The phonetic space of phonological categories in heritage speakers of Mandarin

Charles B. Chang, Erin F. Haynes, Yao Yao, and Russell Rhodes*
University of California, Berkeley

1 Introduction

Though previous linguistic research has produced a wide range of scholarship on second language acquisition, the field has only begun to examine heritage language acquisition in its own right. The few studies that have focused on the phonological competence of heritage speakers – that is, speakers who have had exposure to a particular language as a child, but who have shifted to another language for the majority of their communication needs – have found that childhood experience with a minority language, even if merely overhearing, provides a significant boost to a speaker's pronunciation of the language later in life in comparison to late learners with no prior experience (cf. Au *et al*. 2002, Knightly *et al*. 2003 on Spanish; Oh *et al*. 2002, 2003 on Korean). The results of these studies, which include acoustic measures such as voice onset time, perceptual measures such as presence of lenition, and overall accent ratings, indicate that heritage speakers tend to have a phonological advantage over late learners in production of the heritage language.

Curiously, though, only Godson (2003) explicitly examined the question of categorical neutralization. Though heritage speakers may end up with better accents than late learners overall, do they actually make all the phonological distinctions that a native speaker would? Furthermore, do they realize these contrasts in the same way and to the same degree? Godson found that for heritage speakers of Western Armenian, English appeared to have an influence on their pronunciation of Armenian vowels, but only for those Armenian vowels close to English vowels; furthermore, the influence did not necessarily neutralize contrasts.

In previous work extending this line of inquiry to consonants (Chang *et al*., to appear), we compared fricative production in heritage speakers of Mandarin to that of native Mandarin speakers and native English speakers who had learned Mandarin as a foreign language. We found that while native Mandarin speakers and late learners tend to merge similar Mandarin and English sounds (namely, Mandarin retroflex /ʂ/ and English palato-alveolar /ʃ/, and Mandarin /s/ and English /s/), heritage speakers tend to maintain a contrast between these similar

* We are indebted to Sharon Inkelas, Keith Johnson, participants in a fall 2007 UC Berkeley seminar on phonological learning, eight anonymous reviewers, and audiences at the Berkeley Phonetics and Phonology Forum, PLC 32, and CLS 44 for helpful comments and discussion. We are also grateful to the Mandarin speakers and learners who took part in this study. This work was supported in part by a National Science Foundation Graduate Research Fellowship to the first author. Any errors are our own.

sounds. In this study, we follow up on these findings by examining whether heritage speakers' apparent advantage in maintaining categorical contrast in consonantal place of articulation carries over to other domains of phonological contrast.

Here we report the results of two experiments designed to investigate the realization of two different types of phonemic categories: vowel categories and laryngeal categories. Experiment 1 focused on five rounded vowel categories (English /u/ and /ou/ and Mandarin /u/, /ou/, and /y/), while Experiment 2 focused on four laryngeal categories (English voiced /b, d, g/, English voiceless /p, t, k/, Mandarin unaspirated /p, t, k/, and Mandarin aspirated /p^h, t^h, k^h/).

2 Methods

2.1 Participants

A total of 18 Mandarin speakers and learners participated in this study. Five were native Mandarin speakers who were born and educated in a Mandarin-speaking country; eight were heritage speakers of Mandarin who either were born in or came to the U.S. before the age of 10; and five were native English speakers who were born and educated in the U.S. and had learned Mandarin as a foreign language in high school or college. Speakers were assigned to these groups, as well as rank-ordered within them, based on a detailed questionnaire about their language background, current language use, and comprehension of Mandarin in formal and informal situations. Participants ranged in age from 18 to 40, and none reported any history of speech or hearing impairments.

2.2 Stimuli

Participants were presented with 62 Mandarin words and phrases and 35 English words in random order via individual index cards. English words were written in English orthography, and Mandarin words in Mandarin orthography (traditional or simplified characters) and romanization (pinyin and/or BoPoMoFo). In Experiment 1, critical stimuli contained one of the five rounded vowel categories mentioned above. English /u/ appeared in eleven words, English /ou/ in ten, Mandarin /u/ in ten, Mandarin /ou/ in seven, and Mandarin /y/ in three. In Experiment 2, critical stimuli contained a word-initial plosive of one of the four laryngeal categories mentioned above. There were two words per combination of laryngeal category and place of articulation, for a total of twelve English items and twelve Mandarin items.

All critical English stimuli were of the form CVC, while all critical Mandarin stimuli were of the form CV. In choosing English and Mandarin stimuli, segmental context was matched across language as much as possible, and Mandarin items with falling tones were selected when such words existed (e.g. English *boot* [b̥ut] vs. Mandarin 不 [pu^{51}] 'not'; English *tote* [t^hout] vs. Mandarin 透 [t^hou^{51}] 'transparent'; English *goat* [g̊out] vs. Mandarin 够 [kou^{51}] 'enough').

