MANDARIN LISTENERS' PERCEPTION OF ENGLISH VOWELS: PROBLEMS AND STRATEGIES

Xinchun Wang
Department of Linguistics, California State University, Fresno, California. E-mail: xinw@csufresno.edu

ABSTRACT

Previous research suggests L2 vowel perception problems are often due to the assimilation of L2 sounds to L1 categories. However, there is also evidence for a universal strategy which states listeners will rely on duration cues whenever spectral cues are not sufficient for discriminating L2 vowel contrasts. This study examines Mandarin listeners' perceptual problems with English vowels. In a perception test, a group of adult Mandarin-English bilinguals residing in Canada identified synthesized English /i/-/ɪ/, /u/-/ʊ/, and /æ/-/ɛ/ continua that manipulated vowel spectral and duration cues. Compared with a native English group who responded exclusively to the spectral cues, the majority of native Mandarin listeners failed to show native-like perceptual patterns for the three vowel contrasts. However, they responded heavily and consistently to duration cues for the /i/-/ɪ/ but not for the /u/- /ʊ/ and /æ/-/ɛ/ contrasts. Both group and individual data suggest that native Mandarin listeners used different strategies in perceptual identification of L2 vowel contrasts. Most Mandarin listeners did not appear to have clear category distinctions for /u/-/ʊ/ and /æ/-/ɛ/ pairs and some established incorrect perceptual representation of the /i/-/ɪ/ contrast. The findings did not fully support the universal strategy of using duration cues when non-native vowel contrasts are difficult to perceive.

SOMMAIRE

Les recherches antérieures suggèrent que les problèmes liés à la perception de voyelles dans la langue seconde (L2) sont souvent dus à l'assimilation de sons L2 aux catégories de la langue maternelle (L1). Par contre, il y a également évidence d'une stratégie universelle selon laquelle les auditeurs utilisent les indices temporaux (durée) lorsque les indices spectraux ne sont pas suffisants pour distinguer les voyelles contrastes de L2. Cette étude explore les problèmes perceptuels pour des voyelles anglophones chez des auditeurs mandarins. Dans le cadre d'un test de perception, un groupe d'adultes bilingues mandarin-anglophone résidant au Canada ont identifié des voyelles anglophones synthétisées (/i/-/ɪ/, /u/-/ʊ/ et /æ/-/ɛ/) qui diffèrent selon leur spectre et leur durée. Alors que des auditeurs dont la langue maternelle est l'anglais ont répondu exclusivement aux indices spectraux, la majorité des auditeurs mandarins n'ont pas démontré ce même patron perceptuel pour les voyelles contrastes examinées. Ces derniers ont par contre répondu fortement et uniformément aux indices de durée pour le contraste /i/-/ɪ/, mais non pour les contrastes /u/- /ʊ/ et /æ/-/ɛ/. Les donnÿes individuelles et de groupe suggŸrent que les auditeurs mandarins utilisent des stratÿgies diffÿrentes lors de lùidentification perceptuelle de voyelles contrastes de L2. La majoritÿ dùentre eux ne semblaient pas possÿder de catÿgories distinctives claires pour les paires /u/-/ʊ/ et /æ/-/ɛ/, et certains avaient établi une représentation perceptuelle inexacte du contraste /i/-/ɪ/. Les résultats n'appuient pas l'existence d'une stratégie universelle d'utilisation d'indices de durée lorsque les voyelles contrastes de L2 sont difficiles à distinguer.

1. INTRODUCTION

It is well known that adult second language (L2) learners often have difficulties with the perception and production of non-native phonological segments that either do not occur or are realized differently in their first language (L1). Current L2 speech learning theories have hypothesized that such problems are due to the assimilation of L2 sounds to L1 phones. For example, the two most influential models, Best's Perceptual Assimilation Model (PAM) and Flege's Speech Learning Model (Best, 1994, 1995; Flege, 1995, Flege, Schirru, & MacKsy, 2003) both describe how the L1 segments are related to L2 categories in a systematic way.

The PAM Model (Best, 1994, 1995; Best, McRoberts, & Sithole, 1988; Best & Strange, 1992) predicts the levels of difficulty in differentiating L2 sounds on the basis of how a pair of L2 segments is assimilated to L1 sounds. For instance, two L2 sounds can be assimilated to two different L1 phones that are similar to the nonnative pair, the Two Categories type (TC). The non-native pair can also be assimilated to a single native sound equally well or poorly, the Single Category type

(SC), or to a single non-native sound with different phonetic match which will result in one better assimilated than the other, the Category Goodness type (CG).

Flegeùs (1995, 2003) Speech Learning Model (SLM) states that the success in the perception of L2 phones depends on whether a learner is able to establish phonetic categories for the segments that exist in L2 but not in L1. According to the SLM, a new category fails to be established as an L2 speech sound in spite of the audible differences between the L2 sound and the closest L1 sound or between two closest L2 sounds if the learner fails to perceive such differences. Therefore, L1 and L2 speech sounds interact through a öcategory assimilationð mechanism (Flege et al. 2003). öBy hypothesis, category formation will be blocked if instances of an L2 speech category continue to be identified as instances of an L1 categoryð (Flege et al. 2003, p. 469). The SLM also states that learnersù capacity for speech learning remains intact across the life span and L2 learners will gradually approximate L2 phonetic norms for certain L2 sounds more closely as their L2 experiences increases over time. Therefore, L2 experience plays an important part in the process of learning.

During the past few decades, L2 speakers' problems with L2 segments, mostly on consonants, have been well documented. Cross-linguistic vowel perception and production has received more attention since the last decade and there is increasing evidence that learners' perception of L2 vowels is strongly influenced by their first language experience. Non-native vowels that do not have clear L1 counterparts have been repeatedly shown to be difficult to learn (Bohn & Flege, 1990, 1992, Flege, Bohn, & Jung, 1997; Ingram & Park, 1997; Munro, 1993; Munro, Flege & Mackay, 1995; Strange, Yamada, Kubo, Trent, Nishi, & Jenkins, 1998; Wang, 1997, 2002; Wang & Munro, 1998, 1999, 2004). For example, Munro (1993) found that native Arabic speakers perceived and produced English tense and lax vowel distinctions in terms of long and short vowel category differences as found in Arabic, indicating strong influence of the speakers' L1 experience. Similarly, in L2 vowel production, Wang (1997) found that native Mandarin speakers produced English vowels without clear Mandarin counterparts significantly less intelligible than those with their L1 counterparts. In a most recent study, late Korean-English bilinguals showed evidence of unidirectional influence of the L1 on the L2 and produced English vowels that were heavily "colored" by acoustic properties of their Korean vowels (Baker & Trofimovich, 2005).

