UC Berkeley
UC Berkeley PhonLab Annual Report

Production of phonetic and phonological contrast by heritage speakers of Mandarin

Charles B. Chang,[a)] Yao Yao, Erin F. Haynes, and Russell Rhodes

University of California,

Berkeley

Department of Linguistics

1203 Dwinelle Hall #2650

Berkeley,

CA 94720

(Dated: December 19, 2009)

Abstract

Production of Mandarin and English by heritage speakers of Mandarin was compared to that of native Mandarin speakers and native English-speaking late learners of Mandarin in three experiments. In Experiment 1, a spectral distinction between corresponding back vowels in Mandarin and English was produced by participants in all groups, but the greatest acoustic perceptual separation between similar vowels was achieved by heritage speakers. In Experiment 2, Mandarin unaspirated and English voiced plosives were distinguished in voice onset time by few participants; however, Mandarin aspirated and English voiceless plosives were distinguished by native Mandarin speakers and heritage speakers, both groups putting more acoustic distance between the two categories than late learners. In Experiment 3, the Mandarin retroflex and alveolo-palatal fricatives, as well as the Mandarin alveolo-palatal and English palato-alveolar fricatives, were produced as spectrally distinct by nearly all participants. The Mandarin retroflex and English palato-alveolar fricatives were distinguished by fewer participants, but by more heritage speakers and late learners than native Mandarin speakers. Thus, overall heritage speakers were found to be the most successful at simultaneously maintaining language-internal and cross-linguistic contrasts, a result which may stem from an acute 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 (HL) speakers—that is, individuals whose current primary language differs from the language they spoke or only heard as a child (the 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 (e.g. 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).

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.*, 2002; Knightly *et al.*, 2003; Oh and Au, 2005; Au *et al.*, 2008). The majority of this research has come out of joint work by Au, Jun, Knightly, Oh, and Romo on HL speakers of Korean and Spanish. In this 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

---

[a)]`cbchang@cal.berkeley.edu`

tend to have a phonological advantage over L2 learners in perception and production of the target language.

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, focus on childhood overhearers of Spanish—Spanish speakers who had regular childhood experience with overhearing Spanish, but not with speaking or being spoken to—and find that these childhood overhearers are measurably better than L2 learners at producing individual Spanish phonemes as well as whole Spanish narratives. Similarly, Oh *et al.* (2003) find that HL speakers of Korean have a phonological advantage over L2 learners; however, they examine not only childhood hearers, but also childhood speakers who spoke Korean regularly during childhood. Comparing these two HL groups, they find that while childhood speakers are measurably better than L2 learners in both perception and production of Korean, they also best childhood hearers in production. Childhood hearers, on the other hand, are better than L2 learners only in perception. This discrepancy with the results of Knightly *et al.* (2003) is 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 between the two contrasts examined (a two-way laryngeal contrast in Spanish vs. a three-way laryngeal contrast in Korean).

Although studies on HL phonology have investigated the authenticity of HL speakers' production as compared to native speakers, few have explicitly examined the question of categorical neutralization. This question is somewhat trivial if only HL categories are considered and HL speakers are judged as having native-like accents, since hitting HL phoneme targets in a native-like way most likely means that HL phonemes are not being merged. However, the question becomes non-trivial once the categories of the dominant language are also considered: 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? Godson (2003, 2004) found that for HL speakers

of Western Armenian, English (the dominant language) appeared to have an influence on their pronunciation of Armenian vowels, but only for the Armenian vowels closest to English vowels; furthermore, the influence did not necessarily neutralize contrasts between similar Armenian and English vowels.

The present study also examines the linguistic competence of HL speakers, but from a bilingual perspective. In particular, the production of both Mandarin and English phonological categories by HL speakers of Mandarin (whose dominant language is English) is compared to that of native L1 speakers of Mandarin (who are late L2 learners of English) and late L2 learners of Mandarin (who are 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, stop (plosive) voicing categories, and fricative place categories. Like previous studies on HL phonology, one objective of the current study is to see how closely HL Mandarin speakers pattern with native speakers with respect to production of language-internal phonemic contrasts. However, unlike previous HL studies, another objective is to see the extent to which HL Mandarin speakers effect cross-linguistic contrast between similar categories in their two languages. In this respect, the present study bears similarities to research on bilingual phonology, much of which has also been concerned with this question (for an extensive overview of these studies, see Laeufer, 1996).

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 eliminated via the detailed sort of screening of participants used in previous studies—is 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 (cf. Section II.A) was that their HL experience was with Mandarin, as opposed to another variety of Chinese. This serves to maximize the generalizability of the results, which have emerged in spite of the variability purposefully left within this HL speaker sample.

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. All participants were presented with identical stimuli, described in Section II.C.

For the purposes of analysis, participants were divided into groups according to their responses on a detailed questionnaire about their language background, current language use, and comprehension of Mandarin in formal and informal settings. The six participants in the native Mandarin (NM) group were native Mandarin speakers who were born and educated in mainland China or Taiwan; one came to the U.S. at the age of 14 years, while the other five came after college. HL speaker participants reported speaking English most of the time overall, but they were born to Mandarin-speaking parents; six of them were also born in a Mandarin-speaking region and spent the first 3–10 years of life there. The nine participants in the high-exposure (HE) HL group were heritage speakers who had extensive exposure to Mandarin as children and reported that they used Mandarin to communicate with both parents most or all of the time, while the six participants in the low-exposure (LE) HL group were heritage speakers who had limited exposure to the language and whose

reported frequency of Mandarin use with parents and other family was half of the time or less. The five participants in the second-language (L2) learner group were native English speakers who were born and educated in the U.S., grew up in monolingual English-speaking families, and started to learn Mandarin in high school or college.

B. Procedure

Study participants were recorded reading aloud 59 Mandarin words and phrases and 32 English words 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 Romanization (pinyin and/or BoPoMoFo). The recordings were made in a sound-proof booth with 48-kHz sampling and 16-bit resolution using professional-quality recording equipment and a head-mounted condenser microphone.

