first or both of the syllable(s), revealing some problem of over-fitting.

reflected by the SC between-word pitch distance in the latter case.

Similar to the results in the speaker-wise models, the main effect of the SC tonal relations was significant in all the models including individual backgrounds and it functioned differently with different levels of between-dialect identity. When neither of the two words was identical to its SC counterpart, sharing tones in SC reduced the JM between-word pitch distance and sharing tones on the second syllable reduced the distance more than sharing tones on the first syllable, $F_{\text{neither}} \in [201.90, 1140.22]$, $p < 0.05$. Similarly, when both of the two words were identical to their counterpart in SC, sharing tones in SC also reduced the JM between-word pitch distance. However, under this condition, sharing tones on the first syllable instead of on the second syllable reduced the distance more, $F_{\text{both}} \in [23.11, 521.81]$, $p < 0.05$. When one of the two words was identical to its SC counterpart, this effect was reversed in that sharing tones in SC increased the JM between-word pitch distance, $F \in [21.86, 41, 92]$, $p < 0.05$.

3.3. Neutral tone disrupts systematic correspondence

The JM neutral tone generally disrupted the predicting power of systematic correspondence. In the speaker-wise models, the main effect of neutral tone was significant in 38 of the models, $F \in [6.64, 6106.00]$, $p < 0.05$ but insignificant in 4, $F \in [0.39, 1.85]$, $p > 0.05$. The direction of the main effect of neutral tone was inconsistent across speakers. The two-way interaction of the tonal relation on the first SC syllables and neutral tone was significant in 39 of the models, $F \in [7.53, 2002.00]$, $p < 0.05$ and insignificant in 3, $F \in [0.39, 2.96]$, $p > 0.05$. The two-way interaction of the tonal relation on the second SC syllable and neutral tone was significant in 40 of the models, $F \in [3.93, 2829.00]$, $p < 0.05$, insignificant in 2, $F \in [0.38, 2.85]$, $p > 0.05$. The three-way interaction of the tonal relations on the first and second SC syllables and neutral tone was significant in 19 of the models, $F \in [3.87, 37.11]$, $p < 0.05$ but was removed in 23. Although sharing the tone on the first or second SC syllable generally reduced the between-word pitch distance, JM neutral tones counterweighed these effects.

Taking between-dialect identity into consideration, we found more complex interactions. The two-way interaction of neutral tone and between-dialect identity was significant in 26 of the models, $F \in [4.35, 246.40]$, $p < 0.05$, insignificant in 4, $F \in [0.55, 2.78]$, $p > 0.05$, and removed in 12. The three-way interaction of the tonal relation on the first SC syllables, neutral, and between-dialect identity was significant in 19 of the models, $F \in [50.61, 307.71]$, $p < 0.05$ but removed in 23. The three-way interaction of the tonal relation on the second SC syllable, neutral, and between-dialect identity was significant in 10 of the models, $F \in [4.66, 54.44]$, $p < 0.05$ but removed in 32. The post-hoc analysis showed that neutral tones interacted with the SC tonal relations in different ways depending on the condition of between-dialect identity. When neither or only

one JM word was identical to its SC counterpart, the involvement of neutral tone generally increased the between-word pitch distance and reduced the effect of SC tonal relations. This is consistent with the general finding that neutral tones disrupted the predicting power of systematic correspondence. However, when both of the two words were identical to their SC counterparts, neutral tones enhanced the effect of the tonal relations on the first SC syllables but reduced the effect of the tonal relations on the second SC syllables.

Why did neutral tones disrupt the effect of the SC tonal relation on the second syllable? This is probably because there is no unified realization of neutral tone and the pitch contour of the neutral tone depends on the tonal category of the preceding syllable. When two SC words both carry neutral tones on the second syllables, the two neutral tones can be realized as very different variants, so long as their previous syllables carry different tones. Thus sharing SC neutral tones cannot reduce the between-word pitch distance when both of the two JM words were identical to their SC counterparts. Similarly, when neither of the JM words are identical to their SC counterparts and the SC words carry neutral tones, the systematic correspondence mechanism predicts that the JM words are also more likely to carry neutral tones. The two JM neutral tones are not necessarily similar in pitch contour either. Hence, unlike the case of sharing the other tonal categories, sharing neutral tones on the second syllable does not reduce the between-word pitch distance.

Why did neutral tones also disrupt the effect of the tonal relations on the first SC syllables when neither of the JM words was identical to its SC counterpart? When both second syllables carry neutral tones, the pitch contours and the between-word pitch distance depend on the tones of the first syllables. However, the same JM citation tone is realized as one of several different sandhi forms before a neutral tone and before the other tones. As a result, when one of the second syllables carries a neutral tone and the other does not, sharing citation tones on the first SC syllables cannot reduce between-word pitch distance of the JM words.

Then why did neutral tones instead enhance the effect of the tonal relations on the first SC syllables when both words were identical to their SC counterparts? Unlike the JM tones, the SC tones preceding neutral tones are realized very much like the corresponding citation forms and the other sandhi forms. As a result, the between-word pitch distance depends mostly on the SC tonal categories of the first syllables.

3.4. Effects of JM tonal merging

Tonal merging in JM generally reduces between-word pitch distance. In the final speaker-wise models, the main effect of merge was significant in all the models, $F \in [31.37, 2870.00]$, $p < 0.05$. Potential merging in JM reduced between-word pitch distance. This effect was more robust for the merging of tonal combinations.

However, Merge also showed complex interactions with the SC tonal relations. The two-way interaction of the tonal relation on the first SC syllables and Merge was significant in 16 of the models, $F \in [3.38, 30.22]$, $p < 0.05$, insignificant in 2, $F \in [1.84, 2.06]$, $p > 0.05$, and removed in 24. The two-way interaction of the tonal relation on the second SC syllable and Merge was significant in 26 of the models, $F \in [4.80, 113.40]$, $p < 0.05$, insignificant in 1 (speaker35), $F = 0.87$, $p > 0.05$, and removed in 15. The three-way interaction of the tonal relations on the first and second SC syllables and Merge was significant in 8 of the models, $F \in [11.01, 66.96]$, $p < 0.05$ and removed in 34. The post-hoc tests showed that the effect of merge is more robust for word pairs that do not share any tonal category in SC, in that the word pairs with both syllables undergoing tonal merging had smaller between-word distances and the word pairs with merging of tonal combinations had even smaller distances. However, when the two words shared tonal categories in SC, the effect of Merge was disrupted and only the merging of tonal combinations guaranteed the smallest between-word pitch distance for every speaker.

3.5. Effect of word frequency

The effect of word frequency was generally inconsistent across speakers, although there was weak evidence supporting the idea that two words from the same word frequency groups tend to have relatively smaller pitch distance.

In the final speaker-wise models, the main effect of word frequency was significant in 26 of the models, $F \in [3.59, 55.91]$, $p < 0.05$ and insignificant in 16 of the models, $F \in [0.02, 2.83]$, $p > 0.05$. The direction of word frequency effects was not consistent across speakers, except that in most models the between-word pitch distances were relatively smaller when both words were high frequency words.

The interaction from word frequency was unclear. The two-way interaction of the tonal relation on the first SC syllables and word frequency was significant in 31 of the models, $F \in [4.65, 71.62]$, $p < 0.05$ and insignificant in 10 of the models, $F \in [0.43, 2.70]$, $p > 0.05$, except that the term was removed for Speaker09. The two-way interaction of the tonal relation on the second SC syllable and word frequency was significant in 20 of the models, $F \in [3.32, 85.86]$, $p < 0.05$ and insignificant in 21 of the models, $F \in [0.07, 2.94]$, $p > 0.05$, except that the term was removed for Speaker25. The two-way interaction of Merge and word frequency was significant in all of the models, $F \in [3.87, 44.69]$, $p < 0.05$, except that the term was removed in 3 models. The two-way interaction of neutral and word frequency was significant in 30 of the models, $F \in [3.58, 91.45]$, $p < 0.05$ but insignificant in 12 of the models, $F \in [0.09, 1.95]$, $p > 0.05$. The two-way interaction of between-dialect identity and word frequency was significant in 31 of the models, $F \in [2.52, 37.15]$, $p < 0.05$ and

insignificant in 10 of the models, $F \in [0.28, 2.02]$, $p > 0.05$, except that the term was removed for Speaker05. The three-way interaction of the tonal relation on the first SC syllables, Merge, and word frequency was significant in only one of the models $F = 13.10$, $p < 0.05$ but removed in all the other models. The three-way interaction of the tonal relation on the first SC syllables, between-dialect identity, and word frequency was significant in 22 of the models, $F \in [3.03, 24.58]$, $p < 0.05$, but removed in 20. The three-way interaction of the tonal relation on the first SC syllables, neutral, and word frequency was significant in 24 of the models, $F \in [3.38, 74.36]$, $p < 0.05$, but removed in 18. The three-way interaction of the tonal relation on the second SC syllable, merge, and word frequency was significant in 6 of the models, $F \in [4.11, 11.22]$, $p < 0.05$ but removed in 36. The three-way interaction of the tonal relation on the second SC syllable, between-dialect identity, and word frequency was significant in 25 of the models, $F \in [2.38, 17.14]$, $p < 0.05$ but removed in 17. The three-way interaction of the tonal relation on the second SC syllable, neutral, and word frequency was significant in 29 of the models, $F \in [3.24, 50.36]$, $p < 0.05$, but removed in 13. The three-way interaction of the tonal relations on the first and second SC syllables and word frequency was significant in 16 of the models, $F \in [3.13, 67.38]$, $p < 0.05$, but removed in 26. The three-way interaction of neutral tone, between-dialect identity, and word frequency was significant in 11 of the models, $F \in [3.96, 22.79]$, $p < 0.05$, but removed in 31. The directions of these interactions were inconsistent across speakers except that in most models the effect of sharing tones on the first SC syllables increased when both words were low frequency words.

3.6. Effects of the speakers' age, cognitive, and socio-linguistic backgrounds

Individual backgrounds influence the way systematic correspondence takes effect. In this section we start with the predictor age and investigated how the speakers' cognitive and socio-linguistic backgrounds exert age-mediated and age-unmediated influences on systematic correspondence.

3.6.1. Individual models

Before starting investigating the age-independent effects, we first included each aspect of the individual backgrounds separately in smaller models and investigated their mediated and unmediated effects together.