However, there is also evidence that first language experience influence alone cannot explain all L2 vowel perception problems. For example, Bohn (1995) reported that native Spanish and Mandarin speakers relied heavily on duration cues in their perceptual identification of the English /i/-/ɪ/ contrast although neither Mandarin nor Spanish contrasts long and short vowels in their vowel systems. Based on these findings, Bohn (1995) proposed the Linguistic Desensitization Hypothesis (LDH), which states that L2 listeners, regardless of their L1 experience, will rely on duration cues whenever spectral cues are not sufficient to signal non-native vowel contrasts. Therefore, the use of duration cues reflects a general speech perception strategy that takes over whenever spectral information is not sufficient to signal the non-native vowel contrasts. The model hypothesizes that there is a common strategy for L2 vowel perception that is independent of a speaker's first language experience.

It is important to note that the LDH is based on a study that involved only one pair of L2 vowels, the English /i/-/ɪ/contrast. Whether it applies to other vowel contrasts needs to be tested. In a preliminary study using synthesized English /i/-/ɪ/ and /u/-/ʊ/ contrasts that differed systematically in duration and spectral steps, Wang and Munro (1999) found that while native English speakers responded exclusively to vowel spectral cues for both vowel contrasts, native Mandarin speakers relied heavily on duration cues for the /i/-/ɪ/ but not for the /u/-/ʊ/ contrast. The results of the /i/-/ɪ/ contrast were in agreement with Bohn's findings in that the listeners who failed to perceive the vowel spectral differences all responded exclusively to the duration cues. However, the findings with the /u/-/ʊ/ contrast did not appear to support the LDH because the general speech perception strategy of using the duration cues did not kick in when the listeners failed to perceive the spectral differences for the /u/ and /ʊ/ contrast. It was not clear why the native Mandarin listeners made use of duration cues systematically for only the /i/-/ɪ/ pair but not the /u/ - /ʊ/ contrast as Mandarin has both /i/ and /u/ but lacks /ɪ/ and /ʊ/ categories in a symmetrical way and the synthesized test stimuli manipulated both spectral and duration properties in the same fashion. Obviously, studies with multiple non-native vowel contrasts are needed to test the Linguistic Desensitization Hypothesis as well as different assimilation patterns hypothesized in both the PAM and SLM models.

This study addresses L2 vowel perception problems through testing native Mandarin speakers' perception of three non-native vowel contrasts, the English /i/-/ɪ/, /u/-/ʊ/, and /æ/-/ɛ/ pairs which pose serious problems for native Mandarin speakers in both perception and production (Wang, 1997, Rogers, 1997). Synthesized vowel continua manipulating vowel spectral and temporal cues were used to assess the listeners' perceptual patterns. The goal is to explore the nature of Mandarin listeners' perceptual problems with these vowel contrasts and to identify strategies they implement to differentiate non-native vowel contrasts. With increased number of nonnative vowel contrasts than the previous studies, this study also explores whether native Mandarin listeners use different strategies to identify the three target vowel contrasts when they fail to show native-like perceptual patterns. The specific research questions to be addressed are: 1) Do native Mandarin listeners show native-like perceptual patterns when identifying English vowel contrasts? 2) If they fail to show native-like perception, do they show ability to distinguish the spectral end points for each of the three target

English vowel contrasts under investigation? 3). If the Mandarin listeners fail to respond to vowel spectral cues in a native-like way, will they automatically rely on duration cues to distinguish all three vowel contrasts? In other words, do Mandarin listeners apply the same strategy in perceptual identification of the three target English vowels that do not contrast in their L1?

1.1. Mandarin Vowels

Mandarin is generally described as having five vowel phonemes /i y u e a/ with about a dozen surface forms [i y ɨ ɯ u e ə ɤ o ɛ a]. The exact number of surface forms varies according to different descriptions (Chao, 1968; Cheng, 1966; Howie, 1976; Li & Thompson, 1981; Svantesson, 1984; Wu, 1994). The mid vowel /e/ is realized as [ɛ] in rising diphthongs /jɛ/ and /ɥɛ/, and as [o] before /u/. The low vowel /a/ is centralized in quality and is phonetically different from the English /ɑ/. The Mandarin [u] is described as both higher and more posterior than the English counterpart (Norman, 1998). Although the [ɪ] and [ʊ] symbols sometimes appear in Mandarin diphthongs [eɪ], [aɪ] and [oʊ], [aʊ], they only indicate the direction of the movement and therefore function as glides [j] and [w] (Dow, 1972). Compared

Table 1. Native Mandarin participants' background information

ID	Gender	Age	LOR	AOL	% Use English
M03	f	30	9	11	20
M04	f	33	12	15	50
M05	f	22	69	12	50
M07	f	22	10	11	80
M08	m	33	5	15	10
M09	f	24	44	13	50
M12	f	31	13	9	50
M13	m	18	48	12	60
M14	m	24	14	12	30
M15	f	26	13	13	25
M16	f	34	2	13	100
M17	m	30	6	9	50
M18	f	39	27	13	30
M19	m	38	60	12	90
M20	m	31	5	11	20
M23	f	42	12	23	98
M24	f	28	4	14	70
Mean		29.7	20.8	12.8	51.9

Age = reported in years
LOR = Length of residence in Canada (months)
AOL = Age at which learning English began in home country (years)
% Use English = Subject's estimated % daily use of English outside home.

with the English [ɪ ɛ æ ʊ], which all occur as stressed monophthongs, Mandarin lacks such counterparts in its vowel system. Therefore, it can be viewed that Mandarin has categories comparable to the English [i eɪ u oʊ] but lacks the [ɪ ɛ ɔ æ ʊ ɑ ʌ] categories.

2. METHOD

2.1. Participants

The participants were 17 Mandarin speakers (6 male and 11 female) recruited from the international student population from a western Canadian university in British Columbia. Fifteen of them were born and raised in Mainland China and the remaining two in Taiwan. They ranged in age from 18 - 42 years (mean = 30) at the time of the study. All were advanced Mandarin-English bilinguals and their mean length of residence in Canada was two years (range = 0.5-5.5 years). All participants had studied English in their country of origin, beginning at a mean age of 13 years. Their mean age of arrival in Canada was 28 years (range = 13-34 years). According to self-report, their estimated daily use of English ranged from 20%-100% (mean = 52%). The participants' background information is summarized in Table 1. Six native English speakers (4 male and 2 female, mean age = 29) from the student and faculty population in a university in California took the same test as control subjects. Although the native English speakers were not from the same region in which the L2 speakers resided, they were included because their perceptual patterns in terms of responding to the spectral and duration steps of the three vowel continua were virtually identical to six native Canadian English speakers who were tested on the same synthesized vowel continua during stimuli preparation. Both the Canadian and American English speakers responded exclusively to spectral cues for the target vowel contrasts. Only the American English listeners' data were analyzed as a control group as the Canadian English listeners' data were used for testing the stimuli.