C. Stimuli

In choosing Mandarin and English stimuli, segmental context was matched across language as much as possible, and Mandarin words with falling tones were selected when such words existed (e.g. English *boot* [b̥ut] vs. Mandarin 不 [pu˥˩] 'not'; English *tote* [tʰoᵘt] vs. Mandarin 透 [tʰoᵘ˥˩] 'transparent'; English *shot* [ʃɑt] vs. Mandarin 煞 [ʂa˥˩] 'suddenly' and 下 [ɕa˥˩] 'below'). Critical Mandarin stimuli were of the form CV, while critical English stimuli were generally of the form CVC.

In Experiment 1, critical stimuli contained one of five rounded vowel categories: Mandarin /u/ appeared in ten items, Mandarin /oᵘ/ in seven, Mandarin /y/ in three, English /u/ in eleven, and English /oᵘ/ in ten. 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 /pʰ, tʰ, kʰ/, English voiced /b, d, g/, and English voiceless /p, t, k/. 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 /ɕ/, and English palato-alveolar /ʃ/. These fricatives appeared prevocalically in seven Mandarin words and two English words.

D. Acoustic analysis

All acoustic measurements were taken manually in Praat (Boersma and Weenink, 2008). In Experiment 1, average values of the first (F1), second (F2), and third (F3) formants were measured 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, voice onset time (VOT) was measured for word-initial plosives as time at the onset of periodicity minus time at plosive release. In Experiment 3, peak amplitude frequency (PAF) and centroid frequency (Ladefoged, 2005) 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 F2 region (so as to get a better measure of specifically front cavity resonances varying with place of articulation). The location of the F2 region (the endpoint of the band filter) was estimated for each subject as $\frac{3}{5}$ of the speaker's average F3 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.

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 F1 measurements (81% were less than 25 Hz apart), 24 Hz in F2 measurements (63% were less than 25 Hz apart), and 26 Hz in F3 measurements (62% were less than 25 Hz apart). In Experiment 2, 9% of the total number 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

Differences between Mandarin and English back rounded vowels in terms of F1 and F2 are summarized in Table I (Mandarin figures from Wu and Lin 1989, English figures from Hagiwara 1997, cf. also Labov *et al.* 2006). On average, Mandarin /u/ and English /u/ are quite similar in F1, but differ substantially in F2: the average F2 for English /u/ is approximately 1000 Hz higher than that of Mandarin /u/ for both male and female speakers. On the other hand, Mandarin /oᵘ/ and English /oᵘ/ differ in both F1 and F2, English /oᵘ/ being 100–200 Hz lower in F1 and approximately 475 Hz higher in F2. Thus, to closely approximately these phonetic norms, speakers with experience in both languages should produce a slight difference in F1 and a considerable difference in F2 between the two mid vowels, as well as a large difference in F2 between the two high vowels.

Mean F1 and F2 in participants' productions of the mid rounded vowels are plotted in Figure 1. For each group the Mandarin and English vowels occupy distinct phonetic spaces,
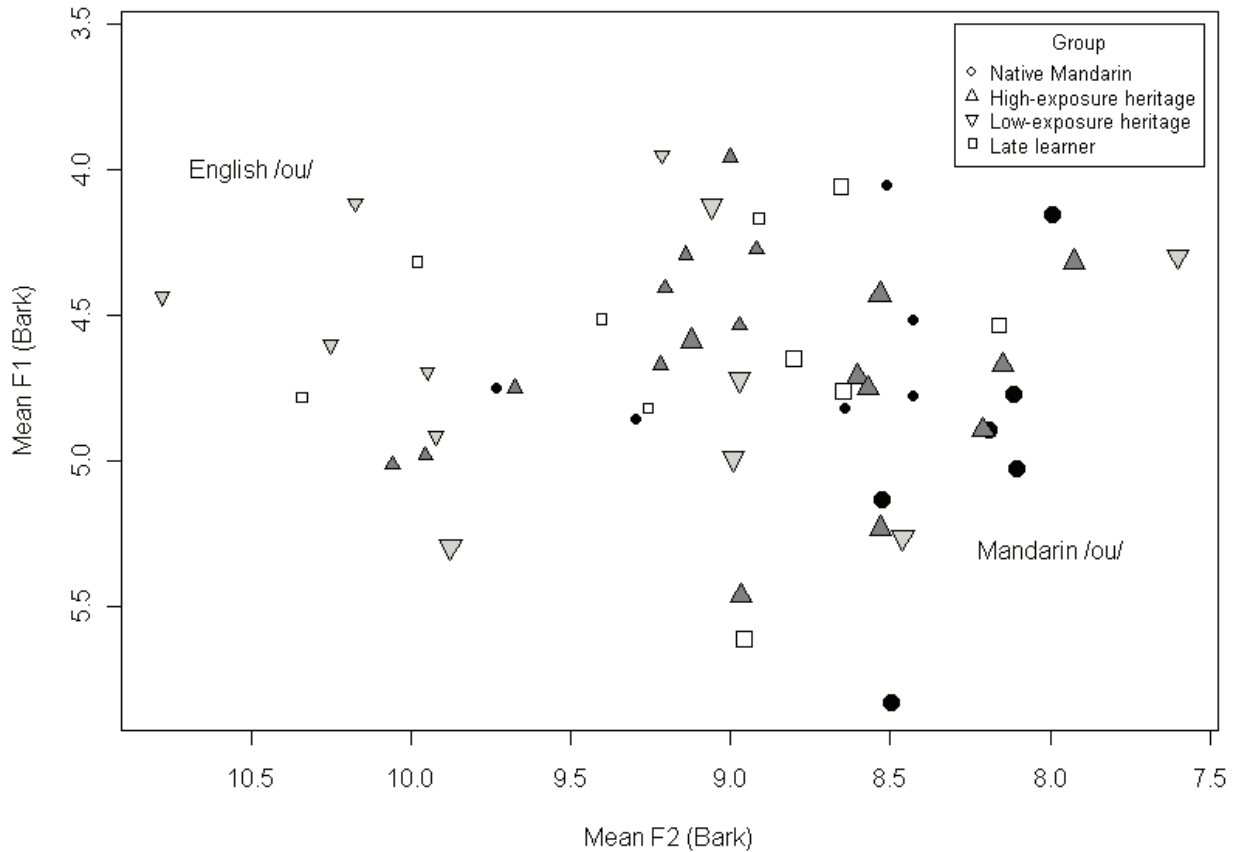
FIG. 1. Bark plot of the first two formants in mean productions of Mandarin /oᵘ/ and English /oᵘ/. NM speakers are plotted in black circles, HE speakers in dark gray triangles, LE speakers in light gray upside-down triangles, and L2 learners in white squares. Mandarin productions are plotted in larger symbols, English productions in smaller symbols.

