

Production of phonetic and phonological contrast by heritage speakers of  
Mandarin

Charles B. Chang<sup>a)</sup>

*University of Maryland,*

*College Park*

*Center for Advanced Study of Language*

*7005 52nd Avenue*

*College Park,*

*MD 20742 and*

*University of California,*

*Berkeley*

*Department of Linguistics*

*1203 Dwinelle Hall #2650*

*Berkeley,*

*CA 94720*

Yao Yao, Erin F. Haynes, and Russell Rhodes

*University of California,*

*Berkeley*

*Department of Linguistics*

*1203 Dwinelle Hall #2650*

*Berkeley,*

*CA 94720*

(Dated: February 28, 2011)

## Abstract

This study tested the hypothesis that heritage speakers of a minority language, due to their childhood experience with two languages, would outperform late learners in producing contrast: language-internal phonological contrast, as well as cross-linguistic phonetic contrast between similar, yet acoustically distinct, categories of different languages. To this end, production of Mandarin and English by heritage speakers of Mandarin was compared to that of native Mandarin speakers and native American English-speaking late learners of Mandarin in three experiments. In Experiment 1, back vowels in Mandarin and English were produced distinctly by all groups, but the greatest separation between similar vowels was achieved by heritage speakers. In Experiment 2, Mandarin aspirated and English voiceless plosives were produced distinctly by native Mandarin speakers and heritage speakers, who both put more distance between them than late learners. In Experiment 3, the Mandarin retroflex and English palato-alveolar fricatives were distinguished by more heritage speakers and late learners than native Mandarin speakers. Thus, overall the hypothesis was supported: across experiments, heritage speakers were found to be the most successful at simultaneously maintaining language-internal and cross-linguistic contrasts, a result that may stem from a close approximation of phonetic norms that occurs during early exposure to both languages.

PACS numbers: 43.70.Kv, 43.70.Fq, 43.70.Bk, 43.70.Ep

## I. INTRODUCTION

Although there exists a wide range of scholarship on the linguistic competence of child first-language (L1) and adult second-language (L2) acquirers, researchers have only begun to examine the linguistic knowledge of heritage-language speakers—that is, individuals whose current primary language differs from the language they spoke or only heard as a child (i.e. the heritage language, or HL). HL speakers are a group of interest because they often have a rich knowledge of their HL, even when they do not actively speak the language. Typical HL re-learners are predicted to have acquired “nearly 90% of the phonological system” and “80% to 90% of the grammatical rules” of the HL—a significantly more extensive command of the language than second-year college L2 learners (Campbell and Rosenthal, 2000, 167). Indeed, studies that have examined the phonological competence of HL speakers have found that childhood experience with a minority language, even if merely overhearing, can provide a significant boost to a speaker’s production and perception of that language later in life in comparison to L2 learners with no prior experience (Tees and Werker, 1984; Knightly *et al.*, 2003; Oh *et al.*, 2003). Similarly, studies that have examined the grammatical competence of HL speakers have found that they tend to be more native-like than L2 learners in their morphosyntax as well, although they nonetheless pattern differently from native speakers (Montrul, 2008; Au *et al.*, 2008; Polinsky, 2008). There seems to be something special about early linguistic experience acquired in childhood, and this point has been made especially clear in studies of HL phonology.

### A. Heritage-language phonology

Studies of HL phonology have been conducted on a number of languages, including Armenian (Godson, 2003, 2004), Korean (Au and Romo, 1997; Oh *et al.*, 2002, 2003; Au and Oh, 2009), Russian (Andrews, 1999), and Spanish (Au and Romo, 1997; Au *et al.*,

---

<sup>a)</sup>cbchang@cal.berkeley.edu

2002; Knightly *et al.*, 2003; Oh and Au, 2005; Au *et al.*, 2008), the majority of this research coming out of joint work by Au, Jun, Knightly, Oh, and Romo on HL speakers of Korean and Spanish. In their series of studies, which include acoustic measures such as voice onset time and degree of lenition, holistic measures such as overall accent ratings, and perceptual measures such as phoneme identification accuracy, the recurring theme is that HL speakers tend to have a phonological advantage over L2 learners. However, whether HL speakers show an advantage over L2 learners just in perception or in both perception and production of the HL seems to be related to the nature of their HL experience. In this regard, Au and colleagues have distinguished between “childhood hearers” and “childhood speakers”.

Knightly *et al.* (2003), for example, focused on childhood overhearers of Spanish—Spanish speakers who had regular childhood experience with overhearing Spanish, but not with speaking or being spoken to—and found that these childhood overhearers were measurably better than L2 learners at producing individual Spanish phonemes as well as whole Spanish narratives. Similarly, Oh *et al.* (2003) found that individuals with HL experience in Korean had a phonological advantage over L2 learners of Korean; however, they examined not only childhood hearers, but also childhood speakers who spoke Korean regularly during childhood. Comparing these two HL groups, they found that while childhood speakers were measurably more native-like than L2 learners in both perception and production of Korean, childhood hearers were more native-like than L2 learners only in perception. This discrepancy with the results of Knightly *et al.* (2003) was attributed to two possible factors: the difference in average duration of HL re-learning (longer in the case of the HL Spanish speakers) and the difference in complexity between the two contrasts examined (a two-way laryngeal contrast in Spanish between voiced and voiceless stops vs. a three-way laryngeal contrast in Korean among lenis, fortis, and aspirated stops/affricates). In short, the findings of Au and colleagues have suggested that previous HL speaking experience confers an advantage in both production and perception of the HL, and that previous HL listening experience confers an advantage in perception of the HL, even when this experience is limited to just the first year of life (Oh *et al.*, 2010).<sup>1</sup> The benefit conferred by HL listening

experience in production of the HL, however, appears to be mediated by additional factors.

Although studies on HL phonology have investigated the authenticity of HL speakers' production, few have explicitly examined the question of categorical merger—that is, whether HL speakers merge different sound categories rather than producing them distinctly. This question merits investigation even if only HL categories are considered, as phonological merger is commonly attested in cases of L1 attrition (Andersen, 1982; Campbell and Muntzel, 1989; Goodfellow, 2005), which bears a number of similarities to L2 and HL acquisition (see, e.g., Montrul 2008). Moreover, HL speakers' production of categories of the dominant language relative to those of the HL has yet to be fully addressed: though HL speakers may make all the phonological contrasts in each of their languages, do they also make phonetic contrasts across their two languages between similar, yet acoustically distinct phones? Suggestive results were obtained by Godson (2003, 2004), who found that HL speakers of Western Armenian showed some influence of English vowels in their pronunciation of the Armenian back vowels closest to English vowels, but this influence did not necessarily result in the merger of similar Armenian and English vowels.

## **B. Second-language phonology**

Whereas few HL studies have investigated the extent to which HL speakers produce cross-linguistic contrast between similar categories in their two languages, this question has long been a subject of inquiry in research on L2 speech and bilingual phonology (see, e.g., Flege, 1995; Laeuffer, 1996), a field which has been informed by two influential models: the Perceptual Assimilation Model (PAM; Best 1994) and the Speech Learning Model (SLM; Flege 1995). The PAM is applicable to the process of L2 phonological acquisition at its very beginning stages. Principally a model of non-native speech perception by naive listeners (i.e. those who have no knowledge of the non-native language), the PAM sets forth a typology of ways in which non-native speech contrasts may be interpreted by naive listeners relative to L1 phonological categories (so-called “perceptual assimilations”). The type of

perceptual assimilation that occurs with members of a non-native contrast predicts the degree of difficulty that learners will have with perceiving that contrast: if the members of the contrast are assimilated to different L1 categories, the contrast will be perceived accurately; if not, the contrast will be perceived less accurately, to a degree depending upon how equally well the members of the contrast are assimilated to the same L1 category. In a more recent version of this model, the PAM-L2 (Best and Tyler, 2007), the connection between non-native speech perception and L2 speech perception is made explicit. The PAM-L2 expands upon the PAM by incorporating the influence of an L2 learner’s developing phonetic and phonological knowledge of L2, thus allowing for perceptual assimilation at the gestural, phonetic, and phonological levels. The novel possibility of assimilation at the phonological level in particular is one of the features of this model that most distinguishes it from the SLM.

The SLM, in contrast to the PAM(-L2), is mainly a model of later stages of L2 speech acquisition, focusing on proficient bilinguals rather than novice learners. The model posits that phonetic categories are continually modified in response to sounds in another language that are identified with these categories. Furthermore, categories of L1 and L2 are said to exist in a common phonological space for bilinguals, who tend to keep them distinct under a general pressure to maintain contrast between different sounds. Central to the SLM is its account of inaccurate production of an L2 sound in terms of the recruitment of a similar L1 category. While a “new” L2 sound—one that has no clear parallel in L1—will motivate the formation of a new phonetic category, a “similar” L2 sound tends to undergo “equivalence classification” with a close L1 counterpart, a phenomenon that becomes increasingly likely as age of L2 learning increases. In this way, an L1 sound and an L2 sound may become linked to each other perceptually. A major way in which the SLM differs from the PAM—and the principal reason the SLM is more relevant to the present study—is that the SLM overtly addresses the connection to L2 production: perceptually linked L1 and L2 sounds are predicted to eventually approximate each other in production. At the same time, however, following from the notion of L1 and L2 sounds existing in the same phonological space, the model

allows for the possibility of an L2 category dissimilating from an L1 category for the sake of maintaining contrast between them. In other words, the existence of L1 and L2 sounds in a shared space may lead to either convergence or divergence between the sounds.

That similar L1 and L2 sounds undergo equivalence classification and influence each other in production has been demonstrated in a number of studies (e.g. Flege and Hillenbrand, 1984; Flege, 1987; Major, 1992; Sancier and Fowler, 1997). In one such investigation focusing on L2 speakers of English and French, Flege (1987) found that native English speakers who had learned French and native French speakers who had learned English produced French /u/ differently from monolingual native French speakers: both groups produced French /u/ with significantly higher second-formant values in approximation to the high second-formant norms of English /u/. Moreover, with regard to the realization of French /t/ (unaspirated) vs. English /t/ (aspirated), speakers did not typically reach the L2 phonetic norm for voicing lag and the L2 phonetic norm had an effect on their L1 /t/, such that both groups ended up over-aspirating French /t/ and under-aspirating English /t/. On the other hand, native English speakers' production of French /y/ (a "new" sound with no counterpart in English) was comparable to native French /y/.

### C. The present study

It remains to be seen whether this sort of subphonemic, bidirectional cross-linguistic influence or even, as alluded to above, categorical merger is found in HL speakers (individuals who, unlike typical adult L2 learners, received some degree of early exposure to both of their languages). Thus, the present study reexamined phonological production by HL speakers—in this case, HL speakers of Mandarin Chinese—in order to address three main questions. First, do HL speakers in fact reliably produce the phonological contrasts in both the HL and the dominant language? Second, do HL speakers produce phonetic contrasts between similar, yet acoustically distinct categories in their two languages? Third, how do HL speakers compare to native speakers and late learners in their production of phonetic and

phonological contrast? Like previous studies on HL phonology, one objective of the current study was to see how closely HL Mandarin speakers would pattern with native speakers versus late learners with respect to production of language-internal phonemic contrasts. However, unlike previous HL studies, another objective was to see how HL Mandarin speakers would compare to these other groups in production of cross-linguistic contrast between similar categories in their two languages (Mandarin and English).

Another way in which the current study differed from previous HL studies was in the treatment of variability among HL speakers. Though HL speakers have been noted to outpace novice L2 learners of a language in a number of ways, the population of language users referred to as HL speakers has also been noted to be an extremely heterogeneous group. Li and Duff (2008, 17), for instance, note of Chinese HL speakers that “even within a proficiency-defined ‘HL’ group, learners generally have a very uneven grasp of the HL, falling along a continuum of having very little HL knowledge to being highly proficient.” In this study, the heterogeneity of the HL group—rather than being artificially reduced via the detailed sort of screening of participants used in previous studies—was instead accepted as representative of the larger population under study. The only requirement for HL speakers to be included in the current HL speaker sample (see Section II.A) was that their primary HL experience was with Mandarin, as opposed to another variety of Chinese. Although inclusion of a wider spectrum of HL experience than examined in previous studies increased the probability of inter-speaker variation within the HL group (and, thus, the probability of obtaining null results), it also served to maximize the generalizability of the results, which emerged in spite of the variability purposefully left within this HL speaker sample. In other words, the main findings are expected to be robust and reproducible with a different pseudorandom sampling of HL Mandarin speakers.

The research questions in this study were addressed via an acoustic investigation of American HL speakers’ production of Mandarin Chinese and American English. The production of both Mandarin and English phonological categories by HL speakers of Mandarin (whose dominant language was English) was compared to that of native L1 speakers of



Mandarin (who were late L2 learners of English) and late L2 learners of Mandarin (who were native L1 speakers of English) in a series of experiments designed to investigate the realization of three different types of phonemic categories: vowel quality categories, plosive voicing (i.e. laryngeal) categories, and fricative place categories. These categories in particular were examined for two reasons. On the one hand, focusing on vowel quality and laryngeal categories facilitated comparison with previous studies on HL and L2 phonology, as both of these category types have figured prominently in earlier work (e.g. Flege and Hillenbrand, 1984; Flege, 1987; Godson, 2003, 2004; Knightly *et al.*, 2003; Oh *et al.*, 2003); on the other hand, extending the domain of inquiry to consonantal place of articulation categories allowed for an examination of whether previous findings would generalize to new dimensions of phonological contrast that have not yet been examined in this regard. As the study was concerned with the production of similar categories, the specific categories chosen for investigation mostly comprised pairs of similar Mandarin and English categories that stood to be identified with each other: rounded vowels (Mandarin and English /o<sup>u</sup>, u/, Mandarin /y/), short-lag stops (Mandarin unaspirated, English voiced), long-lag stops (Mandarin aspirated, English voiceless), and post-alveolar fricatives (Mandarin retroflex /ʂ/ and alveolo-palatal /ç/, English palato-alveolar /ʃ/). The acoustic data comprised measurements of formant resonances (Section III.A), voice onset time (Section III.B), and spectral features such as center of gravity, or centroid (Section III.C). These measurements, as well as the phonetic norms against which they were compared, are described in more detail in Sections II.D and III.A–III.C.