First we started with the age effect. When age was included alone with the linguistic predictors in the model, as shown in Fig. 8, a younger age generally increased the between-word pitch distance between two JM words and enhanced the effect of the tonal relations in SC. The main effect of age and the interaction of age with the SC tonal relations were significant when neither of the words was identical to its SC counterpart, $F_{\text{main}}(40.00) = 5.23$, $p < 0.05$; $F_{\text{interaction}}(119.97) = 12.03$, $p < 0.05$; and

when one word was identical to its SC counterpart, $F_{\text{main}}(40.00) = 18.43$, $p < 0.05$; $F_{\text{interaction}}(119.97) = 5.60$, $p < 0.05$, but insignificant when both of the two words were identical to their counterpart in SC, $F_{\text{main}}(37.53) = 0.06$, n.s.; $F_{\text{interaction}}(116.89) = 0.65$, n.s. Thus the age effect only existed when at least one word occurred exclusively in JM. The proportions of variance accounted for by the final models (R^2) with age are $R_{\text{neither}}^2 = 0.96$, $R_{\text{one}}^2 = 0.79$, $R_{\text{both}}^2 = 0.89$.

When auditory working memory was included alone with the linguistic predictors in the model, we found that a better auditory working memory enhanced the effect of the SC tonal relations. The main effect of auditory working memory was only significant when one word was identical to its SC counterpart, $F(38.00) = 9.95$, $p < 0.05$, but insignificant when neither or both of the words was/were identical to its/their SC counterpart(s), $F_{\text{neither}}(38.00) = 2.66$, n.s.; $F_{\text{both}}(35.59) = 0.06$, n.s. The interaction of auditory working memory and SC tonal relations was significant when neither or one of the words was identical to its SC counterpart, $F_{\text{neither}}(113.97) = 13.88$, $p < 0.05$; $F_{\text{one}}(113.97) = 4.70$, $p < 0.05$, but insignificant when both of the two words were identical to their SC counterparts, $F_{\text{both}}(110.96) = 0.48$, n.s. The direction of the main effect of auditory working memory when one word was identical to its SC counterpart indicates that a better auditory working memory generally increased the between-word pitch distance between a JM form and a common form. Also it is clear that speakers who have better auditory working memories showed greater differences of between-word pitch distance across different levels of SC tonal relations when neither of the two words was identical to its SC counterpart. The proportions of variance accounted for by the final models (R^2) with auditory working memory are $R_{\text{neither}}^2 = 0.96$, $R_{\text{one}}^2 = 0.78$, $R_{\text{both}}^2 = 0.88$.

When tonal awareness was included alone with the linguistic predictors in the model, we found that better tonal

awareness enhanced the effect of the SC tonal relations. The main effect of SC tonal awareness was only marginally significant when one word was identical to its SC counterpart, $F(38.00) = 4.08$, $p = 0.05$, and insignificant when neither or both of the words was/were identical to its/their SC counterpart(s), $F_{\text{neither}}(38.00) = 0.84$, n.s.; $F_{\text{both}}(35.84) = 0.29$, n.s. The interaction of SC tonal awareness and SC tonal relations was significant only when neither of the words was identical to its SC counterpart, $F_{\text{neither}}(113.97) = 4.03$, $p < 0.05$, but insignificant when neither or both of the words was/were identical to its/their SC counterpart(s), $F_{\text{one}}(113.98) = 0.87$, n.s.; $F_{\text{both}}(111.22) = 0.69$, n.s. Similarly, the main effect of JM tonal awareness was only significant when one word was identical to its SC counterpart, $F(38.00) = 4.43$, $p < 0.05$, but insignificant when neither or both of the words was/were identical to its/their SC counterpart(s), $F_{\text{neither}}(38.00) = 2.35$, n.s.; $F_{\text{both}}(35.87) = 0.51$, n.s. The interaction of JM tonal awareness and SC tonal relations was significant when neither or one of the words was identical to its SC counterpart, $F_{\text{neither}}(113.97) = 3.70$, $p < 0.05$; $F_{\text{one}}(113.98) = 2.44$, $p < 0.05$, but insignificant when both of the two words were identical to their SC counterparts, $F_{\text{both}}(111.26) = 0.15$, n.s. The direction of the main effect of tonal awareness when one word was identical to its SC counterpart indicates that better tonal awareness also generally increased the between-word pitch distance between a JM form and a common form. Also, it is clear that speakers with better tonal awareness showed greater differences of between-word pitch distance across different levels of SC tonal relations when neither of the two words was identical to its SC counterpart. The proportions of variance accounted for by the final models (R^2) with JM tonal awareness are $R_{\text{neither}}^2 = 0.95$, $R_{\text{one}}^2 = 0.77$, $R_{\text{both}}^2 = 0.88$.

However, digit naming speed showed no significant main effect with any level of the between-dialect identity, $F_{\text{neither}}(40.00) = 0.04$, n.s., $F_{\text{one}}(40.00) = 0.80$, n.s.,

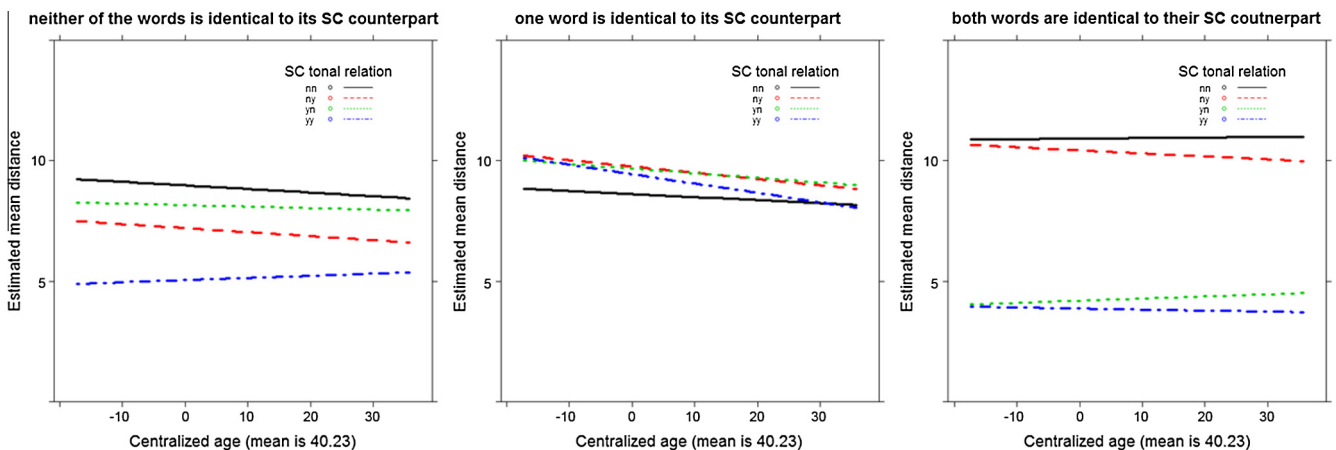


Fig. 8. The interaction of age and SC tonal relation [neither (nn), only the first (yn), only the 2nd (ny), both (yy) syllable(s) are from the same tonal category] on the estimated between-word pitch distance of JM words. The estimates were split into three plots according to between-dialect identity [neither (left), one (middle), both (right) of the two words is/are identical to its/their counterpart(s) in SC].

F_{both} (37.29) = 0.35, n.s., and no significant interaction with the SC tonal relations, F_{neither} (119.97) = 2.21, n.s., F_{one} (119.98) = 0.40, n.s., F_{both} (116.66) = 0.33, n.s. The proportions of variance accounted for by the final models (R^2) with digit naming speed are $R_{\text{neither}}^2 = 0.94$, $R_{\text{one}}^2 = 0.76$, $R_{\text{both}}^2 = 0.88$.

Also, with the pitch range difference across speakers normalized, gender also showed no significant main effect, F_{neither} (40.00) = 0.06, n.s.; F_{one} (40.00) = 0.09, n.s.; F_{both} (38.00) = 0.12, n.s., no significant interaction with the SC tonal relations, F_{neither} (119.97) = 0.67, n.s.; F_{one} (119.98) = 0.22, n.s.; F_{both} (117.40) = 0.19, n.s., and no significant interaction with age, F_{neither} (35.00) = 1.55, n.s.; F_{one} (35.00) = 1.57, n.s.; F_{both} (33.06) = 1.42, n.s. Thus there is no evidence in support of any gender-affected variant choice in JM.

As for the other socio-linguistic predictors, when JM absolute or relative frequency was included alone with the linguistic predictors into the model, we found that a lower frequency in JM enhanced the effects of the SC tonal relations. The main effects of JM relative and absolute frequencies were significant only when one of the words was identical to its SC counterpart, F_{relative} (40.00) = 22.94, $p < 0.05$, F_{absolute} (40.00) = 12.30, $p < 0.05$, but insignificant when neither or both of the words was/were identical to its/their counterpart in SC, $F_{\text{neither_relative}}$ (40.00) = 2.87, n.s., $F_{\text{both_relative}}$ (38.15) = 0.35, n.s., $F_{\text{neither_absolute}}$ (40) = 1.26, n.s., $F_{\text{both_absolute}}$ (38.27) = 0.12, n.s. The interaction of JM relative frequency with SC tonal relations was significant for all the levels of between-dialect identity, F_{neither} (119.96) = 27.36, $p < 0.05$, F_{one} (119.98) = 5.42, $p < 0.05$, F_{both} (117.46) = 3.51, $p < 0.05$. The interaction of JM absolute frequency and SC tonal relations was significant when neither or one of the words was identical to its SC counterpart, F_{neither} (119.97) = 16.58, $p < 0.05$, F_{one} (119.98) = 2.97, $p < 0.05$, but insignificant when both of the two words were identical to their counterpart in SC, F_{both} (117.60) = 2.10, n.s. The direction of the main effect of JM absolute and relative frequencies when one word was identical to its SC counterpart indicates that the increased usage of JM generally reduces the between-word pitch distance between JM and SC forms. Also it is clear that speakers who use JM relatively or absolutely more often showed smaller differences of between-word pitch distance across different levels of SC tonal relations when neither of the two words was identical to its SC counterpart. The proportions of variance accounted for by the final models (R^2) with relative frequency are $R_{\text{neither}}^2 = 0.97$, $R_{\text{one}}^2 = 0.79$, $R_{\text{both}}^2 = 0.90$. The proportions of variance accounted for by the final models (R^2) with absolute frequency are $R_{\text{neither}}^2 = 0.96$, $R_{\text{one}}^2 = 0.78$, $R_{\text{both}}^2 = 0.89$.

