The Effect of Input on the Production of English Tense and Lax Vowels by Chinese Learners: Evidence from an Elementary School in China

Mengrou Li, Ying Chen, Jie Cui**

Nanjing University of Science and Technology, Nanjing, Jiangsu, China

limengrou@njust.edu.cn, ychen@njust.edu.cn, jiecui@njust.edu.cn

Abstract

In both American and British English, tense high vowels /i/ and /u/ show extreme positions of the tongue and lips in articulation rather than their lax counterparts /ɪ/ and /ʊ/. However, the tenseness contrast in English is taught to Chinese learners in classroom by most instructors as duration difference—/i/ and /u/ are longer than /ɪ/ and /ʊ/ respectively. The present study therefore examines English production of /i/ vs. /ɪ/ and /u/ vs. /ʊ/ by Chinese elementary students and investigates how L2 beginners actually realize the target vowels and how their production resembles that of their classroom instructors and talkers who recorded their teaching materials. The results show that the students differentiated /i/ from /ɪ/ and /u/ from /ʊ/ mainly in duration and marginally in F2 but not in F1. Their production was found closer to their English teacher's than the textbook recordings' and native English speakers', suggesting the input from the teachers significantly affects the English production of elementary school students in China.

Index Terms: tense vowel, lax vowel, English, L2 speech, input

1. Introduction

Vowels are categorized based on tongue position (high or low, front or back), lip shape (rounded or unrounded) and glottal gesture (tense or lax). Jacobson and Halle [1] classified vowels into tense and lax vowels based on the amount and spread of the energy in spectrum and duration and the deformation of the vocal tract. The present study is interested in how Chinese elementary school students as L2 beginners acoustically distinguish high front vowels /i/ vs. /ɪ/ and high back vowels /u/ vs. /ʊ/ in their English production.

It has been postulated that the pronunciation can be influenced by the phonological distance between the first language (L1) and the second language (L2) [2]. Theoretical models of L2 speech, such as speech learning model (SLM) [3], have been applied to explore this factor. SLM proposed that L2 learners may fail to identify the phonetic difference between L2 sounds and L1 sounds that are close to L2, however, they may gradually discern the difference with the increasing experience of L2 [3].

In China, English is mostly taught in the classroom as a foreign language. The Acquisition-Learning Hypothesis of Input Hypothesis [4] proposed that "acquisition" is a way for children to acquire their L1 subconsciously, while "learning" is a conscious way to know about a language. Therefore, for most Chinese learners, they learn English instead of acquiring it. Flege and Liu [5] pointed out that the input from the native speakers that the Chinese adults living in United States received

had influenced their progress of learning L2. Similarly, a large amount of high-quality input is also important for foreign language learners [6]. Both the quality and quantity of the input the learners have received impact with their success of long-term L2 learning [7]. In the situation of foreign language teaching and learning, the teachers' speech production in classroom plays a crucial part in the input of English as a foreign language for the students.

Therefore, the current study explores the following four research questions: (1) Do the elementary school students distinguish English tense and lax vowels in their production? (2) Does these learners' production acoustically resemble native speakers'? (3) Does these students' production acoustically resemble their English teachers' production? (4) How do they distinguish /i/ from /ɪ/ and /u/ from /ʊ/ acoustically?

2. Methods

2.1. Participants

Thirty-five fifth graders and their English teacher from X Elementary School in Nanjing participated in this experiment. All the students were born and raised in Nanjing, started learning English from Grade Three (around nine years old), and at the time of testing had been taught by the same female Chinese teacher of English whose native language is Mandarin. All the students had three to five English classes each week during school terms.

2.2. Stimuli

Twenty monosyllabic words (see Table 1) of relatively high frequency were selected from the English textbooks of the third to fifth grade. All the twenty target words are CVC syllable. Two English teachers from X Elementary School confirmed that all words included in the stimuli had been taught in class before the experiment.

Table 1: *Target words.*

/i/	/ɪ/	/u/	/ʊ/
these	this	food	foot
meet	big	shoes	good
teach	give	whose	book
week	miss	soup	cook
read	sit	move	put

The carrier sentences for recording were as follows: *X, I say X. I say X again. I say X four times.* This design resulted in the first production of the target words as isolation, the second as in sentence-final position, the third as in prevocalic context, and the last as in preconsonantal context.