There is little literature on L1 Mandarin speakers’ production of L2 English segmentals as opposed to prosody (e.g. Zhang *et al.*, 2008) and even less literature on L1 English speakers’ production of L2 Mandarin segmentals that offers predictions regarding the sort of patterns one might expect to find in the current study of cross-language production in Mandarin and English. The few studies that have examined Mandarin speakers’ production of English vowels (Jia *et al.*, 2006; Jiang, 2008, 2010) have generally examined accuracy via listener ratings rather than acoustic analysis, although some acoustic data from Chen

(2006) suggest that Mandarin speakers produce English /u/ with second-formant values lower than those of native English speakers, in keeping with the respective phonetic norms of Mandarin and English back rounded vowels (Section III.A). As for Mandarin speakers' production of English laryngeal contrast, Zhang and Yin (2009, 144) noted in a review article that due to the different features at work in the two languages (voicing in English, aspiration in Mandarin), "Chinese learners of English often neglect the differences between voiced and voiceless sounds in English"; however, this statement was not about plosives specifically, and no data were presented in support of this claim. No known studies exist on Mandarin speakers' production of English fricatives or on English speakers' production of the corresponding Mandarin segmentals. In sum, previous studies offer little in the way of specific predictions regarding cross-language production in this study, though the data that do exist suggest that in L2 production there is some influence of the phonetic norms for similar L1 categories, as has been found in many studies of L2 speech including that of Flege (1987).

Specific hypotheses regarding the research questions follow from the principles of the SLM. Since equivalence classification and concomitant linking of similar L1 and L2 categories is thought to occur more often with increasing age of L2 learning, it was hypothesized that heritage speakers of a minority language, due to their early childhood experience with two languages (the dominant language and the heritage language), would outperform late L2 learners in producing contrast between distinct sounds: language-internal phonological contrast between phonemic categories, as well as cross-linguistic phonetic contrast between similar, yet acoustically distinct categories of different languages. On the other hand, HL speakers and L2 learners were both predicted to do well with producing HL/L2 categories that are substantially different from those of the dominant language. Thus, the results were expected to show HL speakers producing contrast between relatively similar categories (e.g. Mandarin /ʂ/ and /ʃ/; Mandarin /o<sup>u</sup>/ and English /o<sup>u</sup>/) more often and more effectively than L2 learners, but not necessarily producing significantly more accurately than L2 learners those categories that would be, for the L2 learner, new categories vis-à-vis L1 (e.g.

Mandarin /y/).

As for the manner in which L2 sounds are deemed “similar” to L1 sounds, it is argued that the PAM-L2 offers the most comprehensive view of how these equivalences may arise. As discussed above, this model (unlike the SLM) is explicit about allowing category equivalence to be based on gestural, phonetic, and phonological considerations, rather than strictly phonetic proximity. Given the abundant evidence that has been found for the role played by phonological information in determining cross-language category equivalence in loanword adaptation (see, e.g., LaCharité and Paradis, 2005; Kang, 2008; Chang, 2009), it follows that phonological information is indeed likely to play an important role in determining category equivalence in L2 acquisition. However, there may be cases where the phonological level conflicts with the phonetic level and/or gestural level with respect to cross-language proximity between categories, and the PAM-L2 does not indicate which level prevails in these cases. In fact, the current study concerned one such case, where phonological considerations are at odds with phonetic ones. Experiment 1 examined two Mandarin high rounded vowels, back /u/ and front /y/, which are each similar to English /u/, but in different ways; consequently, it is unclear which of these vowels should be considered “similar” and liable to be linked to English /u/ in perception/production. When only acoustic measures of vowel quality are considered, English /u/ is more similar to Mandarin /y/ (which is on the order of 3 Bark away from English /u/ in second-formant frequency; see Table I) than to Mandarin /u/ (which is twice as far away). When the phonological statuses of these vowels are considered, however, Mandarin /u/ emerges as the clear counterpart of English /u/, since they both function in their respective vowel inventory as high back rounded vowels. Due to this ambiguity in cross-language proximity, both possible vowel equivalences (English /u/–Mandarin /u/, English /u/–Mandarin /y/) were considered in this study. However, it was predicted that, as with French /y/ in Flege (1987), Mandarin /y/ would constitute a “new” vowel for L2 learners—on the basis of its phonological, rather than phonetic, deviance from English /u/—and, thus, that it would be produced relatively accurately. This point is further discussed in Section IV.

The paper is organized as follows. Section II provides an overview of the characteristics of the speakers who participated in the study, the procedure and stimuli used in the experiments, and the acoustic analyses conducted on participants' productions. Section III presents the results of each of the three experiments (Experiment 1: vowel categories, Experiment 2: laryngeal categories, Experiment 3: fricative categories), and finally, Section IV discusses the findings in light of the hypotheses discussed above.

## II. METHODS

### A. Participants

A total of 28 Mandarin speakers and learners participated, with two excluded from the final analysis due to language backgrounds inconsistent with the focus of the study. All were recruited at the University of California, Berkeley, and paid for a single session that encompassed all three experiments. Participants who were included in the analysis comprised fifteen females and eleven males ranging in age from 18 to 40 years, none of whom reported any history of speech or hearing impairments. They were each presented with the same set of stimuli, described in Section II.C.

Demographic information about all participants is presented in Appendix A, listing each participant's identifier (PID), gender, age at the time of the study (in years), place of birth or residential history (including ages), and where applicable: age of arrival to the U.S. (in years), other languages spoken or exposed to at home, frequency of current Mandarin use, and general experience with Mandarin (including the ages at which the experience occurred). For the purposes of analysis, participants were divided into groups according to their responses on a detailed questionnaire about their life history and family background, language background, current language use, formal language education, and Mandarin proficiency.<sup>2</sup> If participants had not received exposure to Mandarin until the age of 18 years, they were classified as late L2 learners. If, on the other hand, they were born and schooled in a Mandarin-speaking region, reported their current Mandarin proficiency

level to be native-like, and judged Mandarin to be their best language, they were classified as native Mandarin speakers. Anyone with prior Mandarin experience in the home who did not fulfill all of the criteria for the native speaker group was classified as an HL speaker.

HL speakers were divided into high- and low-exposure groups using self-reported frequency of current Mandarin use as the primary consideration and the number of years lived in a Mandarin-speaking region as a secondary consideration. For HL speakers, there was a general trend for people who had lived longer in Mandarin-speaking regions to also use Mandarin more often with their family. Three exceptional cases were participants H9, H13, and H20. Participant H9, who had visited Mandarin-speaking regions many times but was born and educated entirely in the U.S., reported extensive use of Mandarin in her family, both with parents and siblings and with other relatives. Participant H13, on the other hand, did not come to the U.S. until she was 10 years old, yet reported using Mandarin only half of the time with parents and not with anyone else currently. Finally, participant H20 was born and spent the first two years of life in China, but Mandarin was spoken to her at home only by her nanny; though her father was also a Mandarin speaker, both parents spoke to her in English, and she only started to hear and speak Mandarin again when taking a Chinese language class during the semester of recording. These three participants were divided into the high- and low-exposure groups by simultaneously considering both their current use of the language and the amount of time they had spent in Mandarin-speaking areas. In general, however, HL speakers were put into the high-exposure group if they reported using Mandarin at home more than half of the time and into the low-exposure group if they reported using Mandarin at home half of the time or less.

The participants in each resulting group possessed several shared background characteristics. The six participants (four female, two male; mean age 29.8 years) in the native Mandarin (NM) group were all native Mandarin speakers who were born and educated (up to at least seventh grade) in mainland China or Taiwan. The fifteen HL speaker participants reported speaking English most of the time overall, but they were all born to Mandarin-speaking parents. Generally speaking, the nine participants (four female, five male; mean

age 21.0 years) in the high-exposure (HE) HL group were heritage speakers who had extensive exposure to Mandarin as children and reported using Mandarin to communicate with both parents most or all of the time. Most of the HE participants were either born in a Mandarin-speaking region (H8, H10, H11, H12, H13), with a mean age of arrival to the U.S. of 6.9 years, or had otherwise lived for a number of years in a Mandarin-speaking region. In contrast, the six participants (four female, two male; mean age 20.0 years) in the low-exposure (LE) HL group were heritage speakers who had limited exposure to the language and reported using Mandarin with their parents half of the time or less. With the exception of H20, all of the LE participants were born in the U.S. and had never lived in a Mandarin-speaking region. The five participants (three female, two male; mean age 21.6 years) in the second-language (L2) learner group were native English speakers who were born and educated in the U.S., grew up in English-speaking families, and started to learn Mandarin after the age of 18. Three (L22, L23, L24) grew up in a monolingual home environment, while the other two (L25, L26) had some degree of exposure to other languages as well. With the exception of L26, all had received formal Mandarin language instruction, ranging in duration from three months in an immersion environment to two years in an American university setting. Nearly all reported their current Mandarin proficiency to be quite poor and estimated that they understood 10–25% of normal conversational Mandarin; the exception is L25, who had received the most formal instruction and reported understanding 30–50% of conversational Mandarin.

## **B. Procedure**

Study participants were recorded reading aloud 59 Mandarin items and 32 English items presented via individual index cards in random order by language. Each language block was repeated four times in a single session for a total of 364 tokens in all. Participants completed all blocks in one language before moving on to the second language, with the order of the languages (Mandarin-English or English-Mandarin) balanced across participants. English

words were written in English orthography, and Mandarin words in Mandarin orthography (traditional or simplified characters) and phonetic spelling (pinyin, the spelling system used in mainland China, and/or zhuyin/Bopomofo, the spelling system used in Taiwan). The recordings were made in a sound-proof booth with 48-kHz sampling and 16-bit resolution using either a Marantz PMD660 solid-state recorder and an AKG C420 head-mounted condenser microphone, or a Dell desktop computer connected to an M-AUDIO Mobile-Pre USB preamp audio interface and an AKG C520 head-mounted condenser microphone.

### C. Stimuli

In choosing Mandarin and English stimuli, segmental context was matched across language as much as possible, and Mandarin items with falling tones were selected (when such words existed) so as to make the pitch contour of the Mandarin items maximally similar to the falling pitch contour of English words spoken in isolation (e.g. English *boot* [bʊt] vs. Mandarin 不 [pu˥] ‘not’; English *tote* [tʰoʊt] vs. Mandarin 透 [tʰoʊ˥] ‘transparent’; English *shot* [ʃat] vs. Mandarin 煞 [ʃa˥] ‘suddenly’ and 下 [ca˥] ‘below’). In addition, the most common character corresponding to the phonological shape of each Mandarin item was selected, minimizing the possibility of participants being unfamiliar with any of the items. In the end, the stimulus items chosen were all common for native speakers of that language, although L2 learners—particularly the L2 Mandarin learners, who had relatively little experience with Mandarin—were not necessarily familiar with all of the items. Nonetheless, because multiple sources of information were provided about each item, participants were able to complete the task described in Section II.B with little trouble.

The speech stimuli used in Experiments 1–3 are presented in Appendix B. Critical stimuli (i.e. the non-filler items subjected to acoustic analysis; see Section II.D) were generally of the form CV in the case of Mandarin and of the form CVC in the case of English. In Experiment 1, critical stimuli contained one of five rounded vowel categories: Mandarin /u/ appeared in ten items, Mandarin /oʊ/ in seven, Mandarin /y/ in three, English /u/

in eleven, and English /o<sup>u</sup>/ in ten. With the exception of Mandarin /y/, which is phonotactically restricted to coronal contexts, all of these vowels occurred following onsets of several different places of articulation and laryngeal types. In Experiment 2, critical stimuli contained a word-initial plosive of one of four laryngeal categories and three places of articulation: Mandarin unaspirated /p, t, k/, Mandarin aspirated /p<sup>h</sup>, t<sup>h</sup>, k<sup>h</sup>/, English voiced /b, d, g/, and English voiceless /p, t, k/. With one exception (due to the absence of /po<sup>u</sup>/ in Mandarin), all of these plosives preceded back rounded vowels. There were two items per combination of laryngeal category and place of articulation, for a total of twelve Mandarin items and twelve English items. In Experiment 3, critical stimuli contained one of three post-alveolar sibilant fricatives: Mandarin retroflex /ʂ/, Mandarin alveolo-palatal /ç/, and English palato-alveolar /ʃ/. These fricatives appeared prevocally in seven Mandarin items and two English items. All occurred in a low vowel context.

#### D. Acoustic analysis

All acoustic measurements were taken manually in Praat (Boersma and Weenink, 2008) using a 5-ms analysis window and 50-dB dynamic range. In Experiment 1, vowel quality was analyzed by measuring average values of the first ( $F_1$ ) and second ( $F_2$ ) formants (Ladefoged, 2005, 40–43) over the whole duration of the vowel, from the beginning of the first glottal pulse to the end of the last visible glottal pulse (Mandarin tokens) or the beginning of the final consonant constriction (English tokens). In Experiment 2, voicing lag in word-initial plosives was analyzed by measuring voice onset time (VOT) as time at the onset of periodicity minus time at plosive release (Lisker and Abramson, 1964; Ladefoged, 2003, 96–101). In Experiment 3, peak amplitude frequency (PAF) and centroid frequency (Ladefoged, 2003, 156–158) were measured over an average spectrum of the middle 100 ms of the fricative. A low-frequency stop-band filter was applied to this spectrum to remove frequencies from 0 up to the  $F_2$  region (so as to get a better measure of specifically front cavity resonances varying with place of articulation). The location of the  $F_2$  region (the endpoint of the band



filter) was estimated for each subject as  $\frac{3}{5}$  of the speaker’s average third formant in the vowel /a/ (Li *et al.*, 2007). Frequency measurements were later converted to Bark for an acoustic perceptual view of participants’ vowel and fricative productions using the following formula from Traunmüller (1990):  $z = [26.81/(1 + 1960/f)] - 0.53$ .

To ensure that measurements were reliable, 25% of the measurements from each experiment were double-checked by a second researcher in a pseudorandom fashion. Any discrepancy between the two researchers’ measurements in excess of 100 Hz (for formants, PAFs, and centroids) or 5 ms (for VOT) was checked again by a third researcher. In Experiment 1, 8% of formant measurements were triple-checked in this fashion. Final calculations of the differences between researchers’ measurements here revealed an average difference of 13 Hz in  $F_1$  measurements (81% were less than 25 Hz apart) and 24 Hz in  $F_2$  measurements (63% were less than 25 Hz apart). In Experiment 2, 9% of VOT measurements were triple-checked, with an average difference of 1.4 ms between different researchers’ measurements. In Experiment 3, 3% of the measurements were triple-checked. There was an average difference of 12 Hz in PAF measurements (72% were less than 25 Hz apart) and 33 Hz in centroid measurements (41% were less than 25 Hz apart). If after a third measurement there still remained a discrepancy between different researchers’ measurements of greater than 100 Hz/5 ms, all of these measurements were discarded; however, this resulted in the discarding of less than 1% of the total number of measurements.

