Characterizing ESL rhythm produced by Mandarin Speakers Using Both Duration- and Pitch-Based Measures

Hua Lin
University of Victoria, Canada

ABSTRACT

An experiment is conducted to investigate rhythm in the ESL production by native Mandarin speakers. Rhythm is characterized not only by the traditional acoustic rhythmic measures of duration, but pitch-based analysis is also used, which include minimum and maximum F0s, pitch range, number of pitch rises and pitch slope. The results show that the ESL production has not met the native-speaking target on any of the nine acoustic measures, although it is closer to some than others. More specifically, the ESL production on durational measures suggests a consistent trend of a "interlanguage" whose durational values all fall between those of the L1 and L2. The pitch analysis yields varied results, some of which go outside the range between the L2 and L1 values.

Keywords: L2 speech rhythm, ESL, L2 speech production, Mandarin ESL learners, acoustic analysis, gender difference

1. INTRODUCTION

Ever since Arthur Lloyd-James [1](1940) made the observation of a dichotic distinction in the rhythms of the world's languages, phoneticians have been trying to pinpoint what exactly in the human speech is responsible for such a perceptual observation. Many phonetic/acoustic methods have been advanced to classify languages into a stressed-timed group and a syllable-timed group. Earlier efforts included measuring the isochrony of stress intervals in utterances of stress-timed language and the sameness of the syllable durations in the utterances of the syllable-timed language. Unfortunately, in spite of a great deal of research, such stress- versus syllable-timing was not consistently found beyond the auditory perception.

Beginning with Ramus et al.'s [2] 1999 study, the focus came to be on the duration of segmental sequences (e.g., consonantal and vowel sequences as opposed to the syllable or the stress interval). The idea was that those languages of the same rhythm share similar segmental durational characteristics. Within this research paradigm, there was also attempts at examining the rhythmic grouping of interlanguages. Lin and Wang [3] is one earliest attempt at it, which examined the rhythmic properties in the L2 English production by native Mandarin speakers. However more recently, these duration-based methods came to be challenged in such studies as Arvaniti [6] and [5] and Vicenik and Sundara [8]. According to [7], rhythm grouping should be based on pitch, amplitude, and intensity in additional to duration [9]. Newer acoustic approaches have subsequently been proposed which focus on another dimension of speech: pitch (or F0), using methods borrowed from intonation studies. It appears that characterizing rhythm is not complete without taking the parameter of the pitch into consideration.

The purpose of this study is to investigate how second language speech differ in rhythmic measures from either L1 or L2 on *both* durational and pitch measures. To that end, an experiment is conducted to examine the rhythm in the ESL production by native Mandarin speakers, using duration- and pitch-based analyses, and ESL results by Mandarin speakers are compared with those of English (the target) and Mandarin (the mother-tongue). Additionally, the study wishes to gain knowledge as to how the pitch parameters compare with the durational parameters.

2. THE EXPERIMENT

A speech production experiment was conducted to meet the goals of the study. Six native Mandarin speakers (3 male and 3 female) who were ESL speakers, and six native Canadian English speakers (also 3 male and 3 female), all young adults attending college in British Columbia, Canada, participated in the study. Speech production data were collected through recording all 12 participants reading the English passage "the North Wind and Sun." The six ESL speakers were also recorded reading the same story in their native Mandarin.

As a departure from previous studies [3, 4, 5] co-authored by the present author on rhythm, sonorant intervals rather than vowel intervals, and obstruent intervals rather than consonant intervals were segmented out for durational analysis. One of the reasons for the change was that it was often very hard if not impossible to segment sonorant consonants such as glides and liquids from vowels. This is especially true for Mandarin.

On the two types of intervals, four durational measurements were taken: sonorant percentage (%SON), obstruent standard deviation (ΔOBS), sonorant-based nominalized pairewise variability index (nPVI), and obstruent-based raw pairwise variability index (rPVI). The segmentation was done using both audio and visual cues in Praat (Paul Boersma and David Weenink; Version 5.4.04). For pitch analysis, sentences were isolated, and the following measures: minimum F0, maxmum F0, F0 slope and number of pitch rises were taken.