English /oᵘ/ being produced with higher F2 values than Mandarin /oᵘ/. The NM and L2 groups each produce the /oᵘ/ of their non-native language with F2 values approximating the /oᵘ/ of their native language, while HL speakers pattern somewhat in between these two groups. For example, in the case of Mandarin /oᵘ/, most NM speakers have lower F2 values of approximately 8.0–8.5 Bark, whereas most L2 learners have higher F2 values of approximately 8.6–9.0 Bark. Although the majority of HE speakers are located in the same region as NM speakers, the majority of LE speakers are located in the same region as L2 speakers; both these groups, however, span a wide phonetic space that extends across the
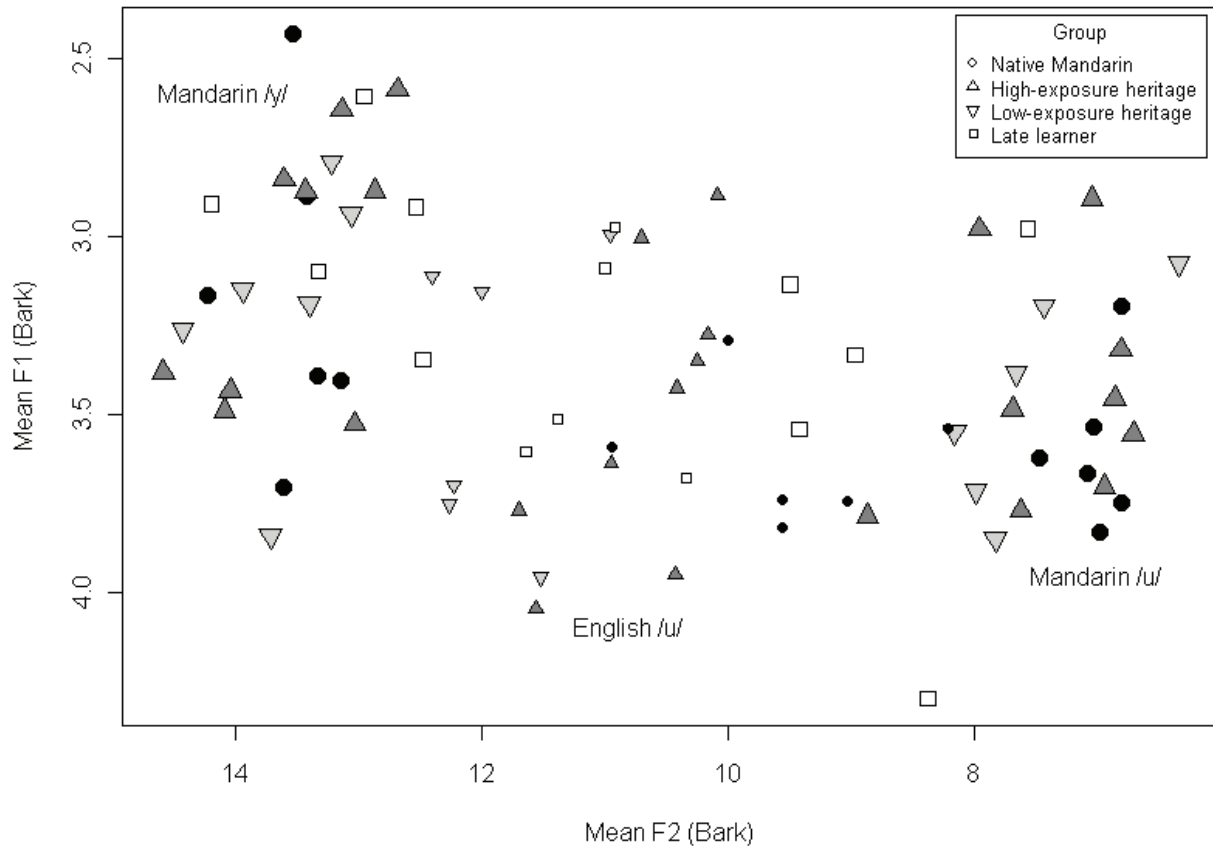
FIG. 2. Bark plot of the first two formants in mean productions of Mandarin /u/, English /u/, and Mandarin /y/. NM speakers are plotted in black circles, HE speakers in dark gray triangles, LE speakers in light gray upside-down triangles, and L2 learners in white squares. Mandarin productions are plotted in larger symbols, English productions in smaller symbols.

regions occupied by NM and L2 speakers. Figure 1 also shows some differentiation of the two vowels in terms of F1. In accordance with the small difference in native F1 targets seen in Table I, for all speaker groups the space for Mandarin /oᵘ/ extends into a higher F1 region than the space for English /oᵘ/.