### III. RESULTS

#### A. Experiment 1: vowels

On the basis of relative acoustic phonetic similarity as well as place in the relevant vowel inventory, the “similar” vowels compared to each other were Mandarin /o<sup>u</sup>/–English /o<sup>u</sup>/ and Mandarin /u/–English /u/, while Mandarin /y/ was predicted to constitute a “new” vowel vis-à-vis English. Differences between formant norms for the Mandarin and English vowels under study are summarized in Table I, converted to Bark according to the formula

given in Section II.D (Mandarin figures from Wu and Lin 1989 and Lin and Wang 1992, English figures from Hagiwara 1997). On average, Mandarin /u/ and English /u/ are quite similar in  $F_1$ , but differ substantially in  $F_2$ : the average  $F_2$  for English /u/ is approximately 6 Bark higher than that of Mandarin /u/ for both male and female speakers. On the other hand, Mandarin /o<sup>u</sup>/ and English /o<sup>u</sup>/ differ in both  $F_1$  and  $F_2$ , English /o<sup>u</sup>/ being 1–1.5 Bark lower in  $F_1$  and approximately 2.5–3 Bark higher in  $F_2$ . Mandarin /y/ is similar to English /u/ in  $F_1$ , but approximately 3 Bark higher in  $F_2$ . Thus, if speakers with experience in both languages closely approximate these phonetic norms, they are expected to produce a slight difference in  $F_1$  and a substantial difference in  $F_2$  between the two mid vowels, as well as a large difference in  $F_2$  between the two high vowels. Furthermore, they are expected to produce the front vowel /y/ with the highest  $F_2$  of all.

Mean  $F_1$  and  $F_2$  in participants’ productions of the mid rounded vowels are plotted in Figure 1. For each group the Mandarin and English vowels occupied distinct phonetic spaces, English /o<sup>u</sup>/ being produced with higher  $F_2$  values than Mandarin /o<sup>u</sup>/ . The NM and L2 groups each produced the /o<sup>u</sup>/ of their non-native language with  $F_2$  values approximating the /o<sup>u</sup>/ of their native language, while HL speakers patterned somewhat in between these two groups. For example, in the case of Mandarin /o<sup>u</sup>/ , most NM speakers had lower  $F_2$  values of approximately 8.0–8.5 Bark, whereas most L2 learners had higher  $F_2$  values of approximately 8.6–9.0 Bark. The majority of HE speakers were located in the same region as NM speakers, while the majority of LE speakers were located in the same region as L2 speakers; both these groups, however, spanned a wide phonetic space that extended across the regions occupied by NM and L2 speakers. Figure 1 also shows some differentiation of the two vowels in terms of  $F_1$ . In accordance with the small difference in native  $F_1$  norms seen in Table I, for all speaker groups the space for Mandarin /o<sup>u</sup>/ extended into a higher  $F_1$  region than the space for English /o<sup>u</sup>/ .

Mean  $F_1$  and  $F_2$  in participants’ productions of the high rounded vowels are plotted in Figure 2. There are several patterns of note here. First, all groups distinguished Mandarin /u/ and English /u/, producing the latter with substantially higher  $F_2$  values. However,

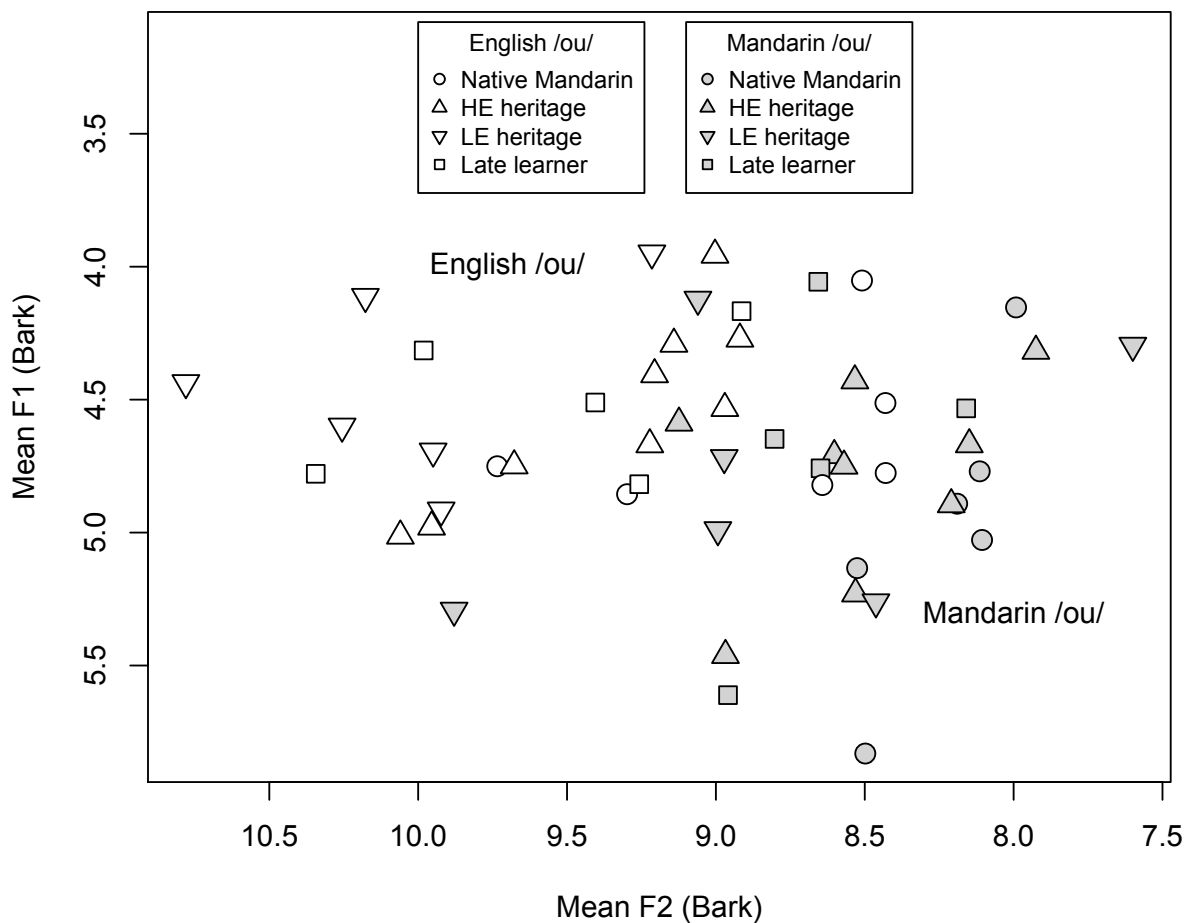


FIG. 1. Bark plot of the first two formants in mean productions of Mandarin / $o^u$ / (gray symbols) and English / $o^u$ / (white symbols). NM speakers are plotted in circles, HE speakers in triangles, LE speakers in upside-down triangles, and L2 learners in squares.

the groups differed in terms of their location in  $F_1$ - $F_2$  space. NM speakers produced both vowels with the lowest  $F_2$  values, while L2 learners (native English speakers) produced both vowels with the highest  $F_2$  values, with HL speakers located somewhat in between these two groups for both vowels. Thus, similar to the case of / $o^u$ /, both NM speakers and L2 learners appeared to be influenced in their pronunciation of the / $u$ / of their second language by the phonetic characteristics of the / $u$ / of their first language: NM speakers produced

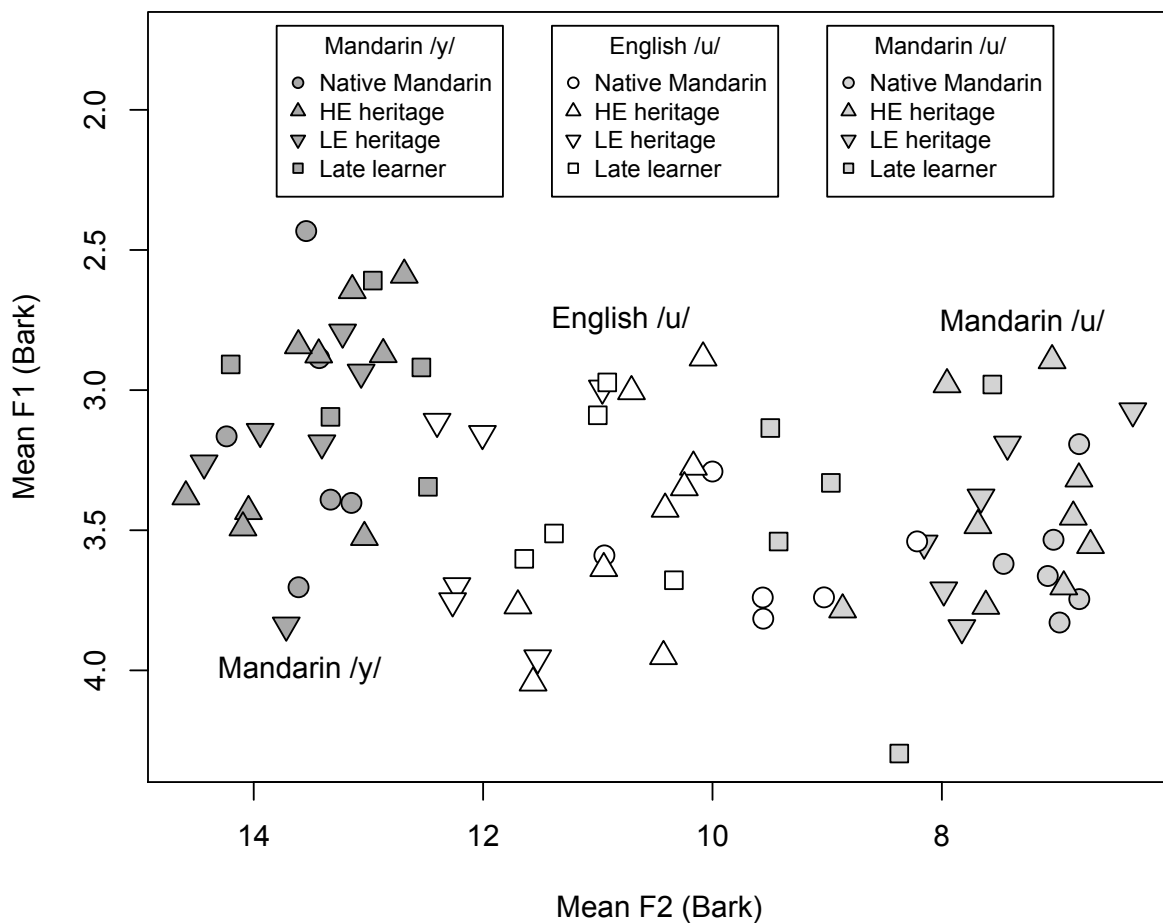


FIG. 2. Bark plot of the first two formants in mean productions of Mandarin /u/ (light gray symbols), English /u/ (white symbols), and Mandarin /y/ (dark gray symbols). NM speakers are plotted in circles, HE speakers in triangles, LE speakers in upside-down triangles, and L2 learners in squares.

English /u/ with a relatively low  $F_2$  approximating the low  $F_2$  of Mandarin /u/, whereas L2 learners produced Mandarin /u/ with a relatively high  $F_2$  approximating the high  $F_2$  of English /u/. On the other hand, HL speakers generally produced Mandarin /u/ and English /u/ with  $F_2$  values that were relatively close to native values. To put it another way, for most HL speakers  $F_2$  for Mandarin /u/ was not as high as it was for L2 learners, nor was

$F_2$  for English /u/ as low as it was for NM speakers.

With regard to the Mandarin high front rounded vowel /y/, all groups produced this vowel in a distinct phonetic space with much higher  $F_2$  values than the Mandarin and English back vowels, and the groups did not differ from each other appreciably with respect to their location in  $F_1$ - $F_2$  space. The results of a two-way analysis of variance (with factors Group and Gender<sup>3</sup>) were consistent with this impression. There was a main effect of Gender on  $F_1$  [ $F(1, 18) = 16.27, p < 0.001$ ] as well as  $F_2$  [ $F(1, 18) = 14.79, p < 0.01$ ], but no main effect of Group on either  $F_1$  or  $F_2$  and no interaction with Gender. In other words, although men and women (unsurprisingly) had different formants for /y/, L2 learners and HL speakers did not differ statistically from NM speakers in their production of Mandarin /y/ as it was measured here.

Formant data for the mid and back rounded vowels were subjected to mixed-model analyses of variance (ANOVAs), with Group and Gender as between-subjects factors and Language, Vowel (/o<sup>u</sup>/ or /u/), and Place<sup>4</sup> (of articulation of the onset consonant) as within-subjects factors. With respect to  $F_1$ , there was no main effect of Language, but there were highly significant main effects of Vowel [ $F(1, 5) = 563.58, p < 0.001$ ], Place [ $F(4, 51) = 115.17, p < 0.001$ ], and Gender [ $F(1, 5) = 20.28, p < 0.01$ ], as expected: /o<sup>u</sup>/ > /u/; velar > alveolar > bilabial > glottal/post-alveolar; and female > male. As one would predict from the formant norms cited in Table I, there was also a two-way interaction between Language and Vowel [ $F(1, 5) = 73.69, p < 0.001$ ], attributable to only Mandarin /u/ and English /u/ not being produced with distinct  $F_1$  values. Males were more successful than females at producing an  $F_1$  difference between the high back vowels, resulting in a three-way interaction of Gender, Language, and Vowel [ $F(1, 5) = 10.99, p < 0.05$ ]. However, Group did not have a main effect on  $F_1$ , nor did it interact significantly with any other factors. In short, though the various vowels were overall produced differently in terms of  $F_1$ , which was moreover affected by the consonantal context in which the vowels occurred, the participant groups did not differ from each other statistically with respect to production of  $F_1$ .

While there was no main effect of Group on  $F_1$ , there was a main effect of Group

on  $F_2$  [ $F(3, 4) = 11.24, p < 0.05$ ], though no main effect of Gender. There were also highly significant main effects of Language [ $F(1, 4) = 704.62, p < 0.001$ ] and Place [ $F(4, 48) = 315.82, p < 0.001$ ]: English > Mandarin, and post-alveolar > alveolar > velar > glottal/bilabial. Though the effect of Vowel was only marginally significant, a two-way interaction between Language and Vowel [ $F(1, 4) = 316.25, p < 0.001$ ] arose from the greater effect of Language on  $F_2$  in the case of /u/ than in the case of /o<sup>u</sup>/. Group not only had a main effect on  $F_2$ , it also interacted with Language [ $F(3, 4) = 11.20, p < 0.05$ ] and with Language and Vowel [ $F(3, 4) = 15.93, p < 0.05$ ]. The Group x Language interaction was attributable to the pattern seen in Figures 1–2: English back vowels were produced with greater  $F_2$  values than Mandarin back vowels in all groups, but this language effect differed across the groups, which produced disparate  $F_2$  values and unequal distances between languages. The Group x Language x Vowel interaction arose from the fact that the Group x Language interaction was more pronounced for /u/ than for /o<sup>u</sup>/.<sup>5</sup> In sum, the  $F_2$  results contrasted with the  $F_1$  results in two main ways: the English vowels were found overall to be produced with higher  $F_2$  values than the Mandarin vowels, and the participant groups (but not the genders) differed from each other significantly in  $F_2$  production.

