



Tone production of Mandarin disyllabic words by Korean learners

Jung-Yueh Tu¹, Yuwen Hsiung², Jih-Ho Cha³, Min-Da Wu³, & Yao-Ting Sung³

¹ International Chinese Education Center, School of Humanities, Shanghai Jiao Tong University

² Department of Teaching Chinese as a Second Language, Chung Yuan Christian University

³ Center of Learning Technology for Chinese, National Taiwan Normal University

jungyueh.tu@outlook.com, hsiungyu@hotmail.com, toddcha@gmail.com, tarzan8720@yahoo.com.tw, & sungtc@ntnu.edu.tw

Abstract

This study investigated the production of Mandarin disyllabic tones by Korean speakers. We focused on disyllabic tones since it is disyllabic words that dominate the vocabulary in modern Mandarin. In particular, we examined the tonal production in Mandarin by Korean speakers, which is somewhat understudied in the previous literature. In our study, there were 25 Korean learners of Mandarin, who were requested to produce 80 Mandarin disyllabic words with all tonal combinations (except for the neutral tone). The overall results showed a level of difficulty: Tone 2 = Tone 3 > Tone 1 = Tone 4. Most errors in the first syllable were found for Tone 2 and Tone 3 when followed by Tone 1 or Tone 4 (both start with a high pitch). In the second syllable, error patterns among those tones were not significantly different. Other errors for specific tones were also analyzed. The findings are discussed from the perspectives of phonetic representations of lexical tones as well as effects of native phonology in the first language.

Index Terms: Mandarin tones, disyllabic words, tonal production, Korean CSL/CFL learners

1. Introduction

Mandarin Chinese is a tone language, and its tone system poses a problem in pronunciation to many Chinese as a second/foreign language (CSL/CFL) learners [1, 2]. Prior research on the learning of Mandarin tones suggests effects of learners' language experience and lexical tones' phonetic nature [3, 4, 5]. Building on our previous study regarding error patterns of Mandarin disyllabic words in Japanese learners [5], the current study intends to expand our understanding of tonal production by investigating CSL learners with experience of another language profile. This study targets disyllabic tones since Modern Chinese contains a high percentage of disyllabic words (69.8% of the total words) [6], which dominates the vocabulary [1, 6, 7]. To further explore a broader picture of phonological representations for lexical tones by second language (L2) speakers, this study focuses on the production of Mandarin disyllabic tones by Korean speakers.

1.1. L2 production models

Theories accounting for the production of L2 speech, particularly L2 intonation, have been proposed. *Prosodic Transfer Hypothesis* [8, 9] assumes that the types of prosodic representations that can be built in the L2 are restricted by the prosodic constraints in the L1, and hypothesizes that L2 learners fail to produce native-like speech because they are not capable of constructing prosodic representations required for the L2 speech production if these representations are disallowed in the first language (L1).

Another relevant model is the *Asymmetry Hypothesis* [10], which hypothesizes that differences of prosodic domain and rule applications between L1 and L2 affect L2 phonological acquisition. When the prosodic domain in the L1 (e.g., word level) is smaller than L2 (e.g., intonational phrase), the learners will have positive evidence for the occurrence of the alternation in the larger prosodic domain, in which the L2 acquisition will eventually take place. On the contrary, if the prosodic domain in the L1 (e.g., intonational phrase) is larger than L2 (e.g., word level), the learners will need negative evidence (e.g., forms and correction) for learning the proper domain setting of the L2 alternation, in which L2 acquisition will not take place.

1.2. Prosodic aspects in Mandarin and Korean

Chinese is a tonal language, which uses pitch to distinguish lexical meaning. It has four lexical tones plus a neutral tone. The first tone is a high-level tone (HH). The second tone is a mid-rising tone (MH). The third tone is a low-dipping tone (MLH). The fourth tone is a high-falling tone (HL). Mandarin is well-known for its *third tone sandhi*, where a third tone is turned into a rising tone, similar to the second tone, when immediately followed by a third tone. Chinese tone is superimposed on monosyllables. In connected speech, it is found that a third tone always appears as a *half third tone*, with only the low-falling contour shape [11].

