L2 Experience and Non-Native Vowel Categorization of L1-Mandarin Speakers

Bo-ren Hsieh, Ho-hsien Pan

Graduate Institute of Foreign Literatures and Linguistics, National Chiao Tung University, Taiwan
borensieh@gmail.com, hhpan@cc.nctu.edu.tw

Abstract

This study investigates the effect of L2-English experience on the perception of the English tense-lax high vowel contrast. Experienced L1-Mandarin, inexperienced L1-Mandarin and L1-English listeners identified and discriminated synthetic heed-bid and who id-hood continua varying in five steps of F1, F2, and F3 variations and seven steps of duration variations. The results show a strong reliance on formant variations by L1-English listeners; reliance on duration variations by the inexperienced L1-Mandarin listeners, and a more dominant reliance on formant cues than on duration cues by the experienced L1-Mandarin listeners.

Index Terms: vowel perception, English, Mandarin

1. Introduction

Several studies have indicated that second language acquisition (SLA) is influenced significantly by first language (L1) of a speakers, especially when speaking skills are involved [1, 2, 3]. The pre-linguistic experience affects not only the production of the second language (L2) sounds, but its perception. For example, Best’s Perceptual Assimilation Model predicts that listeners are likely to perceive two similar non-native sounds as the same by assimilation when there is only one corresponding category in their L1 (a case of single-category assimilation) [1]. However, the influence of L1 on L2 perception varied with the L1 of the listener concerned.

In American English (AE), the tense/lax high vowel pairs /ɪ/-/ʊ/ and /u/-/ʊ/ differ in formant frequency and duration: /ɪ/ has a lower first formant (F1), higher second and third formants (F2 & F3) and an inherently longer duration than /ʊ/; /u/ has a lower F1, F2, and F3 and an inherently longer duration than /ʊ/. L1-English listeners use spectral variations primarily to discriminate /iː/ from /i/ and /u/ from /uː/, and display durational differences [2-8].

It has been suggested that non-native listeners tend to rely predominantly on durational differences to differentiate between English tense and lax vowels when phonemic duration contrasts are in their L1, e.g. Japanese and Finnish [4, 5, 6]. It has been found, moreover, that the duration-based categorical boundary between English /i/ and /u/ is at the identical point with Japanese’s categorical boundary for /iː/ and /uː/. Taken together, the consistent duration-based categorical perception is due to the transfer of the phonemic vowel duration contrasts in the listener’s native language. L1-Japanese listeners are not exposed to phoneme-contrasting vowel spectral differences in Japanese and thus they have difficulties in using spectral quality to discriminate /iː/ from /i/ [4, 5]. Lack of tense/lax vowel distinction, L1-Finnish listeners had difficulties in distinguishing English /iː/ from /i/ when durational differences were equalized. Similarly, durational cues are important to separate the long/short phonemes in Finnish. By measuring the mismatch negativity brain response, the organization of cortex was found to be different between L1-Finnish and L1-English [6].

Similarly, both L1-Spanish and L1-Russian listeners made use of duration cues only, different from L1-English listeners. The vowel duration, however, is not a phonemic feature in Spanish and Russian. To account for this, it has been proposed that vowel duration is an acoustic cue for perceiving word stress or voicing in both languages. Thus, it is possible that both non-native listeners are able to transfer their L1 allophonic experience to the perception of the English tense/lax vowels [7].

The L1 transfer of suprasegmental duration contrast, with Tone 3 inherently being longer than Tone 2, has also been adopted to explain the prominent use of durational properties by L1-Mandarin listeners to differentiate between English /i/ and /u/ [2].

With regard to L2 speech learning, the amount of L2 experience plays a role in filling the gap between L1 and L2 skills. The perception of English vowels by experienced non-native listeners is more like that of a native listener and was found to be significantly different from that of an inexperienced listener [3]. When investigating the developmental stages of English tense/lax vowel contrast, it has been found that L2-Spanish listeners failed to use spectral quality and showed a strong reliance on durational cues at the beginning stage. Though the intermediate listeners began to use spectral cues, durational cues were still stronger. Only at the advanced stage did the L1-Spanish listeners achieve a native-like performance level [8].