2.2. Stimulus Preparation

The English /hid/-/hɪd/, /hud/-hʊd/, and /hæd/-/hɛd/ continua were generated using a Klatt (1980) synthesizer at 20-kHz sampling rate with 16-bit resolution in the cascade mode. All three continua were synthesized with six duration and six spectral steps producing 36 tokens per continuum. The F1-F3 formant frequency values are summarized in Table 2. For the duration steps, the six longest vowels (at step 1) were 250 ms and the six shortest were 125 ms (at step 6) with a 25 ms increment between two steps. A formant contour was incorporated by using different values of F1 and F2 at the beginning and end of the vowel portion for each spectral step. F0 was set at 125 Hz at the beginning and dropped gradually to 105Hz at about the mid point of the vowel and then to 100 Hz toward the end. During the stimulus synthesis phase, three native Canadian English listeners provided feedback in an open-set identification test. They were not told that the stimuli

Table 2. F1-F3 values of synthesized /hid/- /hɪd/, /hud/- hʊd/, and /hæd/-/hɛd/ vowel continua

/hid/- /hɪd/

	F1		F2		F3
Spectral Steps	Start	End	Start	End	
1	300	300	2020	2020	2960
2	316	320	1984	1972	2900
3	332	339	1948	1923	2840
4	348	363	1912	1862	2780
5	364	394	1876	1776	2720
6	380	440	1840	1640	2660

/hud/- hʊd/

	F1		F2		F3
Spectral Steps	Start	End	Start	End	
1	350	320	1250	1000	2300
2	370	355	1220	1035	2310
3	390	390	1190	1070	2320
4	410	425	1160	1105	2330
5	430	460	1130	1140	2340
6	450	500	1100	1180	2350

/hæd/-/hɛd/

	F1		F2		F3	
Spectral Steps	Start	End	Start	End	Start	End
1	640	690	1690	1520	2430	2470
2	610	670	1730	1560	2445	2485
3	580	650	1770	1600	2460	2500
4	550	630	1810	1640	2475	2515
5	520	600	1850	1680	2490	2530
6	490	580	1890	1720	2505	2545

presented to them were synthesized speech and were only told to write down the words they heard in English orthography. When the continua were completed, all endpoint stimuli (steps 1 & 2 and steps 5 & 6) were perfectly identified (as the target vowel) by another three native Canadian English speakers in an open-set identification test; this was observed regardless of the differences in duration. The stimuli were then normalized for peak amplitude using Sound Edit 16 software for playback.

2.3. Procedures

Individual perceptual test sessions were held in a sound-treated room using custom-designed software on a Macintosh computer. The 36 /hVd/ words per vowel continuum were repeated once generating 72 tokens for each vowel contrast. The test stimuli were presented in three separate blocks as two-way forced choice identification tasks. The keyword labels used for the tests were heed/hid, who'd/hood, and had/head respectively for each vowel contrast. The order each listener identified the three vowel blocks was counterbalanced. Each listener completed a trial session to learn the test procedure before taking the test. During the test, each stimulus was played only once through the headphones and the listener identified the word by pressing a labeled button on the computer screen. The listener had the control over the pace of the test by either immediately or delaying as much as they wanted in clicking the button each time after a stimulus was presented. As soon as the listener clicked the button, the next trial was played back. The test data were collected automatically by the computer and saved for subsequent analysis.

3. RESULTS

3.1. Overall Results

The Mandarin and English groups' mean percentage identification as the tense endpoint in the series (hereafter % ID) scores on "heed/hid," who'd/hood," and "had/head" vowel continua were calculated for each of the six spectral and six duration steps and are presented in Figure 1. The % ID scores on each spectral step were pooled over the six duration steps, and the scores on each duration step were pooled over the six spectral steps. To quantify listeners' responses to spectral cues, their % ID scores on "heed/hid", "who'd/hood", and "had/head series" at spectral steps 1 & 2 across all six duration steps (the most "heed", "who'd", and "had" like tokens) and at spectral steps 5 & 6 across the six duration steps, (the most "hid", "hood", and "head" like tokens) were calculated. If the listeners relied exclusively on the spectral cues to contrast the vowel pairs, they should identify the stimuli at spectral steps 1 & 2 (the unambiguous "heed", "who'd", and "had" stimuli) 100% as "heed", "who'd", and "had" respectively across the three vowel continua. Their % ID scores on "heed", "who'd", and "had" should decline along the continua and reach 0% at spectral steps 5 & 6 (the spectrally unambiguous "hid", "hood", and "head" stimuli). Similarly, if the listeners relied exclusively on duration cues to contrast these vowel pairs, they should identify all the longest tokens (at duration steps 1 & 2) 100% and shortest stimuli (at duration steps 5 & 6) 0% as "heed", "who'd", and "had".

Furthermore, listeners' degree of sensitivity to spectral cues to contrast these vowel pairs were assessed by subtracting their mean % ID scores of spectral steps 5 & 6 from the mean % ID scores of spectral steps 1 & 2. The higher the % difference between the two spectral end points, the more the listeners responded to the spectral cues for labeling the vowel contrasts. For example, the spectral end point difference score would be 100% if a listener identified all the spectral step 1 & 2 tokens as "heed" and step 5 & 6 tokens as "hid". Similarly, the higher the % difference scores between the duration end points, the more the listeners relied on duration cues for contrasting the vowel pairs.

For the native English group, the mean spectral end point difference scores were 95%, 99% and 97% for "heed", "who'd", and "head" respectively, indicating almost exclusive