To examine between-group differences in the realization of cross-linguistic contrasts between similar vowel categories, the mean differences in  $F_1$  and  $F_2$  produced between corresponding back vowels (Mandarin and English /o<sup>u</sup>/, Mandarin and English /u/) were calculated for each participant. One-way ANOVAs showed a highly significant main effect of Group on  $F_2$  distances between Mandarin and English /u/ [ $F(3, 22) = 7.85, p < 0.001$ ], but not on  $F_2$  distances between Mandarin and English /o<sup>u</sup>/. These mean  $F_2$  distances are presented in Figure 3, where it can be seen that both the HE group and the LE group put more acoustic distance between Mandarin and English /u/ than did the L2 group [HE vs. L2: Mann–Whitney  $U = 38, n_1 = 9, n_2 = 5, p < 0.05$  two-tailed; LE vs. L2: Mann–Whitney  $U = 30, n_1 = 6, n_2 = 5, p < 0.01$  two-tailed]. The LE group also surpassed the NM group [Mann–Whitney  $U = 34, n_1 = n_2 = 6, p < 0.01$  two-tailed] and the HE group [Mann–Whitney  $U = 45, n_1 = 6, n_2 = 9, p < 0.05$  two-tailed] in this regard. In short, HL

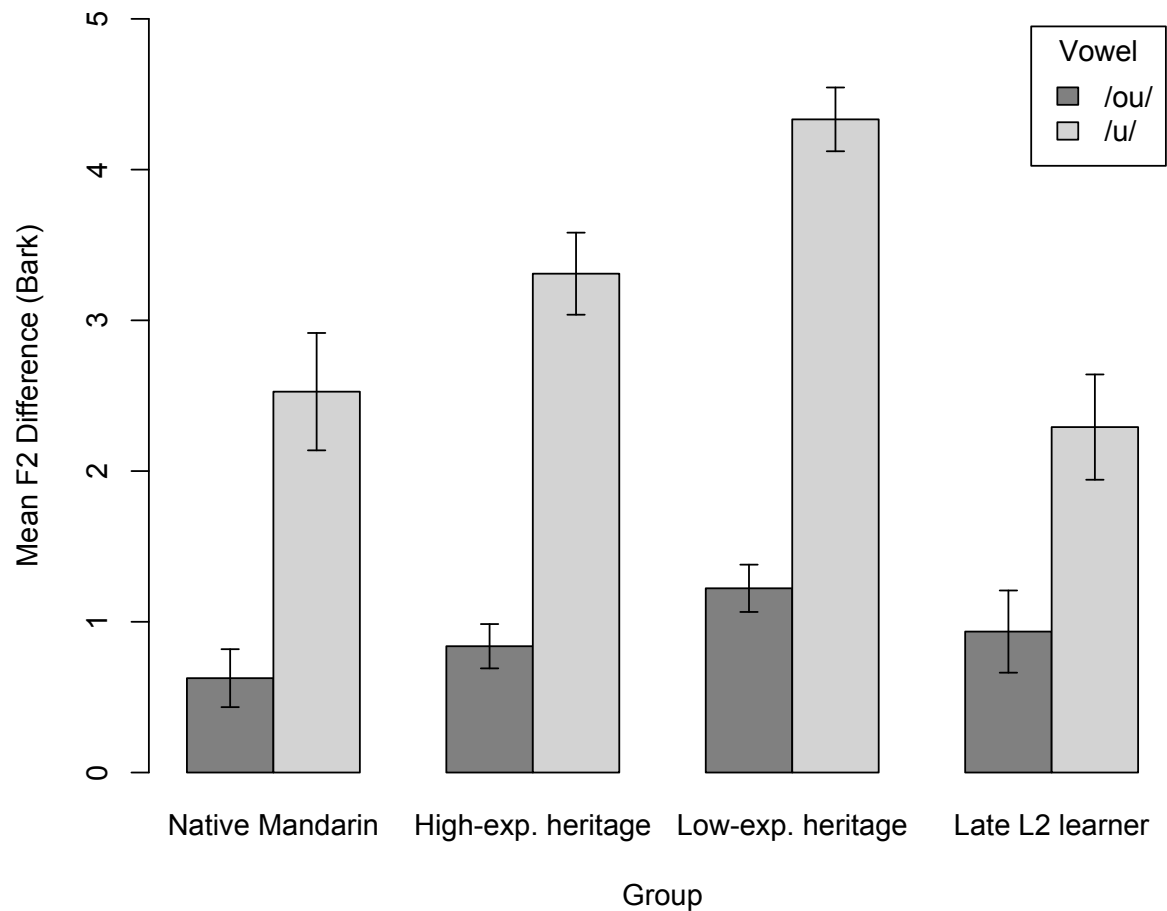


FIG. 3. Mean differences in  $F_2$  produced between Mandarin and English back rounded vowels, by participant group (from left to right: NM, HE, LE, L2). Differences between Mandarin and English /o<sup>u</sup>/ are in dark gray bars, differences between Mandarin and English /u/ in light gray bars. Error bars indicate  $\pm 1$  standard error about the mean.

speakers separated their two high back rounded vowels in  $F_2$  to a greater degree than L2 learners did, and LE speakers in particular also produced greater  $F_2$  separation than NM speakers.

## B. Experiment 2: plosives

Differences between VOT norms for Mandarin and English plosives are summarized in Table II (Mandarin figures from Wu and Lin 1989, English figures from Lisker and Abramson 1964). On the basis of their acoustic phonetic similarity, the categories compared to each other were the two short-lag VOT categories (Mandarin unaspirated and English voiced) and the two long-lag VOT categories (Mandarin aspirated and English voiceless), which in initial position are all typically realized without vocal fold vibration during closure. Of the two short-lag categories, Mandarin unaspirated plosives are on average characterized by the shorter VOT, with the VOT of English voiced plosives being similar, but 2–9 ms shorter or longer at the same place of articulation. With respect to the long-lag categories, Mandarin aspirated plosives are significantly more aspirated than English voiceless plosives, by as much as 48 ms at the same place of articulation. In short, both pairs of similar laryngeal categories differ in VOT, although the difference between the long-lag categories is much greater than that between the short-lag categories. If speakers with some degree of experience in both languages closely approximate these phonetic norms, then, it is expected that they will produce a subtle difference between the short-lag categories and a pronounced difference between the long-lag categories.

As a first step towards testing this prediction, the VOT data collected in Experiment 2 were subjected to a mixed-model ANOVA, with Group and Gender as between-subjects factors and Language, Voicing Type (short-lag vs. long-lag), Place (of articulation), and Vowel<sup>6</sup> (environment) as within-subjects factors. As expected, the ANOVA results showed highly significant main effects of every within-subjects factor: Language [ $F(1, 6) = 46.49, p < 0.001$ ], Voicing Type [ $F(1, 6) = 613.05, p < 0.001$ ], Place [ $F(2, 18) = 93.52, p < 0.001$ ], and Vowel [ $F(1, 6) = 19.49, p < 0.01$ ]. These main effects were all in the expected direction: Mandarin > English; long-lag > short-lag; velar > alveolar > bilabial; and /u/ > /o<sup>u</sup>/. There was also a two-way interaction between Language and Voicing Type [ $F(1, 6) = 18.64, p < 0.01$ ], an effect mostly attributable to there being no significant difference in VOT between



Mandarin unaspirated and English voiced stop productions. While there were no main effects of Group or Gender on VOT, there was a significant six-way interaction between these factors and the four within-subjects factors: Group x Gender x Language x Voicing Type x Place x Vowel [ $F(3, 17) = 3.41, p < 0.05$ ]. This interaction occurred due to between-group differences only for comparisons of a few combinations of the within-subjects factors. For instance, in comparing the HE and L2 groups, Tukey's HSD test showed a reliable difference only between HE and L2 females and only with respect to Mandarin long-lag velar stops preceding /u/ [ $p < 0.05$ ]. In summary, Mandarin VOT was produced as longer overall than English VOT (due to the long-lag VOT of Mandarin aspirated stop productions being longer than the long-lag VOT of English voiceless stop productions), and there were no significant differences among groups with respect to overall VOT levels.

When the VOT data were examined by participant, it was apparent that there was a strong tendency for HL speakers to make a VOT distinction between cross-linguistically similar laryngeal categories. The short-lag categories were produced with reliably distinct VOTs by only six participants. Of these six, half came from the HE or LE groups; the other three were divided between the NM and L2 groups. The long-lag categories, on the other hand, were distinguished by 18 participants (Figure 4). These participants were concentrated in the NM, HE, and LE groups, such that all NM speakers, all but one HE speaker, and half of LE speakers produced a reliable difference in VOT between the long-lag categories. In contrast, all but one L2 learner produced no reliable difference in VOT between these two categories. Note that this pattern still held after adjusting for multiple comparisons. When the Bonferroni correction was applied, five NM speakers and eight HL speakers, but only one L2 learner, were found to produce a significant difference in VOT.

The results of Experiment 2 thus indicated that while all participants reliably produced the language-internal contrasts between Mandarin unaspirated and aspirated plosives and between English voiced and voiceless plosives, the same could not be said of their realization of cross-linguistic contrasts. Few made the cross-linguistic contrast between the short-lag categories of Mandarin and English. On the other hand, many participants produced a

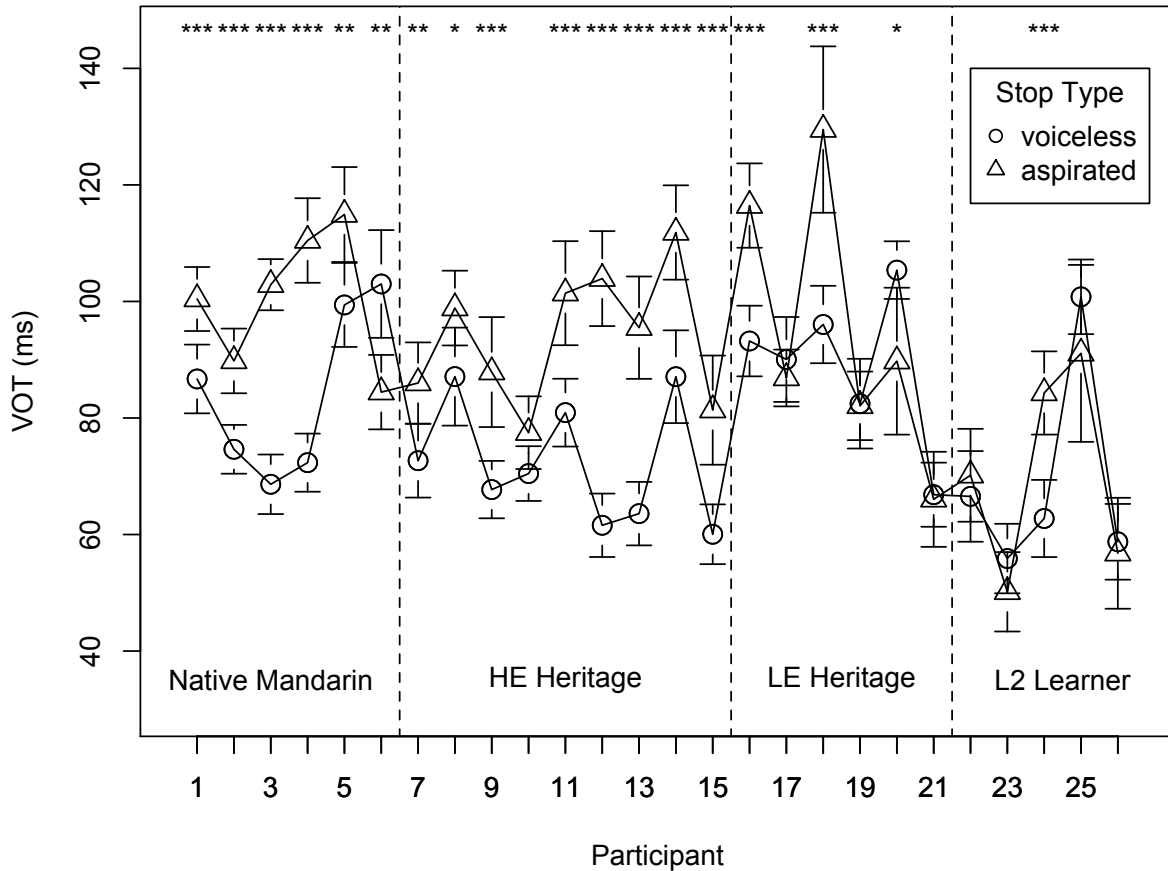


FIG. 4. VOT in Mandarin aspirated plosives (triangles) and English voiceless plosives (circles), by participant. Error bars indicate 95% confidence intervals. Participants who produce reliably different means are marked with stars: \* ( $p < 0.05$ ), \*\* ( $p < 0.01$ ), \*\*\* ( $p < 0.001$ ).

contrast between the long-lag categories. However, these were nearly all participants with the greatest Mandarin experience—namely, NM and HL speakers. Most L2 learners failed to distinguish the long-lag categories.