Mean F1 and F2 in participants' productions of the high rounded vowels are plotted in Figure 2. There are several patterns to note here. First, all groups distinguish Mandarin /u/ and English /u/, producing the latter with substantially higher F2 values. However, the groups differ in terms of their location in F1-F2 space. NM speakers produce both vowels

with the lowest F2 values, while L2 learners (native English speakers) produce both vowels with the highest F2 values, with HL speakers located somewhat in between these two groups for both vowels. Thus, similar to the case of /oᵘ/, both NM speakers and L2 learners appear 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 produce English /u/ with a relatively low F2 approximating the low F2 of Mandarin /u/, whereas L2 learners produce Mandarin /u/ with a relatively high F2 approximating the high F2 of English /u/. On the other hand, HL speakers generally produce Mandarin /u/ and English /u/ with F2 values that are relatively close to the native ones. To put it another way, for most HL speakers F2 for Mandarin /u/ is not as high as it is for L2 learners, nor is F2 for English /u/ as low as it is for NM speakers.

With regard to the Mandarin high front rounded vowel /y/, all groups produce this vowel in a distinct phonetic space with much higher F2 values than the Mandarin and English back vowels, and the groups do not differ from each other appreciably with respect to their location in F1-F2 space. The results of a two-way analysis of variance (with factors Group and Gender) are consistent with this impression. Although there is a main effect of Gender on F1 [$F(1,18)=16.27$, $p<0.001$] as well as F2 [$F(1,18)=14.79$, $p=0.001$], there is no main effect of Group on either F1 or F2 and no interaction with Gender.

Formant data for the mid and back rounded vowels were subjected to repeated-measures analyses of variance (ANOVAs), with Group and Gender as between-subjects factors and Language, Vowel (/oᵘ/ or /u/), and Place (of articulation of the onset consonant) as within-subjects factors. With respect to F1, there is no main effect of Language, but there are 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 [$F(1,5)=20.28$, $p=0.006$], as expected: /oᵘ/ > /u/; velar > alveolar > bilabial > glottal/post-alveolar; and female > male. As one would predict from the formant norms cited in Table I, there is also a two-way interaction between Language and Vowel [$F(1,5)=73.69$, $p<0.001$], which is attributable to only Mandarin /u/ and English /u/ not being produced with distinct F1s. Males do better than females at producing an

F1 difference between the high back vowels, resulting in a three-way interaction of Gender, Language, and Vowel [$F(1,5)=10.99$, $p=0.02$]. However, Group does not have a main effect on F1, nor does it interact significantly with any other factors.

While there is no main effect of Group on F1, there is a main effect of Group on F2 [$F(3,4)=11.24$, $p=0.02$], though no main effect of Gender. There are 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 is only marginally significant, a two-way interaction between Language and Vowel [$F(1,4)=316.25$, $p<0.001$] arises from the greater effect of Language on F2 in the case of /u/ than in the case of /oᵘ/. Group not only has a main effect on F2, it also interacts with Language [$F(3,4)=11.20$, $p=0.02$] and with Language and Vowel [$F(3,4)=15.93$, $p=0.01$]. The Group x Language interaction is attributable to the pattern seen in Figures 1 and 2: English back vowels are produced with greater F2s than Mandarin back vowels in all groups, but this language effect differs across the groups, which produce disparate F2 values and unequal distances between languages. The Group x Language x Vowel interaction arises from the fact that the Group x Language interaction is more pronounced for /u/ than for /oᵘ/.[1]

To examine between-group differences in the realization of cross-linguistic contrasts between similar vowel categories, the mean differences in F1 and F2 produced between Mandarin and English /oᵘ/ and between Mandarin and English /u/ were calculated for each participant. One-way ANOVAs show a highly significant main effect of Group on F2 distances between Mandarin and English /u/ [$F(3,22)=7.85$, $p<0.001$], but not on F2 distances between Mandarin and English /oᵘ/. These mean F2 distances are presented in Figure 3, where it can be seen that both the HE group and the LE group put more acoustic perceptual distance between Mandarin and English /u/ than does the L2 group [HE vs. L2: Mann–Whitney $U=38$, $n_1=9$, $n_2=5$, $p=0.04$ two-tailed; LE vs. L2: Mann–Whitney $U=30$, $n_1=6$, $n_2=5$, $p=0.004$ two-tailed]. The LE group also tops the NM group [Mann–Whitney $U=34$, $n_1=n_2=6$, $p=0.009$ two-tailed] and the HE group [Mann–Whitney $U=45$, $n_1=6$,
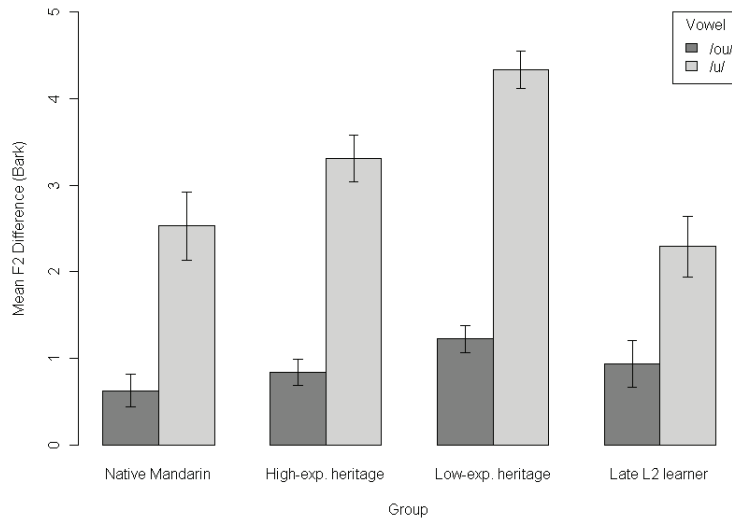
FIG. 3. Mean differences in F2 produced between Mandarin and English back rounded vowels, by participant group (from left to right: NM, HE, LE, L2). Differences between Mandarin and English /oᵘ/ are in dark gray bars, differences between Mandarin and English /u/ in light gray bars. Error bars indicate ±1 standard error about the mean.

n_2=9, p=0.04 two-tailed] in this regard. In short, HL speakers separate their two high back rounded vowels in F2 to a greater degree than L2 learners do, and LE speakers in particular also establish greater F2 distance than NM speakers.