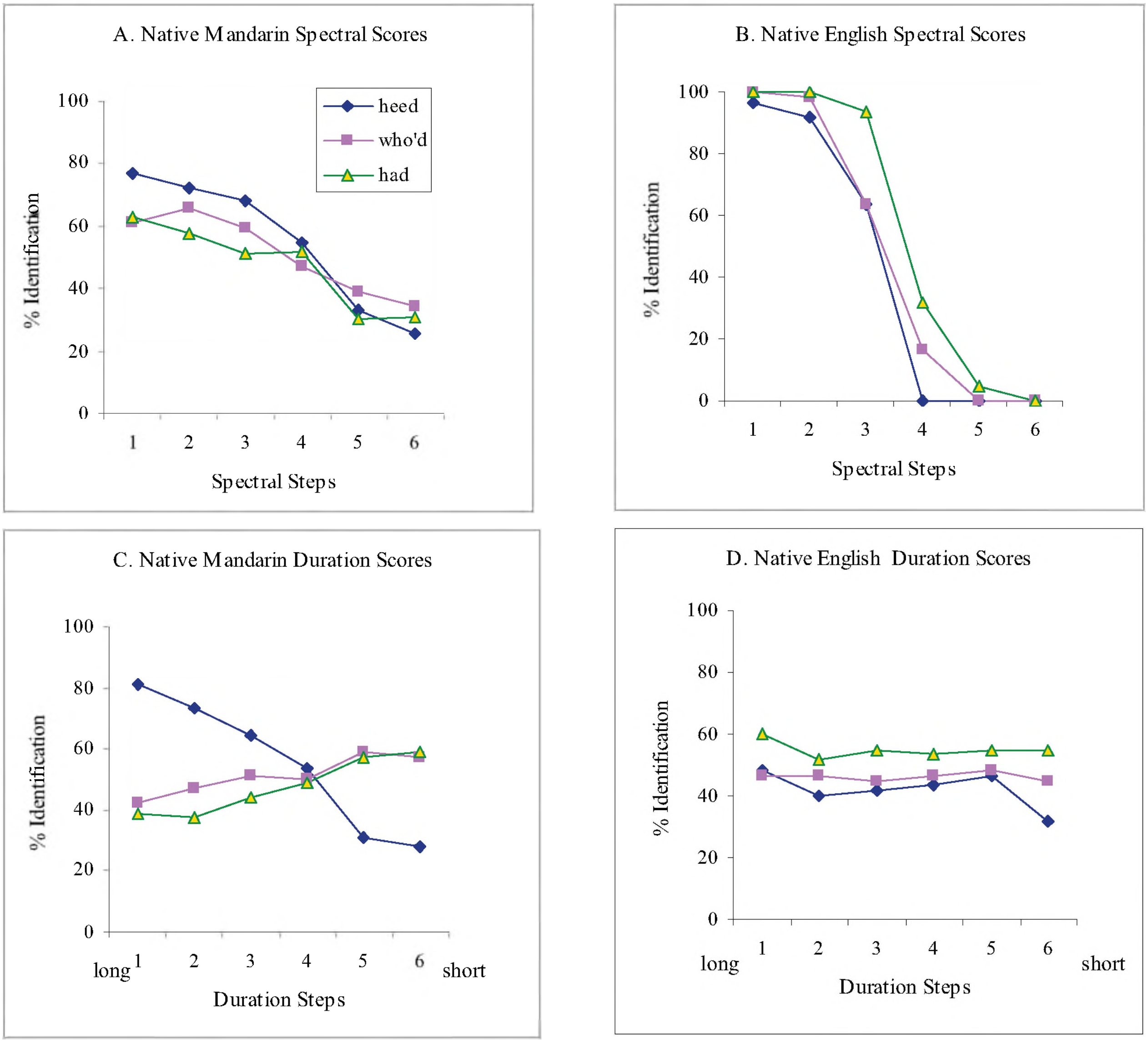

Figure 1. Native Mandarin (Panel A & C) and English (Panel B & D) listeners' mean % ID scores on "heed," who'd," and "had" of the three vowel continua at six spectral steps (Panel A & B) and six duration steps (Panel C & D). The scores on each spectral step were pooled over the six duration steps, and the scores at each duration step were pooled over the six spectral steps.

use of spectral cues in perceptual identification of these three vowel pairs. In contrast, their % duration end point difference scores were 6%, 2%, and 4% respectively. For the native Mandarin group, the mean spectral end point difference scores were 35%, 26%, 29% and the mean duration end point difference scores were 47%, -13%, and -20% for "heed/hid", "who'd/hood", and "had/head series respectively.

Two ANOVAs were conducted to address question 1 raised in the introduction – do Mandarin listeners show native-like perception when identifying the 3 English vowel contrasts. First, spectral endpoint difference scores were analyzed in a two-way ANOVA with group (Mandarin & English) as between subject factor and vowel pair (heed/hid, who'd/hood, had/head) as within subject factor. This analysis established an effect of group [$F(1,21) = 42.428$, $p = .000$] but the factor of vowel pair [$F(2,42) = .203$, $p = .817$] and vowel pair × group interaction [$F(2,42) = .421$, $p = .659$] were not significant. Next, the same two-way ANOVA was conducted on the mean *duration* end point difference scores to assess whether Mandarin listeners responded to the duration cues like native English listeners. This analysis revealed no effect of group [$F(1,21) = .001$, $p = .970$]. However, both vowel pair [$F(2,42) = 4.688$, $p = .015$] and vowel pair × group interaction [$F(2,42) = 3.900$, $p = .028$] were significant. Follow up one-way ANOVAs established a significant difference between Mandarin and English groups on the "heed/hid vowel pair [$F(1,21) = 8.343$, $p = .009$] only. English and Mandarin listeners did not differ in their response to duration end points for the who'd/hood pair [$F(1,21) =$

.597, p = .448] or the "had/head" pair [F (1,21) = 1.459, p = .240]. These results show that the native Mandarin listeners did not show sensitivity to spectral cues that is comparable to that of native English listeners for any of the three target vowel contrasts and they failed to show a native-like response pattern to duration cues when identifying the heed/hid contrast. However, for who'd/hood and the had/head contrasts the Mandarin listeners conformed to the pattern observed in native English listeners showing little sensitivity to the durations cues.

To further explore native Mandarin listeners' sensitivities to spectral and duration cues and to address questions 2 and 3 regarding perceptual strategy, two statistical analyses on the % ID scores were carried out for the Mandarin group only. First, a two-way ANOVA with vowel pair (heed/hid, who'd/hood, had/head) and spectral endpoint (mean steps 1 & 2 vs. mean steps 5 & 6) as within subject factors was conducted. This ANOVA established a significant effect for spectral end point [F(1,28) = 33. 097, p = .000] and for vowel pair [F(2,48)= 4.109, p = .023] but the vowel × spectral end point interaction was not significant F(2,48) = .906, p = .411]. These findings show that Mandarin listeners were able to reliably distinguish the spectral end points of each vowel pair although their perceptual pattern was not native-like. Pairwise comparisons on vowel pair show that overall % ID scores were higher for the heed/hid than for the had/head series.