Between-group differences in the realization of cross-linguistic contrasts were further examined by calculating for each participant the mean difference in VOT produced between similar laryngeal categories. The mean VOT distances produced by all groups are presented

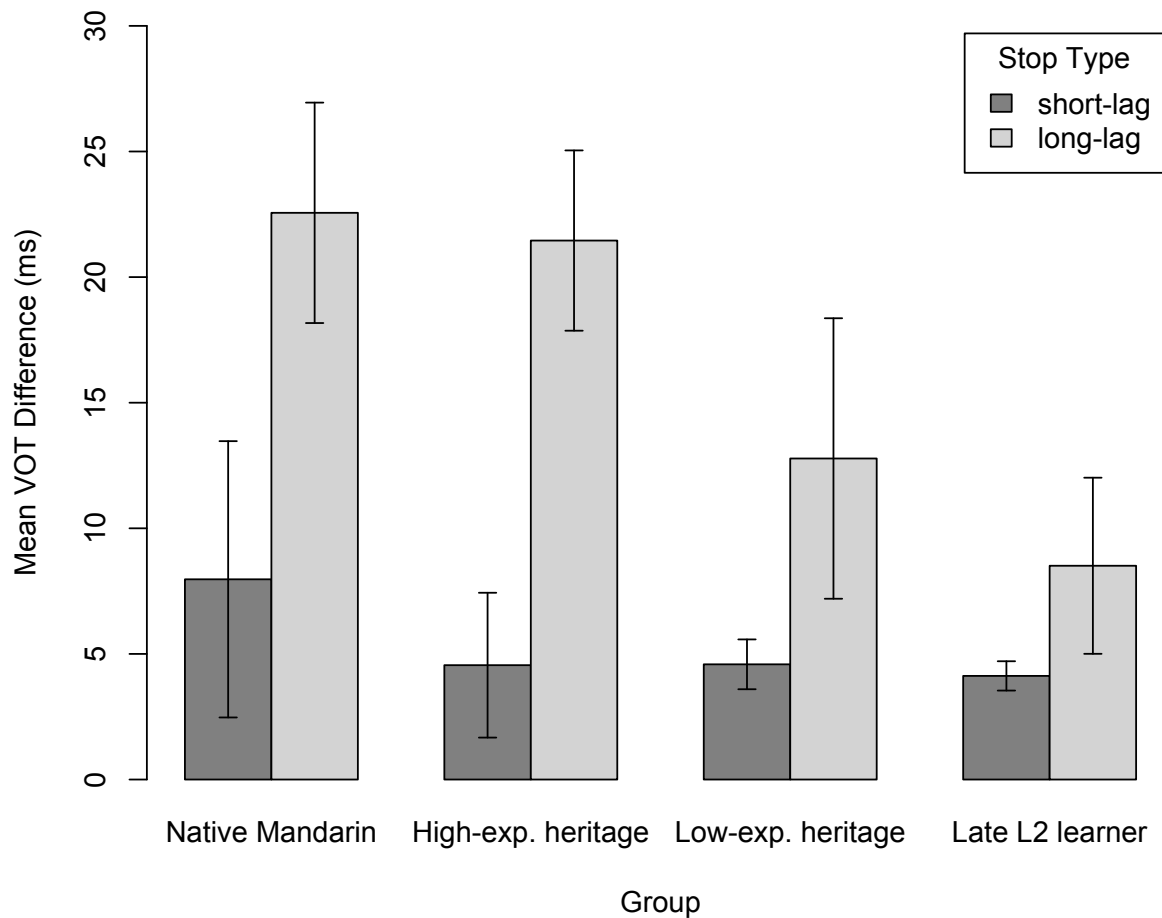


FIG. 5. Mean differences in VOT produced between Mandarin and English plosives, by participant group (from left to right: NM, HE, LE, L2). Differences between Mandarin unaspirated and English voiced plosives are in dark gray bars, differences between Mandarin aspirated and English voiceless plosives in light gray bars. Error bars indicate  $\pm 1$  standard error about the mean.

in Figure 5. A one-way ANOVA showed no main effect of Group on the VOT distances established between the short-lag categories, but a marginally significant main effect of Group on the VOT distances established between the long-lag categories [ $F(3, 22) = 2.27, p = 0.1$ ]. Here the HE group produced reliably greater distance between the two categories than the

L2 group [Mann–Whitney  $U = 26, n_1 = 9, n_2 = 6, p = 0.05$  two-tailed], as did the NM group [Mann–Whitney  $U = 38, n_1 = n_2 = 6, p < 0.05$  two-tailed]. These results were consistent with the findings of the participant analysis described above: NM speakers and HL speakers—HE speakers, in particular—established a greater acoustic distance between the long-lag VOT categories of Mandarin and English than did L2 learners of Mandarin.

### C. Experiment 3: fricatives

Before the results of Experiment 3 are presented, the phonetics of the three post-alveolar fricatives under investigation are first reviewed. These fricatives have been described in detail by Ladefoged and Maddieson (1996, 148–154). While the English palato-alveolar /ʃ/ is described as “domed” (i.e. with the front of the tongue raised) and rounded, the Mandarin retroflex /ʂ/ is described as “flat” (i.e. without the front of the tongue raised), laminal, and not truly retroflexed, having a location and width of constriction that are “very comparable with those for English ʃ”. The Mandarin alveolo-palatal /ç/, on the other hand, is described as significantly “palatalized”, with a long, flat constriction formed by a greater degree of raising of the blade and front of the tongue. These descriptions suggest that, compared to /ç/, /ʂ/ is closer phonetically to /ʃ/ and, consequently, that /ʂ/ is more likely to be merged with /ʃ/ in production. Note, however, that merger of /ç/ and /ʃ/ has been found before (Young, 2007). Thus, it is possible that both Mandarin fricatives might be merged with the English fricative, though the influence of phonological knowledge in perceptual assimilation (i.e. that the Mandarin sounds serve to distinguish words; see the PAM-L2) is likely to prevent such dual merger from happening.

In fact, differences between the centroids of Mandarin and English post-alveolar fricatives suggest that, at least with respect to centroid frequency, both Mandarin fricatives differ significantly from the English palato-alveolar. The centroid norms are summarized in Table III, converted to Bark according to the formula given in Section II.D (Mandarin figures averaged from Svantesson 1986, English figures from Jongman *et al.* 2000). Mandarin /ç/ is

characterized by the highest centroid frequency, followed by English /f/ and then Mandarin /ʃ/. Taking into account that the average centroid for /f/ is likely to be slightly higher than the figure given in Table III (an average that includes the corresponding voiced fricative /ʒ/, whose centroid will be drawn down by the lower frequencies of voicing), one can see that the centroid of /ʃ/ is slightly closer to that of /f/ than is the centroid of /ɛ/, but each Mandarin centroid lies on the order of 1 Bark away from the English centroid. Thus, if speakers with some degree of experience in both languages closely approximate these phonetic norms, it is expected that they will produce a three-way contrast in centroid among these fricatives. Conversely, if there is merger of any two of these categories, it is predicted that the fricatives merged will be the more similar /ʃ/ and /f/.

The centroid and PAF data<sup>7</sup> collected in Experiment 3 were subjected to mixed-model ANOVAs, with Group and Gender as between-subjects factors and Fricative as a within-subjects factor. As expected, the ANOVA results showed a highly significant main effect of Fricative on both centroid [ $F(2, 36) = 52.33, p < 0.001$ ] and PAF [ $F(2, 36) = 87.47, p < 0.001$ ]: in both cases, /ɛ/ > /ʃ/ > /f/ (in contrast to the predictions of Table III). There was also a main effect of Gender on PAF [ $F(1, 14) = 13.99, p < 0.01$ ]: female > male. There was no main effect of Group on centroid or PAF, although there was a significant three-way interaction between Group, Gender, and Fricative with respect to PAF [ $F(6, 36) = 3.07, p < 0.05$ ], an effect due mainly to between-group differences only for comparisons of particular combinations of Gender and Fricative. For example, in comparing the HE and L2 groups, Tukey’s HSD test showed no reliable difference between HE and L2 males or between HE and L2 females with respect to /ɛ/ or /f/, but did show a reliable difference between HE and L2 males with respect to /ʃ/ [ $p < 0.001$ ]. In summary, although the fricatives were produced as spectrally distinct in general, the groups did not differ from each other significantly in terms of overall centroid or PAF.

Between-group differences in the realization of cross-linguistic contrasts between similar fricative categories (especially Mandarin /ʃ/ and English /f/) were examined by calculating for each participant the mean distances in centroid and PAF established between each pair

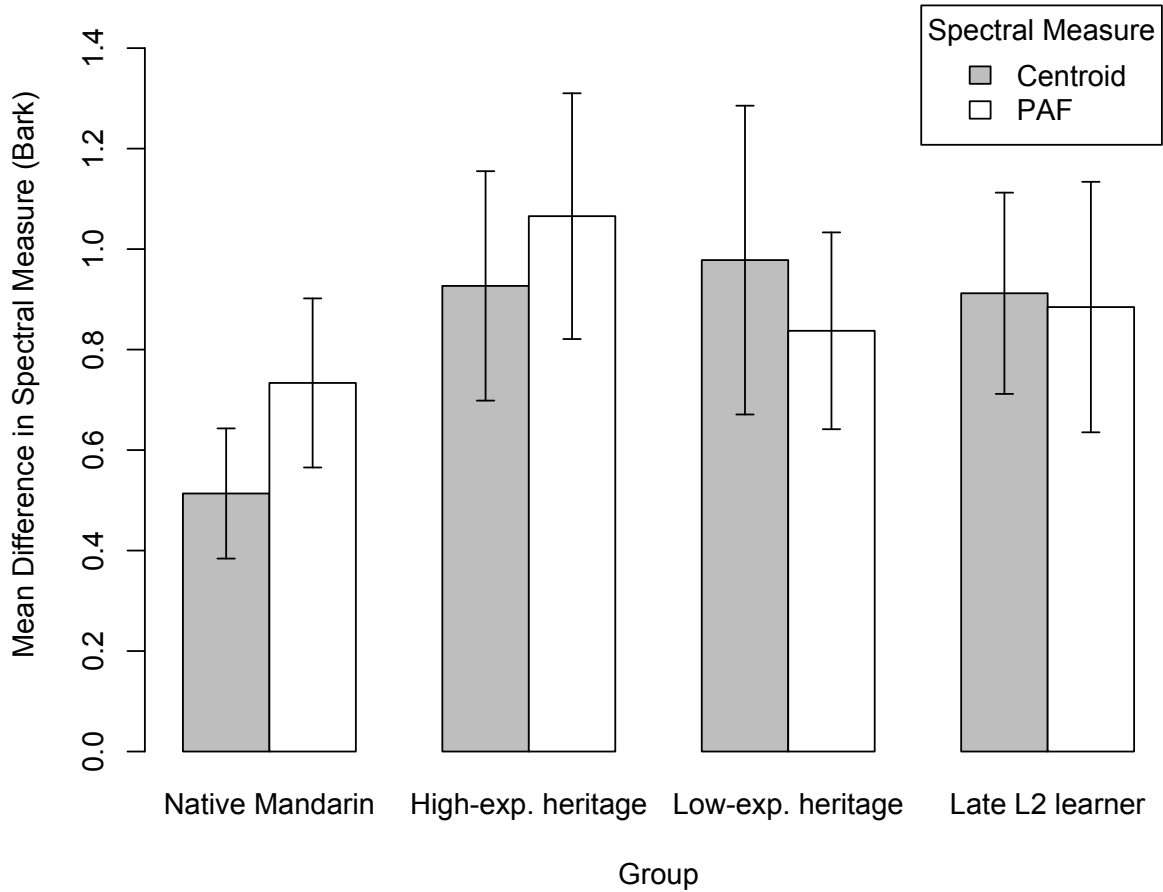


FIG. 6. Mean differences in centroid and peak amplitude frequency produced between Mandarin /ʂ/ and English /f/, by participant group (from left to right: NM, HE, LE, L2). Differences in centroid are in gray bars, differences in peak amplitude frequency in white bars. Error bars indicate  $\pm 1$  standard error about the mean.

of fricatives. In contrast to the results of Young (2007), which suggested that HL speakers might tend to merge Mandarin /ɕ/ with English /f/, the HL speakers in this study did not differ from other groups with respect to producing acoustic distance between /ɕ/ and /f/: all produced a robust contrast between these two fricatives. With respect to /ʂ/ and /f/, on the other hand, the HL and L2 groups appeared to separate these categories to a greater

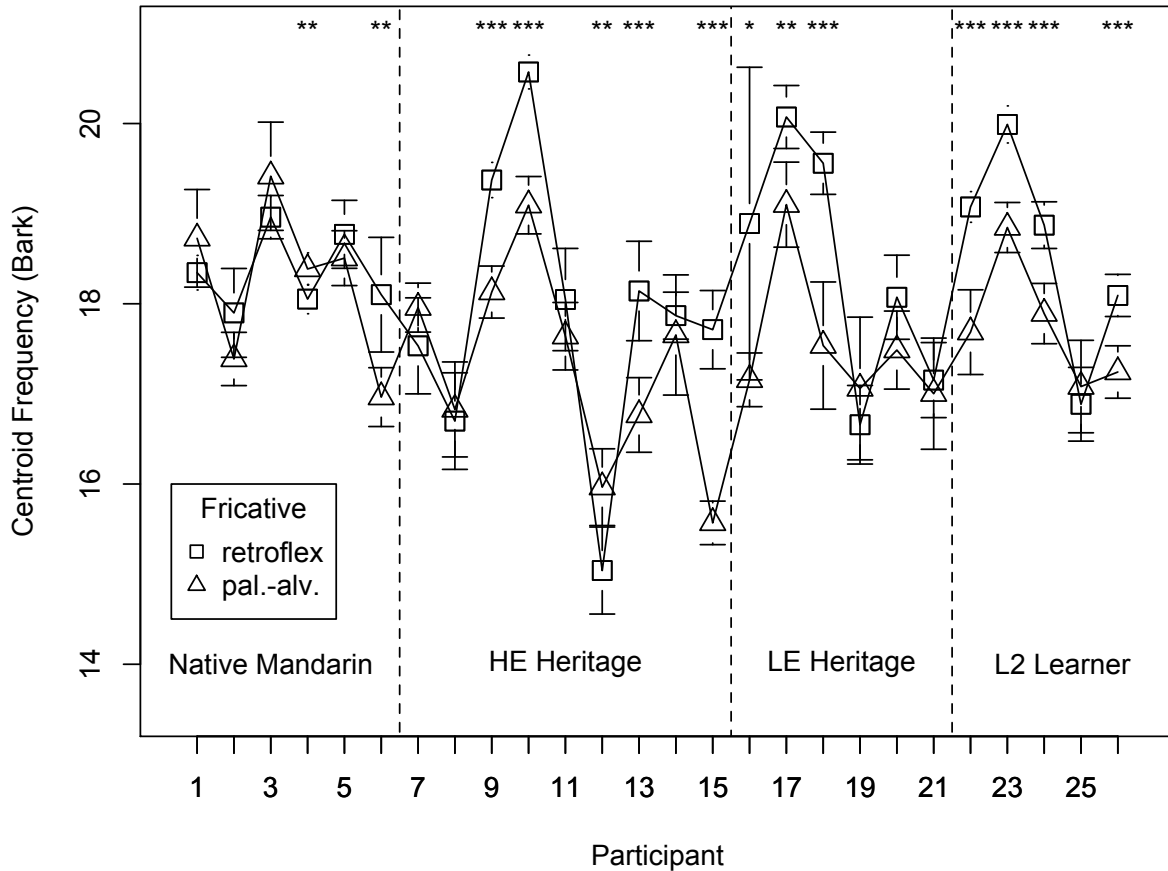


FIG. 7. Centroids in Mandarin retroflex /ʂ/ (squares) and English palato-alveolar /ʃ/ (triangles), by participant. Error bars indicate 95% confidence intervals. Participants who produce reliably different means are marked with stars: \* ( $p < 0.05$ ), \*\* ( $p < 0.01$ ), \*\*\* ( $p < 0.001$ ).

degree than the NM group, particularly with respect to centroid (Figure 6). Nevertheless, one-way ANOVAs showed no main effect of Group on centroid distances or PAF distances.