2.3 Recording

Recording was done in a sound-proof booth at 48 kHz and 16 bps. The equipment used was either a Marantz PMD660 solid-state recorder with an AKG C420 head-mounted condenser microphone, or an M-AUDIO Mobile-Pre USB preamp audio interface with an AKG C520 head-mounted condenser microphone. Stimuli were recorded in eight blocks (four blocks of Mandarin and four blocks of English), resulting in a total of four tokens of each item. Blocks were grouped by language such that participants completed all blocks in one language before moving on to blocks in the other language, with the order of the languages (Mandarin-English or English-Mandarin) balanced across participants.

2.4 Acoustic Analysis

All measurements were taken by hand 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. In Experiment 2, voice onset time (VOT) was measured for word-initial plosives.

To ensure that the measurements taken were reliable, 25% of each of the VOT, F1, F2, and F3 measurements were double-checked by a second researcher in a pseudorandom fashion. Any discrepancy between the two researchers' measurements in excess of 5 ms for VOT or 100 Hz for formants was checked again by a third researcher, resulting in 8% of the total number of VOT measurement checks and 9% of the total number of formant checks being triple-checked. Final calculations of the differences between researchers' measurements revealed an average difference of 1.3 ms in VOT measurements (84% within 2 ms of each other), 12 Hz in F1 measurements (83% less than 25 Hz apart), 24 Hz in F2 measurements (64% less than 25 Hz apart), and 24 Hz in F3 measurements (63% less than 25 Hz apart). If after a third measurement there still remained a discrepancy between different researchers' measurements of greater than 5 ms/100 Hz, all of these measurements were discarded; however, this resulted in the disposal of less than 1% of the total number of VOT measurements and formant measurements.

3 Results

3.1 Experiment 1: vowel contrast

The differences between Mandarin and English back rounded vowels with respect to F1 and F2 are summarized in Table 1 (figures from Wu and Lin 1989 and Hagiwara 1997, cf. also Labov *et al*. 2006). On average, Mandarin /u/ and English /u/ are quite similar in F1, but very different 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, both pairs of vowels differ significantly in F2, and the two mid vowels also differ in F1, although to a lesser degree than they differ in F2.

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

Table 1: Native F1 and F2 targets (in Hz) for back rounded vowels in Mandarin and English.

Mean F1 vs. mean F2 for the high rounded vowels is graphed in Figure 1 below. There are several patterns present here. First, all three speaker groups distinguish Mandarin /u/ vs. English /u/, articulating the latter with significantly higher F2 values than the former. However, the groups differ in terms of their location in F1 vs. F2 space. Native speakers of Mandarin articulate both Mandarin /u/ and English /u/ with the lowest F2 values, while late learners of Mandarin (that is, native English speakers) articulate both vowels with the highest F2 values, with heritage speakers located somewhat in between these two groups in both cases. Thus, it appears that both native Mandarin speakers and late learners are influenced in their pronunciation of the /u/ of their second language by the phonetic characteristics of the /u/ of their first language. Specifically, native Mandarin speakers produce English /u/ with a relatively low F2 approximating that of the low-F2 /u/ of Mandarin, whereas late learners produce Mandarin /u/ with a relatively high F2 approximating that of the high-F2 /u/ of English. On the other hand, heritage speakers generally produce Mandarin /u/ and English /u/ with F2 values that are closer to the native ones. To put it another way, for most heritage speakers F2 for Mandarin /u/ is not as high as it is for most late learners, nor is F2 for English /u/ as low as it is for native Mandarin speakers.

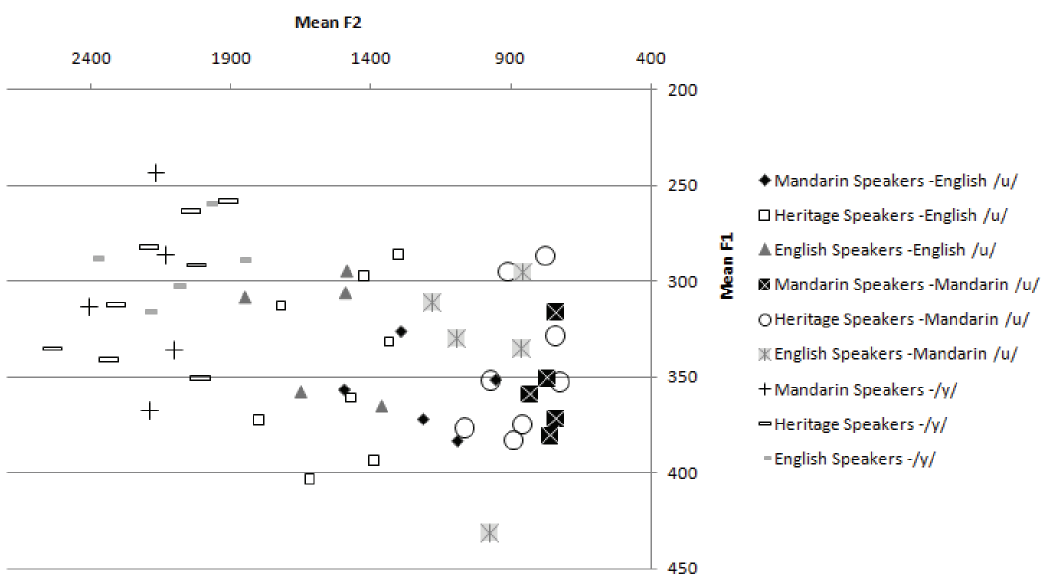

Figure 1: Mean F1 vs. mean F2 (in Hz) for Mandarin /u/, English /u/, and Mandarin /y/.