3. RESULTS AND DISCUSSION

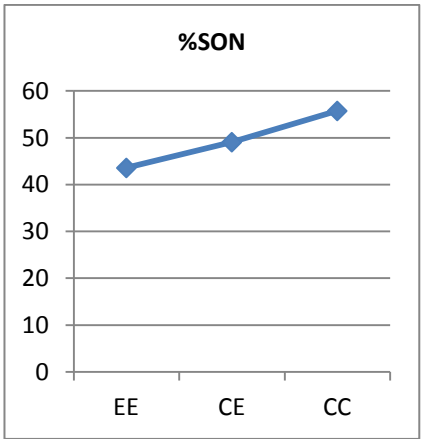
In this section, all results are provided in mean values.

3.1 From duration-based measures

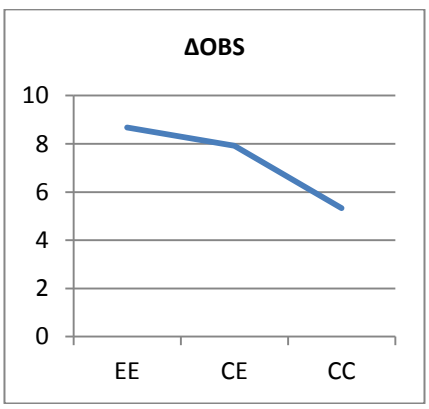
Figure 1 shows results for the groups of participants: English speakers reading the English passage (EE), Mandarin speakers reading the Chinese passage (CC), and the same Mandarin speakers reading the English passage (CE, or ESL).

Figure 1: Duration-based results

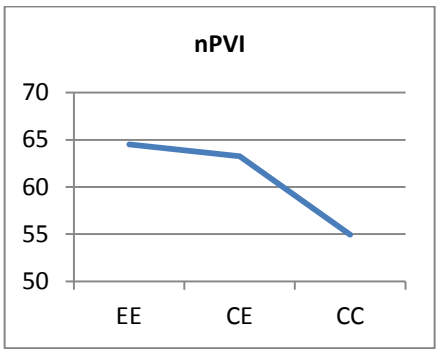
(A)



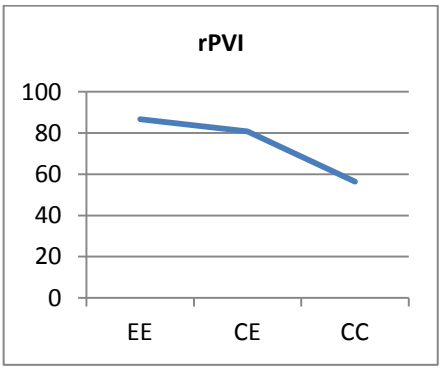
(B)



(C)



(D)



The duration data in Figure 1 show three remarkable consistencies: First, *all* the four measures exhibit the same trend that the ESL (i.e., the CE) results all fall in the middle. For instance, the measurement of the sonorant percentage (A) yielded the results of 43.58% for EE, 55.75% for CC and an intermittent 49.08% for the ESL. For the obstruent standard deviation measure (B), the ESL production value of 7.92 falls between 8.68 of the native English value and 5.33 of the native Mandarin value. The nPVI results (C) of 64.5 for EE, 63.26 for CE and 54.93 for CC also place the ESL result in the middle, and the same trend is repeated in the rPVI values (D) with 86.68 for EE, 80.76 for ESL and 56.44 for CC.

Another consistency Figure 1 displays is that the three variance values (standard deviation, rPVI and nPVI) are all higher for English and lower for Mandarin, suggesting the higher variability of the obstruent and the sononant intervals in duration in English as opposed to those of Mandarin. A third consistency which is related to the above is that the ESL values in the three variance measures, although they all fall between the two native results, are not equally spaced between the two, but they are all closer to English than Mandarin, reflecting perhaps the fairly high English proficiency levels of the participants.