In standard (Seoul) Korean, a pitch pattern can be represented by two pitches, high (H) and low (L), and their combinations, such as falling (HL) or rising (LH). Jun [12, 13] proposes a model of intonation, which defines two prosodic units above the phonological word: Intonation Phrase (IP) and Accentual Phrase (AP). An IP is the highest prosodic units and can have more than one AP. An IP is marked by a boundary tone, realized on the phrase-final syllable. An AP is marked by a tonal pattern, phrase-initial rise and phrase-final rise (i.e., LH-LH or HH-LH). An AP begins with either a rising tone (LH) or a high plateau (HH) on the first two syllables. An IP has phrase-final lengthening which is not found in an AP. When an AP has less than four syllables, the medial L or H or both is often not realized, leading to three tone patterns: a simple rise (LH), an early rise (LHH), and a late rise (LLH) for L-initial APs.

1.3. Research questions

The goal of this study is to understand how L1 experience with prosodic features affects L2 speech production at the suprasegmental level by examining the Mandarin tone patterns of disyllabic words in Korean learners. The current study aims to address two research issues: a.) the general error patterns of Mandarin disyllabic tones by Korean learners, and b.) how Korean prosodic system influences tonal production of Mandarin.

2. Method

2.1. Participants

There were 25 Korean learners participating in this study (15 females, 10 males; age range: 21-47 years old; mean age: 27.04 years old). The average length of learning Mandarin for participants was 2.5 years. They came from different regions of Korea and their native language was Korean. None had difficulty in hearing and speaking.

2.2. Stimuli

Eighty disyllabic Mandarin words were used for the stimuli, which consisted of four tones in the first syllable, four tones in the second syllable, and five disyllabic words in each tonal combination (4 tones * 4 tones * 5 disyllabic words = 80). Those words included all Mandarin tonal combinations, except for the neutral tone (which is not the research target in this study). The four tones were arranged into disyllabic words, and then 16 tonal combinations can be retrieved (4 tones * 4 tones = 16 pairs). The numbers 1, 2, 3, 4 represent Tone 1, Tone 2, Tone 3, and Tone 4, respectively. Note that the tonal combination 3-3 should be pronounced as 2-3 due to the third tone sandhi. The design of word chart is illustrated below.

1-1, 2-1, 3-1, 4-1
 1-2, 2-2, 3-2, 4-2
 1-3, 2-3, 3-3, 4-3
 1-4, 2-4, 3-4, 4-4

Consider that tonal production depends on participants' knowledge of the words with their tonal specifications. In the stimuli, the disyllabic words were mostly taken from the *800 Chinese Words for Beginners* [14], which were familiar to participants. The presentation order of the words was randomized to avoid participants' expectation of a pattern. Every word was presented with Mandarin phonetic symbols (Hanyu Pinyin) and traditional Chinese characters.

2.3. Procedures

The participants did the production experiment in a quiet room. Their utterances of the stimuli were recorded by a stand-alone microphone with a sampling frequency of 16 kHz and a resolution of 16 bit on a desktop. The recording was attained through using a recording program released by Speech and Machine Intelligence Laboratory of National Taiwan Normal University. First, the participants were familiarized with the practice section. Then, the participants were requested to produce and record the stimuli. They saw a disyllabic word on the screen, followed by clicking the recording icon when they were ready to produce the word. The participants were asked to read out those disyllabic words with the correct lexical tones at a normal rate.

2.4. Analysis

The recordings were judged by three phonetically trained native speakers of Mandarin, who further identified the tonal errors made by the participants. The 3 native speakers evaluated the recordings and labelled the tone of each syllable in the disyllabic words with a choice among the four lexical tones. When a production was too ambiguous and could not be

categorized as any lexical tone, it would be labelled as the category 'other.' When there was any disagreement among them, the decision made by the majority was selected. Then the acoustic analysis was carried out using the software PRAAT with visual pitch contour to decide the label of the tone.

3. Results

3.1. Overall accuracy and syllable effect

Two-way repeated-measures ANOVA were conducted for the accuracy within syllable and cross syllable. The main effect of tone was significant [$F(1.94, 46.47) = 14.50$, $MSE = .019$, $p < .001$, $\eta^2 = .38$]. The accuracy of Tone 2 ($M = .850$, $SD = .187$) was significantly lower than Tone 1 ($M = .975$, $SD = .050$) and Tone 4 ($M = .967$, $SD = .065$), as was Tone 3 ($M = .897$, $SD = .186$) (see Figure 1).