With respect to the front rounded vowel /y/, all speaker groups articulate this Mandarin vowel in a distinct phonetic space with much higher F2 values than the back vowels in both Mandarin and English, and the groups do not differ from each other appreciably with respect to their location in F1 vs. F2 space.

Graphs of mean F2 by speaker for all three of these high vowels are given in Figure 2 below, separated by speaker gender (Speakers 1-5 are native Mandarin speakers; 6-13, heritage speakers; and 14-18, late learners). Note that the difference between Mandarin /u/ and English /u/, the difference between Mandarin /u/ and /y/, and the difference between English /u/ and /y/ are highly significant in nearly all cases ($p < .01$ by the Wilcoxon matched pairs signed-rank test, cf. Table 2 below), the only exception being English /u/ vs. /y/ for Speaker 14.

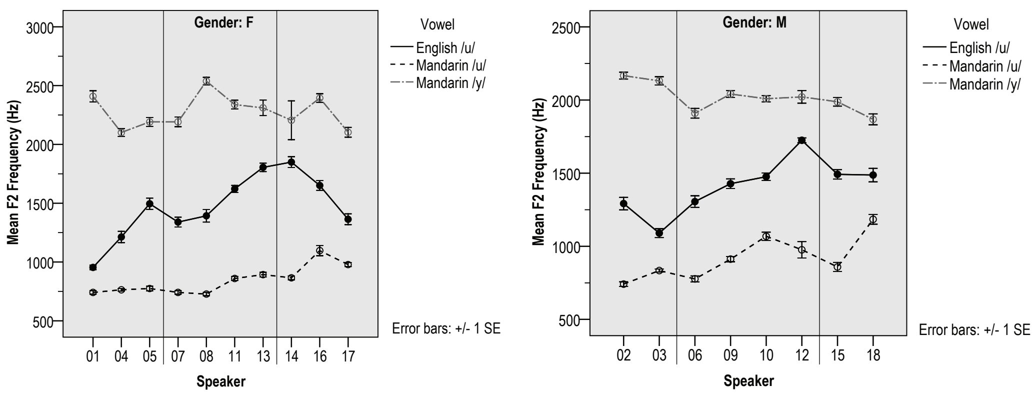

Figure 2: Mean F2 frequency for Mandarin /u/ vs. English /u/ vs. Mandarin /y/.

Mean F1 vs. mean F2 for the mid rounded vowels is graphed in Figure 3 below. The results for Mandarin /ou/ vs. English /ou/ are similar to those for Mandarin /u/ vs. English /u/. For each speaker group the Mandarin and English vowels occupy distinct phonetic spaces (English /ou/ being articulated with higher F2 values than Mandarin /ou/), and the native Mandarin speaker group and late learner group each produce the /ou/ of their non-native language with F2 values approximating the /ou/ of their native language. Here again, heritage speakers pattern somewhat in between native Mandarin speakers and late Mandarin learners. For example, in the case of Mandarin /ou/, the majority of native Mandarin speakers (three out of five) have low F2 values of approximately 900-950 Hz, the majority of late learners (four out of five) have high F2 values of approximately 1025-1100 Hz, and half of the heritage speakers are clustered around an intermediate F2 value of 1000 Hz.

Figure 3 also appears to show some differentiation of the two vowels by height. In accordance with the slight difference in native F1 targets seen in Table 1, for all speaker groups the space for Mandarin /ou/ extends into a higher F1 region than the space for English /ou/, although on the low frequency end the phonetic space for both vowels enters the 400-450 Hz region.

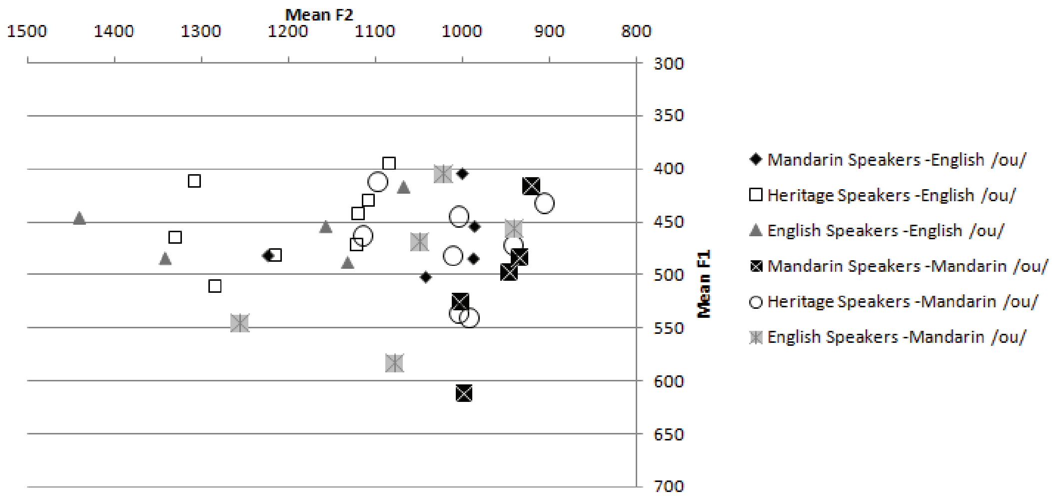

Figure 3: Mean F1 vs. mean F2 (in Hz) for Mandarin /ou/ and English /ou/.