3.2 From pitch-based measures

Five pitch based results were generated from the analysis: minimum F0, maximum F0, range of F0, number of F0 rises and F0 slope. All calculations were made separately for male and female participants in view of the natural fact that males and females have distinctive pitch range and characteristics.

Figure 2 Minimum F0

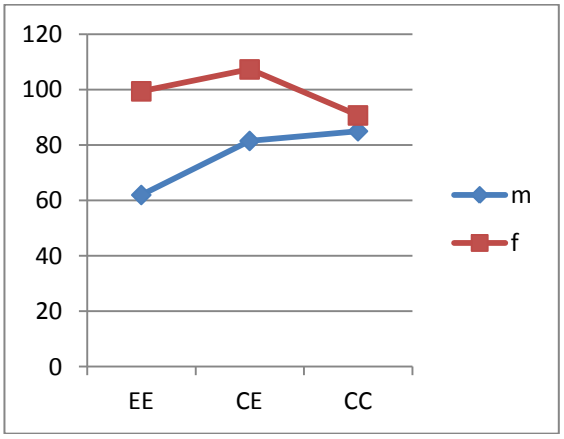


Figure 2 give the results of the minimum F0. An interesting observation that can be made is that Mandarin male and female speakers do not differ as much in minimum F0 (85 and 97 respectively) as their native English counterparts do (62 and 99.4 respectively). (A question that came to mind is: Could this have been caused by the difference between the phonetics and phonology of the two languages?) What is of interest to us is that the ESL production resembles native production in that there is a difference between males and females, with ESL males having lower minimum pitch than ESL females. What is also interesting is that both Mandarin male and female speakers use higher minimum F0 than their English counterparts.

Figure 3 Maximum F0

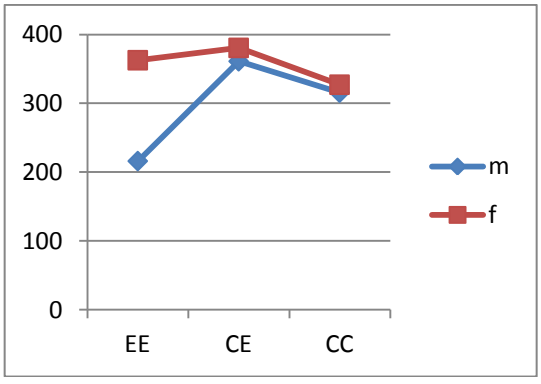
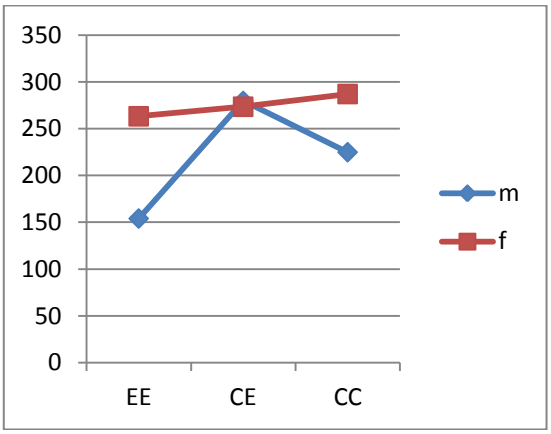


Figure 3 gives the maximum F0. A remarkable difference can be seen between the native English results (216 for males and 362.6 for females) on the one hand and the native Mandarin (male: 315.5 and female: 327) and ESL values on the other (male: 361.1 and female: 380.5). The former shows a much larger gap between males and females, while the latter two are much less gender distinguishing. Another observable fact is that there is no "interlanguage" trend here: Both ESL male and female values show higher maximum F0 than both their EE and CC cocunterparts, falling outside of the values for the mother tongue and the target language.