Next, the same two-way ANOVA was conducted on % ID scores with *duration* end point (mean steps 1 & 2, mean steps 5 & 6) and vowel pair (heed/hid, who'd/hood, had/head) as within-subject factors. This ANOVA showed no effect of duration end point [F(1, 48) = .562, p = .457] or vowel pair [F(2, 48) = 1.836, p = .170]. However, the vowel pair× duration endpoint interaction was significant [F(1, 48) = 12.241, p = .000]. Three follow up t-tests were conducted to examine the simple effect of vowel duration endpoint for each vowel pair. The results statistically confirm the pattern plotted in Figure 1C, showing that percentage identification scores were significantly different for long vowel tokens (steps 1&2) compared to shirt vowel tokens (steps 5&6) for the heed/hid series [t(16) = 5.749, p = .000] but not for the who'd/hood series [t(16) = 1.153, p = .266] and had/head series [t(16) = 1.729, p = .103]

Overall, the analyses show that while Mandarin listeners did not demonstrate native-like perceptual patterns in responding to vowel spectral cues across the target vowel contrasts, they did reliably distinguish the spectral end points for each vowel series indicating they had some sensitivity to vowel spectral cues. However, they differed in their response to the duration end points across the three vowel pairs, they reliably distinguished the duration cues for the heed/hid contrast but not for the other two vowel contrasts.

Table 3. Native English listeners' end point difference scores (mean % scores of steps 1&2–mean % scores of steps 5&6)

ID	Contrasts	Spectral*	Duration*	Absolute D*
E01		100	13	13
E02		96	0	0
E03		96	0	0
E04	heed/hid	88	17	17
E05		96	13	13
E06		96	-4	4
E01		100	13	13
E02		100	-8	8
E03		96	8	8
E04	who'd/hood	100	8	8
E05		100	-17	17
E06		100	8	8
E01		92	21	21
E02		100	8	8
E03		100	17	17
E04	head/had	92	-8	8
E05		100	-4	4
E06		96	-8	8
Mean/SD		97(3.7)	10 (5.7)	10 (5.7)
- 2 SD		90	-1.7	
+ 2 SD		104	21.1	

Spectral = Spectral end point difference scores
Duration = Duration end point difference scores
Absolute D = Absolute value of duration end point difference scores

3.2. Individual Differences

Individual listener responses were also examined to address the same research questions. Individual data were evaluated in two ways: 1) to determine whether any native Mandarin listeners established native-like perceptual patterns in responding to spectral and duration cues to label the three target vowel pairs and 2) to assess how individual listeners responded to spectral and duration cues when perception was not native-like. For the first type of analysis, the mean and standard deviations of native English group's end point difference scores were used as references. A Mandarin listener would be considered to have established native-like perceptual pattern for contrasting each vowel pair if their spectral and duration end point difference scores fall within two plus and minus standard deviations of the native English mean scores, i.e. 90% or above for spectral cues and 21% or less for duration cues.

Individual native English listeners' spectral and duration end point difference scores with the group mean and standard deviations are presented in Table 3. The absolute values of native English listeners' end point difference scores were taken for computing the mean and standard deviations as one or two listeners had negative duration end point difference scores for each vowel pair (see Table 3). Individual Mandarin listener's % ID scores for heed/hid, who'd/hood and had/head at six spectral and six duration steps as well as the end point difference scores (mean % scores of steps 1 & 2 – mean % scores of steps 5 & 6) are presented in Table 4 through Table 6.

As seen from Tables 4-6, across the three vowel pairs, very few listeners' spectral end point scores fell within two standard deviations (highlighted in bold) of the native English mean score, which is 90% and above. Two out of the 17

Table 4. Mandarin listeners' % ID scores of "heed" at each spectral and duration step and the % end point difference scores (steps 1&2-steps 5&6). The numbers in bold indicate those fall within 2 standard deviations of the native English mean. Listeners whose end point difference scores meet the 70% criterion for "spectral" (S), or "duration" (D) cues , or neither (N) are presented in the last row.

Individual Listeners and % Perceptual Scores for "heed"																		
Steps	M03	M04	M05	M07	M08	M09	M12	M13	M14	M15	M16	M17	M18	M19	M20	M23	M24	Mean
Spectral 1	100	58	83	50	100	67	100	75	92	83	33	58	92	100	92	50	75	77
Spectral 2	100	58	92	67	100	50	83	58	83	92	50	58	50	83	100	58	42	72
Spectral 3	92	58	75	58	67	50	67	67	75	50	67	58	75	83	92	67	58	68
Spectral 4	75	50	25	50	58	50	67	58	50	25	75	58	58	75	42	58	50	54
Spectral 5	17	42	0	67	0	25	17	25	17	25	50	58	50	33	17	67	58	33
Spectral 6	0	50	0	58	0	33	8	0	33	17	67	42	33	0	0	50	42	25
1&2-5&6	**92**	12	88	-4	**100**	30	79	54	63	67	-17	8	30	75	88	-5	9	45
Steps	M03	M04	M05	M07	M08	M09	M12	M13	M14	M15	M16	M17	M18	M19	M20	M23	M24	Mean
Duration 1	75	100	58	100	58	100	58	67	92	67	83	100	100	75	67	92	83	81
Duration 2	67	67	58	92	50	92	58	67	67	42	67	100	92	75	58	100	92	73
Duration 3	58	67	50	83	50	42	75	50	67	42	92	92	83	67	50	58	67	64
Duration 4	67	67	42	50	67	42	75	58	58	50	50	33	42	42	50	67	50	53
Duration 5	58	8	42	17	58	0	50	17	33	42	33	8	25	42	42	17	33	31
Duration 6	58	8	25	8	42	0	25	25	33	50	17	0	17	75	75	17	0	28
1&2-5&6	**13**	76	**16**	84	**4**	96	**21**	46	47	**9**	50	96	75	**17**	**4**	79	71	48
70% S. D. N.	S	D	S	D	S	D	S	N	N	N	N	D	D	S	S	D	D	

Table 5. Mandarin listeners' % ID scores of "who'd" at each spectral and duration step and the % end point difference scores (steps 1&2-steps 5&6). The numbers in bold indicate those fall within 2 standard deviations of the native English mean. Listeners whose end point difference scores meet the 70% criterion for "spectral" (S), or "duration" (D) cues , or neither (N) are presented in the last row.