Graphs of mean F2 by speaker for both mid vowels are given in Figure 4 below, separated by speaker gender. Unlike in the case of the high vowels, where all vowel pairs are distinguished in F2 by nearly all speakers, the mid vowels are distinguished in F2 by only 12 out of 18 speakers (cf. Table 2). These 12 speakers are not evenly distributed across the three speaker groups; instead, the majority of these "distinguishers" are clustered in the heritage speaker group, with the result that half the native Mandarin speakers and late Mandarin learners are not found to distinguish Mandarin /ou/ vs. English /ou/ in F2, whereas all heritage speakers but one do make an F2 distinction.

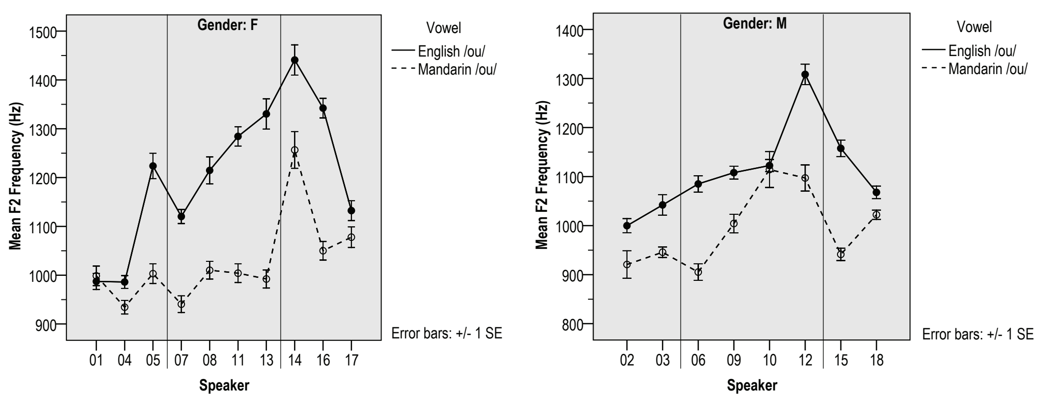

Figure 4: Mean F2 frequency for Mandarin /ou/ vs. English /ou/.

However, given that these mid vowels differ in F1 as well as F2, it is possible that the speakers who apparently merge the two in F2 nonetheless distinguish them in F1. Graphs of mean F1 by speaker for both mid vowels are given in Figure 5 below, separated by speaker gender.

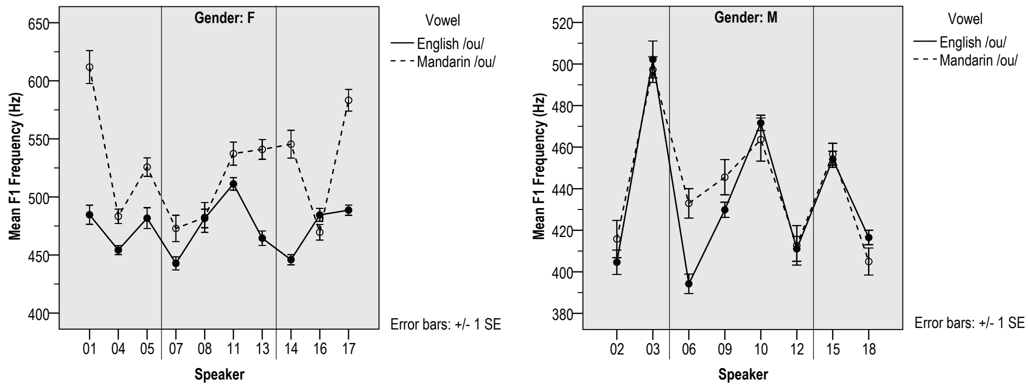

Figure 5: Mean F1 frequency for Mandarin /ou/ vs. English /ou/.

Fewer speakers distinguish the mid vowels in F1 than in F2, with only 8 speakers making an F1 distinction (cf. Table 2). However, three of these speakers (Speakers 1, 14, and 17) are among those who do not distinguish the vowels in F2. Thus, in the end, all groups do rather well in distinguishing the mid vowels: four out of five native Mandarin speakers, eight out of nine heritage speakers, and four out of five late learners distinguish the two in F1 and/or F2. The results of planned comparisons of all vowel pairs examined using the Wilcoxon matched pairs signed-rank test are summarized in Table 2 below.