However, when the centroid data for /ʂ/ and /ʃ/ were examined by participant, it was apparent that HL speakers and L2 learners more often made a distinction between the two fricatives than NM speakers. These fricatives were distinguished in centroid by a total of 14 participants, who were unevenly distributed across groups (Figure 7). Whereas the majority

of HL speakers (five of nine HE speakers and half of LE speakers) and the majority of L2 speakers (four of five) produced a reliable difference in centroid, the majority of NM speakers (four of six) did not. Again, the pattern held after Bonferroni correction, in which case six HL speakers and four L2 learners, but no NM speakers, were found to produce a reliable difference in centroid between /ʂ/ and /ʃ/ .

In short, the results of Experiment 3 showed no overall differences between groups in the realization of contrast between Mandarin and English post-alveolar fricatives (as it was measured here). However, on an individual level HL speakers and L2 learners were found more often to achieve a reliable distinction between Mandarin /ʂ/ and English /ʃ/ than NM speakers.

#### IV. DISCUSSION AND CONCLUSIONS

To summarize, in Experiments 1–3 evidence was found that HL speakers were more successful than NM speakers and L2 learners at producing cross-language contrasts simultaneously with language-internal contrasts. In Experiment 1, participants in all groups were found to make an  $F_2$  distinction between Mandarin and English back vowels, with NM speakers’ back vowels having lower  $F_2$  values in both languages than those of HL speakers and L2 learners. However, HL speakers—in particular, LE speakers—clearly outperformed both NM speakers and L2 learners in achieving acoustic separation between similar vowel categories. In Experiment 2, few participants distinguished Mandarin unaspirated and English voiced plosives, but HL and NM speakers distinguished Mandarin aspirated and English voiceless plosives; furthermore, they put more acoustic distance between these categories than L2 learners, who mostly failed to distinguish them. In Experiment 3, HL speakers produced a contrast between the two Mandarin post-alveolar fricatives and were also more likely to produce a contrast between Mandarin /ʂ/ and English /ʃ/ than NM speakers.

Thus, it was found that HL speakers maintained not only language-internal “functional” contrast (that is, contrast that functions to distinguish words, e.g. English /u/ vs. /o<sup>u</sup>/),



but also cross-linguistic “non-functional” contrast (that is, contrast that has no function in distinguishing words by virtue of the members of the contrast belonging to different languages, e.g. English /u/ vs. Mandarin /u/). On the first point, HL speakers did not differ significantly from other groups, as almost no speaker in any group failed to distinguish the phonemic categories of their L1 and L2. HL speakers did not all realize categories in the same way as more L1-dominant native speakers (e.g.  $F_2$  values for Mandarin /u/ were slightly higher for several HL speakers than those of NM speakers), but on average they came very close—much closer than L2 learners—and this close approximation of phonetic norms seems to lie at the heart of why HL speakers were more successful than L2 learners at maintaining contrasts between similar L1 and L2 categories, which for the most part they would never need to distinguish for the purposes of being understood.

### A. Approximation of phonetic norms

In the present study, it is somewhat difficult to tell how closely speakers approached the phonetic norms of Mandarin and English, given the amount of inter-speaker variation and the limited nature of the acoustic norms available in the literature (e.g. the Mandarin figures provided by Wu and Lin 1989 are based on only a few speakers). However, if the numbers cited in Tables I–III are indeed representative of the relevant speech communities, then it seems that at least some of the current data show the same sort of bidirectional cross-linguistic influence found in Flege (1987). For example, the phonetic norm for  $F_2$  in Mandarin /u/ is cited as approximately 450–650 Hz (equivalent to 4.5–6.2 Bark), but speakers in this study produced this vowel with  $F_2$  values of approximately 6.9–9.7 Bark. Similarly, the phonetic norm for VOT in Mandarin unaspirated plosives is estimated at 7–15 ms, but speakers in this study produced these with VOTs of approximately 15–25 ms. What is most significant about the findings of this study, however, is that when taken together, the results of Experiments 1–3 showed HL speakers to have been the most successful at approximating the phonetic norms of both of their languages.

As for why HL speakers would tend to be more successful than late learners at maintaining contrasts between similar categories in their two languages, there are two possible explanations. First, early exposure to both languages might simply make HL speakers better able to hit close targets accurately, due to the existence of more fine-grained, less language-specific perceptual capabilities early in life (see, e.g., Werker and Tees 1984; Kuhl *et al.* 1992). Alternatively, similar categories that are acquired early may interact with each other in a shared phonological system and dissimilate. The results of Experiments 2–3 are more consistent with the former hypothesis, as similar laryngeal and place categories in these experiments were not produced by the HL groups as “too native” with respect to the productions of the NM and L2 groups (e.g. Mandarin unaspirated stops were not produced with VOTs that were even shorter than native Mandarin VOTs). On the other hand, the results of Experiment 1 showed signs that some HL speakers had dissimilated similar vowel categories, resulting in a “polarized” phonetic space that went past native targets (Laeuffer, 1996). In both Figures 1 and 2, it can be seen that there were HL speakers who went lower in  $F_2$  for their Mandarin vowels than the NM group, as well as HL speakers who went higher in  $F_2$  for their English vowels than the L2 group.

Thus, there are two ways to arrive at the patterns observed among HL speakers in this study, but it should be noted that these accounts of how HL speakers come to produce cross-linguistic phonetic contrasts are not mutually exclusive. Perhaps, as the data suggest, close approximation of native phonetic norms occurs generally during HL speakers’ relatively early exposure to both languages, but the pressure to keep categories distinct within a speaker’s phonological system (regardless of which language they come from) is what serves to keep similar L1 and L2 categories apart—close to the native phonetic norms—and prevent them from merging on a “compromise” value. Apparently this pressure may even push the categories further apart than they need to be, although the present results suggest that this is very much the minority case.

The ways in which the linguistic input received by NM speakers in mainland China and Taiwan differs from the linguistic input received by the other two groups in the U.S. must

also be considered. In particular, NM speakers' initial English input is likely to have been accented, making it possible that the amount of non-approximation to English phonetic norms seen for a given NM speaker, rather than being attributable to that one speaker, had actually accumulated over a chain of L2 acquirers. For that matter, one wonders whether the early Mandarin input received by HL speakers born in the U.S. (e.g. the Mandarin spoken by their parents, who had for the most part been living in the U.S. for a considerable period of time prior to their birth) would have differed significantly from the Mandarin input they would have received in a country where English is not so widely spoken. These are questions that will require more detailed study of the relevant acquisition situations to be able to answer, but there is reason to believe that if there were such an effect of inaccurate input here, it would stand to be the strongest in the NM speakers, who might have been exposed to heavily accented L2 English, whereas HL speakers were probably exposed to no worse than native Mandarin that had "drifted" (Sancier and Fowler, 1997; Chang, 2010) in an English-speaking environment.

Finally, it should be observed that although in Experiment 1, HL speakers seemed to outperform both the NM group and the L2 group in producing cross-linguistic contrast, in Experiments 2–3 they appeared to pattern together with one other group in outperforming the third group. In Experiment 2, both the HL group and the NM group surpassed the L2 group in producing acoustic distance (in terms of VOT) between the similar Mandarin aspirated and English voiceless plosives; likewise, in Experiment 3, both the HL group and the L2 group surpassed the NM group in distinguishing the similar fricatives /ʃ/ and /f/. Why did this occur?

As for why NM speakers, themselves L2 learners of English, outperformed L2 learners of Mandarin in distinguishing the two long-lag VOT categories in Experiment 2, one possibility is that NM speakers, being accustomed to very long VOTs for their native long-lag VOT category, are attuned to picking out VOTs that are too short to qualify as Mandarin aspirated stops, thus leading them to perceive English voiceless stops as significantly less aspirated than Mandarin aspirated stops. On the other hand, L2 learners of Mandarin might simply be

focused on whether a VOT is long enough to be an exemplar of English voiceless as opposed to English voiced, in which case they may be relatively insensitive to the difference between Mandarin aspirated and English voiceless, since in initial position both are aspirated enough to pass the VOT boundary that is salient for them.

The explanation for L2 learners of Mandarin outperforming NM speakers in distinguishing Mandarin /ʃ/ and English /f/ is likely quite different. Here it should be noted that these two types of L2 learners face very different tasks: Mandarin-speaking L2 learners of English, who already have two L1 post-alveolar fricatives, need to learn just one L2 post-alveolar fricative category, while English-speaking L2 learners of Mandarin, who have only one L1 post-alveolar fricative, need to learn to distinguish two L2 post-alveolar fricative categories. This one-to-many vs. many-to-one contrast does not in itself account for why the two groups seem to have reached disparate learning outcomes, but it does suggest a possible difference in learning strategies and perhaps instructional input as well. Whereas NM speakers can afford to produce English /f/ relatively inaccurately (e.g. as /ʃ/) with no serious consequences for intelligibility, L2 learners of Mandarin cannot similarly afford to produce the Mandarin fricatives inaccurately because there is a real chance they will be misunderstood, due to the relative crowdedness of the Mandarin fricative inventory. As a result of this pressure, they are probably highly conscious of their pronunciation of these segments, and formal instruction might serve to amplify their efforts to differentiate these fricatives from each other and from their L1 inventory by, for example, exoticizing the pronunciation of /ʃ/ as “retroflex” (which, as mentioned in Section III.C, is actually a misnomer). Thus, the role played by explicit knowledge and instruction in L2 production may be largely responsible for this latter result.

## **B. “New” vs. “similar” categories**

Just as the cross-linguistic influence seen in Experiments 1–2 was consistent with the cross-linguistic influence documented in Flege (1987), the results obtained in Experiment

1 for the production of Mandarin /y/ were also consistent with Flege’s (1987) results for French /y/: L2 learners did not differ appreciably from NM speakers in their phonetic space for /y/, suggesting that it was indeed perceived as a “new” sound. However, an alternative explanation for L2 learners’ relatively accurate production of /y/ exists—namely, that they were not producing anything new at all. In other words, they may have simply been drawing upon vowel tokens that were already in their repertoire: particularly fronted versions of English /u/. It has been proposed that due to the fact that American English /u/, which is already relatively front on average, is further fronted in the context of alveolars, the front vowel /y/ may not actually constitute a “new” vowel for English-speaking L2 learners, but instead a “similar” vowel, at least in the context of alveolars (Strange *et al.*, 2007; Levy, 2009; Levy and Law, 2010). Documented perceptual assimilation patterns in which English speakers identify German and French /y/ as closest to English /u/ (e.g. Polka and Bohn, 1996; Strange *et al.*, 2004) are consistent with this idea. Nevertheless, this proposal does not provide a convincing account of the current findings for three reasons.

First, it is unclear whether the requisite pattern of perceptual assimilation occurs for /y/ to become categorically linked to English /u/. Although German /y/ in the context of alveolars, for example, is assimilated to English /u/ in cross-language labeling tasks, German /u/ in the same context is also assimilated to English /u/ and is, moreover, rated as a better exemplar of English /u/ than /y/ is (Polka and Bohn, 1996). These data suggest that, if L2 German learners are to maintain contrast between German /u/ and German /y/, only one of these vowels will be perceptually linked to English /u/, and that vowel will be German /u/, not German /y/ (which is actually the acoustically closer vowel). Such a linkage, in which the L2 category linked to the L1 category is not the phonetically closest L2 category, is plausible in light of findings showing that perceptual assimilation of vowels does not necessarily follow from strict phonetic proximity (e.g. French /y/ is acoustically closer to English /i/ than to English /u/, yet is consistently assimilated to /u/; see Strange *et al.* 2004).

The account of /y/ as a “similar” vowel is also not supported by the distributional

characteristics of L2 learners' vowel productions in this study. For L2 learners' accurate production of /y/ to be an artifact of the existence of phonetically similar, fronted realizations of English /u/, there should be significant overlap in the  $F_2$  distributions of these two categories, yet most L2 learner participants showed little to no overlap between these two distributions. For example, participant L22 had an  $F_2$  range of 8.63–12.33 Bark for English /u/ vs. 12.30–13.39 Bark for Mandarin /y/; similarly, the relevant  $F_2$  ranges for participants L24 and L25 were, respectively, 9.02–13.17 vs. 13.47–14.59 and 7.74–12.78 vs. 12.66–14.01. Thus, it is not the case that L2 learners produced /y/ by simply recruiting their most fronted exemplars of English /u/; in fact, they produced /y/ with significantly higher  $F_2$  values than those of their most fronted /u/ productions.

Finally, given the phonetic norms of Mandarin /y/ and English /u/, the hypothetical classification of /y/ as a vowel “similar” to English /u/ is inconsistent with how /y/ was produced by the L2 learners in this study. If Mandarin /y/ were treated as a similar vowel, according to the SLM it would be perceptually linked to English /u/ and, therefore, produced inaccurately (with too-low  $F_2$  values) under influence from English /u/, since English /u/ is on average characterized by an  $F_2$  that is lower than that of Mandarin /y/ (Table I), even if its most fronted realizations might show  $F_2$  values that overlap with those of /y/. What was found instead, however, was that /y/ was produced by L2 learners relatively *accurately*—right in the same region with the /y/ productions of native Mandarin speakers, rather than retracted in the  $F_2$  dimension.

Thus, the findings of this study support the initial prediction that Mandarin /y/ would constitute a “new” vowel for English speakers, confirming that—consistent with the PAM-L2—category equivalence in L2 speech acquisition can be based on phonological proximity. Furthermore, the findings suggest that when phonetic and phonological considerations conflict in these situations, the higher-level phonological considerations override the lower-level phonetic ones. However, it may not always be the case that higher-level proximity preempts lower-level proximity. The exact nature of their interaction with respect to determining category equivalence remains a question for future research.