Similarly, when JM absolute or relative proficiency was included alone with the linguistic predictors into the model, we found that a lower proficiency in JM enhanced the effect of the SC tonal relations. The main effect of JM relative

proficiency was significant only when one word was identical to its SC counterpart, F (40.00) = 13.20, $p < 0.05$, but insignificant when neither or both of the words was/were identical to its/their SC counterpart(s), F_{neither} (40.00) = 1.60, n.s.; F_{both} (37.25) = 0.17, n.s. The proportions of variance accounted for by the final models (R^2) with relative proficiency are $R_{\text{neither}}^2 = 0.95$, $R_{\text{one}}^2 = 0.77$, $R_{\text{both}}^2 = 0.89$. The interaction of JM relative proficiency and SC tonal relations was significant when neither or one of the words was identical to its SC counterpart, F_{neither} (119.97) = 15.38, $p < 0.05$; F_{one} (119.98) = 3.53, $p < 0.05$, but insignificant when both of the two words were identical to their SC counterpart, F_{both} (116.58) = 1.59, n.s. The effect of JM absolute proficiency was similar in that the interaction of JM absolute proficiency and SC tonal relations was significant when neither of the words was identical to its SC counterpart, F_{neither} (119.97) = 3.29, $p < 0.05$, but insignificant when one or both of the two words were identical to their SC counterparts, F_{one} (119.98) = 1.40, n.s., F_{both} (116.83) = 0.32, n.s. However, the main effect of JM absolute proficiency was insignificant for all the levels of between-dialect identity, F_{neither} (40.00) = 0.01, n.s.; F_{one} (40.00) = 1.95, n.s.; F_{both} (37.46) = 0.02, n.s. The direction of the main effect of JM relative proficiency when one word was identical to its SC counterpart indicates that the increased dominance of JM generally reduced the between-word pitch distance between JM and SC forms. Similar to what was found for JM frequencies, speakers who are more proficient in JM showed smaller differences of between-word pitch distance across different levels of SC tonal relations when neither of the two words was identical to its SC counterpart. The proportions of variance accounted for by the final models (R^2) with absolute proficiency are $R_{\text{neither}}^2 = 0.96$, $R_{\text{one}}^2 = 0.78$, $R_{\text{both}}^2 = 0.89$.

Excluding the only speaker who did not receive formal education and including education level or language of literacy education alone with the linguistic predictors into the model, the main effect of education level was only significant when one word was identical to its SC counterpart, F (37.00) = 4.47, $p < 0.05$, but insignificant when neither or both of the words were identical to their SC counterpart (s), F_{neither} (37.00) = 0.00, n.s.; F_{both} (34.26) = 1.27, n.s. The interaction of education level and SC tonal relations was significant when neither or one of the words was identical to its SC counterpart, F_{neither} (110.97) = 9.14, $p < 0.05$; F_{one} (110.98) = 4.75, $p < 0.05$, but insignificant when both of the two words were identical to their SC counterpart, F_{both} (107.61) = 0.36, n.s. Similarly, the main effect of the language of literacy education was only significant when one word was identical to its SC counterpart, F (36.00) = 11.16, $p < 0.05$, but insignificant when neither or both of the words were identical to their SC counterpart, F_{neither} (36.00) = 0.59, n.s.; F_{both} (33.56) = 0.05, n.s. The interaction of the language of literacy education and SC tonal relations was significant when neither or one of the words was identical to its SC counterpart, F_{neither}

(107.97) = 7.38, $p < 0.05$; F_{one} (107.98) = 3.13, $p < 0.05$, but insignificant when both of the two words were identical to their SC counterparts, F_{both} (104.90) = 1.14, n.s. When one of the words was identical to its SC counterpart, a higher education or a literacy education in SC increased the between-word pitch distance and changed the way SC tonal relations influenced acoustic distance. When neither of the words was identical to its SC counterpart, a higher education or a literacy education in SC enhanced the effect of the SC tonal relations. When both of the words were identical to their SC counterparts, the education level and literacy education had no effect. The proportions of variance accounted for by the final models (R^2) with education are $R_{\text{neither}}^2 = 0.95$, $R_{\text{one}}^2 = 0.78$, $R_{\text{both}}^2 = 0.88$. The proportions of variance accounted for by the final models (R^2) with literacy education are $R_{\text{neither}}^2 = 0.96$, $R_{\text{one}}^2 = 0.79$, $R_{\text{both}}^2 = 0.89$.

Taken together, any one of these speaker-related variables proved useful when included alone in the modeling. Systematic correspondence had a stronger effect with younger ages, better tonal awareness, larger auditory working memories, literacy educations in SC, higher education levels, and lower frequencies and proficiencies of JM usage, when neither of the words was identical to its SC counterpart. Such backgrounds also increased the between-word pitch distance when only one of the JM words was produced identical to its SC counterpart. However, when both of the JM words were produced identically to their SC counterparts, these individual background predictors had little effect.

However, which individual backgrounds have age-independent effects? We addressed this question by including these predictors (after residualization) together with age to see to what extent they still played a significant role in accounting for the between-word distances.

3.6.2. Residualized models

We first analyzed how the other interval predictors of individual backgrounds are related to age and between each other. As expected, complex multi-collinearity exists across the individual backgrounds predictors. Taking all the individual backgrounds together, the Kappa value was 7.77, indicating that the collinearity problem was not very serious in general (Kappa smaller than 10 indicates reasonable collinearity). However, pair-wise Pearson correlations showed that most of the other predictors strongly correlated with age (absolute r ranges from 0.38 to 0.73). The other predictors also showed strong correlations between each other (absolute $r > 0.27$), except that the digit naming speed seemed to be more independent (absolute r ranged between 0.04 and 0.38). Moreover, the education level and language of literacy education also co-varied with age, in the way that younger speakers tend to receive a higher education and receive their literacy education in SC. Only gender was independent of age.

Accordingly, we residualized all the other interval predictors against age. We also assumed that the proficiency

effects are mediated by frequency effects and that the effects of JM tonal awareness are mediated by the effects of SC tonal awareness. Thus the JM proficiency was residualized against the combination of JM frequency and age and the tonal awareness of JM was residualized against the combination of SC tonal awareness and age. Instead of the relative frequency and proficiency, the absolute values were kept because they had smaller correlations with age. The initial model included Speaker as the random intercept and the following terms as fixed effects: SC tonal relations, standardized age, education level, language of literacy education, all the possible two-, three-, and four-way interactions of the above mentioned terms, residualized JM absolute proficiency, residualized JM absolute frequency, residualized JM auditory working memory, residualized JM tonal awareness, residualized SC tonal awareness, and all the possible two- and three-way interactions of each residualized predictor with SC tonal relations and education level.

After statistically removing the collinearity, we managed to test all the individual backgrounds in one bigger model. The result showed that only the lower frequency of JM still had an unmediated effect of strengthening the effect of systematic correspondence.

When neither of the two words was identical to its SC counterpart, the main effects of SC tonal relations, F (86.97) = 201.91, $p < 0.05$, and education level, F (27.00) = 5.53, $p < 0.05$ were significant. The main effect of residualized JM frequency, F (27.00) = 0.14, n.s., was insignificant in the final model and the main effect of residualized JM proficiency, F (27.00) = 3.71, $p = 0.06$, was only marginally significant. All the terms including residualized auditory working memory and residualized tonal awareness were removed in the final model. The two-way interaction of SC tonal relations and residualized JM frequency, F (86.97) = 3.94, $p < 0.05$ and the two-way interaction of education level and residualized JM proficiency, F (27.00) = 7.47, $p < 0.05$ were significant. The tree-way interaction of the SC tonal relations, age, and the language of literacy education was significant, F (86.97) = 2.24, $p < 0.05$. The three-way interaction of the SC tonal relations, education level, and the language of literacy education was also significant, F (86.97) = 4.39, $p < 0.05$. Similar to the results from the other types of models, sharing tones in SC reduced the JM between-word pitch distance under the present condition. As shown in Fig. 9, after removing the mediated effect of age, a higher JM frequency reduced the differences induced by the SC tonal relations and the between-word pitch distance increased with a higher level of education. The interaction of age and the SC tonal relations was modulated by the type of literacy education. If the speaker had received literacy education in JM, a younger age reduced the differences induced by the SC tonal relations. However, if the speaker had received literacy education in SC, a younger age enhanced the differences induced by the SC tonal relations. If the speaker had received literacy education in a mixed

condition, age did not have much effect. The proportion of variance accounted for by this final model (R^2) is $R_{\text{neither}^2} = 0.98$.

When only one of the two words was identical to its SC counterpart, only the main effects of SC tonal relations, $F(107.98) = 25.74$, $p < 0.05$, the main effect of the language of literacy education, $F(36.00) = 11.16$, $p < 0.05$, and their interaction, $F(107.98) = 3.13$, $p < 0.05$ were significant. All the other terms were insignificant and removed in the final model. As shown in Fig. 10, sharing tonal categories in SC did not reduce but increased between-word pitch distance, because in this case we are comparing corresponding tonal categories in JM and SC. Receiving literacy education in SC increased the between-word pitch distance between a JM form and a common form but receiving literacy education in JM reduced such pitch distance. Literacy education in SC seems to tear the corresponding tonal categories in the two dialects apart and the literacy education in JM seems to make the corresponding tonal categories acoustically more similar. The proportion of variance accounted for by this final models (R^2) is $R_{\text{one}^2} = 0.79$.

When both of the words were identical to their counterpart in SC, only the main effect of SC tonal relations was significant, $F(3) = 245.87$, $p < 0.05$ in the final model (the

random intercept of Speaker was removed in the model trimming). All other predictors and interactions were kept but insignificant. This effect was consistent with the results of the speaker-wise models. Sharing tonal categories in SC reduced the JM between-word pitch distance. The proportion of variance accounted for by this final model (R^2) is $R_{\text{both}^2} = 0.92$.

3.6.3. Stratified models

The nominal predictors of socio-linguistic backgrounds (except gender) also co-varied with age. Younger speakers tend to receive a higher education and receive the literacy education in SC. This type of collinearity was not considered in the model mentioned above. We further stratified the data, and fitted separate models to subsets of data to see whether the effects found in the bigger model persist within each group of speakers. (The scale predictors were normalized and residualized again for each subset of data.)