Speaker	M /u/-E /u/, F2	M /y/-M /u/, F2	M /y/-E /u/, F2	M /ou/-E /ou/, F2	M /ou/-E /ou/, F1
1	**	**	**	n.s.	**
2	**	**	**	n.s.	n.s.
3	**	**	**	**	n.s.
4	**	**	**	**	**
5	**	**	**	**	**
6	**	**	**	**	**
7	**	**	**	**	n.s.
8	**	**	**	**	n.s.
9	**	**	**	**	n.s.
10	**	**	**	n.s.	n.s.
11	**	**	**	**	*
12	**	**	*	**	n.s.
13	**	**	**	**	**
14	**	**	n.s.	n.s.	**
15	**	**	**	**	n.s.
16	**	**	**	**	n.s.
17	**	**	**	n.s.	**
18	**	**	**	n.s.	n.s.

Table 2: Distinctions made between vowels in formants, by speaker and contrast.
Speakers 1-5: native Mandarin speakers; 6-13: heritage speakers; 14-18: late learners.
'M': Mandarin; 'E': English. *: $p < .05$; **: $p < .01$; n.s.: p not significant.

Thus far, the results of Experiment 1 show that most speakers produce Mandarin back vowels with lower F2 values than English back vowels, especially in the case of Mandarin /u/ vs. English /u/. Furthermore, native Mandarin speakers' back vowels have lower F2 values than those of heritage speakers and late learners in both languages. However, not only does speaker group help predict F2 values, speaker rank does as well. Spearman's correlations of speaker rank (based on experience with Mandarin) with F2 in each vowel are summarized in Table 3. There is a weak, but significant negative correlation of speaker rank with F2 in the front vowel /y/, and stronger positive correlations of speaker rank with F2 in each back vowel. In other words, as a speaker's level of experience with Mandarin increases, F2 of the front rounded vowel tends to go up, while F2 of the back rounded vowels tends to go down.

	Man. /y/	Man. /u/	Eng. /u/	Man. /ou/	Eng. /ou/
Spearman's *r*	-.143	.535	.453	.312	.435
p-value	.037	< .001	< .001	< .001	< .001

Table 3: Non-parametric correlations between speaker rank and F2, by vowel.

These data help to locate different speakers' vowel categories in phonetic space, but do not indicate how spread out this space is. The question remains: how much distance do speakers put between these distinct vowel categories? To get at an answer to this question, F2 differences between similar word pairs in Mandarin and English containing the vowel /u/ (e.g. English *boot* [b̥ut] vs. Mandarin 不 [pu^{51}] 'not') were computed from the mean F2 of a word across all four tokens. The mean F2 differences for all speaker groups are presented in Figure 6.

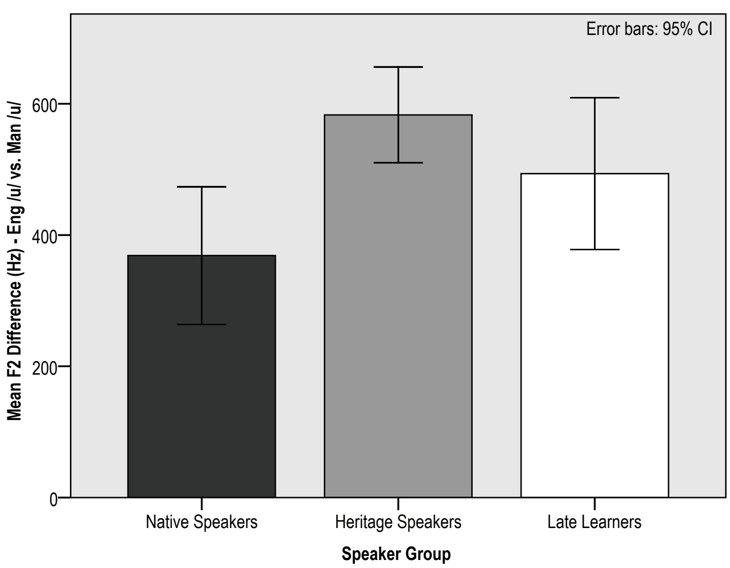

Figure 6: Mean F2 differences between Mandarin /u/ and English /u/, by speaker group.

Figure 6 suggests that heritage speakers put more acoustic distance between Mandarin /u/ and English /u/ than do native speakers and late learners. The results of a Mann-Whitney test comparing the groups indicate that the difference between native Mandarin speakers and heritage speakers is highly significant ($U = 371.000$, $Z = -3.584$, $p < .001$). However, neither the difference between native Mandarin speakers and late learners nor that between heritage speakers and late learners reaches significance (native speakers vs. late learners: $U = 326.000$, $Z = -1.833$, $p = .067$; heritage speakers vs. late learners: $U = 574.000$, $Z = -1.500$, $p = .134$). In short, heritage speakers put a significantly greater F2 separation between their two high back rounded vowels than do native Mandarin speakers, though this separation is not significantly greater than that achieved by late learners of Mandarin.

3.2 Experiment 2: laryngeal contrast

The differences between Mandarin and English laryngeal categories in terms of VOT are summarized in Table 4 (figures from Wu and Lin 1989, Byrd 1993, and Lisker and Abramson 1964). Of the two short-lag VOT categories, 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, Mandarin aspirated plosives are significantly more aspirated than English voiceless plosives, by as much as 48 ms at the same place of articulation. Thus, both pairs of similar laryngeal categories differ in VOT, although the difference between Mandarin aspirated plosives and English voiceless plosives is much greater than that between Mandarin unaspirated plosives and English voiced plosives.