Individual Listeners and % Perceptual Scores for "who'd"																		
Steps	M03	M04	M05	M07	M08	M09	M12	M13	M14	M15	M16	M17	M18	M19	M20	M23	M24	Mean
Spectral 1	8	100	50	58	100	100	17	33	58	100	25	92	50	58	100	33	58	61
Spectral 2	25	100	33	42	92	100	33	42	58	92	42	92	67	92	100	58	50	66
Spectral 3	17	83	33	50	100	100	42	33	42	75	50	92	58	67	83	42	42	59
Spectral 4	33	58	50	42	50	50	58	67	33	42	50	42	58	33	25	58	50	47
Spectral 5	58	25	42	67	17	42	58	50	58	25	42	8	50	33	0	42	42	39
Spectral 6	83	0	58	58	0	0	100	67	33	0	33	0	33	0	0	42	75	34
1&2-5&6	-54	88	-9	-13	88	79	-54	-21	13	84	-4	88	17	59	**100**	4	-5	**27**
Steps	M03	M04	M05	M07	M08	M09	M12	M13	M14	M15	M16	M17	M18	M19	M20	M23	M24	Mean
Duration 1	25	58	100	92	58	58	33	17	25	58	17	50	0	42	58	25	0	42
Duration 2	58	50	75	92	50	67	42	42	33	58	25	58	8	33	42	42	25	47
Duration 3	42	58	42	83	58	50	42	67	42	67	0	58	58	58	58	58	33	51
Duration 4	25	50	33	50	58	75	42	33	50	58	25	67	67	50	50	50	67	50
Duration 5	50	75	17	0	67	67	67	67	67	58	83	50	83	42	58	58	92	59
Duration 6	25	75	0	0	67	75	83	67	67	33	92	42	100	58	42	42	100	57
1&2-5&6	**4**	**-21**	79	92	**-13**	**-9**	-38	-38	-38	**13**	-67	**8**	-88	**-13**	**0**	**-17**	-84	-13
70% S. D. N.	N	S	D	D	S	S	N	N	N	S	N	S	D	N	S	N	D	

Table 6. Mandarin listeners' % ID scores of "who'd" at each spectral and duration step and the % end point difference scores (steps 1&2-steps 5&6). The numbers in bold indicate those fall within 2 standard deviations of the native English mean. Listeners whose end point difference scores meet the 70% criterion for "spectral" (S), or "duration" (D) cues , or neither (N) are presented in the last row.

Individual Listeners and % Perceptual Scores for "had"																		
Steps	M03	M04	M05	M07	M08	M09	M12	M13	M14	M15	M16	M17	M18	M19	M20	M23	M24	Mean
Spectral 1	100	58	83	58	33	75	25	83	83	58	58	58	42	50	58	50	100	100
Spectral 2	75	58	58	58	50	67	25	50	58	75	58	83	58	25	58	33	92	75
Spectral 3	50	42	42	67	58	50	33	50	50	33	75	58	58	25	33	50	92	50
Spectral 4	58	42	17	50	50	50	58	42	50	75	58	58	50	50	42	42	92	58
Spectral 5	0	42	8	42	50	8	25	25	17	25	67	42	17	33	58	50	8	0
Spectral 6	0	33	25	25	58	25	42	50	8	8	58	42	67	8	50	25	0	0
1&2-5&6	88	21	54	25	-13	55	-9	29	58	50	-5	29	8	17	4	4	**92**	30
Steps	M03	M04	M05	M07	M08	M09	M12	M13	M14	M15	M16	M17	M18	M19	M20	M23	M24	Mean
Duration 1	58	8	33	67	25	67	0	33	42	50	83	25	0	0	92	8	67	58
Duration 2	50	8	33	67	42	50	8	17	42	42	67	33	0	33	67	8	67	50
Duration 3	33	33	58	67	33	50	8	58	50	42	75	42	50	42	50	8	50	33
Duration 4	33	42	33	58	33	42	33	50	58	58	75	50	58	58	33	50	67	33
Duration 5	50	83	50	33	75	33	67	75	33	50	42	100	83	25	33	75	67	50
Duration 6	58	100	25	8	92	33	92	67	42	33	33	92	100	33	25	100	67	58
1&2-5&6	**0**	-84	**-5**	47	-50	26	-76	-46	**5**	**5**	38	-67	-92	**-13**	51	-80	**0**	-20
70% S. D. N.	S	D	N	N	N	N	D	N	N	N	N	N	D	N	N	D	S	

listeners (M03 and M08) met the criteria for the heed/hid contrast. Only one listener each met this standard for the who'd/hood (M20) and for the had/head contrast (M24). None of the Mandarin listeners demonstrated native-like perceptual patterns to identify all three target vowel contrasts.

A more complicated pattern was observed with individual Mandarin listeners' duration end point difference scores across the three target vowel contrasts. First, while none of the listeners had negative duration end point difference score for the heed/hid pair, over half of the listeners had it for the who'd/hood and had/head pairs (see Tables 4-6). Second, several listeners had substantially high negative end point duration difference scores (ranging from -76% to -92%) suggesting that these listeners switched the labels in responding to duration cues as they consistently identified short stimuli as "who'd" and "had" and long stimuli as "hood' and "head" respectively. Because of these negative scores, the absolute values of the spectral and duration end point scores were taken for judging whether each individual listener met the native-like criteria. However, the negative end point difference scores are still presented in their original value in Tables 4-6 to facilitate identification the label switching listeners.

Applying the 21% or less native-like standard for the duration end point difference scores, seven, nine, and six listeners (highlighted in bold in Tables 4-6) fell within the 2 standard deviation criterion for the heed/hid, who'd/hood and had/head pair respectively. These results show that while only one or two listeners met the native-like standard for responding to spectral cues, substantially more listeners met the native-like criterion for not using duration cues to contrast these target vowel pairs.

For the second type of analysis on individual data, a 70% or higher end point difference score was used as criterion to assess whether an individual listener showed some sensitivity to spectral or duration cues (even though most did not show native-like perceptual patterns). This criterion requires the listeners to identify each endpoint at least 20% above chance level. Applying this criterion, listeners whose spectral end point difference score was 70% or higher were categorized as +S for their sensitivity to the spectral cues. Similarly, those whose duration end point difference scores was 70% (absolute value) or higher were categorized as +D for their sensitivity to duration cues. Listeners whose spectral and duration endpoint difference scores both fell below the 70% criterion were classified as +N for using neither spectral nor duration cues. The classification of each individual listener into these three categories based on this standard is presented in the last row of Table 4 through Table 6 for the three vowel contrasts respectively.

The total number of listeners fell into each of the above three categories, +S, +D, and +N for each vowel contrast are summarized in Table 7. Applying this standard, as seen in Table 7, more listeners (7) responded to duration cues (+D) for the heed/hid pair than for the who'd/hood (4) and had/head

Table 7. Number of listeners classified as +S, +D, and +N for each target vowel pair

Classified	Heed/hid	Who'd/hood	Had/head
+S	6	6	2
+D	7	4	4
+N	4	7	11

pairs (4). Furthermore, substantially more number of listeners respond to neither spectral nor duration cues (+N) for the had/head (11) and who'd/hood (7) pairs than for the heed/hid (4) pair.

To summarize, only one or two out of the 17 Mandarin listeners met the native-like criteria to respond to spectral cues to identify each vowel contrast and none met the native-like criteria to identify all three vowel contrasts. The individual data also showed that listeners used different strategies in differentiating the three target vowel pairs when they failed to show native-like perceptual patterns. More listeners responded to duration cues for the heed/hid contrast than for the other two contrasts. Furthermore, the majority of individual listeners did not appear to respond to either spectral or duration cues to contrast the had/head and who'd/hood pairs. These listeners did not appear to have clear separate categories for these vowel pairs.