Although all six possible combinations of language of literacy education and education level (middle-school-JM, college-JM, middle-school-SC, college-SC, middle school-mix, and college-mix) existed in the data, only two subsets of data, namely middle school-JM and college-SC, had enough data points for fitting separate models. Here we

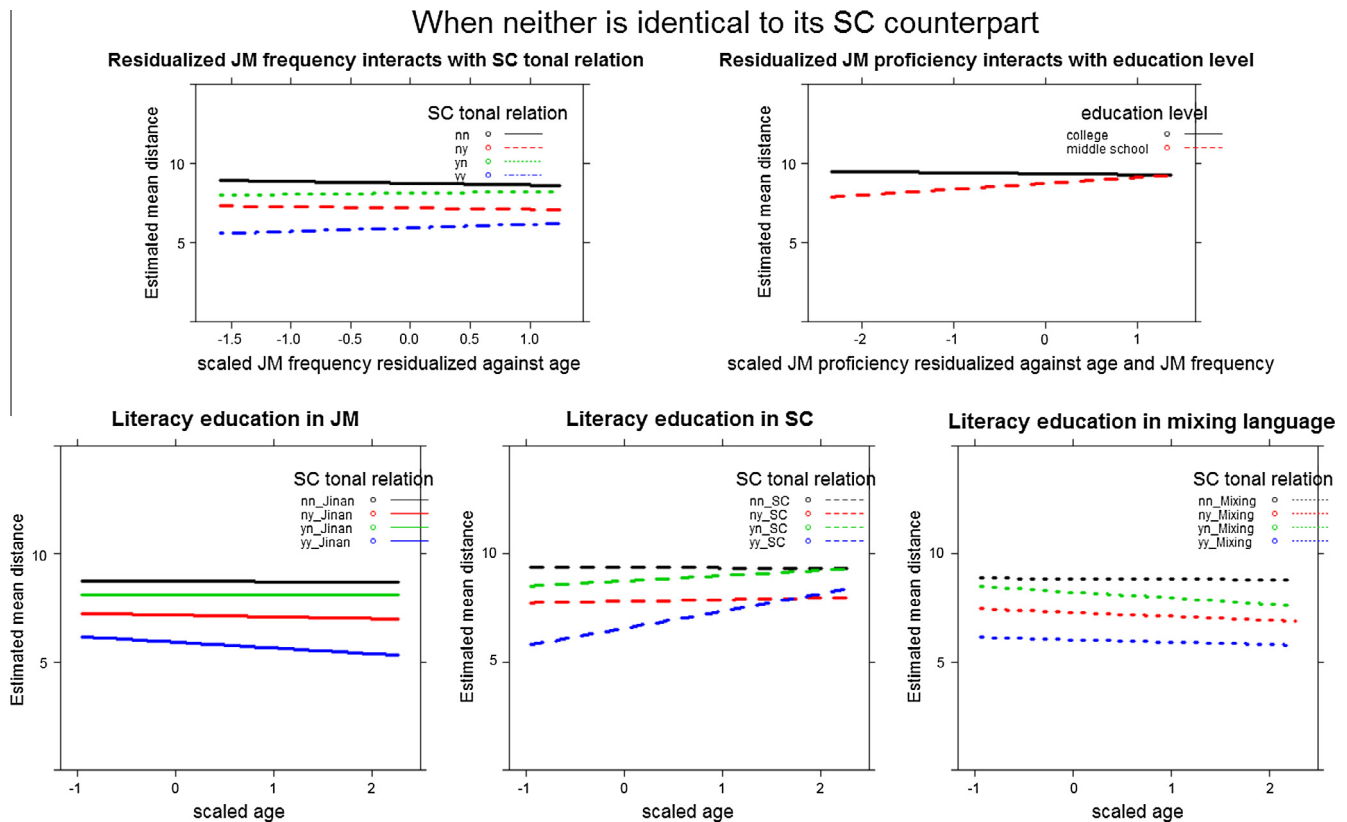


Fig. 9. (1) Top-left: the interaction of the speaker's residualized JM frequency and their interaction with SC tonal relation [neither (nn), only the first (yn), only the 2nd (ny), both (yy) syllable(s) are from the same tonal category] on the estimated between-word pitch distance of JM words. (2) Top-right: the interaction of the speaker's residualized JM proficiency and the speaker's education level (middle school, college) on the estimated between-word pitch distance of JM words. (3) Bottom: the interaction of age, literacy education, and SC tonal relation [neither (nn), only the first (yn), only the 2nd (ny), both (yy) syllable(s) are from the same tonal category] on the estimated between-word pitch distance of JM words. The estimates were split into three plots according to the language of literacy education.

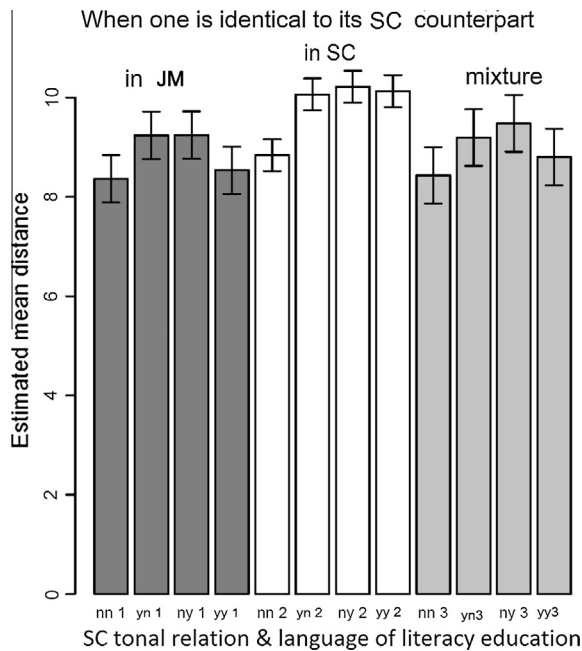


Fig. 10. The interaction of the language of literacy education and tonal relation in SC [neither (nn), only the first (yn), only the 2nd (ny), both (yy) syllable(s) are from the same tonal category] on the estimated between-word pitch distance of JM words.

report and discuss the results of these two sets of models. The effects found in the bigger model persisted and more unmediated effects emerged in the middle-school-JM group.

First we studied the speakers who achieved middle school level education and received their literacy education in JM. As shown in Fig. 11, they were mostly middle aged. When neither of the two words was identical to its SC counterpart, the main effects of SC tonal relations was significant, $F(17.99) = 248.09$, $p < 0.05$. Although the main effect of residualized absolute proficiency of JM was insignificant, its interaction with SC tonal relations was significant, $F(17.99) = 3.61$, $p < 0.05$. The main effects of age, residualized JM absolute frequency, normalized JM auditory working memory, and normalized JM tonal awareness were kept in the final model but proved insignificant. All the other interaction terms were removed. When residualized relative frequency and proficiency of JM were used instead of the absolute values in the modeling, the effect of residualized JM proficiency became insignificant, whereas the interactions with normalized age, $F(14.99) = 3.29$, $p < 0.05$, and residualized auditory working memory, $F(14.99) = 5.14$, $p < 0.05$, became significant. The increase of age and residualized auditory working memory enhanced the differences induced by the SC tonal relations. This suggests that the effect of systematic correspondence is stronger for the oldest speakers and the speakers who have better auditory working memories. The advantage of the oldest speakers was also found in the model depicted in Fig. 9 and may be due to the philology training the oldest speakers received. As shown in Fig. 11, the residualized JM

absolute proficiency enhanced the differences induced by the SC tonal relations, which is opposite to the direction of the general effect of JM proficiencies. This suggests that after removing the influences from the times and JM frequency, middle-school-JM speakers with higher JM proficiencies reflect the systematic correspondence between their two dialects more closely. The proportions of variance accounted for by these final models (R^2) by the middle-school-JM group are $R_{\text{neither}}^2 = 0.98$, $R_{\text{one}}^2 = 0.96$, $R_{\text{both}}^2 = 1.00$.

For this middle-school-JM group, when one of the two words was identical to its SC counterpart, the main effects of SC tonal relations was significant, $F(14.99) = 33.71$, $p < 0.05$. Although the main effect of residualized JM absolute proficiency was insignificant, its interaction with SC tonal relations was significant, $F(14.99) = 8.33$, $p < 0.05$. The interaction between SC tonal relations and SC tonal awareness was also significant, $F(14.99) = 6.10$, $p < 0.05$. For word pairs sharing tones on the second SC syllables, the between-word pitch distance increased with the increase of residualized JM absolute proficiency. For word pairs sharing tones on the second SC syllables but not on the first, the residualized JM absolute proficiency increased with tonal awareness. These effects reflect the pitch differences between the corresponding SC and JM tonal categories. The results indicate that speakers with higher JM proficiency and better tonal awareness maintain greater pitch differences between the corresponding SC and JM tonal categories.

For this middle school-JM group, when both of the two words were identical to their SC counterparts, the random intercept of speaker was removed in the final model and here we report the results of the full model. The main effects of SC tonal relations were significant, $F(3.99) = 521.80$, $p < 0.05$. The main effect of residualized JM frequency was significant in the F -statistics, $F(3.99) = 8.00$, $p < 0.05$, but insignificant in the t -statistics. The interaction between SC tonal relations and SC tonal awareness was insignificant in the F -statistics, but sharing the tonal category on the second syllable but not the first syllable significantly interacted with SC tonal awareness in the t -statistics, $t(3.99) = 3.30$, $p < 0.05$, in that a higher SC tonal awareness increased the between-word pitch distance. This result indicates that the speakers with a better SC tonal awareness focus more on the tonal contrasts on the first SC syllables.

Second we studied the speakers who had achieved college level education and received the literacy education in SC. As shown in Fig. 12, they were mostly in their twenties, i.e. younger than the middle school-JM group. When neither of the two words was identical to its SC counterpart, the main effect of SC tonal relations was significant, $F(53.99) = 529.45$, $p < 0.05$. Although the main effect of residualized absolute frequency of JM was insignificant, its interaction with SC tonal relations was significant, $F(53.99) = 3.04$, $p < 0.05$. All the other terms were insignificant and removed. We found the same result when residualized relative