## Acknowledgments

This work was supported in part by a National Science Foundation Graduate Research Fellowship to the first author. The authors are grateful to Susanne Gahl, Sharon Inkelas, Keith Johnson, participants in a fall 2007 UC Berkeley seminar on phonological learning, several anonymous reviewers, and audiences at the University of California, Berkeley, the University of Pennsylvania, and the University of Chicago for helpful comments and discussion. Portions of these data are also discussed in the University of Pennsylvania Working Papers in Linguistics (Chang *et al.*, 2009) and the Proceedings from the Annual Meeting of the Chicago Linguistic Society (Chang *et al.*, 2010).

## APPENDIX A: PARTICIPANT INFORMATION

See Tables IV, V, and VI.

## APPENDIX B: LIST OF STIMULI

See Tables VII, VIII, and IX.

## ENDNOTES

1. Note that contradictory null results have emerged from the work of Pallier and colleagues (Pallier *et al.*, 2003), who, examining subjects from a different HL situation (Korean adoptees in France), have failed to find a perceptual or even low-level neural advantage for the HL in individuals with early HL exposure, suggesting that some re-learning of the HL, or at least intermittent exposure to it, may be necessary for distant HL experience to become readily accessible.
2. An anonymous reviewer questioned the analysis of participants in terms of groups, as well as the basis for dividing participants into the groups described in this section. It should be noted that there are a number of ways the data in this study could have

been analyzed; however, a group approach was the most appropriate because the focus was on differences between HL speakers (i.e., individuals with early exposure to two languages) and individuals without early exposure to two languages. The advantages of a group approach over, e.g., a correlational approach were twofold. First, a group approach obviated the need to arbitrarily define one background characteristic as the independent variable or to compute a holistic index of Mandarin experience based on several background characteristics as the independent variable. This was beneficial because HL speakers often have complex language backgrounds and residential histories (as seen in Table V) that make it unclear how two relatively non-proficient HL speakers should stack up relative to one another. Second, a group approach made it possible to include variability in the analysis of HL speakers, in keeping with the characteristics of this population. In the interest of yielding more generalizable results, the decision was made not to arbitrarily examine one specific level of HL experience. Instead, HL speakers with a range of experience were included to more accurately represent the population of HL Mandarin speakers (as described in Section I.C), and they were divided into subgroups in a manner that was straightforward and highly replicable (as described in this section).

3. Gender was entered as a factor in all of the statistical analyses in Experiments 1–3 because it has been shown to have an effect on all of the acoustic dimensions investigated (an effect that is, moreover, usually perceptible to listeners): vowel formants (Whiteside, 1998b,c), voice onset time (Swartz 1992; Whiteside and Irving 1998; though see Morris *et al.* 2008), and spectral properties of frication (Whiteside, 1998a). For a recent review of gender differences in speech, see Simpson (2009). Thus, even though gender was not the focus of any analysis, its potential effect on the data was accounted for explicitly, rather than ignored, so that any possible interactions with other factors of interest could be accounted for.
4. The segmental context in which a vowel occurs—in particular, the place of articulation



of flanking consonants—has been shown to have a significant, predictable effect on the quality of the vowel. American English /u/, for example, has been shown to be significantly fronted in the context of alveolars (Hillenbrand *et al.*, 2001; Strange *et al.*, 2007). Place of articulation has also been shown to have a predictable effect on VOT (e.g. Lisker and Abramson, 1967; Nearey and Rochet, 1994; Liu *et al.*, 2007), which was measured in Experiment 2. Thus, although place was not the focus of any analysis, given its probable effect on the data it was entered into the analyses so that any possible interactions with other factors could be accounted for.

5. For the dependent variables of  $F_1$  and  $F_2$  in Experiment 1 as well as the dependent variable of VOT in Experiment 2 (Section III.B), there were significant two- and three-way interactions involving Place: Language x Place, Place x Vowel, and Language x Place x Vowel. However, these interactions are not of concern here, so they will not be discussed further.
6. Vowel environment, like place of articulation, was entered into the analyses because it has been shown to have a significant effect on VOT (Klatt, 1975; Nearey and Rochet, 1994).
7. Formant transition data showed that alveolo-palatal /ç/, having by far the lowest  $F_1$  onset (approximately 1 Bark lower than that of /ʃ/ and /j/ in all groups) and highest  $F_2$  onset (approximately 2 Bark higher than that of /ʃ/ and /j/ in all groups), was clearly the most “palatalized” of the three post-alveolar fricatives. However, the formant data did not clearly differentiate /ʃ/ and /j/. Thus, the focus in this study was on centroid frequency data, which were supplemented with PAF data as described in Section II.D.

## REFERENCES

- Andersen, R. W. (1982). "Determining the linguistic attributes of language attrition", in *The Loss of Language Skills*, edited by R. D. Lambert and B. F. Freed, 83–118 (Newbury House, Rowley, MA).
- Andrews, D. R. (1999). *Sociocultural Perspectives on Language Change in Diaspora: Soviet Immigrants in the United States*, 1–200 (John Benjamins Publishing, Amsterdam).
- Au, T. K., Knightly, L. M., Jun, S.-A., and Oh, J. S. (2002). "Overhearing a language during childhood", *Psychological Science* **13**, 238–243.
- Au, T. K. and Oh, J. S. (2009). "Korean as a heritage language", in *Handbook of East Asian Psycholinguistics, Volume III: Korean*, edited by C. Lee, G. B. Simpson, and Y. Kim, 268–275 (Cambridge University Press, Cambridge).
- Au, T. K., Oh, J. S., Knightly, L. M., Jun, S.-A., and Romo, L. F. (2008). "Salvaging a childhood language", *Journal of Memory and Language* **58**, 998–1011.
- Au, T. K. and Romo, L. F. (1997). "Does childhood language experience help adult learners?", in *The Cognitive Processing of Chinese and Related Asian Languages*, edited by H.-C. Chen, 417–443 (Chinese University Press, Hong Kong).
- Best, C. T. (1994). "The emergence of native-language phonological influences in infants: A perceptual assimilation model", in *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*, edited by J. C. Goodman and H. C. Nusbaum, 167–224 (MIT Press, Cambridge, MA).
- Best, C. T. and Tyler, M. D. (2007). "Nonnative and second-language speech perception: Commonalities and complementarities", in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, edited by O.-S. Bohn and M. J. Munro, 13–34 (John Benjamins Publishing, Amsterdam, The Netherlands).
- Boersma, P. and Weenink, D. (2008). "Praat: Doing phonetics by computer", Version 5.0.26. <http://www.praat.org> (date last viewed 6/15/08).
- Campbell, L. and Muntzel, M. C. (1989). "The structural consequences of language death",

- in *Investigating Obsolescence: Studies in Language Contraction and Death*, edited by N. C. Dorian, 181–196 (Cambridge University Press, Cambridge, UK).
- Campbell, R. N. and Rosenthal, J. W. (2000). “Heritage languages”, in *Handbook of Undergraduate Second Language Education*, edited by J. W. Rosenthal, 165–184 (Lawrence Erlbaum Associates, Mahwah, NJ).
- Chang, C. B. (2009). “English loanword adaptation in Burmese”, *Journal of the Southeast Asian Linguistics Society* **1**, 77–94.
- Chang, C. B. (2010). “First language phonetic drift during second language acquisition”, Ph.D. dissertation, University of California, Berkeley, Berkeley, CA.
- Chang, C. B., Haynes, E. F., Yao, Y., and Rhodes, R. (2009). “A tale of five fricatives: Consonantal contrast in heritage speakers of Mandarin”, *University of Pennsylvania Working Papers in Linguistics* **15**, 37–43.
- Chang, C. B., Haynes, E. F., Yao, Y., and Rhodes, R. (2010). “The phonetic space of phonological categories in heritage speakers of Mandarin”, in *Proceedings from the 44<sup>th</sup> Annual Meeting of the Chicago Linguistic Society: The Main Session*, edited by M. Bane, J. Bueno, T. Grano, A. Grotberg, and Y. McNabb, 31–45 (Chicago Linguistic Society, Chicago, IL).
- Chen, Y. (2006). “Production of tense-lax contrast by Mandarin speakers of English”, *Folia Phoniatica et Logopaedica* **58**, 240–249.
- Flege, J. E. (1987). “The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalence classification”, *Journal of Phonetics* **15**, 47–65.
- Flege, J. E. (1995). “Second language speech learning: Theory, findings, and problems”, in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange, 233–272 (York Press, Baltimore, MD).
- Flege, J. E. and Hillenbrand, J. (1984). “Limits on phonetic accuracy in foreign language speech production”, *Journal of the Acoustical Society of America* **76**, 708–721.
- Godson, L. (2003). “Phonetics of language attrition: Vowel production and articulatory setting in the speech of Western Armenian heritage speakers”, Ph.D. dissertation, University

of California, San Diego, San Diego, CA.

Godson, L. (2004). “Vowel production in the speech of Western Armenian heritage speakers”, *Heritage Language Journal* **2**, 44–69.

Goodfellow, A. M. (2005). *Talking in Context: Language and Identity in K<sup>w</sup>ak<sup>w</sup>aka’wak<sup>w</sup> Society*, 1–219 (McGill-Queen’s University Press, Montréal, Canada).

Hagiwara, R. (1997). “Dialect variation and formant frequency: The American English vowels revisited”, *Journal of the Acoustical Society of America* **102**, 655–658.

Hillenbrand, J. M., Clark, M. J., and Nearey, T. M. (2001). “Effects of consonant environment on vowel formant patterns”, *Journal of the Acoustical Society of America* **109**, 748–763.

Jia, G., Strange, W., Wu, Y., Collado, J., and Guan, Q. (2006). “Perception and production of English vowels by Mandarin speakers: Age-related differences vary with amount of L2 exposure”, *Journal of the Acoustical Society of America* **119**, 1118–1130.

Jiang, H. (2008). “Effect of L2 phonetic learning on L1 vowels”, Ph.D. dissertation, Simon Fraser University, Vancouver, Canada.

Jiang, H. (2010). “Effect of L2 phonetic learning on the production of L1 vowels: A study of Mandarin-English bilinguals in Canada”, in *New Sounds 2010: Proceedings of the 6<sup>th</sup> International Symposium on the Acquisition of Second Language Speech*, edited by K. Dziubalska-Kořaczyk, M. Wrembel, and M. Kul, 227–232 (Adam Mickiewicz University, Poznań, Poland).

Jongman, A., Wayland, R., and Wong, S. (2000). “Acoustic characteristics of English fricatives”, *Journal of the Acoustical Society of America* **108**, 1252–1263.

Kang, Y. (2008). “Interlanguage segmental mapping as evidence for the nature of lexical representation”, *Language and Linguistics Compass* **2**, 103–118.

Klatt, D. H. (1975). “Voice onset time, frication, and aspiration in word-initial consonant clusters”, *Journal of Speech and Hearing Research* **18**, 686–706.

Knightly, L. M., Jun, S.-A., Oh, J. S., and Au, T. K. (2003). “Production benefits of childhood overhearing”, *Journal of the Acoustical Society of America* **114**, 465–474.

- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (1992). “Linguistic experience alters phonetic perception in infants by 6 months of age”, *Science* **255**, 606–608.
- LaCharité, D. and Paradis, C. (2005). “Category preservation and proximity versus phonetic approximation in loanword adaptation”, *Linguistic Inquiry* **36**, 223–258.
- Ladefoged, P. (2003). *Phonetic Data Analysis*, 96–101, 156–158 (Blackwell Publishing, Malden, MA).
- Ladefoged, P. (2005). *Vowels and Consonants*, 2nd edition, 40–43 (Blackwell Publishing, Malden, MA).
- Ladefoged, P. and Maddieson, I. (1996). *The Sounds of the World’s Languages*, 148–154 (Blackwell Publishers, Oxford, UK).
- Laeufer, C. (1996). “Towards a typology of bilingual phonological systems”, in *Second-Language Speech: Structure and Process*, edited by A. James and J. Leather, 325–342 (Mouton de Gruyter, Berlin).
- Levy, E. S. (2009). “Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French”, *Journal of the Acoustical Society of America* **125**, 1138–1152.
- Levy, E. S. and Law, F. F. (2010). “Production of French vowels by American-English learners of French: Language experience, consonantal context, and the perception-production relationship”, *Journal of the Acoustical Society of America* **128**, 1290–1305.
- Li, D. and Duff, P. A. (2008). “Issues in Chinese heritage language education and research at the postsecondary level”, in *Chinese as a Heritage Language: Fostering Rooted World Citizenry*, edited by A. W. He and Y. Xiao, 13–32 (University of Hawaii, National Foreign Language Resource Center, Honolulu, HI).
- Li, F., Edwards, J., and Beckman, M. (2007). “Spectral measures for sibilant fricatives of English, Japanese, and Mandarin Chinese”, in *Proceedings of the 16<sup>th</sup> International Congress of Phonetic Sciences*, edited by J. Trouvain and W. J. Barry, 917–920 (Pirrot, Dudweiler, Germany).

- Lin, T. and Wang, L. (1992). *Yu Yin Xue Jiao Cheng (in Chinese)*, 1–208 (Beijing University Press, Beijing, China).
- Lisker, L. and Abramson, A. S. (1964). “A cross-language study of voicing in initial stops: Acoustical measurements”, *Word* **20**, 384–422.
- Lisker, L. and Abramson, A. S. (1967). “Some effects of context on voice onset time in English stops”, *Language and Speech* **10**, 1–28.
- Liu, H., Ng, M. L., Wan, M., Wang, S., and Zhang, Y. (2007). “Effects of place of articulation and aspiration on voice onset time in Mandarin esophageal speech”, *Folia Phoniatica et Logopaedica* **59**, 147–154.
- Major, R. C. (1992). “Losing English as a first language”, *The Modern Language Journal* **76**, 190–208.
- Montrul, S. A. (2008). *Incomplete Acquisition in Bilingualism: Re-examining the Age Factor*, 1–312 (John Benjamins Publishing, Amsterdam).
- Morris, R. J., McCrea, C. R., and Herring, K. D. (2008). “Voice onset time differences between adult males and females: Isolated syllables”, *Journal of Phonetics* **36**, 308–317.
- Nearey, T. M. and Rochet, B. L. (1994). “Effects of place of articulation and vowel context on VOT production and perception for French and English stops”, *Journal of the International Phonetic Association* **24**, 1–18.
- Oh, J., Jun, S.-A., Knightly, L., and Au, T. (2003). “Holding on to childhood language memory”, *Cognition* **86**, B53–B64.
- Oh, J. S. and Au, T. K. (2005). “Learning Spanish as a heritage language: The role of sociocultural background variables”, *Language, Culture and Curriculum* **18**, 229–241.
- Oh, J. S., Au, T. K., and Jun, S.-A. (2002). “Benefits of childhood language experience for adult L2 learners’ phonology”, in *Proceedings of the 26<sup>th</sup> Annual Boston University Conference on Language Development*, edited by B. Skarabela, S. Fish, and A. H.-J. Do, volume 2, 464–472 (Cascadilla Press, Somerville, MA).
- Oh, J. S., Au, T. K., and Jun, S.-A. (2010). “Early childhood language memory in the speech perception of international adoptees”, *Journal of Child Language* **37**, 1123–1132.