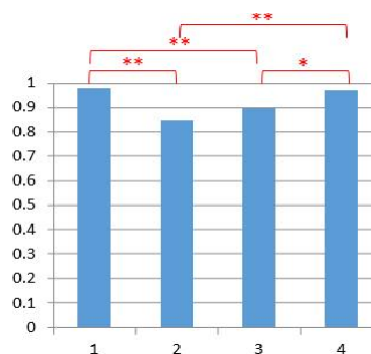


Figure 1. Accuracy of production for individual tone (* means $p < .05$; ** means $p < .01$).

The main effect of syllable was significant [$F(1, 24) = 35.79$, $MSE = .012$, $p < .001$, $\eta^2 = .60$]. The interaction between syllable and tone was also significant [$F(1.82, 43.64) = 11.56$, $MSE = .023$, $p < .001$, $\eta^2 = .33$]. A simple main effect test showed that in the first syllable, the accuracy of production for Tone 2 ($M = .720$, $SD = .186$) was significantly lower than that for other tones. In the second syllable, the accuracy of production for Tone 3 ($M = .902$, $SD = .229$) was significantly lower than that for Tone 1 ($M = .996$, $SD = .014$) and Tone 4 ($M = .996$, $SD = .020$). For Tone 2, the accuracy was significantly higher in the second syllable than that in the first syllable (see Figure 2).

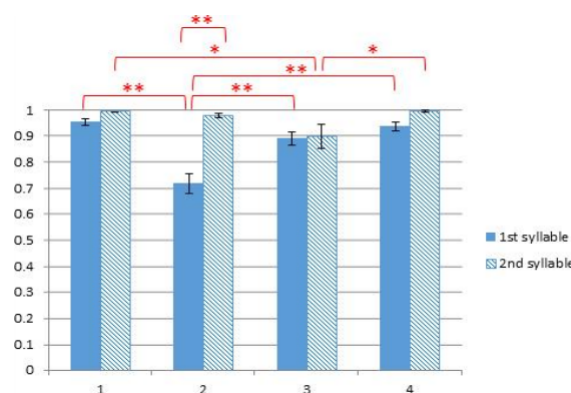


Figure 2. Accuracy for the four tones in the two syllables (* means $p < .05$; ** means $p < .01$).

3.2. First syllable

One-way repeated-measures ANOVAs were conducted for accuracy of tonal production in the first syllable. When Tone 1 in the first syllable was followed by different tones in the second syllable, its accuracy was not significantly different [$F(3, 72) = 2.04$, $MSE = .010$, $p = .115$, $\eta^2 = .08$]. When Tone 2 in the first syllable was followed by different tones in the second syllable, its accuracy was significantly different [$F(1.88, 45.20) = 38.85$, $MSE = .088$, $p < .001$, $\eta^2 = .62$]. When Tone 3 in the first syllable was followed by different tones in the second syllable, its accuracy showed marginally significant difference [$F(1.67, 40.05) = 3.31$, $MSE = .061$, $p = .055$, $\eta^2 = .12$]. When Tone 4 in the first syllable was followed by different tones in the second syllable, its accuracy was not significantly different [$F(3,72) = 0.82$, $MSE = .021$, $p = .488$, $\eta^2 = .03$] (see Figure 3). As shown in Figure 3, the accuracy of production for Tone 2 followed by Tone 1 ($M = .432$, $SD = .374$) and Tone 4 ($M = .504$, $SD = .384$) was significantly lower than that for Tone 2 followed by Tone 2 ($M = .976$, $SD = .065$) and Tone 3 ($M = .968$, $SD = .073$).

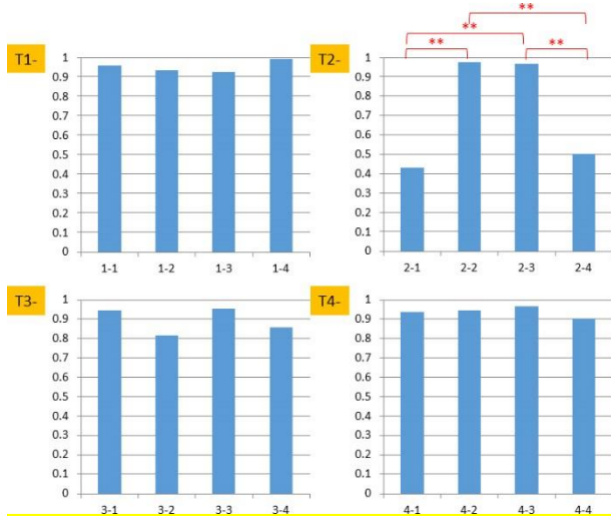


Figure 3. Accuracy of production for individual tone in the first syllable (* means $p < .05$; ** means $p < .01$).