Mandarin category		VOT
unaspirated	/p/	10
	/t/	7
	/k/	15
aspirated	/pʰ/	106
	/tʰ/	113
	/kʰ/	116

English category		VOT
voiced	/b/	18
	/d/	24
	/g/	27
voiceless	/p/	58
	/t/	70
	/k/	80

Table 4: Native VOT targets (in ms) for laryngeal categories in Mandarin and English.

Mean VOTs by speaker for all laryngeal categories are graphed in Figure 7 below, with the results of the Wilcoxon signed-rank test given in Table 5. Mandarin unaspirated and English voiced are distinguished by only four speakers, most of whom distinguish the two in the expected direction, with the VOT of English voiced being longer; the other 14 speakers do not produce a significant difference in VOT between these two categories. In contrast, Mandarin aspirated and English voiceless are distinguished by 13 speakers. These speakers are concentrated in the native Mandarin speaker and heritage speaker groups, such that all of the native Mandarin speakers and all but one of the heritage speakers produce a significant difference in VOT between the two, while all but one of the late learners do not. Of these speakers, the vast majority (11 of 13) distinguish the categories in the expected direction, with the VOT for Mandarin aspirated being longer.

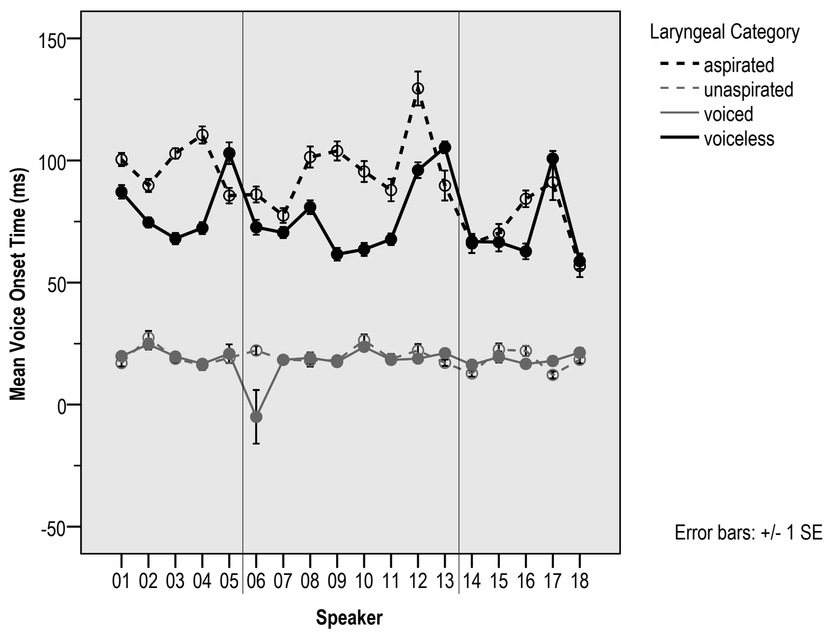

Figure 7: Mean voice onset time for Mandarin vs. English laryngeal categories.

Speaker	M *unaspirated* - E *voiced*	M *aspirated* - E *voiceless*
1	*	**
2	n.s.	**
3	n.s.	**
4	n.s.	**
5	n.s.	**
6	n.s.	**
7	n.s.	n.s.
8	n.s.	**
9	n.s.	**
10	n.s.	**
11	n.s.	**
12	n.s.	**
13	*	*
14	n.s.	n.s.
15	n.s.	n.s.
16	*	**
17	**	n.s.
18	n.s.	n.s.

Table 5: Distinctions made in VOT between laryngeal categories, by speaker and contrast. Speakers 1-5: native Mandarin speakers; 6-13: heritage speakers; 14-18: late learners. ‘M’: Mandarin; ‘E’: English. *: $p < .05$, **: $p < .01$; n.s.: p not significant.

The results of Experiment 2 thus show that all speakers distinguish Mandarin unaspirated vs. aspirated and English voiced vs. voiceless, while few distinguish Mandarin unaspirated vs. English voiced. Many speakers distinguish Mandarin aspirated vs. English voiceless, but these are nearly all speakers with the greatest Mandarin experience – namely, native Mandarin speakers and heritage speakers, most of whom have longer VOTs for Mandarin aspirated than for English voiceless. Most of the late learners fail to distinguish Mandarin aspirated vs. English voiceless. Consistent with these trends, speaker rank is negatively correlated with VOT for Mandarin aspirated. Spearman's correlations of speaker rank with VOT for each laryngeal category are summarized in Table 6. There is a weak, but highly significant negative correlation of speaker rank with VOT in Mandarin aspirated, meaning that as a speaker's level of experience with Mandarin increases, they tend to more heavily aspirate the Mandarin aspirated plosives.

	Man. *unasp*.	Man. *asp*.	Eng. *vcd*.	Eng. *vcls*.
Spearman's *r*	-.082	-.323	-.064	-.073
p-value	.090	< .001	.182	.131

Table 6: Non-parametric correlations between speaker rank and VOT, by laryngeal category.