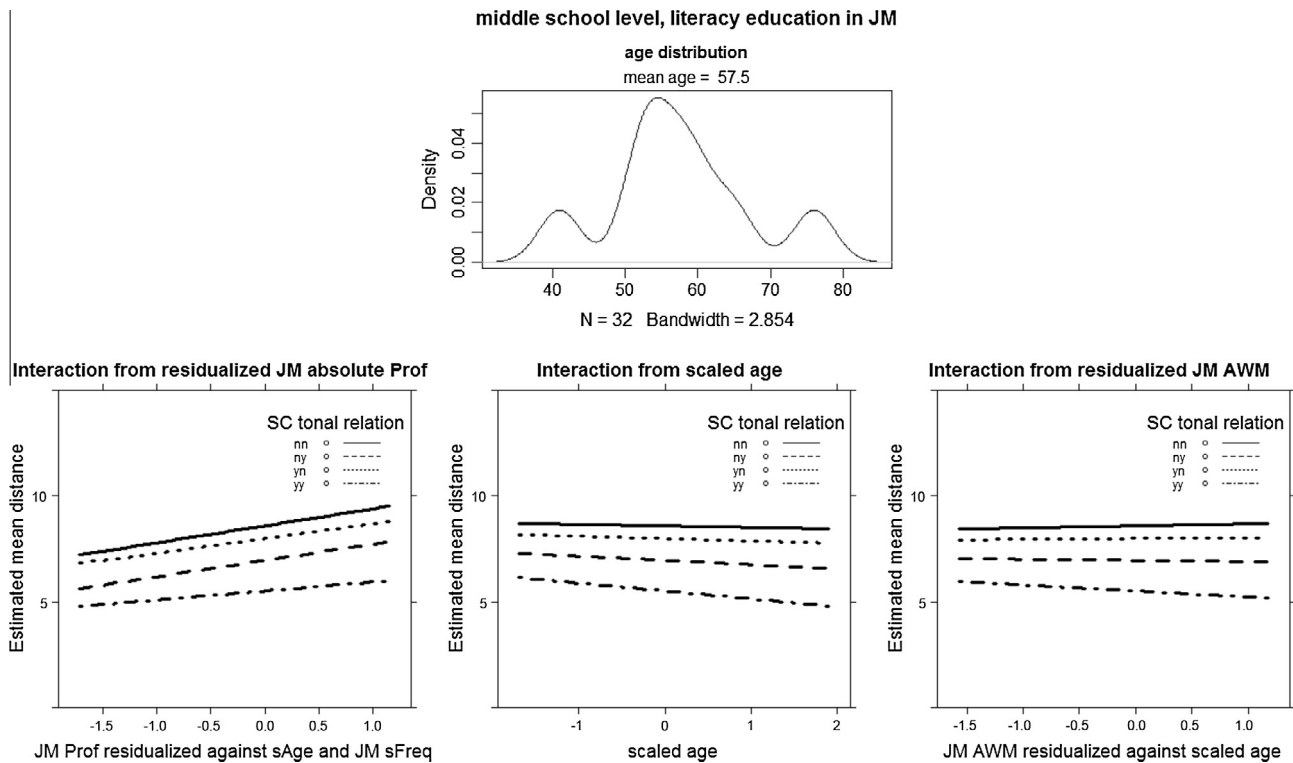


Fig. 11. The models were built for the middle-school-JM group when neither of the two words was identical to its SC counterpart. (1) Top: the age distribution. (2) Bottom left: the interaction of the speaker's residualized JM absolute proficiency and its interaction with SC tonal relation [neither (nn), only the first (yn), only the 2nd (ny), both (yy) syllable(s) are from the same tonal category] on the estimated between-word pitch distance of JM words. (3) Bottom middle: the interaction of the speaker's scaled age and its interaction with SC tonal relation [neither (nn), only the first (yn), only the 2nd (ny), both (yy) syllable(s) are from the same tonal category] on the estimated between-word pitch distance of JM words. (4) Bottom right: the interaction of the speaker's residualized auditory working memory and its interaction with SC tonal relation [neither (nn), only the first (yn), only the 2nd (ny), both (yy) syllable(s) are from the same tonal category] on the estimated between-word pitch distance of JM words.

frequency and proficiency of JM were used instead of the absolute values in the modeling. As shown in Fig. 12, after removing the age effect a higher JM frequency reduced the differences induced by the SC tonal relations for these speakers. Thus the effect of JM frequency on systematic correspondence is not just mediated by the age effect. Within this young group of bilinguals, those who use JM more frequently are less affected by the systematic correspondence. The proportions of variance accounted for by these final models by the college-SC group (R^2) are $R_{\text{neither}}^2 = 0.96$, $R_{\text{one}}^2 = 0.68$, $R_{\text{both}}^2 = 0.87$.

For this college-SC group, when one of the two words was identical to its SC counterpart, only the main effects of SC tonal relations were significant, $F(56.99) = 21.86$, $p < 0.05$. When both of the two words were identical to their SC counterparts, the random intercept of speaker was removed in the final model and in the full model also only the main effects of SC tonal relations were significant, $F(37.70) = 123.02$, $p < 0.05$.

Taken together, the age effect interacts and covaries with the language of literacy education. Considering all the speakers in this corpus, when the speaker received literacy education in SC, a younger age strengthens the effect of systematic correspondence. However, when the speaker received literacy education in JM, a younger age reduces the effect of systematic correspondence. One explanation

is that the oldest people (who received their literacy education in JM) received some training in philology and the youngest people (who received their literacy education in SC) received better education in pinyin. These interactions only occur within the JM system. When considering a pair of JM words that are both realized almost identically to their SC counterparts, only the SC tonal relations count.

However, most of the participants are either middle-aged people who achieved middle school education and received their literacy educations in JM or young people who achieved college education and received their literacy educations in SC. This demographic distribution is shaped by the social progress in the past 30 years. Analyzing these two main groups of bilinguals separately, more cognitive and sociolinguistic effects emerged with the older group. After statistically removing the mediated effects, we found that the older bilinguals with higher JM proficiencies and better auditory working memories more closely reflect the systematic correspondence. Also the older bilinguals with higher JM proficiency and better tonal awareness are better at distinguishing SC and JM tonal categories. We did not find such effects with the younger group. Instead we only found an unmediated effect of JM absolute frequency for these younger bilinguals. Those who use JM more often are less influenced by systematic correspondence.

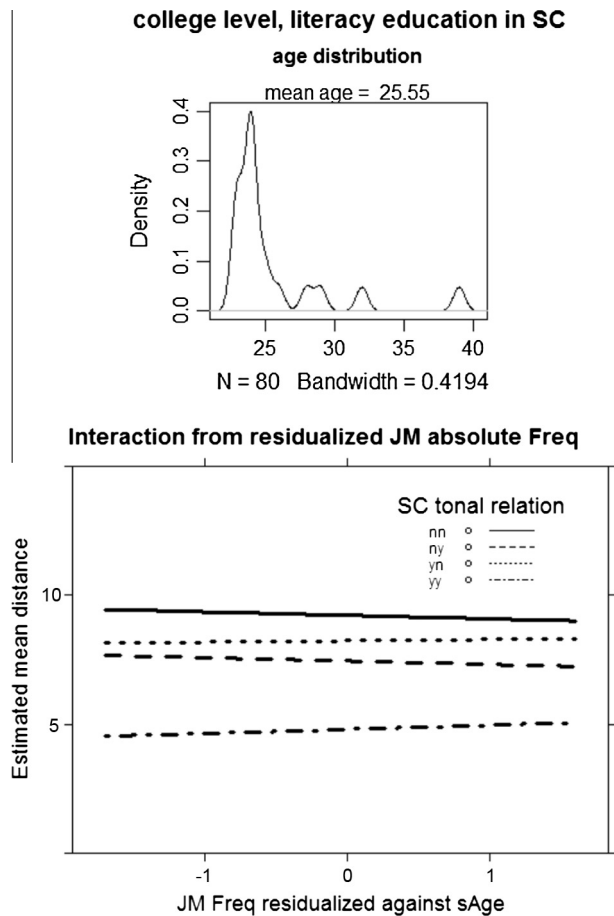


Fig. 12. The models built for the college-SC group when neither of the two words was identical to its SC counterpart. (1) Top: the age distribution. (2) Bottom: the interaction of the speaker's residualized JM absolute frequency and its interaction with SC tonal relation [neither (nn), only the first (yn), only the 2nd (ny), both (yy) syllable(s) are from the same tonal category] on the estimated between-word pitch distance of JM words.

4. General discussion

4.1. Main findings

We have shown that SC tonal relations, between-dialect identity, and their interactions are significant in predicting the between-word pitch distance in JM. It is not surprising that the present results are in line with the basic principle of historical linguistics, namely that phonemes correspond systematically in related words between two related languages/dialects (Meillet and Ford, 1967). We deduced from this well-known phenomenon that the tonal realizations of JM words could be predicted from the tonal categories of their SC counterparts (see Table 1). We also reproduced this implication via statistical modeling of JM between-word pitch distance.

There is always a certain proportion of words that do not fit the systematic correspondence rule. These are usually treated as exceptions by historical linguists. However, these “exceptions” need to be handled if we want to apply

the systematic correspondence rules in engineering. We have shown that systematic correspondence is not randomly violated. Many linguistic and paralinguistic factors influence the strength of systematic correspondence.

First, the “exceptions” may reflect another potentially useful type of relation between the two languages/dialects, namely phonological similarity. JM has been in close and long-term contact with SC and most JM speakers are bilinguals with SC as L2. As a result, a large proportion of the JM “exception” words are almost identical to their SC counterparts, which are sometimes, though neither necessarily nor sufficiently, a result from borrowing (see Sections 1.3 and 1.5). In the present study, we have analyzed how this mechanism interacts with the effects of systematic correspondence in predicting JM between-word pitch distance.

Second, neutral tones disrupt systematic correspondence. Neutral tones disrupted the effect of the SC tonal relation on the second syllable. The effect of the SC tonal relation on the second syllable was also disrupted by neutral tones when neither of the two words was identical to its SC counterpart. Only when both words were identical to their SC counterparts, did neutral tones enhance the effect of the tonal relations on the first SC syllables. These findings have a linguistic basis. Neutral tones, whether in JM or SC, are realized differently depending on the preceding tonal categories. Thus simply sharing neutral tones does not guarantee a smaller between-word pitch distance. Nevertheless, neutral tone sandhi works differently in SC than in JM. JM morphemes followed by neutral tones carry different tonal variants than their citation forms and other sandhi forms. However, SC morphemes followed by neutral tones be realized similarly to the corresponding citation forms. These findings taken together suggest that, in order to improve the predicting power of systematic correspondence when converting an SC pronunciation dictionary to a JM version, it may be better to distinguish subtypes of neutral tones by their preceding tones and also include the information of whether the current syllable is followed by a neutral tone.

Third, we examined how cognitive and sociolinguistic backgrounds affect the way systematic correspondence works, where we differentiated the age-mediated and age-independent effects of these individual backgrounds.

We started with the general (mediated and unmediated together) effects of individual backgrounds. Including each aspect of individual backgrounds separately into the model, we found that the younger, less frequent, less proficient speakers of JM, who received higher education, received their literacy educations in SC, and have better working memories, were generally more sensitive to the SC tonal relations and produced the JM and SC corresponding tonal categories with greater differences. These findings reflect the change of the JM tonal system across different generations: the tonal relations across JM words are increasingly aligned with the SC tonal relations but the acoustic realizations of the JM tonal categories are less and less similar to their corresponding tonal categories in

SC. With the changes in society, the systematic correspondence mechanism has been strengthened and the surface tonal similarity has been reduced.

However, after the age-mediated effects were statistically removed from the individual backgrounds and different speaker-related predictors were modeled together, more complex interactions emerged. First, JM frequency showed age-independent effects. The more the speakers use JM, the less sensitive they are to the SC tonal relations compared with their peers. Second, the more highly-educated speakers and the more proficient speakers generally maintain a greater JM between-word pitch distance than their peers. Third, for the speakers who received their literacy education in JM, age increases their sensitivity to SC tonal relations and the pitch distances between JM and SC corresponding tonal categories increase. However, for the speakers who received their literacy education in SC, age reduces their sensitivity to SC tonal relations. This discontinuity in the age effect may be attributed to changes in education. The younger speakers received pinyin education and the oldest speakers received traditional philology training. Only the middle-aged speakers suffered from their deprivation of formal primary education. Taken together, with age-mediated effects excluded, the frequency of language usage and the educational experiences influence the bilingual individuals' sensitivity to systematic correspondence and the way they maintain the tonal contrasts.