3.3. Second syllable

One-way repeated-measures ANOVAs were conducted for accuracy of tonal production in the second syllable. When Tone 1 in the second syllable was preceded by different tones in the first syllable, its accuracy was not significantly different [$F(3, 72) = 0.66$, $MSE = .001$, $p = .58$, $\eta^2 = .03$]. When Tone 2 in the second syllable was preceded by different tones in the first syllable, its accuracy was not significantly different [$F(3, 72) = 1.65$, $MSE = .029$, $p = .19$, $\eta^2 = .08$]. When Tone 3 in the second syllable was preceded by different tones in the first syllable, its accuracy showed marginally significant difference [$F(2.18, 52.23) = 4.53$, $MSE = .004$, $p = .051$, $\eta^2 = .11$]. When Tone 4 in the second syllable was preceded by different tones in the first syllable, its accuracy was not significantly different [$F(3,72) = 0.89$, $MSE = .008$, $p = .45$, $\eta^2 = .04$] (see Figure 4).

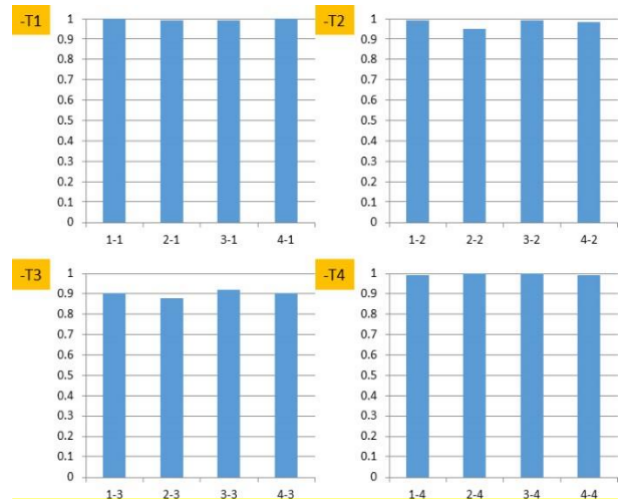


Figure 4. Accuracy of production for individual tone in the second syllable (* means $p < .05$; ** means $p < .01$).

3.4. Tonal error in each combination

The tonal error matrix for the first syllable and the second syllable in each combination are given in Table 1 and Table 2, respectively. Note that realizations of T3 as T2 in the first syllable when followed by another T3 in the second syllable were presented as correct productions based on *third tone sandhi* rule. (The correct response is highlighted in dark grey. When the percent of correct response is lower than 70%, the wrong response is highlighted in light grey.)

Table 1. Tonal error matrix- 1st syllable.

T1		1	2	3	4	Others
	11	96	0	0	3.2	0.8
12	93.6	0.8	0	0.8	4.8	
13	92.8	2.4	0	1.6	3.2	
14	99.2	0	0	0	0.8	

T2		1	2	3	4	Others
	21	0	43.2	54.4	0	2.4
22	0.8	97.6	1.6	0	0	
23	0.8	96.8	1.6	0	0.8	
24	0	50.4	47.2	0	2.4	

T3		1	2	3	4	Others
	31	0	4.8	94.4	0	0.8
32	0	16	81.6	0	2.4	
33	0.8	95.2	3.2	0.8	0	
34	0.8	10.4	85.6	0	3.2	

T4		1	2	3	4	Others
	41	4.8	0	0	93.6	1.6
42	5.6	0	0	94.4	0	
43	2.4	0	0	96.8	0.8	
44	6.4	0	0	90.4	3.2	

Table 2. Tonal error matrix- 2nd syllable.