B. Experiment 2: plosives

Differences between Mandarin and English plosives in terms of VOT are summarized in Table II (Mandarin figures from Wu and Lin 1989, English figures from Lisker and Abramson 1964 and Byrd 1993). Of the two short-lag VOT categories (unaspirated and voiced), Mandarin unaspirated plosives are on average characterized by the lower VOT, with the VOT of English voiced plosives being similar, but 8–17 ms greater at the same place of articulation. With respect to the long-lag VOT categories (aspirated and voiceless), 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 realize a slight 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 repeated-measures ANOVA, with Group and Gender as between-subjects factors and Language, Voicing Type (short-lag vs. long-lag), Place (of articulation), and Vowel (environment) as within-subjects factors. As expected, the ANOVA results show 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=0.004]. These main effects are all in the expected direction: Mandarin > English; long-lag > short-lag; velar > alveolar > bilabial; and /u/ > /oᵘ/. There is also a two-way interaction between Language and Voicing Type [$F(1,6)$=18.64, p=0.005], an effect mostly attributable to there being no significant difference in VOT between Mandarin unaspirated and English voiced stop productions.

While there are no main effects of Group or Gender on VOT, there is 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=0.04]. This interaction occurs because there are 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 shows a reliable difference only between HE and L2 females and only with respect to Mandarin long-lag velar stops preceding /u/ [p=0.03].

When the VOT data are examined by participant, it is apparent that there is a strong tendency for HL speakers to make a VOT distinction between cross-linguistically similar laryngeal categories. The short-lag categories are produced with reliably distinct VOTs by only six participants. Of these six, half come from the HE or LE groups; the other three are divided between the NM and L2 groups. The long-lag categories, on the other hand,
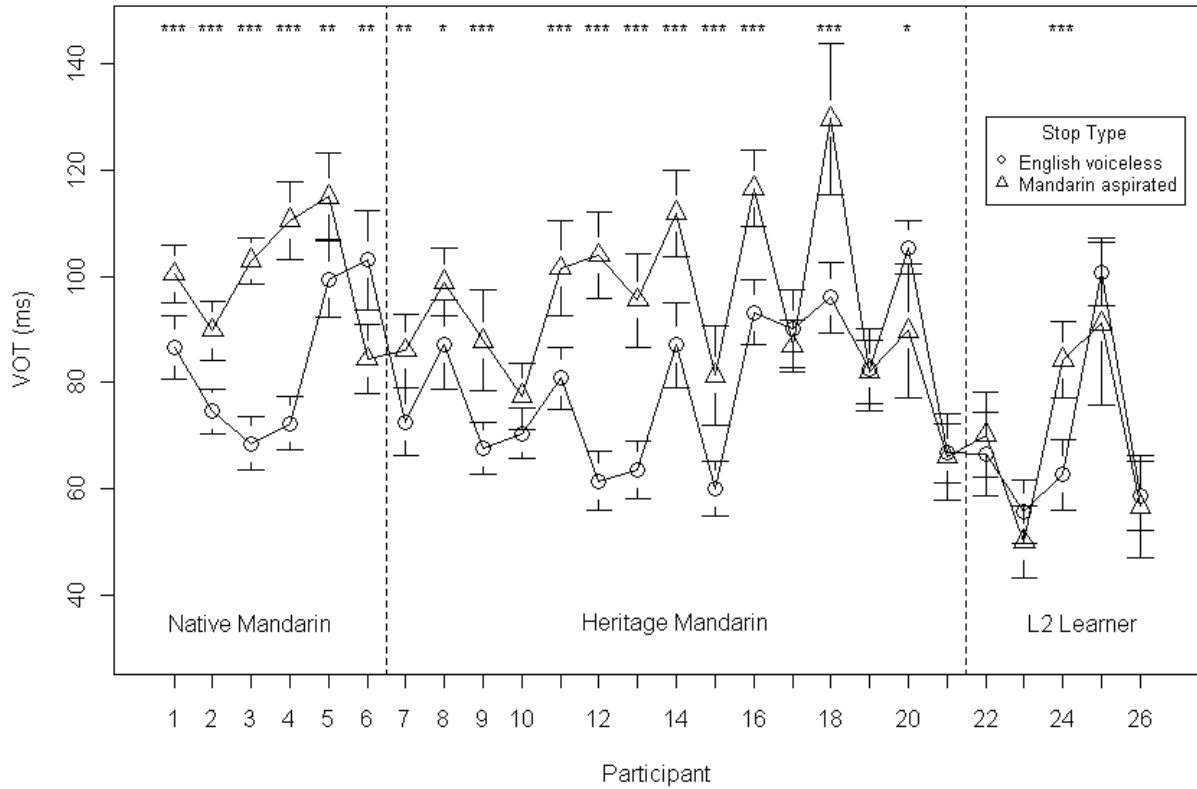
FIG. 4. VOT in Mandarin aspirated plosives (triangles) and English voiceless plosives (circles), by participant. Error bars indicate 95% confidence intervals. Participants who produce reliably different means are marked with stars: $p<0.05$ (*), $p<0.01$ (**), $p<0.001$ (***).

are distinguished by 18 participants (cf. Figure 4). These participants are concentrated in the NM, HE, and LE groups, such that all NM speakers, all but one HE speaker, and half of LE speakers produce a reliable difference in VOT between the long-lag categories; in contrast, all but one L2 learner produces no reliable difference in VOT between the two categories. Note that this pattern still holds after adjusting for multiple comparisons. When the Bonferroni correction is applied, five NM speakers and eight HL speakers, but only one L2 learner, are found to produce a significant difference in VOT.