2.3. Procedures

Parents of the fifth graders were invited to participate in an online survey. The survey was designed to collect demographic information of the students, the input of English after class, and dialect-speaking environment outside of the classroom.

Recordings took place in a broadcasting studio in X Elementary School. Participants were asked to read aloud the items in the printed stimuli listed in pre-determined random orders. A Marantz professional solid state recorder PMD661 and a Shure professional unidirectional head-worn dynamic microphone MDR-7506 were used for recording with a sampling rate of 44,100 Hz. Stimuli were directly recorded into an SD card and saved on a computer.

2.4. Data analysis

Praat version 6.0.43 [8] and the Praat script FormantPro version 1.4 [9] were used to process the data. Acoustic measurements, including the first (F1) and second (F2) formants and raw duration of the vowels, were extracted for the targeted vowels. F1 and F2 of vowels are the indicators of vowel height and backness [10]. Relative duration (duration of the vowel segment divided by the length of the entire syllable) was calculated to normalize the participants' speech rate. Measurements of the students' vowel production of /i/, /ɪ/, /u/ and /ʊ/ were analyzed in repeated measures ANOVAs and *Post hoc* paired samples *t*-tests. Dependent variables included F1, F2, raw duration and relative duration. Vowel (/i/ vs. /ɪ/ and /u/ vs. /ʊ/) and phonetic context (isolation, final, prevocalic and preconsonantal) were the two within-subjects factors. If there was an interaction between vowel and phonetic context or main effect of vowel or phonetic context, *post hoc* paired samples *t*-tests were performed to compare tense and lax vowels in each phonetic context and to compare the same vowels in different contexts. The four acoustic measurements of the students' production were plotted by vowels and phonetic contexts in graphs.

Because of correlated differences of the formants in vowel acoustics, the Euclidean Distance (see Formula 1) was applied to compare the formant structure, i.e. vowel quality, between tense and lax vowels produced respectively by the students, the English teacher and textbook recordings.

$$\text{Distance} = \sqrt{(F1_{\text{tense}} - F1_{\text{lax}})^2 + (F2_{\text{tense}} - F2_{\text{lax}})^2} \quad (1)$$

F1 and F2 of /i/ vs. /ɪ/ and /u/ vs. /ʊ/ were plotted to show vowel space in order to compare the production of the students, the English teacher and the textbook talkers.

3. Results

3.1. Production of the students

The results of repeated measures ANOVAs (Figure 1) indicate no interaction between vowel and phonetic context and no main effect of vowel or phonetic context on F1 of both vowel pairs.
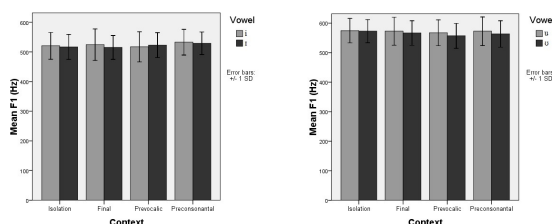

Figure 1: *Mean F1 (Hz) of /i/-/ɪ/and /u/-/ʊ/ by phonetic context.*

The results of ANOVAs on F2 show an interaction between vowel and phonetic context of /i/ and /ɪ/ ($F(3,102) = 4.318, p = .007$), but not of /u/ and /ʊ/. The main effects of vowel /i/ and /ɪ/ ($F(1,34) = 9.361, p = .004$) and of /u/ and /ʊ/ ($F(1,34) = 25.353, p < .001$) and phonetic context of /i/ and /ɪ/ ($F(3,102) = 3.435, p = .020$) and of /u/ and /ʊ/ ($F(3,102) = 3.063, p = .031$) were both significant.

Post hoc paired samples *t*-tests show that F2 of /i/ and /ɪ/ in prevocalic ($t(34) = -2.650, p = .012$) and preconsonantal ($t(34) = -4.441, p < .001$) contexts were significantly different and that F2 of /u/ and /ʊ/ in isolation ($t(34) = 4.209, p < .001$), final ($t(34) = 2.915, p = .006$) and preconsonantal ($t(34) = 3.777, p = .001$) contexts were significantly different. A Bonferroni correction was applied to adjust the threshold of *p* value to 0.0125 (0.05/4) [11]. The F2 of /i/ and /ʊ/ was significantly lower than that of /ɪ/ and /u/.