T1		1	2	3	4	Others
	11	100	0	0	0	0
21	99.2	0	0	0.8	0	
31	99.2	0	0	0	0.8	
41	100	0	0	0	0	

T2		1	2	3	4	Others
	12	0	99.2	0.8	0	0
22	0	95.2	1.6	0.8	2.4	
32	0	99.2	0.8	0	0	
42	0	98.4	0.8	0	0.8	

T3		1	2	3	4	Others
	13	0	7.2	90.4	0	2.4
23	0	9.6	88	0.8	1.6	
33	0	8	92	0	0	
43	0	9.6	90.4	0	0	

T4		1	2	3	4	Others
	14	0	0	0	99.2	0.8
24	0	0	0	100	0	
34	0	0	0	100	0	
44	0	0	0.8	99.2	0	

4. Discussion

4.1. Tonal error patterns

From the overall accuracy for individual tone, the results showed that the percent correct of Tone 2 and 3 is lower than that of Tone 1 as well as Tone 4. It indicates the level of difficulty among the four lexical tones: Tone 2 and Tone 3 > Tone 1 and Tone 4. In the first syllable, the accuracy of production for Tone 2 was significantly lower than that for other tones. In the second syllable, the accuracy of production for Tone 3 was significantly lower than that for Tone 1 and Tone 4. In addition, the syllable effect was found for Tone 2, indicating that the accuracy of Tone 2 in the second syllable was significantly higher than that in the first syllable. The participants were better at producing Tone 2 correctly in the second syllable.

As for the tonal combination, most errors in the first syllable were found in the tone pairs 2-1, 2-4, 3-1, and 3-4. That is, Tone 2 and Tone 3 were mostly misproduced when followed by Tone 1 or Tone 4. Tone 2 and Tone 3 were mostly misproduced as each other. This result was similar to the previous findings for the Japanese learners of Mandarin [5]. It is probably due to the similarity of phonetic properties between Tone 2 and Tone 3, which both contain a rising portion. It also showed that the phonological environment was at play, that is, the misproduced tone in the first syllable was followed by either Tone 1 or Tone 4, which both start with a high pitch. In the second syllable, the percent correct for each of the four tones was all higher than 90%. With the high accuracy rate, no significant difference was found among the four tones in the second syllable.

The findings showed that confusion of Tone 2 and Tone 3 was a common error for Korean learners. Production of Tone 2 and Tone 3 were relatively more difficult and mostly confused with each other.

4.2. Language transfer from Korean phonology

As reviewed in 2.1, Mandarin has four lexical tones while Korean has only two pitches. From the production data by Korean learners, we observed that Korean learners generally have a narrower pitch range, which was similar to the results from Japanese learners [5]. Thus, Korean learners often produced Tone 1 and Tone 4 in Mandarin with a lower starting pitch in terms of tone register. With regard to the tone contour, production of Tone 2 and Tone 3 by Korean learners was somewhat different from that by native speakers. In the production of Tone 2, the pitch rising was not high enough so that it may sound like Tone 3. Hence, the tonal production with a narrower pitch range by Korean learners may lead to their tonal errors. The observation that Korean speakers have a compressed pitch range is also consistent with other studies on L2 speech sounds [5, 15, 16, 17, 18]. The current findings further support the claim that there is a general tendency for L2 speakers to compress pitch range.

Moreover, the transfer from Korean prosody was revealed in the confusion of contour tones, i.e. Tone 2 and Tone 3. One major type of errors was that Tone 2 was produced as a rising tone with an obviously initial dipping portion, which may sound like Tone 3. It is probably because in Korean, when an AP has three or fewer syllable, the medial L or H or both is undershot, resulting in a simple rise (LH), an early rise (LHH), or a late rise (LLH) for the case of L-initial AP [12, 13], where an obviously initial dipping portion emerges. Tone 2 in the first syllable within a Mandarin disyllabic word may be produced as L tone by Korean speakers, which was perceived as Tone 3 by Mandarin speakers.

4.3. Implications for production models

The misproduction of contour tones in Mandarin by Korean learners provides support for *Prosodic Transfer Hypothesis*. The contour tones do not occur in Korean inventory so that Korean learners were unlikely to produce native-like Tone 2 and Tone 3. The results were also in favor of the Asymmetry Hypothesis that lexical tone in Mandarin is smaller than prosodic phrase in Korean. Hence, the Korean learners require negative evidence for success in tonal acquisition.

5. Conclusion

This study examined the production of Mandarin disyllabic tones by Korean learners. The results showed that among the four lexical tones in Mandarin, Tone 2 and Tone 3 were the most difficult and mostly misproduced as each other. It was also observed that confusion of Tone 2 and Tone 3 was a common error for Korean learners. In the first syllable, most errors were found for Tone 2 and Tone 3 when followed by Tone 1 or Tone 4. In the second syllable, no significant difference was found among the four tones. The results suggested effects of phonetic realizations of lexical tones in Mandarin interacting with language transfer from Korean phonology.