The results of Experiment 2 thus indicate that while all participants reliably produce
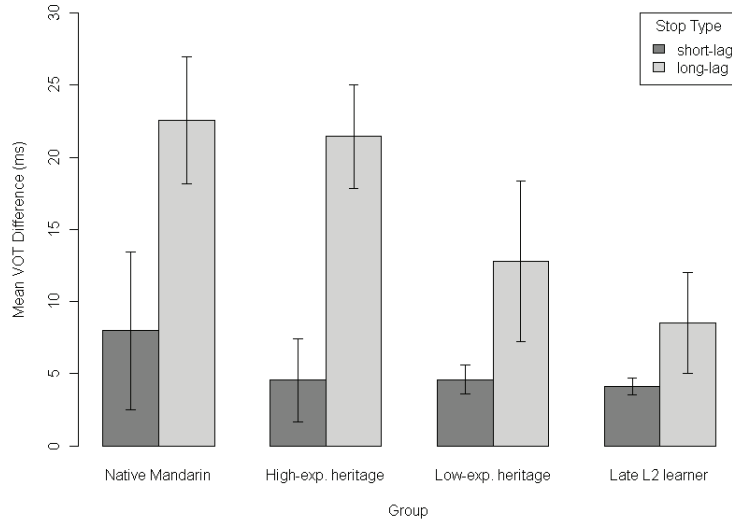
FIG. 5. Mean differences in VOT produced between Mandarin and English plosives, by participant group (from left to right: NM, HE, LE, L2). Differences between Mandarin unaspirated and English voiced plosives are in dark gray, differences between Mandarin aspirated and English voiceless plosives in light gray. Error bars indicate ±1 standard error about the mean.

the language-internal contrasts between Mandarin unaspirated and aspirated plosives and between English voiced and voiceless plosives, the same cannot be said of their realization of cross-linguistic contrasts. Few make the cross-linguistic contrast between the short-lag categories of Mandarin and English. On the other hand, many participants produce a contrast between the long-lag categories. However, these are nearly all participants with the greatest Mandarin experience—namely, NM and HL speakers. Most L2 learners fail 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 in Figure 5. A one-way ANOVA shows 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 produces 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 does the NM group [Mann–Whitney $U=38$, $n_1=n_2=6$, $p=0.04$ two-tailed]. These results are consistent with the findings of the participant analysis described above: NM speakers and HL speakers—HE ones, in particular—establish a greater acoustic distance between the long-lag VOT categories of Mandarin and English than do L2 learners of Mandarin.

C. Experiment 3: fricatives

Differences between Mandarin and English post-alveolar fricatives in terms of centroid frequency are summarized in Table III (Mandarin figures averaged from Svantesson 1986, English figures from Jongman *et al.* 2000).[2] Mandarin /ɕ/ is characterized by the highest centroid frequency, followed by English /ʃ/ and then Mandarin /ʂ/. 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.

Centroid and PAF data[3] were subjected to repeated-measures ANOVAs, with Group and Gender as between-subjects factors and Fricative as a within-subjects factor. As expected, the ANOVA results show 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, /ɕ/ > /ʂ/ > /ʃ/ (in contrast to the predictions of Table III). There is also a main effect of Gender on PAF [$F(1,14)=13.99$, $p=0.002$]: female > male. There is no main effect of Group on centroid or PAF, although there is a significant three-way interaction between Group, Gender, and Fricative with respect to PAF [$F(6,36)=3.07$, $p=0.02$], 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 shows no reliable difference between HE and L2 males or between HE and L2 females with respect to /ɕ/ or /ʃ/, but does show a reliable difference between HE and L2 males with respect to /ʂ/ [$p<0.001$].

Between-group differences in the realization of cross-linguistic contrasts between similar
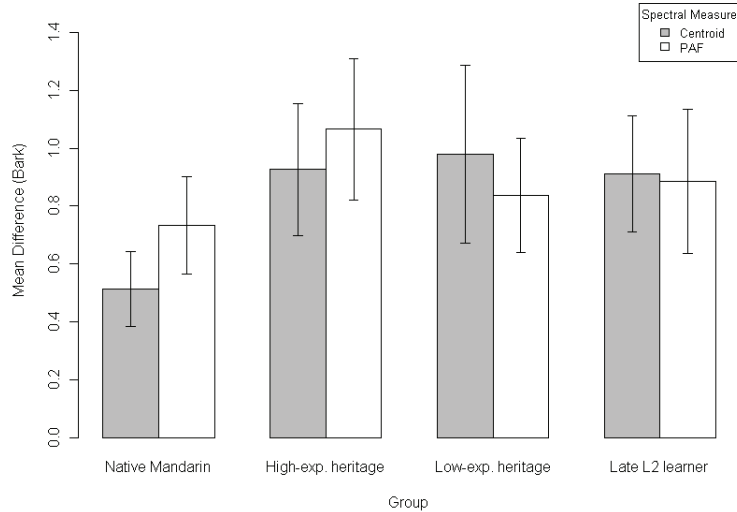
FIG. 6. Mean differences in centroid and PAF produced between Mandarin /ʂ/ and English /ʃ/, by participant group (from left to right: NM, HE, LE, L2). Differences in centroid are in gray bars, differences in PAF in white bars. Error bars indicate ±1 standard error about the mean.

fricative categories (especially Mandarin /ʂ/ and English /ʃ/) were examined by calculating for each participant the mean distances in centroid and PAF established between each pair of fricatives. In contrast to the results of Young (2007), which suggest that HL speakers might tend to merge Mandarin /ɕ/ to English /ʃ/, the HL speakers in this study do not differ from other groups with respect to producing acoustic perceptual distance between /ɕ/ and /ʃ/; all produce a robust contrast between these two fricatives. With respect to /ʂ/ and /ʃ/, on the other hand, the HL and L2 groups appear to separate these categories to a greater degree than the NM group, particularly with respect to centroid (cf. Figure 6). Nevertheless, one-way ANOVAs show no main effects of Group on centroid distances or PAF distances.