4. DISCUSSION

Revisiting the three research questions raised earlier, the current data provide negative answer to question 1 – do native Mandarin listeners show native-like perceptual patterns when identifying the English /i/-/ɪ/, /u/-/ʊ/, and /æ/-/ɛ/contrasts. The listeners were able to reliably distinguish the spectral end points of each vowel pair, indicating they were somewhat sensitive to vowel spectral differences. Therefore, the data provide positive answer to question 2 - if they fail to show native-like perception, do they show ability to distinguish the spectral end points for each of the three target English vowel contrasts under investigation? The data provide negative answer to question 3 – if the Mandarin listeners fail to respond to vowel spectral cues in a native-like way, will they automatically rely on duration cues to distinguish all three vowel contrasts? The results show that the listeners relied on duration cues to distinguish the /i/-/ɪ/ but not /u/-/ʊ/, and /æ/-/ɛ/contrasts. These differences suggest that Mandarin listeners applied different strategies in identifying the three vowel contrasts that had parallel acoustic/phonetic properties.

The analysis of individual listener data further supported the above findings. First, very few listeners met the native like criterion for relying on spectral cues to distinguish the target vowel pairs. Second, when the listeners failed to respond to spectral cues, the majority of them automatically relied on duration cues to contrast the heed/hid contrast but not who'd/hood or had/head contrasts. Third, the majority of the listeners responded to neither spectral nor duration cues systematically to distinguish the who'd/hood or had/head contrasts. The findings suggest that these listeners did not have clear two-category distinctions for the who'd/hood or had/head contrasts.

Although the acoustic structures of the three vowel pairs are parallel and the spectral and duration cues were presented in a systematic manner in the synthesized stimuli, the Mandarin listeners responded differently across the three vowel continua. Are such perceptual differences due to assimilation? What are some other factors that influence the use of different strategies in distinguishing these vowel pairs? To answer these questions, it is necessary to examine the nature of Mandarin speakersù perceptual problems from the perspective of current perceptual assimilation theories, in particular, Bestùs Perceptual Assimilation Model and Flegeùs Speech Learning Model and the Linguistic Desensitization Hypothesis.

The current data did not provide consistent support for the Linguistic Desensitization Hypothesis, which states that L2 speakers will rely on duration cues to differentiate the non-native vowel contrasts whenever vowel spectral differences are not sufficient to signal the contrasts (Bohn, 1995). Apparently, the majority of Mandarin speakers who had problems discriminating the three target vowel contrasts did not automatically and consistently respond to duration cues for the English /u/-/ʊ/ and /æ/-/ɛ/ contrasts, although they did for the /i/-/ɪ/ contrast. In fact, the majority of listeners did not respond to either spectral or duration cues to distinguish the /u/-/ʊ/ and /æ/-/ɛ/. Therefore, the use of the duration cues does not always take over as a general strategy whenever spectral cues are not sufficient to signal the non-native vowel contrasts. Obviously, Mandarin speakers used different strategies for different vowel contrasts in the current study and there is no such universal strategy for relying on duration cues for contrasting all three target vowel contrasts.

As discussed earlier in the introduction, Linguistic Desensitization Hypothesis was based on the L2 listenersù perception of the English /i/-/ɪ/ contrast only. It is interesting to note that both Bohnùs study (1995) and the current data show that speakers whose L1 does not contrast duration for its vowel system relied on duration cues to distinguish the English /i/-/ɪ/ contrast. Future studies need to test other L2 vowel contrasts that have parallel structures as the English /i/-/ɪ/, /u/-/ʊ/, and /æ/-/ɛ/pairs.

Applying the PAM Model, Mandarin listenersù perceptual problems with the English /u/-/ʊ/ and /ɛ/-/æ/ contrasts seem to fit the Single Category assimilation type, the assimilation of two L2 sounds to a single L1 category. However, the PAM model cannot explain the majority of the Mandarin listeners' use of the duration cues to perceptually distinguish the English /i/ - /ɪ/ contrast as the listeners did not simply "assimilate" the /ɪ/ sound to the L1 /i/, the single category in this area of the vowel space where English has /i/ and /ɪ/.

According to the Speech Learning Model, a new category fails to be established as an L2 speech sound in spite of the audible differences between the L2 sound and the closest L1 sound or between two closest L2 sounds if the learner fails to perceive such differences (Flege, 1995). Under this view, Mandarin speakers who had problems in perceiving the English /u/-/ʊ/ and /æ/-/ɛ/contrasts did not seem to have established phonetic categories for /ʊ/ and /æ/. Listeners who distinguished the /i/-/ɪ/ contrast based on temporal cues seemed to have established an *inaccurate* /ɪ/ category, because their perceptual representations of /ɪ/ were different from those of native English speakers. Therefore, at least some Mandarin speakersù perceptual problem with the /i/-/ɪ/ contrast suggests *modification* rather than assimilation.

Obviously, the PAM, the SLM and the LDH all partly explained some of Mandarin speakersù perceptual problem with English vowel contrast but none can fully account for the different strategies Mandarin speakers used in their perception of English vowels. What are some of the possible causes for such differences? In searching answers to this question, some possible factors are examined. First, Mandarin listenersù category representations of the target vowels might be influenced by L2 instruction. In China, the /i/-/ɪ/ contrast is commonly taught as two vowel categories that differ in length (Wang & Munro, 1999) while the /u/-/ʊ/, and /æ/-/ɛ/ pairs are not taught in the same manner. It is important that future studies examine learnersù L2 experience not only in terms of length of residence in the L1 environment but also in terms of formal pronunciation instruction in their country of origin.

Another possible explanation for the listenersù differences in perceptual patterns may be related to the difference in functional load each of the three target vowel contrasts bears. The /i/-/ɪ/ distinction bears more functional load and has more minimal pairs in English than does the /u/-/ʊ/ contrast (Brown, 1988). Hardly any commonly used vocabulary items are distinguished by the /u/-/ʊ/ contrast in English as minimal pairs (Brown, 1988; Kucera & Francis, 1967). Therefore, in real communication, the confusion caused by the substitution of /i/ for /ɪ/ would cause more problems than the substitution of /u/ for /ʊ/. The greater need to contrast /i/and /ɪ/ in communication probably results in more painstaking efforts from learners in finding the acoustic properties that signal the difference between the vowels in order to understand and to be understood. This might provide an explanation for the single category perception of the /u/-/ʊ/ but not for the i/-/ɪ/ pair.