The effects of cognitive background emerged after we focused on the two representative groups of speakers. The young speakers who received their literacy education in SC and completed higher education are only sensitive to JM frequency and did not show any age-independent effects from their cognitive backgrounds. Neither did they show any age effect. However, with the middle-aged speakers, whose aging has already started, we found significant effects of JM proficiency independent of age and JM frequency. The unmediated effect of JM proficiency may reflect some influence from their language aptitude. The middle-aged individuals who can maintain relatively higher JM proficiencies usually are also more responsive to the systematic correspondence across the two tonal systems. Similarly, the middle-aged speakers who maintain a better auditory working memory are also more sensitive to SC tonal relations compared with their peers. Individuals are shaped by their experiences when they are young. When cognitive aging happens, different individuals age at different speeds and in different ways, which endow their tonal systems with different colors.

The effects of individual backgrounds are in line with earlier studies of JM. Researchers have studied the age-related and style-related segmental variations in JM for more than 20 years (Cao, 1991; Qian, 1997), although they did not include the tonal variability in their discussion. Our findings are also in line with the basic principles of sociolinguistics, in that social contexts influence the choice of linguistic variants and speech performances within the language community (Labov, 2006; Weinreich et al.,

1968). However, most phonological and phonetic sociolinguistic studies focused on how individual backgrounds affect specific variants, while the current study explored how individual backgrounds affect the general mechanisms which control the alignment of two tonal systems. The cognitive effects are related to previous findings on bilinguals (Bialystok et al., 2008, 2014; Signorelli et al., 2011). However, the effect of auditory working memory on the strength of systematic correspondence is relatively new.

We did not focus on the acoustic variability [as shown in Fig. 1(a)] which can be handled via speech normalization and changing the parameters in the computational models (Leggetter and Woodland, 1995; Woodland, 2001). We managed to improve the analysis of tonal pattern variability [as shown in Fig. 1(b)] by including individual backgrounds. Note that the tonal pattern variability also exists within the same speaker, which cannot be directly handled by grouping the speakers or building different accent models (Huang et al., 2000; Woodland, 2001). The current results suggest that individual backgrounds perhaps do not work in a “yes-no” way in regulating the selection of tonal patterns. Individual backgrounds may instead affect the relative probability of different tonal patterns of the same word considering how the JM words are related to their SC counterparts.

Thus, with quantified data, we have shown that the strength of systematic correspondence between SC and JM disyllabic vocabularies varies with regularity, which is influenced by linguistic, cognitive, and sociolinguistic factors.

4.2. Limitations

The present study has several limitations. First, the study only involved isolated disyllabic words. Although the majority of modern Chinese words are disyllabic and the uncertainty of disyllabic tonal realizations has partly been explained by individual backgrounds, how disyllabic JM words are realized in connected speech needs further investigation. Researchers have done fruitful studies on contextual tonal realizations and their interactions with sentential prosodies (Chen, 2010; Chen and Gussenhoven, 2008; Xu, 1997, 1999; Xu and Prom-on, 2014; Xu and Wang, 2001). However, how to transfer this knowledge from SC to the other Chinese dialects and how the predictors we investigated work in context still open questions.

Second, the JM pronunciations were only obtained via a word reading task. Reading biases speakers towards a more formal and literal style. As a result, their frequency of some variants in our corpus may diverge from the real probability in everyday speech. Moreover, we do not know how some speakers can read the words in JM without being taught to read in JM. Without enough background knowledge on this question, we do not know the exact effect this has on their JM pronunciation.

Third, whether a rendition of a word was produced almost identically to its SC counterpart was manually marked. This information is necessary for incorporating

the phonological similarity mechanism in predicting the JM tonal realization. However, systematic correspondences make predictions in some cases: the JM neutral tone sandhi with high-level as the first citation tone usually results in sandhi forms almost identical to their SC counterpart; thus, we can predict that the SC words with low-rising + neutral tones are more likely to have identical counterparts in JM. Nevertheless, the rest of the cases are either conventional (most speakers do so for this word but other words from the same tonal categories in SC are not treated as the same) or incidental (only a few speakers do so). The rule-based between-dialect identity can be easily predicted and the conventional between-dialect identity (borrowing) needs to be marked manually. How to build a recognizer that can handle the incidental between-dialect identity (code-mixing) is still an open question.

Fourth, every part of one pitch contour received the same weight in the calculation of Euclidean distance. Euclidean distance serves as an objective measurement for the difference between two pitch contours. However, human perception is more complex. On the one hand, different parts of pitch contours contribute differently to speech perception (Whalen and Xu, 1992). On the other hand, tonal perception is incremental (Shen et al., 2013). Tones can be retrieved quite early (Lee, 2000) and different tones have different isolation points (Lai and Zhang, 2008), so that including the interaction between tonal category and the weights on different parts of the contour may benefit the modeling. Further studies may consider incorporating this knowledge into the weighting of different parts of the pitch contour.

Fifth, the onset condition was not taken into consideration. Previous studies (Chen, 2011; Howie, 1974; Xu and Xu, 2003) have shown that pitch contours carried by the rhymes are influenced by the onset condition. For instance, given two word pairs, both involving SC words from the same tonal categories, the pair with the same type of onset would be judged more similar in pitch contour than the pair with different onsets. This bias would also be applicable to the JM translation equivalents, since they are almost identical to their SC counterparts on the segmental level. This predictor can be taken into consideration in further studies.

Sixth, the Middle Chinese tonal system is known but not taken into consideration in this study. The major systematic discrepancy between SC and JM (and also between SC and other Mandarin dialects) comes from the different redistribution of the historical “Rusheng” (checked tone) category (Collective_work, 1989). This knowledge, although not applicable to other languages, can be potentially beneficial for the modeling of Chinese dialects.

Finally, the methodology used in this study involves two statistical issues. The usage of the distance matrix as the dependent variable for Linear Mixed Modeling introduced an inherent autocorrelation problem. As a result, the model may be over-confident. We carried out post-hoc analysis of model estimates, with the hope that they are more trustworthy for evaluating the effects of the predictors. Another

issue is residualization. Recent study suggests that residualization, counter to the common recognition in the literature, does not change the result for the residualized variable but the variable which is residualized against (Wurm and FisiCaro, 2014). This indicates that more attention should be given to the predictor age, against which we did the residualization. However, the interpretation of our findings remains the same. Note that the present study did not include any model with all the original aspects of individual backgrounds fitted together. Hence, neither did we make comparisons between such a model and the residualized models. Different aspects of individual backgrounds were first modeled alone and then integrated together in residualized models. It is because of this integration that the age-independent effects emerged, regardless of whether the models are residualized or not. Finally, of course, other methods are to be further explored in the future to address our research questions.

4.3. Applications

Using the systematic correspondence between the JM and SC tonal systems, with the between-dialect identity, JM neutral tone and tonal merging, and individual backgrounds under consideration, we have shown that JM lexical tone realizations can be predicted from the tonal categories of their SC counterparts. This finding has implications for both linguists and engineers. On the one hand, we have shown that individual backgrounds not only affect the selection of specific tonal variants but also contribute to how the general systematic correspondence mechanism takes effect. On the other hand, based on the present findings, further implicational models can more economically transfer SC tonal information in order to build pronunciation dictionaries for speech recognizers and synthesizers of JM. This approach can probably be generalized to other related northern Chinese dialects and help multi-accent/lingual modeling.

Acknowledgements

We would like to thank Prof. Shengli Cheng, Prof. Xiufang Du, Jia Li, Lulu Zhou, and the Dianliu Neighbourhood Committee for the recruitment of participants. We also would like to thank Martijn Wieling for his advice on statistics. J. Wu’s work was supported by a Ph.D. Studentship sponsored by Talent and Training China–Netherlands Program. We would like to thank the support to Yiya Chen from the European Research Council (ERC-Starting Grant 206198). The fieldtrip was sponsored by the Leiden University Foundation.

Appendix. A

Model estimates averaged over the speaker-wise models

| Predictors | Mean est. | Mean SE | Mean df | Mean t | Mean p |
|--|-----------|---------|-----------|----------|----------|
| (Intercept) | 3.89 | 0.16 | 602.35 | 24.96 | 0.00 |
| BDidentity one | 0.06 | 0.09 | 74845.14 | 2.20 | 0.09 |
| BDidentity one: WFreq both high | 0.01 | 0.09 | 74792.72 | 0.50 | 0.16 |
| BDidentity one: WFreq both low | -0.01 | 0.11 | 74813.59 | -0.21 | 0.15 |
| BDidentity both | 1.73 | 1.05 | 74771.55 | 2.01 | 0.23 |
| BDidentity both: WFreq both high | 0.02 | 0.57 | 74796.19 | 0.06 | 0.27 |
| BDidentity both: WFreq both low | 0.13 | 1.03 | 74785.24 | 0.04 | 0.34 |
| WFreq both high | -0.43 | 0.13 | 74884.92 | -3.52 | 0.08 |
| WFreq both low | 0.03 | 0.15 | 74894.17 | 0.45 | 0.20 |
| Merge neither | 5.59 | 0.09 | 74702.31 | 60.05 | 0.00 |
| Merge neither: WFreq both high | 0.36 | 0.13 | 75418.47 | 2.78 | 0.12 |
| Merge neither: WFreq both low | -0.07 | 0.15 | 75429.54 | -0.49 | 0.19 |
| Merge 2nd Syl | 4.02 | 0.13 | 73576.72 | 31.31 | 0.00 |
| Merge 2nd Syl: WFreq both high | 0.28 | 0.17 | 75414.59 | 1.63 | 0.22 |
| Merge 2nd Syl:WFreq both low | -0.02 | 0.18 | 75423.47 | -0.08 | 0.30 |
| Merge 1st Syl | 4.74 | 0.10 | 73180.79 | 46.85 | 0.00 |
| Merge 1st Syl: WFreq both high | 0.34 | 0.13 | 75418.58 | 2.55 | 0.13 |
| Merge 1st Syl: WFreq both low | -0.21 | 0.15 | 75428.50 | -1.40 | 0.24 |
| Merge both | 1.55 | 0.17 | 72791.94 | 9.30 | 0.03 |
| Merge both: WFreq both high | 0.13 | 0.23 | 75413.42 | 0.56 | 0.39 |
| Merge both: WFreq both low | -0.01 | 0.23 | 75420.54 | -0.03 | 0.35 |
| NtrlTone yes | 0.12 | 0.04 | 74035.05 | 2.81 | 0.08 |
| NtrlTone yes: BDidentity one | 1.42 | 0.11 | 74059.00 | 13.75 | 0.02 |
| NtrlTone yes: BDidentity one: WFreq both low | -0.31 | 0.16 | 76045.24 | -2.16 | 0.04 |
| NtrlTone yes: BDidentity both | 0.32 | 1.25 | 74153.33 | 0.66 | 0.11 |
| NtrlTone yes: BDidentity both: WFreq both high | 0.49 | 1.02 | 76032.87 | 0.51 | 0.47 |
| NtrlTone yes: BDidentity both: WFreq both low | -0.72 | 1.61 | 76026.85 | -0.50 | 0.45 |
| NtrlTone yes: WFreq both high | 0.06 | 0.06 | 74983.80 | 1.10 | 0.18 |
| NtrlTone yes:WFreq both low | -0.03 | 0.06 | 74987.41 | -0.35 | 0.18 |
| rel.SC1 same | -1.27 | 0.37 | 2886.25 | -4.31 | 0.11 |
| rel.SC1 same: BDidentity one | 0.36 | 0.14 | 73480.24 | 1.17 | 0.01 |
| rel.SC1 same: BDidentity one: WFreq both low | -0.17 | 0.18 | 73672.32 | -0.94 | 0.24 |
| rel.SC1 same: BDidentity both | -5.46 | 0.94 | 74592.06 | -10.58 | 0.10 |
| rel.SC1 same: BDidentity both: WFreq both high | -0.28 | 0.57 | 73636.59 | -0.32 | 0.19 |
| rel.SC1 same: BDidentity both: WFreq both low | 0.05 | 1.09 | 73617.88 | -0.01 | 0.27 |
| rel.SC1 same: WFreq both high | -0.01 | 0.08 | 74880.07 | 0.37 | 0.14 |
| rel.SC1 same: WFreq both low | -0.12 | 0.08 | 74883.00 | -2.54 | 0.12 |
| rel.SC1 same: Merge neither | -0.92 | 0.46 | 72838.08 | -2.20 | 0.26 |
| rel.SC1 same: Merge neither: WFreq both high | 1.95 | 0.65 | 78317.57 | 2.98 | 0.00 |
| rel.SC1 same: Merge neither: WFreq both low | -2.47 | 0.70 | 78314.53 | -3.52 | 0.00 |
| rel.SC1 same: Merge 2nd Syl | -1.65 | 0.48 | 72937.14 | -3.64 | 0.09 |
| rel.SC1 same: Merge 2nd Syl: WFreq both high | 2.51 | 0.68 | 78317.32 | 3.70 | 0.00 |
| rel.SC1 same: Merge 2nd Syl: WFreq both low | -3.51 | 0.72 | 78315.19 | -4.88 | 0.00 |
| rel.SC1 same: Merge 1st Syl | -1.39 | 0.47 | 72869.11 | -3.16 | 0.17 |
| rel.SC1 same: Merge 1st Syl: WFreq both high | 1.46 | 0.69 | 78318.53 | 2.13 | 0.03 |
| rel.SC1 same: Merge 1st Syl: WFreq both low | -2.55 | 0.75 | 78322.31 | -3.40 | 0.00 |
| rel.SC1 same: Merge both | 1.74 | 1.55 | 72825.42 | 1.21 | 0.33 |
| rel.SC1 same: Merge both: WFreq both high | -0.55 | 1.72 | 78323.32 | -0.32 | 0.75 |
| rel.SC1 same: Merge both: WFreq both low | -3.26 | 1.14 | 78312.09 | -2.86 | 0.00 |
| rel.SC1 same: NtrlTone yes | 1.65 | 0.08 | 73124.22 | 21.09 | 0.00 |
| rel.SC1 same: NtrlTone yes: BDidentity one | -3.00 | 0.19 | 71622.46 | -16.21 | 0.00 |
| rel.SC1 same: NtrlTone yes: BDidentity both | -6.10 | 1.47 | 72334.85 | -5.17 | 0.05 |