In this study, the English /i/-/u/ and /u/-/ʊ/ perception of L1-Mandarin listeners has been examined. Without phonemic or allophonic segmental duration contrast, Mandarin has one tonal duration contrast. With regard to the Mandarin high vowels /i/, /u/, and /ʊ/, the spectral difference between /i/ and /ʊ/ was found to be different from that between /i/ and /u/. It is, therefore, useful to investigate whether or not the L1-Mandarin experience with vowel spectra influences the perception of English /i/-/ʊ/ and /u/-/ʊ/. Two dimensional stimuli (spectrum and duration) were conducted to determine how L1-Mandarin listeners make use of acoustic cues to discriminate /i/ from /u/ and /u/ from /ʊ/.

2. Method

2.1. Participants

Twenty-two L1-Mandarin listeners were recruited from National Chiao Tung University. Eleven of them (3 men, 8 women) were experienced L2-English listeners, and they were either foreign literature or linguistics undergraduate or graduate students. In addition, the graduate students also majored in English for their degree. Their ages ranged from 21 to 25 with a mean age of 23. The other eleven students (5 men, 6 women) were not from the Department of Foreign Languages and Literatures and they were self-reported with...
limited English-language experience (during the six-year high-school education). This inexperienced group had a mean age of 22, ranging from 20 to 25 years. The perception experiments were also given to five L1-English listeners (2 men, 3 women). Their ages ranged from 24 to 46 with a mean of 31 years. All of them had lived in English-speaking countries until the age of 18; they came to Taiwan either to work or to learn Mandarin Chinese.

2.2. Stimuli

Perceptual stimuli were edited based on the natural speech, /hid/ and /hud/ tokens, produced by one native English speaker, who did not participate in the perception tests. The steady-state portions of vowels, /i/ and /u/, were replaced by the synthesized vowel tokens using the KLSYN: A Formant Synthesis Program [9]. The formant values for high front vowels /i/- and high back vowels /u/- were adopted from the suggested values of KLSYN (F1, F2, and F3 being 579, 1341, 1748 mels for /i/; 391, 1550, 1864 mels for /i'/; 607, 1113, 1673 mels for /u'/; and 424, 932, 1602 mels for /u/). Five spectral steps, equally spaced in the mel scale, were calculated for the /i/-i' and /u/-u' continua (with Step 1 being most lax vowel-like and Step 5 most tense vowel-like), respectively. Regarding the duration dimension, the values of natural tokens were adopted as the longest points, 264 ms for /i/ and 211 ms for /u/, with the seven-step continuum equally spaced logarithmically: a 20% decrease in vowel duration (with Step 1 being most lax vowel-like and Step 7 most tense vowel-like).

In addition, for each synthesized tokens, the intonation contours were also manipulated using KLSYN to ensure that the tokens would be perceived more naturally; /0/ values were queried using Praat, and the specific time points for entering /0/ values were calculated proportionally, i.e. a 20% decrease with every step. For instance, both the time point “25 ms” for the 7th step of /i/ and the time point “20 ms” for the 6th step of /i/ would be entered with the same /0/ values. The steady-state vowel replacements were manipulated at the zero crossing point using Praat.