However, when the centroid data for /ʂ/ and /ʃ/ are examined by participant, it is apparent that HL speakers and L2 learners more often make a distinction between the two fricatives than NM speakers. These fricatives are distinguished in centroid by a total of 14 participants, who are unevenly distributed across groups (cf. Figure 7). Whereas the
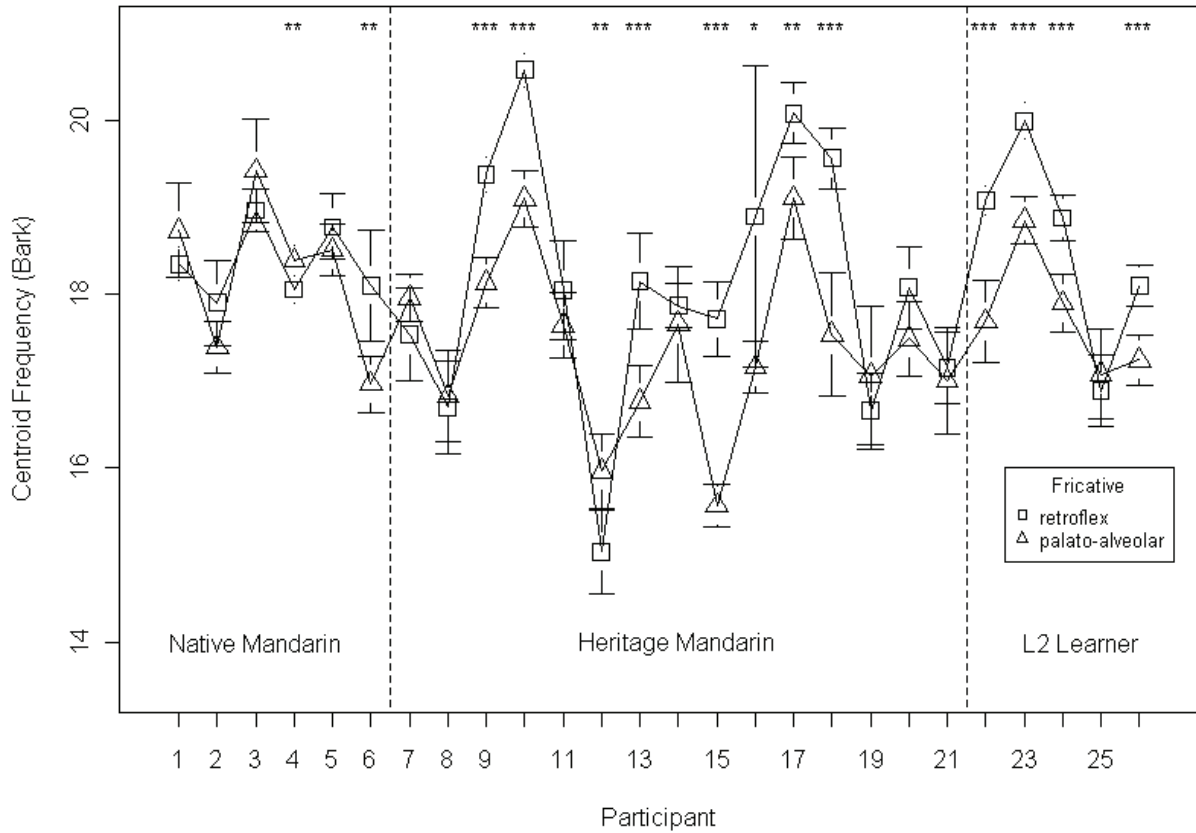
FIG. 7. Centroids in Mandarin /ʂ/ (squares) and English /ʃ/ (triangles), by participant. Error bars indicate 95% confidence intervals. Participants who produce reliably different means are marked with stars: $p<0.05$ (*), $p<0.01$ (**), $p<0.001$ (***).

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

In short, the results of Experiment 3 show no overall differences between groups in the realization of contrast between Mandarin and English post-alveolar fricatives with respect to centroid and PAF. However, on an individual level HL speakers and L2 learners are found

more often to achieve a reliable distinction between /ʂ/ and /ʃ/ than NM speakers.

IV. DISCUSSION AND CONCLUSIONS

To summarize, in Experiments 1–3 evidence was found that HL speakers are better than NM speakers and L2 learners at producing within- and cross-language contrasts. In Experiment 1, participants in all groups were found to make an F2 distinction between Mandarin and English back vowels, with NM speakers' back vowels in both languages having lower F2 values than those of HL speakers and L2 learners. Here HL speakers—in particular, LE speakers—clearly outperformed both NM speakers and L2 learners in achieving acoustic perceptual separation between similar vowel categories. In Experiment 2, few participants distinguished Mandarin unaspirated and English voiced plosives, but HL and NM speakers did distinguish Mandarin aspirated and English voiceless plosives and put more acoustic distance between these categories than L2 learners, who mostly failed to distinguish them.[4] 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 maintain not only language-internal "functional" contrast, but also cross-linguistic "non-functional" contrast. On the first point, HL speakers do not differ significantly from other groups, as almost no speaker in any group fails to distinguish the phonemic categories of their L1 and L2. HL speakers do not all realize categories in the same way as more L1-dominant native speakers (e.g. F2 values for Mandarin /u/ are slightly higher for several HL speakers than those of NM speakers), but on average they come very close—much closer than L2 learners—and this acute approximation of phonetic norms seems to lie at the heart of why HL speakers do better than L2 learners at maintaining contrasts between similar L1 and L2 categories, which for the most part they do not have to distinguish for the purposes of being understood.

To what extent then do L1 and L2 categories that are phonetically close to each other

undergo "equivalence classification" (Flege, 1987, 1995) or "perceptual assimilation" (Best, 1994) and interact with each other? Flege (1987) argues that bidirectional cross-linguistic influence is based on the classification of "similar" sounds as belonging to the same category. L2 phones that are "similar" to L1 phones have "an easily identifiable counterpart in L1", in contrast to "new" sounds, which "have no counterpart in L1" (Flege, 1987, 48). In a study of French and English speakers, 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 a significantly fronted French /u/ in approximation to the fronted realization of English /u/. Moreover, with regard to the realization of French vs. English /t/, speakers did not typically reach the L2 phonetic norm for VOT, 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 actually very similar to native French /y/.