Appendix A (continued)

| Predictors | Mean est. | Mean SE | Mean <i>df</i> | Mean <i>t</i> | Mean <i>p</i> |
|--|-----------|---------|----------------|---------------|---------------|
| rel.SC1 same: NtrlTone yes: WFreq both high | -0.33 | 0.11 | 75338.28 | -3.14 | 0.05 |
| rel.SC1 same: NtrlTone yes: WFreq both low | 0.41 | 0.11 | 75343.70 | 3.82 | 0.06 |
| rel.SC1 same: rel.SC2 same | -1.51 | 0.75 | 8280.33 | -2.77 | 0.09 |
| rel.SC1 same: rel.SC2 same: BDidentity one | 1.94 | 0.25 | 67587.54 | 7.67 | 0.02 |
| rel.SC1 same: rel.SC2 same: BDidentity both | 2.46 | 0.73 | 65456.72 | 3.39 | 0.07 |
| rel.SC1 same: rel.SC2 same: WFreq both high | 0.17 | 0.13 | 75594.58 | 1.27 | 0.14 |
| rel.SC1 same: rel.SC2 same: WFreq both low | -0.42 | 0.13 | 75594.34 | -3.09 | 0.09 |
| rel.SC1 same: rel.SC2 same: Mergenn | -0.09 | 1.84 | 76552.13 | -0.03 | 0.64 |
| rel.SC1 same: rel.SC2 same: Merge 2nd Syl | 3.10 | 2.22 | 76564.02 | 1.57 | 0.22 |
| rel.SC1 same: rel.SC2 same: Merge 1st Syl | 2.70 | 1.87 | 76577.84 | 1.61 | 0.22 |
| rel.SC1 same: rel.SC2 same: Merge both | -0.44 | 2.72 | 76533.48 | -0.38 | 0.54 |
| rel.SC1 same: rel.SC2 same: NtrlTone yes | 0.56 | 0.18 | 68082.55 | 3.17 | 0.01 |
| rel.SC2 same | -2.47 | 0.33 | 488.72 | -7.68 | 0.00 |
| rel.SC2 same: BDidentity one | 0.08 | 0.14 | 73928.97 | 0.12 | 0.16 |
| rel.SC2 same: BDidentity one: WFreq both high | -0.42 | 0.16 | 75341.35 | -2.56 | 0.07 |
| rel.SC2 same: BDidentity one: WFreq both low | 0.36 | 0.20 | 75379.07 | 1.79 | 0.15 |
| rel.SC2 same: BDidentity both | -1.00 | 0.86 | 74652.89 | -1.89 | 0.24 |
| rel.SC2 same: BDidentity both: WFreq both high | -0.28 | 0.53 | 75378.10 | -0.51 | 0.48 |
| rel.SC2 same: BDidentity both: WFreq both low | 0.10 | 0.99 | 75208.04 | 0.08 | 0.37 |
| rel.SC2 same: WFreq both high | 0.15 | 0.12 | 74791.89 | 0.27 | 0.20 |
| rel.SC2 same: WFreq both low | -0.29 | 0.12 | 74793.66 | -1.95 | 0.14 |
| rel.SC2 same: Mergenn | 0.30 | 0.21 | 73864.77 | 1.32 | 0.14 |
| rel.SC2 same: Mergenn: WFreq both high | -0.87 | 0.40 | 73006.28 | -2.14 | 0.15 |
| rel.SC2 same: Mergenn: WFreq both low | 0.91 | 0.40 | 73005.69 | 2.30 | 0.04 |
| rel.SC2 same: Merge 2nd Syl | 2.52 | 0.83 | 73838.07 | 3.46 | 0.10 |
| rel.SC2 same: Merge 2nd Syl: WFreq both high | -0.74 | 0.98 | 73008.28 | -0.75 | 0.49 |
| rel.SC2 same: Merge 2nd Syl: WFreq both low | 1.12 | 0.70 | 73005.59 | 1.46 | 0.17 |
| rel.SC2 same: Merge 1st Syl | -0.64 | 0.23 | 73224.64 | -3.19 | 0.10 |
| rel.SC2 same: Merge 1st Syl: WFreq both high | -0.49 | 0.41 | 73008.19 | -1.17 | 0.42 |
| rel.SC2 same: Merge 1st Syl: WFreq both low | 0.38 | 0.40 | 73007.09 | 0.97 | 0.42 |
| rel.SC2 same: Merge both | 2.50 | 1.31 | 73802.31 | 2.70 | 0.20 |
| rel.SC2 same: Merge both: WFreq both high | -0.67 | 1.59 | 72997.62 | -0.44 | 0.63 |
| rel.SC2 same: Merge both: WFreq both low | 2.99 | 1.31 | 73001.22 | 2.14 | 0.12 |
| rel.SC2 same: NtrlTone yes | 2.61 | 0.09 | 73647.17 | 29.40 | 0.00 |
| rel.SC2 same: NtrlTone yes: BDidentity one | -0.78 | 0.21 | 72818.36 | -3.61 | 0.04 |
| rel.SC2 same: NtrlTone yes: BDidentity both | -2.87 | 1.74 | 74365.52 | -2.33 | 0.27 |
| rel.SC2 same: NtrlTone yes: WFreq both high | 0.01 | 0.12 | 74318.13 | -0.02 | 0.04 |
| rel.SC2 same: NtrlTone yes: WFreq both low | 0.07 | 0.13 | 74326.34 | 0.46 | 0.07 |

* BDidentity: Between-dialect identity [neither (baseline), one, or both]. Merge: Merge [the combination (baseline), neither, 2nd syllable, 1st syllable, or both syllables]. rel.SC1: Tonal relation on the first SC syllables [different (baseline) or same]. rel.SC2: Tonal relation on the second SC syllables [different (baseline) or same]. NtrlTone: Neutral tone [no (baseline) or yes]. WFreq: Word Frequency [different (baseline), both high, or both low].

Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.specom.2015.10.006>.

References

- Baayen, R.H., 2011. languageR: Data Sets and Functions with “Analyzing Linguistic Data: A Practical Introduction to Statistics” R Package Version 1.4 <<http://CRAN.R-project.org/package=languageR>>.
- Baayen, R.H., Feldman, L.B., Schreuder, R., 2006. Morphological influences on the recognition of monosyllabic monomorphemic words. *J. Mem. Lang.* 55, 290–313.
- Bates, D., Maechler, M., Bolker, B., Walker, S., 2013. lme4: Linear Mixed-Effects Models Using Eigen and S4 (Version R package 1.0-4) <<http://CRAN.R-project.org/package=lme4>>.
- Belsley, D.A., Kuh, E., Welsch, R.E., 2005. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. John Wiley & Sons.
- Besacier, L., Barnard, E., Karpov, A., Schultz, T., 2014. Automatic speech recognition for under-resourced languages: a survey. *Speech Commun.* 56, 85–100.