Post hoc paired samples *t*-tests also show that F2 of /i/ in final and preconsonantal contexts ($t(34) = 4.393, p < .001$) was significantly different and that F2 of /ʊ/ in isolation and prevocalic context ($t(34) = -4.277, p < .001$), and in prevocalic and preconsonantal contexts ($t(34) = 3.627, p = .001$) were significantly different (Figure 2). A Bonferroni correction was applied to adjust the threshold of *p* value to 0.004 (0.05/12) [11].
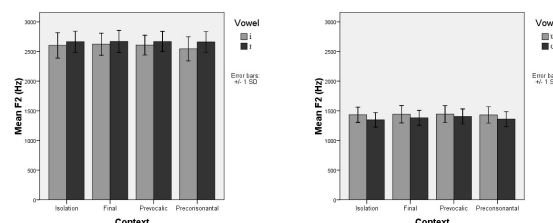

Figure 2: *Mean F2 (Hz) of /i/-/ɪ/and /u/-/ʊ/ by phonetic context.*

The results of ANOVAs on raw duration indicate no interaction between vowel and phonetic context of both vowel pairs, but the main effects of vowel of /u/ and /ʊ/ ($F(1,34) = 108.985, p < .001$) and phonetic context of /u/ and /ʊ/ ($F(3,102) = 17.060, p < .001$) and of /i/ and /ɪ/ ($F(3,102) = 8.071, p < .001$) were both significant.

Post hoc paired samples *t*-tests show that the raw duration between /u/ and /ʊ/ in isolation ($t(34) = 7.367, p < .001$), final ($t(34) = 5.696, p < .001$), prevocalic ($t(34) = 6.896, p < .001$) and preconsonantal ($t(34) = 7.527, p < .001$) contexts were all significantly different. The raw duration of /u/ was longer than that of /ʊ/.

Post hoc paired samples *t*-tests show that the raw duration of /i/ in isolation and final context ($t(34) = 5.151, p < .001$) and in isolation and prevocalic context ($t(34) = 5.201, p < .001$) were significantly different and that the raw duration of /u/ in isolation and final context ($t(34) = 7.262, p < .001$), in isolation and prevocalic context ($t(34) = 6.043 p < .001$), and in prevocalic and preconsonantal contexts ($t(34) = -4.150, p < .001$) were significantly different.

Post hoc paired samples *t*-tests also show that the raw duration of /ɪ/ in isolation and prevocalic context ($t(34) = 5.611, p < .001$), in final and prevocalic contexts ($t(34) = 3.440, p = .002$), and in prevocalic and preconsonantal contexts ($t(34) = -3.167, p = .003$) were significantly different and that the raw duration of /ʊ/ in isolation and final context ($t(34) = 9.194, p < .001$), and in isolation and prevocalic context ($t(34) = 5.750, p < .001$), were significantly different (Figure 3).
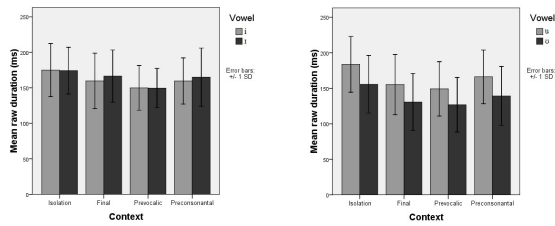
Figure 3: *Mean of the raw duration (ms) of /i/-/ɪ/ and /u/-/ʊ/ by phonetic context.*

The results of ANOVAs on relative duration indicate an interaction between vowel and phonetic context of /i/ and /ɪ/ ($F(3,102) = 6.164$, $p = .001$) and of /u/ and /ʊ/ ($F(3,102) = 6.232$, $p = .001$). The main effects of vowel of /i/ and /ɪ/ ($F(1,34) = 122.976$, $p < .001$) and of /u/ and /ʊ/ ($F(1,34) = 42.214$, $p < .001$) and phonetic context of /i/ and /ɪ/ ($F(3,102) = 7.101$, $p < .001$) and of /u/ and /ʊ/ ($F(3,102) = 19.338$, $p < .001$) were significant.