VOT differences between word-initial consonants in similar Mandarin and English words were also computed from the mean VOT over all four tokens of a word. The mean VOT differences for all speaker groups are presented in Figure 8.

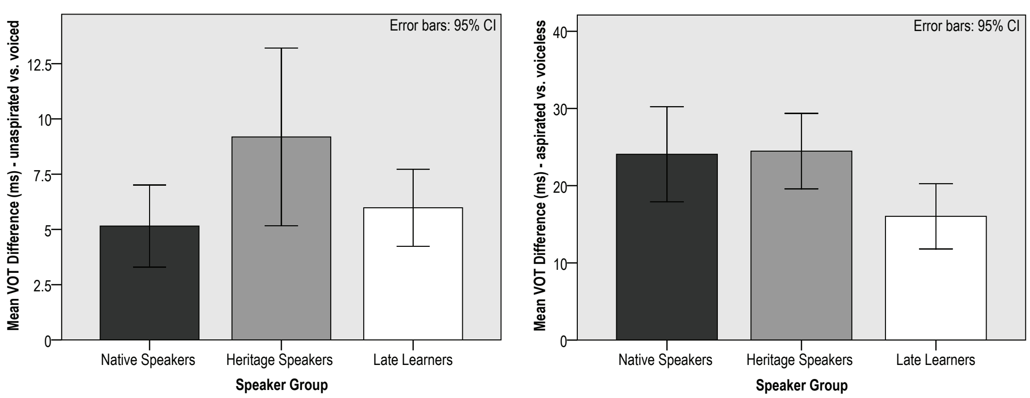

Figure 8: Mean VOT differences, by speaker group – Mandarin unaspirated vs. English voiced (on the left), Mandarin aspirated vs. English voiceless (on the right).

The results of a Mann-Whitney test indicate that with respect to VOT differences between Mandarin unaspirated and English voiced, the three groups do not differ from each other significantly (native speakers vs. heritage speakers: $U = 413.000$, $Z = -1.174$, $p = .240$; native speakers vs. late learners: $U = 265.000$, $Z = -.922$, $p = .356$; heritage speakers vs. late learners: $U = 483.000$, $Z = -.229$, $p = .819$). The higher mean VOT difference for the heritage speaker group seen in the left panel of Figure 8 is mostly due to some large VOT differences achieved by one speaker (Speaker 6), who often prevoices his English voiced plosives. With respect to VOT differences between Mandarin aspirated and English voiceless, native Mandarin speakers do not differ significantly from heritage speakers (U = 714.500, Z = -.056, p = .955); however, heritage speakers differ significantly from late learners (U = 501.500, Z = -2.244, p = .025), and the difference between native speakers and late learners approaches significance (U = 320.000, Z = -1.922, p = .055). Thus, the main result here is that heritage speakers effect a significantly greater separation in VOT between Mandarin aspirated and English voiceless than do late learners of Mandarin.

4 Discussion

To summarize, in Experiment 1 we found that speakers in all groups make an F2 distinction between Mandarin and English back vowels, with native Mandarin speakers' back vowels in both languages having lower F2 values than those of heritage speakers and late learners. However, heritage speakers were found to achieve the greatest acoustic separation between similar vowel categories. In Experiment 2, we found that while few speakers distinguish Mandarin unaspirated vs. English voiced, native Mandarin speakers and heritage speakers distinguish Mandarin aspirated vs. English voiceless, with level of Mandarin experience helping to predict VOT in the Mandarin aspirated plosives.[1] Here as well heritage speakers, along with native Mandarin speakers, put the most acoustic distance between similar laryngeal categories.

Thus, we found that heritage speakers maintain not only language-internal "functional" contrast, but also cross-linguistic "non-functional" contrast. On the first point, heritage speakers do not differ significantly from other groups, as almost no speaker in any group fails to distinguish the phonemic categories of their first (L1) and second (L2) languages. Heritage speakers do not realize categories in quite the same way as more L1-dominant native speakers (cf. F2 values for Mandarin /u/ that are somewhat higher than those of native speakers), but they come very close – much closer than late learners – and this acute approximation of phonetic norms seems to lie at the heart of why heritage speakers are better than late learners at maintaining contrasts between similar L1 and L2 categories that they do not have to distinguish for the purposes of being understood.

[1] As for why native Mandarin speakers should pattern differently from late learners in distinguishing the two long-lag VOT categories as heritage speakers do, one possibility is that native Mandarin 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, thus leading them to perceive English voiceless as significantly less aspirated than Mandarin aspirated. On the other hand, late learners (= native English speakers) might simply be focusing 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.

To what extent then do L1 and L2 categories that are phonetically close 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 at some level 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 furthermore, 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 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 1 and 4 are indeed representative of the relevant speech communities, then it seems that at least some of our 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, but speakers in this study produce this vowel with F2 values of approximately 750-1200 Hz. 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. Our results for the production of Mandarin /y/ are also consistent with those of Flege (1987) for French /y/: late learners do not differ appreciably from native speakers in their phonetic space for this "new" sound. What we find most significant (though perhaps not unexpected) is that when both vowel quality and VOT are considered, heritage speakers appear to be the most successful at approximating the phonetic norms of both of their languages. In accordance with our previous findings on fricatives, these results also suggest a correspondence between heritage speakers' linguistic performance and their level of experience with the heritage language: the higher the level of experience, the closer the degree of approximation to native acoustic norms.