- Pallier, C., Dehaene, S., Poline, J.-B., LeBihan, D., Argenti, A.-M., Dupoux, E., and Mehler, J. (2003). "Brain imaging of language plasticity in adopted adults: Can a second language replace the first?", *Cerebral Cortex* **13**, 155–161.
- Polinsky, M. (2008). "Gender under incomplete acquisition: Heritage speakers' knowledge of noun categorization", *Heritage Language Journal* **6**, 40–71.
- Polka, L. and Bohn, O.-S. (1996). "A cross-language comparison of vowel perception in English-learning and German-learning infants", *Journal of the Acoustical Society of America* **100**, 577–592.
- Sancier, M. L. and Fowler, C. A. (1997). "Gestural drift in a bilingual speaker of Brazilian Portuguese and English", *Journal of Phonetics* **27**, 421–436.
- Simpson, A. P. (2009). "Phonetic differences between male and female speech", *Language and Linguistics Compass* **3**, 621–640.
- Strange, W., Levy, E., and Lehnholz, Jr., R. (2004). "Perceptual assimilation of French and German vowels by American English monolinguals: Acoustic similarity does not predict perceptual similarity", *Journal of the Acoustical Society of America* **115**, 2606.
- Strange, W., Weber, A., Levy, E. S., Shafiro, V., Hisagi, M., and Nishi, K. (2007). "Acoustic variability within and across German, French and American English vowels: Phonetic context effects", *Journal of the Acoustical Society of America* **122**, 1111–1129.
- Svantesson, J.-O. (1986). "Acoustic analysis of Chinese fricatives and affricates", *Journal of Chinese Linguistics* **14**, 53–70.
- Swartz, B. L. (1992). "Gender difference in voice onset time", *Perceptual and Motor Skills* **75**, 983–992.
- Tees, R. C. and Werker, J. F. (1984). "Perceptual flexibility: Maintenance or recovery of the ability to discriminate non-native speech sounds", *Canadian Journal of Psychology* **38**, 579–590.
- Traunmüller, H. (1990). "Analytical expressions for the tonotopic sensory scale", *Journal of the Acoustical Society of America* **88**, 97–100.
- Werker, J. F. and Tees, R. C. (1984). "Cross-language speech perception: Evidence for

perceptual reorganization during the first year of life”, *Infant Behavior and Development* **7**, 49–63.

Whiteside, S. P. (1998a). “Identification of a speaker’s sex: A fricative study”, *Perceptual and Motor Skills* **86**, 587–591.

Whiteside, S. P. (1998b). “Identification of a speaker’s sex: A study of vowels”, *Perceptual and Motor Skills* **86**, 579–584.

Whiteside, S. P. (1998c). “The identification of a speaker’s sex from synthesized vowels”, *Perceptual and Motor Skills* **87**, 595–600.

Whiteside, S. P. and Irving, C. J. (1998). “Speakers’ sex differences in voice onset time: A study of isolated word production”, *Perceptual and Motor Skills* **86**, 651–654.

Wu, Z. and Lin, M. (1989). *Shi Yan Yu Yin Xue Gai Yao (in Chinese)*, 1–347 (Higher Education Press, Beijing, China).

Young, N. (2007). “A case study of phonological attrition of Taiwanese Mandarin in California”, *UC Berkeley Phonology Lab Annual Report* **2007**, 71–126.

Zhang, F. and Yin, P. (2009). “A study of pronunciation problems of English learners in China”, *Asian Social Science* **5**, 141–146.

Zhang, Y., Nissen, S. L., and Francis, A. L. (2008). “Acoustic characteristics of English lexical stress produced by native Mandarin speakers”, *Journal of the Acoustical Society of America* **123**, 4498–4513.



TABLE I. Native  $F_1$  and  $F_2$  norms (in Bark) for rounded vowels in Mandarin and English. The vowels compared are Mandarin and English /o<sup>u</sup>/, Mandarin and English /u/, and Mandarin /y/.

		$F_1$		$F_2$	
		Mandarin	English	Mandarin	English
/o <sup>u</sup> /	male	5.38	4.36	6.61	9.59
	female	6.72	5.06	8.02	10.60
/u/	male	3.54	3.26	4.51	10.72
	female	4.12	3.97	6.06	11.92
/y/	male	2.93	—	13.53	—
	female	3.23	—	14.71	—

TABLE II. Native VOT norms (in ms) for plosives in Mandarin and English. The laryngeal categories compared are Mandarin unaspirated, Mandarin aspirated, English voiced, and English voiceless.

	short-lag		long-lag	
	Mandarin unaspirated	English voiced	Mandarin aspirated	English voiceless
labial	10	1	106	58
coronal	7	5	113	70
dorsal	15	21	116	80
AVERAGE	11	9	112	69

TABLE III. Native centroid norms (in Bark) for post-alveolar fricatives in Mandarin and English. The places of articulation compared are Mandarin retroflex, Mandarin alveolo-palatal, and English palato-alveolar.

Mandarin		English
retroflex	alveolo-palatal	palato-alveolar
/ʂ/	/ç/	/ʃ, ʒ/
16.80	19.12	17.79

TABLE IV. Background information, native Mandarin (NM) speaker group. The NM group comprised native Mandarin speakers who were born and educated up to at least seventh grade in mainland China (MC) or Taiwan (TW).

PID	Gender	Age	Place of birth	Age of arrival	Other languages spoken
N1	F	33	MC	30	English
N2	M	32	TW	26	English
N3	M	40	MC	26	Shanghainese, English
N4	F	35	MC	34	English
N5	F	20	MC	16	Cantonese, English
N6	F	19	TW	13	English

TABLE V. Background information, heritage language (HL) speaker group. The HL group comprised Chinese Americans who were born to Mandarin-speaking parents and had prior experience with Mandarin in the home. Participants in the high-exposure (HE) subgroup had extensive exposure to Mandarin, while participants in the low-exposure (LE) subgroup had more limited exposure to Mandarin.

	PID	Gender	Age	Places lived	Language exposure at home	Current use of Mandarin		
						to grand-parents	to parents	to siblings
high exposure (HE)	H7	M	24	USA (0–6); TW (6–18); USA (18–24)	Mandarin	all the time	mostly	half the time
	H8	M	21	TW (0–9); Singapore (9–13); USA (13–21)	Mandarin, Japanese, Taiwanese	all the time	mostly	half the time
	H9	F	19	USA (0–19)	Mandarin	—	mostly	half the time

H10	F	20	TW (0-5); USA (5-20)	Mandarin, Tai- wanese	all the time	mostly	seldom
H11	M	20	TW (0-3); USA (3-20)	Mandarin, En- glish	all the time	all the time (mixed with En- glish)	seldom
H12	M	20	MC (0-3.5); USA (3.5-20)	Mandarin	all the time	mostly	—
H13	F	23	MC (0-10); USA (10-23)	Mandarin, occa- sional Fuzhounese	—	half the time	—
H14	F	20	USA (0-20)	Mandarin, occa- sional Tai- wanese	all the time	mostly	seldom

	H15	M	22	USA (0–5); Singa- pore (5–11.5); Qatar (11.5–14); TW (14–15); USA (15–22)	Mandarin	mostly	mostly	seldom
low expo- sure (LE)	H16	F	18	USA (0–18)	English, Man- darin	—	half the time	seldom
	H17	F	20	USA (0–20)	Mandarin, occa- sional Tai- wanese	all the time	seldom	seldom
	H18	M	21	USA (0–21)	Mandarin	all the time	sometimes	never
	H19	M	20	USA (0–20)	Mandarin, En- GLISH	all the time	rarely	never

H20	F	21	MC (0-2); USA (2-21)	Mandarin, En- glish	never	never	never
H21	F	20	USA (0-20)	English, occa- sional Man- darin	sometimes	never	never

---



TABLE VI. Background information, late second-language (L2) learner group. The L2 group comprised native English speakers who were born and educated in the U.S. and started to learn Mandarin after the age of 18.

PID	Gender	Age	Other languages spoken in family	Mandarin experience
L22	M	19	—	2 semesters of college-level Mandarin (18–19)
L23	F	19	—	1 intensive summer session (= 2 regular semesters) of college-level Mandarin (18)
L24	F	27	—	1 year living in Beijing (26–27), including a 3-month conversation course
L25	F	24	Cebuano	2 years of college-level Mandarin (20–23)
L26	M	19	Korean	2.5-week trip to Taiwan

TABLE VII. Critical stimuli in Experiment 1 (vowel quality). The vowels of interest were Mandarin and English /o<sup>u</sup>/, Mandarin and English /u/, and Mandarin /y/.

mid back rounded	Mandarin /o <sup>u</sup> /		English /o <sup>u</sup> /	
	豆	/to <sup>u</sup> \/	‘bean’	<i>dote</i> /do <sup>u</sup> t/
	够	/ko <sup>u</sup> \/	‘enough’	<i>goat</i> /go <sup>u</sup> t/
	剖	/p <sup>h</sup> o <sup>u</sup> ɿ/	‘to dissect’	<i>pope</i> /po <sup>u</sup> p/
	透	/t <sup>h</sup> o <sup>u</sup> \/	‘through’	<i>tote</i> /to <sup>u</sup> t/
	扣	/k <sup>h</sup> o <sup>u</sup> \/	‘button’	<i>coat</i> /ko <sup>u</sup> t/
	欧	/o <sup>u</sup> ɿ/	‘Europe’	<i>oat</i> /o <sup>u</sup> t/
	肉	/ɿo <sup>u</sup> \/	‘meat’	<i>wrote</i> /ɿo <sup>u</sup> t/
				<i>boat</i> /bo <sup>u</sup> t/
				<i>cope</i> /ko <sup>u</sup> p/
				<i>host</i> /ho <sup>u</sup> st/
high back rounded	Mandarin /u/		English /u/	
	不	/pu\	‘no; not’	<i>boot</i> /but/
	肚	/tu\	‘belly’	<i>dupe</i> /dup/
	顾	/ku\	‘to take care of’	<i>goose</i> /gus/
	瀑	/p <sup>h</sup> u\	‘waterfall’	<i>poop</i> /pup/
	兔	/t <sup>h</sup> u\	‘rabbit’	<i>toot</i> /tut/
	库	/k <sup>h</sup> u\	‘garage’	<i>coup</i> /ku/
	户	/xu\	‘household’	<i>hoot</i> /hut/
	入	/ɿu\	‘to enter’	<i>root</i> /rut/
	无	/uɿ/	‘not have’	<i>choose</i> /tʃuz/
	暑	/ʂuɿ/	‘summer’	<i>shoe</i> /ʃu/
				<i>shoot</i> /ʃut/
high front rounded	Mandarin /y/			
	绿	/ly\	‘green’	

女	/nyʌ/	‘female’
遇	/yʌ/	‘encounter’

---

---

TABLE VIII. Critical stimuli in Experiment 2 (laryngeal contrast). The laryngeal categories of interest were Mandarin unaspirated, Mandarin aspirated, English voiced, and English voiceless.

short-lag stops	Mandarin unaspirated			English voiced	
	爸	/pa\	‘father’	<i>boat</i>	/bo <sup>u</sup> t/
不	/pu\	‘no; not’	<i>boot</i>	/bu <sup>u</sup> t/	
豆	/to <sup>u</sup> \	‘bean’	<i>dote</i>	/do <sup>u</sup> t/	
肚	/tu\	‘belly’	<i>dupe</i>	/du <sup>u</sup> p/	
够	/ko <sup>u</sup> \	‘enough’	<i>goat</i>	/go <sup>u</sup> t/	
顾	/ku\	‘to take care of’	<i>goose</i>	/gu <sup>u</sup> s/	
long-lag stops	Mandarin aspirated			English voiceless	
	剖	/p <sup>h</sup> o <sup>u</sup> \	‘to dissect’	<i>pope</i>	/po <sup>u</sup> p/
瀑	/p <sup>h</sup> u\	‘waterfall’	<i>poop</i>	/pu <sup>u</sup> p/	
透	/t <sup>h</sup> o <sup>u</sup> \	‘through’	<i>tote</i>	/to <sup>u</sup> t/	
兔	/t <sup>h</sup> u\	‘rabbit’	<i>toot</i>	/tu <sup>u</sup> t/	
扣	/k <sup>h</sup> o <sup>u</sup> \	‘button’	<i>coat</i>	/ko <sup>u</sup> t/	
库	/k <sup>h</sup> u\	‘garage’	<i>coup</i>	/ku/	

TABLE IX. Critical stimuli in Experiment 3 (place of articulation). The places of articulation of interest were the post-alveolar places: Mandarin retroflex, Mandarin alveolo-palatal, and English palato-alveolar.

Mandarin			English				
retroflex			alveolo-palatal			palato-alveolar	
煞	/ʂa˨˥/	‘suddenly’	下	/ɕa˨˥/	‘below’	<i>shot</i>	/ʃɑt/
沙	/ʂa˨˥/	‘sand’	瞎	/ɕa˨˥/	‘shrimp’	<i>shop</i>	/ʃɑp/
啥	/ʂa˨˥/	‘what’	辖	/ɕa˨˥/	‘govern’		
傻	/ʂa˨˥/	‘stupid’					

## LIST OF FIGURES

FIG. 1	Bark plot of the first two formants in mean productions of Mandarin /o <sup>u</sup> / and English /o <sup>u</sup> / . . . . .	19
FIG. 2	Bark plot of the first two formants in mean productions of Mandarin /u/, English /u/, and Mandarin /y/ . . . . .	20
FIG. 3	Mean differences in $F_2$ produced between Mandarin and English back rounded vowels, by participant group . . . . .	23
FIG. 4	VOT in Mandarin aspirated plosives and English voiceless plosives, by participant . . . . .	26
FIG. 5	Mean differences in VOT produced between Mandarin and English plosives, by participant group . . . . .	27
FIG. 6	Mean differences in centroid and peak amplitude frequency produced between Mandarin /ʂ/ and English /f/, by participant group . . . . .	30
FIG. 7	Centroids in Mandarin /ʂ/ and English /f/, by participant . . . . .	31