Post hoc paired samples *t*-tests show that the relative duration between /i/ and /ɪ/ in isolation ($t(34) = -5.081$, $p < .001$), final ($t(34) = -8.753$, $p < 0.001$), prevocalic ($t(34) = -6.070$, $p < .001$) and preconsonantal ($t(34) = -7.549$, $p < .001$) contexts were all significantly different and that the relative duration between /u/ and /ʊ/ in final ($t(34) = -2.721$, $p = .010$), prevocalic ($t(34) = -6.579$, $p < .001$) and preconsonantal ($t(34) = -6.295$, $p < .001$) contexts were all significantly different. The relative duration of the lax vowel /ɪ/ and /ʊ/ was respectively greater than that of the tense vowel /i/ and /u/.

Post hoc paired samples *t*-tests show that the relative duration of /i/ in isolation and final context ($t(34) = 5.726$, $p < .001$) and in final and prevocalic contexts ($t(34) = -4.370$, $p < .001$) were significantly different and that the relative duration of /u/ in isolation and final context ($t(34) = 5.032$, $p < .001$), and in final and preconsonantal contexts ($t(34) = -4.559$, $p < .001$) were significantly different.

Post hoc paired samples *t*-tests also show that the relative duration of /ɪ/ in final and preconsonantal contexts ($t(34) = -3.117$, $p = .004$), and in prevocalic and preconsonantal contexts ($t(34) = -3.953$, $p < .001$) were significantly different and that the relative duration of /ʊ/ in isolation and final context ($t(34) = 6.521$, $p < .001$), in final and prevocalic contexts ($t(34) = -6.112$, $p < .001$), and in final and preconsonantal contexts ($t(34) = -6.281$, $p < .001$) were significantly different (Figure 4).
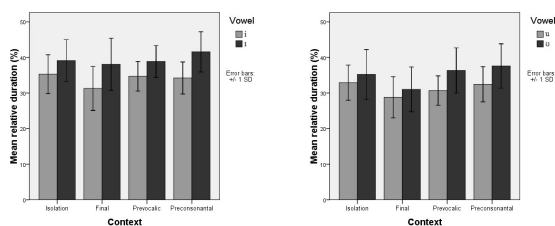

Figure 4: *Mean of the relative duration (%) of /i/-/ɪ/ and /u/-/ʊ/ by phonetic context.*

In summary, the differences between /i/ and /ɪ/ produced by the students were mainly found in relative duration and F2 in prevocalic and preconsonantal contexts. The differences between /u/ and /ʊ/ produced by students were mainly found in raw and relative durations of vowels and F2 in isolation, final and preconsonantal contexts. The relative duration of tense and lax vowels was contrary to the teachers' instruction.

3.2. The students' production vs. the input

The top panel of Figure 5 illustrates the vowel space of /i/ and /ɪ/ of the students, the English teacher and the male and female textbook recordings. The bottom panel in Figure 5 shows the vowel space of /u/ and /ʊ/ based on productions of the students, the English teacher, and textbook recordings respectively.
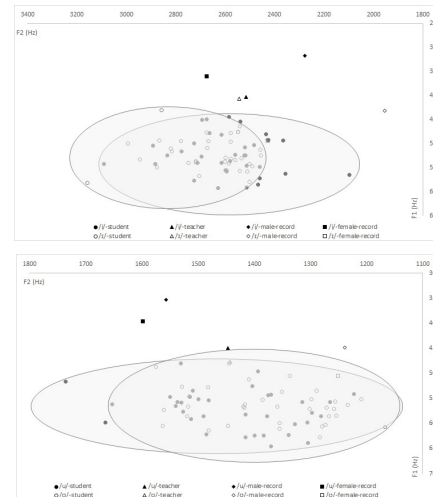

Figure 5: *Vowel spaces of the students, English teacher and textbook recordings of /i/-/ɪ/ (top) and /u/-/ʊ/ (bottom).*

Figure 6 shows the measurements of Euclidean Distance of /i/ and /ɪ/ (top panel) and of /u/ and /ʊ/ (bottom panel) by the students, the English teacher and the textbook recordings.