As for why heritage speakers tend to be better than late learners at maintaining contrasts between similar categories in their two languages, in Chang *et al*. (to appear) we observed that there are two possible explanations. First, early exposure to both languages might simply make heritage 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. Our previous fricative data more strongly supported the former hypothesis, as the acoustic distance between fricative categories did not appear to increase for heritage speakers specifically. The vowel and VOT data in the present study also more strongly support the former hypothesis, since there is no clear sign of dissimilation such as a “polarized” phonetic space that goes past native targets (cf. Laeufer 1997). However, it should be noted that we need better estimates of native targets in order to make conclusive statements about how close heritage speakers come to hitting these targets. It is possible that the absence of “polarized” phonetic spaces in heritage speakers is due at least in part to floor effects and the avoidance of intruding into the phonetic space of other categories. For instance, it may be that heritage speakers do not go lower than native F2 targets for Mandarin /u/ because they have already achieved some maximum degree of tongue backness and lip rounding. However, as mentioned above, in this case heritage speakers actually appear to have undershot the native targets by a good margin, so with respect to vowels at least, it is rather unlikely that the absence of polarization is due to the presence of a phonetic floor.

Finally, we also need to consider the ways in which the English input received by native Mandarin speakers in mainland China and Taiwan differs from the English input received by the other two groups in the United States. In particular, native Mandarin speakers’ initial English input might have been accented, making it possible that the (non-)approximation of English phonetic norms we see for a given native Mandarin speaker, rather than being attributable to that one speaker, has actually accumulated over a chain of second language acquirers. For that matter, one wonders whether the early Mandarin input received by heritage 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 native Mandarin speakers: they might have been exposed to bona fide L2 English, whereas heritage speakers were probably exposed to no worse than native Mandarin that had “drifted” (Sancier and Fowler 1997) in an English-speaking environment.

5 Conclusion

The results of two experiments on vowel and laryngeal categories in Mandarin and English indicate that heritage speakers are better than late learners at approximating the phonetic norms of their two languages and maintaining cross-linguistic contrasts between similar categories. Moreover, the data suggest that there is a correspondence between heritage speakers’ linguistic performance and amount of exposure to the heritage language and are consistent with an explanation in terms of early bilingual exposure nurturing bilingual phonetic accuracy.

References

Au, T. K., L. M. Knightly, S.-A. Jun, & J. S. Oh. 2002. Overhearing a language during childhood. *Psychological Science* 13(3).238–243.

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*, ed. J. Goodman & H. C. Nusbaum, 167–224. Cambridge, MA: MIT Press.

Boersma, P., & D. Weenink. 2008. Praat: doing phonetics by computer. http://www.praat.org.

Byrd, D. 1993. 54,000 American stops. *UCLA Working Papers in Phonetics* 83.97–116.

Chang, C. B., E. F. Haynes, Y. Yao, & R. Rhodes. To appear. A tale of five fricatives: consonantal contrast in heritage speakers of Mandarin. In *University of Pennsylvania Working Papers in Linguistics* 15(1). Philadelphia, PA: Penn Linguistics Club.

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(1).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*, ed. W. Strange, 233–272. Baltimore, MD: York.

Godson, L. 2003. *Phonetics of Language Attrition: Vowel Production and Articulatory Setting in the Speech of Western Armenian Heritage Speakers*. PhD dissertation, University of California, San Diego.

Hagiwara, R. 1997. Dialect variation and formant frequency: the American English vowels revisited. *Journal of the Acoustical Society of America* 102(1).655–658.

Knightly, L. M., S.-A. Jun, J. S. Oh, & T. K. Au. 2003. Production benefits of childhood overhearing. *Journal of the Acoustical Society of America* 114(1).465–474.

Laeufer, C. 1997. Towards a typology of bilingual phonological systems. In *Second-Language Speech: Structure and Process*, ed. A. R. James & J. Leather, 325–342. Berlin: Mouton de Gruyter.

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

Lisker, L., & A. Abramson. 1964. A cross-language study of voicing in initial stops: acoustical measurements. *Word* 20(3).384–422.

Oh, J. S., T. K. Au, & S.-A. Jun. 2002. Benefits of childhood language experience for adult L2 learners' phonology. In *Proceedings of the 26th Annual Boston University Conference on Language Development*, Vol. 2, ed. B. Skarabela, S. Fish, & A. H.-J. Do, 464–472. Somerville, MA: Cascadilla Press.

Oh, J., S.-A. Jun, L. Knightly, & T. Au. 2003. Holding on to childhood language memory. *Cognition* 86(3).B53–B64.

Sancier, M. L., & C. A. Fowler. 1997. Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics* 27(4).421–436.

Wu, Z., & C. Lin. 1989. Shi yan yu yin xue gai yao (实验语音学概要). Beijing: Higher Education Press.