- Bialystok, E., Craik, F., Luk, G., 2008. Cognitive control and lexical access in younger and older bilinguals. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 859.
- Bialystok, E., Poarch, G., Luo, L., Craik, F.I.M., 2014. Effects of bilingualism and aging on executive function and working memory. *Psychol. Aging* 29, 696.
- Boersma, P., Weenink, D., 2001. Praat, a system for doing phonetics by computer. *Glott International* 1, 341–345.
- Breunig, M.M., Kriegel, H.-P., Ng, R.T., Sander, J., 2000. LOF: identifying density-based local outliers. *ACM Sigmod Record. ACM*, 93–104.
- Cai, Q., Brysbaert, M., 2010. SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE* 5, e10729.
- Cao, Z., 1991. Jinan Fangyan Ruogan Shengmu De Fenbu He Yanbian – Jinan Fangyan Dingliang Yanjiu Zhiyi (The distribution and evolution of certain initial consonants in Jinan, quantified study on Jinan I). *Lang. Study* 2.
- Carter, D., Deuchar, M., Davies, P., María del Carmen, P.t.C., 1991. A systematic comparison of factors affecting the choice of matrix language in three bilingual communities. *J. Language Contact* 4, 153–183.
- Chen, M., 1973. Cross-dialectal comparison: a case study and some theoretical considerations. *J. Chin. Ling.* 1, 38–63.
- Chen, Y., 2004. Focus and international phrase boundary in standard Chinese. *Chinese Spoken Language Processing, 2004 International Symposium on*. IEEE, pp. 41–44.
- Chen, Y., 2010. Post-focus F0 compression—Now you see it, now you don't. *J. Phonetics* 38, 517–525.
- Chen, Y., 2011. How does phonology guide phonetics in segment–f0 interaction? *J. Phonet.* 39, 612–625.
- Chen, Y., Gussenhoven, C., 2008. Emphasis and tonal implementation in Standard Chinese. *J. Phonetics* 36, 724–746.
- Chen, Y., Shen, R., Schiller, N.O., 2011. Representation of allophonic tone sandhi variants. In: *Proceedings of Psycholinguistics Representation of Tone, Satellite Workshop to ICPHS. Hongkong*, pp. 38–41.
- Chen, M., Wang, S.-Y.W., 1975. Sound change: actuation and implementation. *Language* 51, 255–281.
- Collective work, 1989. “Hanyu Fangyin Zihui” (Phonetic Dictionary of Chinese Dialects). Dept of Chinese Language and Literature, Peking University, Beijing.
- Dell, G.S., 1990. Effects of frequency and vocabulary type on phonological speech errors. *Lang. Cognitive. Proc.* 5, 313–349.
- Deza, M.M., Deza, E., 2009. *Euclidean Distance, Encyclopedia of Distances*. Springer, p. 94.
- Dyen, I., 1963. Why phonetic change is regular. *Language* 39, 631–637.
- Feng, S., 2001. Cong yunlu kan hanyu ciyu fenliu zhi dajie (Prosodically-determined distinction between word and phrase in Chinese). *Zhongguo Yuwen (Chinese Language)* (1), 27–37.
- Gathercole, S.E., Willis, C.S., Baddeley, A.D., Emslie, H., 1994. The children's test of nonword repetition: a test of phonological working memory. *Memory* 2, 103–127.
- Grainger, J., 1990. Word frequency and neighborhood frequency effects in lexical decision and naming. *J. Mem. Lang.* 29, 228–244.
- Hamed, M.B., 2005. Neighbour-nets portray the Chinese dialect continuum and the linguistic legacy of China's demic history. *Proc. R. Soc. B: Biol. Sci.* 272, 1015–1022.
- Harper, D., 2001. *Online Etymology Dictionary. Online Etymology Dictionary*. URL: <http://www.etymonline.com/index.php>.
- Howie, J.M., 1974. On the domain of tone in Mandarin. *Phonetica* 30, 129–148.
- Huang, C., Chang, E., Zhou, J., Lee, K.-F., 2000. Accent modeling based on pronunciation dictionary adaptation for large vocabulary Mandarin speech recognition. In: *Proc. Interspeech 2000*. Citeseer, pp. 818–821.
- Imseang, D., Motlicek, P., Boulard, H., Garner, P.N., 2014. Using out-of-language data to improve an under-resourced speech recognizer. *Speech Commun.* 56, 142–151.
- Iverson, P., Kuhl, P.K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., Siebert, C., 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87, B47–B57.
- Jaeger, T.F., 2010. Redundancy and reduction: speakers manage syntactic information density. *Cogn. Psychol.* 61, 23–62.
- Kamper, H., Muamba Mukanya, F.J., Niesler, T., 2012. Multi-accent acoustic modelling of South African English. *Sp. Commun.* 54, 801–813.
- Kuznetsova, A., Brockhoff, P.B., Christensen, R., 2013. lmerTest: tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package).
- Labov, W., 2006. A sociolinguistic perspective on sociophonetic research. *J. Phonet.* 34, 500–515.
- Lai, Y., Zhang, J., 2008. Mandarin Lexical Tone Recognition: The Gating Paradigm. *Kansas Working Papers in Linguistics*, 30.
- Lee, C.Y., 2000. Lexical tone in spoken word recognition: a view from Mandarin Chinese. *J. Acoust. Soc. Am.* 108, 2480.
- Leggetter, C.J., Woodland, P.C., 1995. Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models. *Comput. Speech Lang.* 9, 171–185.
- Levelt, W., 1999. Models of word production. *Trends Cogn Sci.* 3, 223–232.
- Li, R., 1988. *The Language Atlas of China*. Longman, London.
- Li, A., 2002. Chinese prosody and prosodic labeling of spontaneous speech. In: *Proc. International Conference on Speech Prosody 2002, Aix-en-Provence*, pp. 39–46.
- Li, Q., Chen, Y., 2016. An acoustic study of contextual tonal variation in Tianjin Mandarin. *J. Phonetics* 54, 123–150.
- Lobanov, B.M., 1971. Classification of Russian vowels spoken by different speakers. *The J. Acoust. Soc. Am.* 49, 606–608.
- Meillet, A., Ford, G.B., 1967. *The Comparative Method in Historical Linguistics*. H. Champion, Paris.
- Nixon, J.S., Chen, Y., Schiller, N.O., 2014. Multi-level processing of phonetic variants in speech production and visual word processing: evidence from Mandarin lexical tones. *Lang., Cognit. Neurosci.* 30, 491–505.
- Norman, J., 2003. The Chinese dialects: phonology. *Sino-Tibet. Lang.*, 72–83.
- Oldfield, R.C., Wingfield, A., 1965. Response latencies in naming objects. *Q. J. Exp. Psychol.* 17, 273–281.
- Peng, S.H., 2000. Lexical versus ‘phonological’ representations of Mandarin Sandhi tones. In: Broe, M.B., Pierrehumbert, J.B. (Eds.), *Papers in laboratory phonology*. Cambridge University Press, New York, pp. 152–167.
- Peng, G., Zhang, C., Zheng, H.-Y., Minett, J.W., Wang, W.S.-Y., 2012. The effect of intertalker variations on acoustic-perceptual mapping in Cantonese and Mandarin tone systems. *J. Speech. Lang. Hear. R.* 55, 579–595.
- Qian, Z.-Y., 1997. “Jinan Fangyan Cidian” (Jinan Dialect Dictionary). In: Li, R. (Ed.), “Xiandai hanyu fangyan da cidian” (Dictionary of Modern Chinese Dialects). Jiangsu Education Press, Nanjing.
- Qian, Y., Soong, F.K.-P., 2012. HMM-based bilingual (Mandarin–English) TTS techniques. Google Patents.
- Qian, Z.-Y., Zhu, G.-Q., 1998. “Jinanhua Yindang” (The Sound System of Jinan Dialect). Shanghai Educational Publishing House, Shanghai.
- R_Core_Team, 2013. R: A Language and Environment For Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, Version 2.15.
- Schultz, T., 2002. Globalphone: a multilingual speech and text database developed at Karlsruhe University. In: *Proc. Interspeech. Denver*, pp. 345–348.
- Shen, J., Deutsch, D., Rayner, K., 2013. On-line perception of Mandarin Tones 2 and 3: evidence from eye movements. *J. Acoust. Soc. Am.* 133, 3016.
- Shu, H., Peng, H., McBride-Chang, C., 2008. Phonological awareness in young Chinese children. *Develop. Sci.* 11, 171–181.
- Signorelli, T.M., Haarmann, H.J., Obler, L.K., 2011. Working memory in simultaneous interpreters: effects of task and age. *Int. J. Bilingual.* 16, 198–212.

- Sproat, R., Zheng, F., Gu, L., Jurafsky, D., Shafran, I., Li, J., Tsakalidis, S., Su, Y., Zheng, Y., Zhou, H., 2004. Dialectal Chinese Speech Recognition: Final Report, Johns Hopkins University, MD, Tech. Rep. Baltimore.
- Tang, C., van Heuven, V.J., 2009. Mutual intelligibility of Chinese dialects experimentally tested. *Lingua* 119, 709–732.
- Torgesen, J.K., Davis, C., 1996. Individual difference variables that predict response to training in phonological awareness. *J. Exp. Child Psychol.* 63, 1–21.
- Van Heerden, C., Kleynhans, N., Barnard, E., Davel, M., 2010. Pooling ASR data for closely related languages. Universiti Sains Malaysia, School of Computer Sciences.
- Wang, W.S.Y., 1969. Competing changes as a cause of residue. *Language* 45, 9–25.
- Weinreich, U., Labov, W., Herzog, M.I., 1968. *Empirical Foundations for a Theory of Language Change*. University of Texas Press, Austin.
- Whalen, D.H., Xu, Y., 1992. Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica* 49, 25–47.
- Woodland, P.C., 2001. Speaker adaptation for continuous density HMMs: a review. In: *ISCA Tutorial and Research Workshop (ITRW) on Adaptation Methods for Speech Recognition*, Sophia Antipolis, pp. 11–19.
- Wu, J., Chen, Y., 2014. Tonal variants in the bilingual mental lexicon. In: *The Fourth International Symposium on Tonal Aspects of Languages (TAL 2014)*. Nijmegen, pp. 152–155.
- Wu, J., Chen, Y., Van Heuven, V.J., Schiller, N.O., 2014. Tonal variability in lexical access. *Language, Cognition and Neuroscience* 29, 1317–1324.
- Wu, A., Jiang, Z., 2000. Statistically-enhanced new word identification in a rule-based Chinese system. In: *Proceedings of the Second Workshop on Chinese Language Processing: Held in Conjunction with the 38th Annual Meeting of the Association for Computational Linguistics*, vol. 12. Association for Computational Linguistics, pp. 46–51.
- Wurm, L.H., Fisiocar, S.A., 2014. What residualizing predictors in regression analyses does (and what it does not do). *J. Mem. Lang.* 72, 37–48.
- Xu, Y., 1994. Production and perception of coarticulated tones. *J. Acoust. Soc. Am.* 95, 2240–2253.
- Xu, Y., 1997. Contextual tonal variations in Mandarin. *J. Phonetics*. 25, 61–83.
- Xu, Y., Prom-on, S., 2014. Toward invariant functional representations of variable surface fundamental frequency contours: Synthesizing speech melody via model-based stochastic learning. *Speech. Commun.* 57, 181–208.
- Xu, Y., Wang, E.Q., 2001. Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech. Commun.* 33, 319–337.
- Xu, C.X., Xu, Y., 2003. Effects of consonant aspiration on Mandarin tones. *J. Int. Phonet. Assoc.* 33, 165–181.
- Yuan, J.H., Chen, Y.Y., 2014. 3rd tone sandhi in standard Chinese: a corpus approach. *J. Chin. Linguist.* 42, 218–237.