6. Acknowledgements

This research is partially supported by the "Aim for the Top University Project" and "Center of Learning Technology for Chinese" of National Taiwan Normal University (NTNU), sponsored by the Ministry of Education, Taiwan, R.O.C. and the "International Research-Intensive Center of Excellence Program" of NTNU and Ministry of Science and Technology, Taiwan, R.O.C. under Grant no. NSC 103-2911-I-003-301, as well as International Chinese Education Center, School of Humanities, Shanghai Jiao Tong University.

7. References

- [1] Chang, K.J., Chen, L.M., & Lee, N.C., "Tonal errors of Japanese students learning Chinese: A study of disyllabic words." *Computational Linguistics and Chinese Language Processing*, 11(3): 281-296, 2006.
- [2] Guo, L & Tao, L., "Tone production in Mandarin Chinese by American students: A case study." In K.M. Chan and H. Kang (Eds.), *Proceedings of the 20th North American Conference on Chinese Linguistics (NACCL-20)*, Vol. 1. (pp. 123-138). Columbus, Ohio: The Ohio State University, 2008.
- [3] Ding, H., "Perception and production of Mandarin disyllabic tones by German learners." *Speech Prosody*, 2012.
- [4] Jongman, A., Wang, Y., Moore, C., & Sereno, J.A., "Perception and production of Mandarin Chinese tones." In P. Li, E. Bates, L.H. Tan, & O. Tseng (Eds.), *The Handbook of East Asian Psycholinguistics* (pp.209-217). Cambridge University Press, 2006.
- [5] Tu, J., Hsiung, Y., Wu, M., & Sung, Y., "Error patterns of Mandarin disyllabic tones by Japanese learners." *Proceedings of the 15th Annual Conference of the International Speech Communication Association (Interspeech-2014)*, Singapore, 2014.
- [6] Duanmu, S., "Stress and the development of disyllabic words in Chinese." *Diachronica* XVI, 1, 1-35, 1999.
- [7] Wang, Y., Jongman, A., & Sereno, J., "Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training." *Journal of the Acoustical Society of America*, 113: 1033-1043, 2003.
- [8] Goad, H., White, L. & Steele, J., "Missing inflection in L2 acquisition: Defective syntax or L1-constrained prosodic representations?" *Canadian journal of linguistics*, 48: 243-263, 2003.
- [9] Goad, H. & White, L., "Prosodic transfer: L1 effects on the production L2 determiners." In *Proceedings of the 30th Annual Boston University Conference on Language Development*, D. Bamman, T. Magnitskaia and C. Zaller (eds), 231-224. 2006.
- [10] Young-Scholten, M., "On positive evidence and ultimate attainment in L2 phonology." *Second language research*, 10: 193-214, 1994.
- [11] Chao, Y. R., "A system of tone letters." *La Maître phonétique* 45, 24-27, 1930.
- [12] Jun, S.A., "Prosody in sentence processing: Korean vs. English." *UCLA Working Papers in Phonetics*, 104: 26-45, 2005.
- [13] Jun, S. A. "The accentual phrase in the Korean prosodic hierarchy." *Phonology*, 15(2): 189-226, 1998.
- [14] Steering Committee for the Test of Proficiency-Huayu (SC-TOP), "800 Chinese Words for Beginners." Online: http://www.sc-top.org.tw/download/800Words_Beginners.pdf, accessed on 20 March, 2013.
- [15] Ullakonoja, R., "Comparison of pitch range in Finnish (L1) and Russian (L2)." *Proc. 16th International Congress of Phonetic Sciences (ICPhS XVI)*, Saarbrücken, 1701-1704, 2007.
- [16] Traunmüller, H. & Eriksson, A., "F0-excursions in speech and their perceptual evaluation as evidenced in liveliness estimations." *Phonetic Experimental Research*, Institute of Linguistics, University of Stockholm (PERILUS), 17, 1-34, 1993.
- [17] Hincks, R., "Processing the prosody of oral presentations." *Proc. InSTIL/ICALL2004 NLP and Speech Technologies in Advanced Language Learning*, Venice, 63-66, 2004.
- [18] Tu, J., "Word Prosody in Loanword Phonology: Focus on Japanese borrowings into Taiwanese Southern Min." Doctoral dissertation, Indiana University, 2013.