The measurements of /i/ and /ɪ/ of nearly 71% of students were closer to the measurements of their English teacher than to the pronunciation in textbook recordings. More than 91% of students' measurements of /u/ and /ʊ/ were closer to those of their English teacher than to the textbook recording.
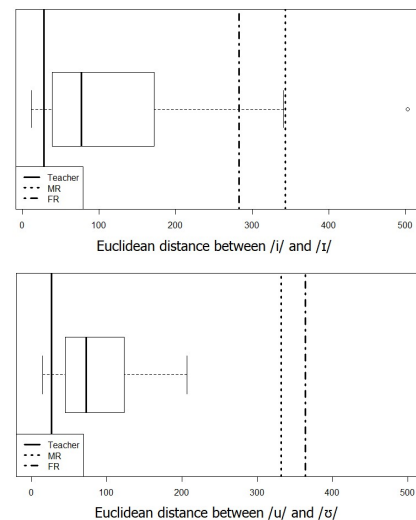

Figure 6: *Boxplot of Euclidean Distance between /i/ and /ɪ/ (top) and between /u/ and /ʊ/ (bottom) of the students, English teacher, male recordings (MR) and female recordings (FR).*

In summary, compared with the textbook recordings, F1 and F2 of the vowels produced by the students mainly resembled their English teacher's measurements.

4. Discussion

The statistical results show that the fifth-grade students in X Elementary School mainly used duration to differentiate tense and lax vowels in their English production. The acoustic measures of F1 and F2 in these learners' vowel production resembled that of their teacher rather than the teaching materials and native speaker's samples.

In reality, vowel quality is more significant than duration in the differentiation of tense-lax pairs [2]. Sample formants of a student, the teacher, textbook recordings were taken from the midpoint of the vowel duration with 50 ms on each side and illustrated in Figures 7 and 8. The statistical results and sample spectrograms show similar F1 values between /i/ and /ɪ/ and between /u/ and /ʊ/ produced by the students and the teacher. The F2 of /i/ produced by the students is lower than that of /ɪ/ (Figure 7), which resembled the pattern of the teacher (Figure 7) and was contrary to the pattern of the textbook recordings (Figure 8) and of a native speaker of American English (Figure 9) [10]. The F2 of /ʊ/ produced by the students is lower than that of /u/ (Figure 7), which resembled the pattern of the teacher (Figure 7) and textbook recordings (Figure 8), and was contrary to the pattern of the native speaker of American English (Figure 9) [10]. The statistical analyses also reveal that the students used F2 to distinguish /i/ vs. /ɪ/ in prevocalic and preconsonantal contexts and /u/ vs. /ʊ/ in isolation, final and preconsonantal contexts. These F2 patterns in the students' production deviate from those of the textbook recordings (Figure 8) and the native speaker of English (Figure 9) but similar to the teachers' patterns as Figure 7 shows.
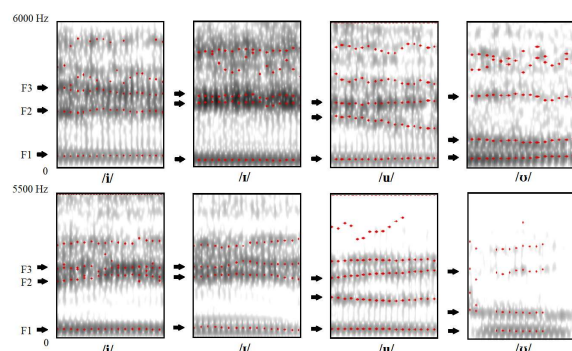

Figure 7: *Vowel formants (F1, F2, F3) of /i/-/ɪ/ and /u/-/ʊ/ by a student (top) and the English teacher (bottom).*
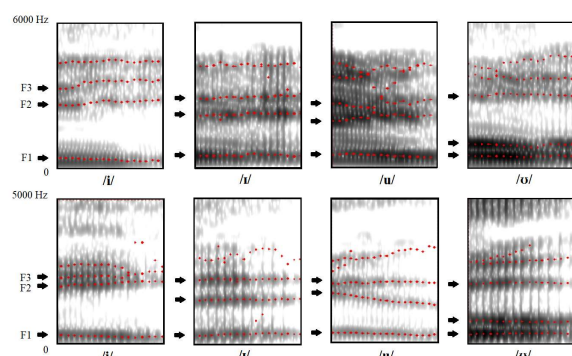

Figure 8: *Vowel formants (F1, F2, F3) of /i/-/ɪ/ and /u/-/ʊ/ by the textbook recordings of female voice (top) and male voice (bottom).*
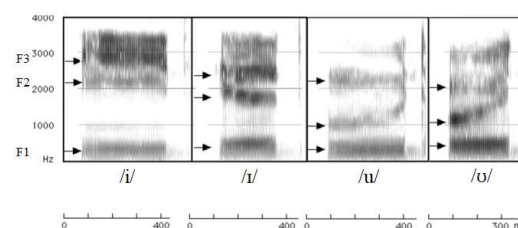