However, the current data also showed that the majority of the listeners who had difficulties with the /æ/-/ɛ/ pair did not attend to the duration cues. This phenomenon cannot be explained by the functional load differences because, compared with the /u/-/ʊ/ pair, the /æ/-/ɛ/contrast actually has more minimal pairs and therefore bears a much greater functional load (Brown, 1988). The need to distinguish the /æ/-/ɛ/ contrast did not appear to help the learners (at least some listeners in the current study) to find strategies to distinguish this vowel pair. In fact, the /æ/-/ɛ/contrast appeared to be more resistant to perceptual learning as more Mandarin listeners were found to have perceptual difficulties with this pair (see also Wang, 2002).

One possible reason for the listenersù difficulties with the /æ/-/ɛ/ pair may be related to the high degree of spectral overlapping for the /æ/-/ɛ/ contrast. Evidence from L1 vowel perception tests suggests that the American English /æ/-/ɛ/contrast is intrinsically more spectrally confusing than the /i/-/ɪ/ and /u/- /ʊ/ contrasts. Hillenbrand & Clark (2000) reported that native American English speakersù identifications of English /i/-/ɪ/ and /u/-/ʊ/ contrasts were not affected by distorted duration (edited long and short vowels) differences because spectral differences were always sufficient for the differentiation of these vowel pairs. In contrast, their identifications of the /æ/-/ɛ/ contrast was affected by the distorted duration differences for the /æ/-/ɛ/ contrast. The authors speculated that the /æ/-/ɛ/ contrast showed a greater degree of spectral overlap than the /i/-/ɪ/ and /u/-/ʊ/ contrasts. Based on these previous findings on the English /æ/-/ɛ/ contrast, it might also be speculated that, although the physical steps of the synthesized vowel continua were controlled as equal, listeners might not *perceive* them as equal because the perceived duration differences might be affected by or interact with spectral properties differently across the three vowel pairs. Future studies need to test these differences with different L2 vowel contrasts.

Overall, the findings suggest that L2 vowel learning is a complex issue. A full understanding of the nature of the L2 vowel perception and production problems must consider not only the sound systems and assimilation patterns but factors such as learner characteristics, learning experience, and the influence of pronunciation teaching in foreign language classrooms. The current study took a step in this direction and future studies should take into consideration of these multiple factors when analyze non-native speakersù perception and production problems in learning L2 vowel contrasts. Future studies should include more vowel contrasts in different target languages, and different types of L2 learners.

ACKNOWLEDGEMENTS

The author appreciates many helpful comments from Linda Polka and two anonymous reviewers. The author also thanks all individuals for their participation.

REFERENCES

Baker, W., & Trofimovich, P. (2005). Interaction of native- and second language vowel system(s) in early and late bilinguals. *Language and Speech, 48*(1), 1-27.

Best, C.T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. Goodman & H. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167-244). Cambridge: MIT Press.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-204). Baltimore: York Press.

Best, C. T, McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 345-360.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics, 20,* 305-330.

Bohn, O.-S. (1995). Cross-language speech perception in adults: first language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 279-304). Baltimore: York Press.

Bohn, O.-S., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics, 11,* 303-328.

Bohn, O.-S., & Flege, J. E. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition, 14,* 131-158.

Brown, A. (1988). Functional load and the teaching of pronunciation. *TESOL Quarterly, 22,* 593-606.

Chao, Y. (1968). *A grammar of spoken Chinese.* Berkeley and Los Angeles: University of California Press

Cheng, R. (1966). Mandarin phonological structure. *Journal of Linguistics, 2,* 135-262.

Dow, F. (1972). *An outline of Mandarin phonetics.* Canberra: Australian National University Press.

Flege, J. E. (1995). Second language speech learning: theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). Baltimore: York Press.

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). The effect of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics, 25,* 437-470.

Flege, J. E., & MacKay, I. R. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition, 26*, 1-26.

Flege, J. E., Mackay, I. R., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America, 106*, 2973-2987.

Flege, J. E., Schirru, C., & MacKay, I. R. A. (2003). Interaction between the native and second language phonetic systems. Speech Communication, 40, 467-491.

Hillenbrand, J. M., & Clark, M. J. (2000). Some effects of duration on vowel recognition. *Journal of the Acoustical Society of America, 108,* 3013 - 3022.

Howie, J. (1976). *Acoustical studies of Mandarin vowels and tones.* London: Cambridge University Press.

Ingram, J.C., & Park, S.-G. (1997). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics, 25,* 343-370.

Klatt, D. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America, 67,* 971-995.

Kucera, H., & Francis, N. W. (1967). *Computational Analysis of present-day American English.* Providence: Brown University Press.

Li, C., Thompson, S. (1981). *Mandarin Chinese: A functional reference grammar.* Berkeley: University of California Press.

Munro, M. (1993). Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness rating. *Language and Speech, 36,* 39-66.

Munro, M., Flege, J., & MacKay, I. (1996). The effects of age of second language learning on the production of English vowels. *Applied Psycholinguistics, 17*, 313-334.

Norman, J. (1988). *Chinese.* Cambridge: Cambridge University Press.

Rogers, C. L. (1997). *Intelligibility of Chinese-Accented English.* Unpublished Ph.D. Dissertation, Indiana University, Indiana.

Strange, W., Yamada, R. A., Kubo, R., Trent, S. A., Nishi, K., & Jenkins, J. J. (1998). Perceptual assimilation of American English vowels by Japanese listeners. *Journal of Phonetics, 26,* 311-344.

Svantesson, J.-O. (1984). Vowels and diphthongs in standard Chinese. *Lund University Working Papers, 27*, 209-235.

Wang, X. (1997). *The acquisition of English vowels by Mandarin ESL learners: A study of production and perception.* Unpublished MA Thesis, Simon Fraser University.

Wang, X., & Munro, M. (1998). *English vowel production by Beijing Mandarin speakers.* Paper presented at the 16th International Congress on Acoustics and the 135th Meeting of the Acoustical Society of America, Seattle.

Wang, X., & Munro, M. J. (1999). *The perception of English tense-lax vowel by native Mandarin speakers: The effect of training on attention to temporal and spectral cues.* Paper presented at the 14th International Congress of Phonetic Sciences, San Francisco.

Wang, X. (2002). *"Training Mandarin and Cantonese Speakers to Identify English Vowel Contrasts: Long-term retention and effects on production".* Unpublished Ph.D. Dissertation, Simon Fraser University, Canada.

Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System*, 32, 539-552

Wang, X., & Munro, M. J. (1999). The perception of English tense-lax vowel pairs by native Mandarin speakers: The effect of training on attention to temporal and spectral cues. *Proceedings of the 14th International Congress of Phonetic Sciences, Vol.3,* (pp. 125-128). San Francisco, August 1-7, 1999.

Wu, Y. (1994). *Mandarin segmental phonology.* Unpublished doctoral dissertation, University of Toronto.