In the present study, it is difficult to tell how well speakers approach the phonetic norms of Mandarin and English given the amount of inter-speaker variation and the limited nature of the acoustic targets 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, II, and 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 F2 in Mandarin /u/ is cited as approximately 450–650 Hz (equivalent to 4.5–6.2 Bark), but speakers in this study produce this vowel with F2 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 produce these with VOTs of approximately 15–25 ms. Results for the production of Mandarin /y/ are also consistent with those of Flege (1987) for French /y/: L2 learners do not differ appreciably from NM speakers in their phonetic space for this "new" sound. What is most significant (though perhaps not unexpected) is that when taken together, the

results of Experiments 1–3 show HL speakers to be the most successful at approximating the phonetic norms of both of their languages.

As for why HL speakers tend to be better 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. Alternatively, similar categories that are acquired early may interact with each other in a shared phonological system and dissimilate. The results of Experiments 2 and 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 show signs that some HL speakers have dissimilated similar vowel categories, resulting in a "polarized" phonetic space that goes past native targets (cf. Laeufer, 1996). In both Figures 1 and 2, it can be seen that there are HL speakers who have gone lower in F2 for their Mandarin vowels than the NM group, as well as HL speakers who have gone higher in F2 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, acute 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.

Finally, 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, has 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 bona fide L2 English, whereas HL speakers were probably exposed to no worse than native Mandarin that had "drifted" (Sancier and Fowler, 1997) in an English-speaking environment.

Endnotes

1. For the dependent variables of F1 and F2 in Experiment 1 as well as the dependent variable of VOT in Experiment 2 (cf. Section III.B), there are 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.

2. Note that the average centroid for /ʃ/ specifically is likely to be slightly higher than the figure given in Table III, which is an average that includes the corresponding voiced fricative /ʒ/ (whose centroid will be drawn down by the lower frequencies of f_0).

3. Formant transition data show that alveolo-palatal /ɕ/, having by far the lowest F1 onset (approximately 1 Bark lower than those of /ʂ/ and /ʃ/ in all groups) and highest F2 onset (approximately 2 Bark higher than those of /ʂ/ and /ʃ/ in all groups), is clearly the most "palatalized" of the three post-alveolar fricatives. However, the formant data do not clearly differentiate /ʂ/ and /ʃ/. Thus, the focus in this study is on centroid frequency data, which are supplemented with PAF data as described in Section II.D.

4. As for why L2 learners of English (NM speakers) should pattern differently from L2 learners of Mandarin in distinguishing the two long-lag VOT categories, 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 ones. 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 both are aspirated enough to pass the VOT boundary that is salient for them.

References

Andrews, D. R. (**1999**). *Sociocultural Perspectives on Language Change in Diaspora: Soviet Immigrants in the United States* (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).

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).

Byrd, D. (**1993**). "54,000 American stops", UCLA Working Papers in Phonetics **83**, 97–116.

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., 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.

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: Theoretical and Methodological Issues in Cross-Language Speech Perception*, edited by W. Strange, 233–272 (York, Baltimore, MD).

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.

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

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

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.

Labov, W., Ash, S., and Boberg, C. (**2006**). *The Atlas of North American English: Phonetics, Phonology and Sound Change* (Mouton de Gruyter, Berlin).

Ladefoged, P. (**2005**). *Vowels and Consonants*, 2nd edition (Blackwell Publishing, Malden, MA).

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).

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 16th International Congress*

of Phonetic Sciences, edited by J. Trouvain and W. J. Barry, 917–920 (Pirrot, Dudweiler, Germany).

Lisker, L. and Abramson, A. S. (**1964**). "A cross-language study of voicing in initial stops: Acoustical measurements", Word **20**, 384–422.

Montrul, S. A. (**2008**). *Incomplete Acquisition in Bilingualism: Re-examining the Age Factor* (John Benjamins Publishing, Amsterdam).

Oh, J., Jun, S.-A., Knightly, L., and Au, T. (**2003**). "Holding on to childhood language memory", Cognition **86**, B53–B64.

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 26th Annual Boston University Conference on Language Development: Volume 2*, edited by B. Skarabela, S. Fish, and A. H.-J. Do, 464–472 (Cascadilla Press, Somerville, MA).

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.

Polinsky, M. (**2008**). "Gender under incomplete acquisition: Heritage speakers' knowledge of noun categorization", Heritage Language Journal **6**, 40–71.

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.

Svantesson, J.-O. (**1986**). "Acoustic analysis of Chinese fricatives and affricates", Journal of Chinese Linguistics **14**, 53–70.

Wu, Z. and Lin, M. (**1989**). *Introduction to Experimental Phonetics (in Chinese)* (Higher Education Press, Beijing).

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

TABLE I. Native F1 and F2 norms (in Hz) for back rounded vowels in Mandarin and English.

	/u/		/oᵘ/	
	male	female	male	female
F1, Mandarin	351	411	554	726
F1, English	323	395	437	516
F2, Mandarin	454	639	711	917
F2, English	1417	1700	1188	1391

TABLE II. Native VOT norms (in ms) for plosives in Mandarin and English.

	labial	coronal	dorsal	AVERAGE
Mandarin unaspirated	10	7	15	11
English voiced	18	24	27	23
Mandarin aspirated	106	113	116	112
English voiceless	58	70	80	69

TABLE III. Native centroid norms (in Hz) for post-alveolar fricatives in Mandarin and English.

Mandarin		English
retroflex	alveolo-palatal	palato-alveolar
/ʂ/	/ɕ/	/ʃ, ʒ/
3585	5381	4229

List of Figures