Figure 9: *Vowel formants (F1, F2, F3) of /i/-/ɪ/ and /u/-/ʊ/ by a native speaker of American English [10].*

The resemblance of the students' and their teacher's production can be explained by the Speech Learning Model (SLM) [3] that long-term success of L2 pronunciation needs a large quantity of high-quality input [7]. In this study, vowel production of the English teacher deviates from the patterns of the native speaker of American English (Figure 7 vs. Figure 9), suggesting inaccurate input in the classroom. Nonetheless, the teacher's in-class production is the major input to the students. Therefore, the lack of high-quality input hinders nativelike production of tense and lax vowels in English for the students.

Previous studies show that many L2 learners of English view the length of vowels as the main difference between /i/ and /ɪ/ [2][12]. Although the teacher in this study also taught the students that lax vowels were shorter than the corresponding tense vowels, the students' production was contrary to the English teachers' instruction. It is possible that the teacher emphasized more on the pronunciation of the lax vowels /ɪ/ and /ʊ/ than the tense vowels /i/ and /u/ because there is no tense vs. lax vowel contrast and there are only /i/ and /u/ in Mandarin. The teacher might have demonstrated the lax vowels with focus and in low speech rate. According to the literature, it is easier for learners to understand L2 speech in a lower speech rate [13]. Lower speech rate refers to longer raw duration. The students might have perceived the demos of especially long duration of lax vowels and produced them as what they had heard. Also, they might have followed the teacher and produced the lax vowels more slowly than the tense vowels.

Therefore, it is plausible that improper teaching methods and non-nativelike input from the English teacher resulted in the students' inaccurate production of tense and lax vowels in English. The students, on the other hand, did not immerse themselves in high-quality input, such as textbook recordings or other native speakers' speech production.

5. Conclusions

The present study has explored the production of /i/ vs. /ɪ/ and /u/ vs. /ʊ/ by fifth graders in an elementary school in China. The results reveal that these students mainly used duration to distinguish these tense-lax vowel pairs and F2 in some phonetic contexts. The students' production resembled their English teacher's production in F2 but in an opposite direction of vowel duration. This study reconfirms that a large quantity of high-quality input is required to the learners [7] for successful acquisition of L2 speech.

6. References

[1] R. Jacobson and M. Halle, *Fundamentals of Language*. The Hague: Mouton, 1956.

[2] T. M. Derwing, "Curriculum issues in teaching pronunciation to second language learners," *Phonology and Second Language Acquisition*, pp. 347-369, 2008.

[3] J. E. Flege, "Second Language Speech Learning: Theory, Findings, and Problems," *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Timonium, MD, USA: York Press, 1995.

[4] S. Krashen, *The input hypothesis: Issues and implications*. London, UK: Longman, 1985.

[5] J. E. Flege and Liu, "The effect of experience on adults' acquisition of a second language," *Studied in second language acquisition*, vol. 23, no. 4, pp. 527-552, 2001.

[6] T. Piske, "Implications of James E. Flege's research for the foreign language classroom," *Language Experience in Second Language Speech Learning: In honor of James Emil Flege*. pp. 301-314, 2007.

[7] J. E. Flege, "A non-critical period for second-language learning," *A Sound Approach to Language Matters*. Department of English, Aarhus University, 2018.

[8] P. Boersma and D. Weenink, "Praat: doing phonetics by computer. Version 6.0.43," *http://www.fon.hum.uva.nl/praat/*, 2018.

[9] Y. Xu, "FormantPro. praat. Version1.4," *http://www.homepages.ucl.ac.uk/~uclyyix/FormantPro*, 2016.

[10] K. Johnson and P. Ladefoged, *A Course in Phonetics*. Cengage Learning, 2015.

[11] T. V. Perneger, "What's wrong with Bonferroni adjustments," *Bmj*, vol. 316, no. 7139, pp. 1236-1238, 1998.

[12] X. Wang and M. J. Munro, "Computer-based training for learning English vowel contrasts," *System 32*, pp. 539-552, 2004.

[13] T. M. Derwing and M. J. Munro, *Pronunciation Fundamentals. Evidence-based Perspectives for L2 Teaching and Research.* Amsterdam/Philadelphia, PA: John Benjamins Publishing Company, 2015.