2.3. Procedure

The perception experiments took place in a sound-attenuated room at National Chiao Tung University. Listeners were required to complete a questionnaire about their language background in advance. The experiment consisted of two identification tests, one for heed-hid, the other for who’d-hood; and each part contained five blocks. The order of the two tests was randomized for each subject. They were only told that all of the stimuli will sound like heed and hid, or who’d and hood. The experiment was run on a laptop using Paradigm 1.0 software. A total of 35 re-synthesized stimuli (5 Duration Steps × 7 Spectrum Steps) were randomized for the ten blocks. During each trial, a cursor appeared at the middle of the screen for 500 ms to assist listeners to concentrate their attention. Each stimulus was played once and the word options appeared on the screen simultaneously. The position (left or right) of the word options was randomly assigned for each trial. Subjects had to use a mouse to click on the option within 5 seconds after hearing the stimulus. The software would automatically move on to the next trial when either the subject had made a response or when the response time had run out. The inter-trial interval was 1 second. Prior to the experiment, there was a six-trial practice section for each test. The stimuli for the practice section were the same, heed-hid and who’d-hood, but they were natural speech produced by the same speaker. The responses were excluded for analysis. After completing the first test, there was a five-minute break for the subjects. The whole experiment took around 40 minutes.

3. Results

3.1. heed-hid

As shown in Figure 1, the proportion of heed responses was higher at step 4 and 5 than at step 1 and 2 for L1-English. Similar observation can be made for the experienced L1-Mandarin, but the proportion of heed responses was always higher for long duration than for short duration. As for inexperienced, by contrast, the proportion of heed responses was less affected as the spectra varied.

A Group (3) × Spectrum (5) two-way repeated measures on the proportion of heed responses was performed. Results showed that no significant effect of group [F(2, 24) = 0.897, p < .421], but a significant effect of spectrum [F(4, 96) = 87.287, p < .01] and a significant interaction [F(8, 96) = 16.273, p < .01]. Five one-way ANOVAs were used to further determine the effect of spectrum among the three groups. Results showed a significant effect of group for step 1 [F(2, 24) = 22.733, p < .01], step 2 [F(2, 24) = 14.897, p < .01], step 4 [F(2, 24) = 9.614, p < .01], and step 5 [F(2, 24) = 19.636, p < .01]. Post-hoc Scheffe pairwise comparisons concerning the group main effect revealed that both the L1-English and experienced groups differed significantly from the inexperienced group (p < .01), with no significant difference between the L1-English and experienced groups.

Figure 1: Proportion of heed responses over the spectrum (dur: duration).
Comparing across the two endpoints (step 1 & 5), post-hoc paired t-tests indicated that the proportion of heed responses was significantly different between step 1 and 5 for L1-English \( (t(20) = 82.947, p < .01) \) and experienced \( (t(20) = 9.912, p < .01) \).

Figure 2: Proportion of heed responses over the duration.

As shown in Figure 2, the proportion of heed responses was less affected as duration varied for both L1-English and experienced L1-Mandarin. With reference to inexperienced, however, the proportion of heed responses was higher at step 6 and 7 than at step 1 and 2.

A Group (3) × Duration (7) two-way repeated measures on the proportion of heed responses was performed. Results showed that no significant effect of group \( F(2, 24) = 2.467, p = .106 \), but a significant effect of duration \( F(6, 144) = 36.655, p < .01 \) and a significant interaction \( F(12, 144) = 15.376, p < .01 \). Seven one-way ANOVAs were used to further determine the effect of duration among the three groups. Results showed a significant effect of group for step 1 \( F(2, 24) = 11.75, p < .01 \), step 2 \( F(2, 24) = 9.122, p < .01 \), step 5 \( F(2, 24) = 12.901, p < .01 \), step 6 \( F(2, 24) = 29.007, p < .01 \), and step 7 \( F(2, 24) = 24.975, p < .01 \). Post-hoc Scheffe pairwise comparisons revealed that the proportion of heed responses was significantly different between step 1 and 7 only for the inexperienced \( (t(20) = 36.273, p < .01) \).

3.2. who’d-hood

As shown in Figure 3, for L1-English and experienced, the proportion of who’d responses was higher at step 4 and 5 than at step 1 and 2. By contrast, the proportion of who’d responses was less affected as the spectra varied for inexperienced.

A Group (3) × Spectrum (5) two-way repeated measures on the proportion of who’d responses was performed. Results showed that no significant effect of group \( F(2, 24) = .343, p = .713 \), but a significant effect of spectrum \( F(4, 96) = 89.74, p < .01 \) and a significant interaction \( F(8, 96) = 10.269, p < .01 \). Five one-way ANOVAs were used to further determine the effect of spectrum among the three groups. Results showed a significant effect of group for step 1 \( F(2, 24) = 13.039, p < .01 \), step 2 \( F(2, 24) = 1.217, p < .01 \), step 4 \( F(2, 24) = 11.171, p < .01 \), and step 5 \( F(2, 24) = 11.534, p < .01 \). Post-hoc Scheffe analysis revealed that both the L1-English and experienced groups differed significantly from the inexperienced group \( (p < .01) \), with no significant difference between the L1-English and experienced groups.

Comparing across the two endpoints (step 1 & 5), post-hoc paired t-tests showed that the proportion of who’d responses was significantly different between step 1 and 5 for L1-English \( (t(8) = 60.104, p < .01) \), experienced \( (t(20) = 9.037, p < .01) \), and inexperienced \( (t(20) = 4.371, p < .01) \).
As shown in Figure 4, the proportion of *who'd* responses was less affected for L1-English and experienced. As for inexperienced, the proportion of *who'd* responses was strongly affected as the duration varied.

A Group (3) × Duration (7) two-way repeated measures on the proportion of *who'd* responses was performed. Results showed that no significant effect of group \([F(2, 24) = .343, p = .713]\), but a significant effect of duration \([F(6, 144) = 19.806, p < .01]\) and a significant interaction \([F(12, 144) = 13.521, p < .01]\). Seven one-way ANOVAs were used to further determine the effect of duration among the three groups. Results showed a significant effect of group for step 1 \([F(2, 24) = 13.295, p < .01]\), step 2 \([F(2, 24) = 9.331, p < .01]\), step 6 \([F(2, 24) = 12.984, p < .01]\), and step 7 \([F(2, 24) = 18.832, p < .01]\). Post-hoc Scheffe analysis revealed that both the L1-English and experienced groups differed significantly from the inexperienced group \((p < .01)\), with no significant difference between the L1-English and experienced groups.

Comparing across the two endpoints (step 1 & 7), post-hoc paired t-tests showed that the proportion of *who d* responses was significantly different between step 1 and 7 only for the inexperienced \((t(20) = 13.785, p < .01)\).

### 4. Discussion and conclusion

The results in this study reveal that the extent of language experience can lead to variations in performance for non-native listeners from the same L1 background. Specifically, while the L1-English listeners predominantly differentiated between tense and lax vowels using spectral quality, the inexperienced listeners showed a strong reliance on durational cues. However, the experienced listeners made greater use of spectral cues and fewer durational cues to distinguish between the tense-lax vowel contrasts after gaining more English-language experience. Moreover, it was found that the inexperienced listeners made significant spectral distinction for the *who'd*-hood contrast; though, their performance was still significantly different from L1-English and experienced listeners.

Furthermore, these findings are consistent with the hypothesis that the inexperienced L1-Mandarin listeners perceived the English /i/-/ɪ/ via Single-Category Assimilation and thus had difficulties using spectral cues to differentiate /i/ from /ɪ/ when durational differences were ambiguous. With regards to the experienced L1-Mandarin listeners, it is possible to suggest that they perceive the tense-lax vowel pairs via Category-Gooodness Difference or Two-Category Assimilation after gaining a significant amount of English-language experience and that they become more native-like. The limitation of this study concerns the synthesized stimuli used in the perception experiments; therefore, the generalization of the results to natural speech might be limited. Future work should focus on why both experienced and inexperienced L1-Mandarin listeners made more uses of durational cues than did L1-English listeners. It is possible to suggest, due to the L1 transfer, that L1-Mandarin listeners were sensitized to the durational differences. However, the exclusive use of durational cues by inexperienced listeners might be attributed to other factors.

### 5. References