Figure 4 F0 range



F0 range results are given in Figure 4. For native speakers, Mandarin or English, females produced wider pitch range than males. This is especially true for native English speakers, among whom the females produced 106 wider F0 range than males (i.e., between 154 for males and 263.2 for females). The F0 range difference for native Mandarin production is 62.2 (between 224.8 for males and 287 for females). In contrast, a fairly puzzling result is seen in the ESL pitch range which has little or no difference between the males and the females (279.6 for males and 273.4 for females). A closer look at the results reveals that the ESL males used much wider F0 range speaking the second language of English than both the native English speakers and themselves speaking their L1 Mandarin. ESL F0 range for females show an "interlanguage" characteristic in that it falls between those for English females and themselves speaking Mandarin. Also revealed in the analysis is that Mandarin speakers use wider pitch range than their English counterparts. Mandarin males, especially, used 70.8 wider pitch range than the English males.

Figure 5 The number of pitch rises

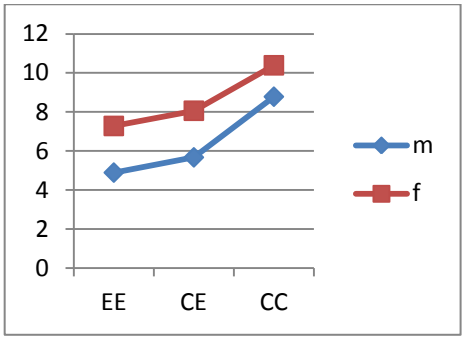


Figure 5 presents the results of the pitch rise numbers. Comparing English and Mandarin, it is clear that Mandarin has more such rises than English (9.5 versus 6.1). Looking at ESL production, we can see a familiar picture of a trend seen earlier in the durational analysis. The ESL value (6.9) falls in between the target and the L1 values (6.1 and 9.5 respectively), while once again the ESL value is closer to the target than to the L1 Mandarin. Gender wise, the male production has consistently fewer pitch rises than the female production across the native and non-native speaking groups (4.89 for EE males versus 7.28 for EE females, 5.67 for CE males versus 8.06 for CE females, and 8.78 for CC males versus 10.39 for CC females).

Figure 6 Average pitch slop

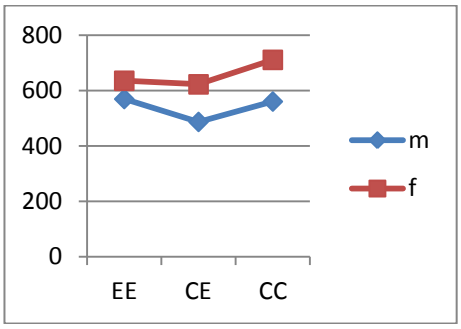


Figure 6 gives the results of pitch slope. The results indicate that the average pitch slope in the ESL production (554.33) is less than both those in L1 Mandarin (635.35) and those in L2 English (602.1). Also, there is a gender difference across the three groups (569.15 versus 635.06 for EE, 486.56 versus 622.09 for CE and 560.11 versus 710.58 for CC) with females producing higher values in pitch slope than males. This is true for ll the three production sets of data. This could be explained as that L1 has no effect on the ESL and that pitch slope is flattened in second language production by more of a general mechanism.

4. REFERENCES

[1] Lloyd James, A. 1940. *Speech signals in telephony*. London: Pitman & Sons.

[2] Ramus, F., Nespor, M. & Mehler, J. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73. 265–292.

[3] Lin, H., & Wang, Q. 2005. Vowel quantity and consonant variance: A comparison between Chinese and English. In Proceedings of Between Stress and Tone. Leiden, June 2005.

[4] Lin, H., & Wang, Q. 2007. Mandarin rhythm: An acoustic study. *Journal of Chinese Linguistics and Computing*, 17 (3), 127-140.

[5] Lin, H., & Wang Q. 2008. Interlanguage rhythm in the English production of Mandarin speakers. In Proceedings of the 8th Phonetic Conference of China and the International Symposium on Phonetic Frontiers, Beijing, April 18-20, 2008.

[6] Arvaniti, A. 2009. Rhythm, timing and the timing of rhythm. *Phonetica*, 66(1-2), 46-63.

[7] Arvaniti, A. 2012. The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics*, 40(3), 351-373.

[8] Vicenik, C., & Sundara, M. 2008. The role of segmental and intonational cues in dialect discrimination. *Journal of the Acoustical Society of America*, 123(5), 3883.

[9] Cumming, R. E. 2011. The interdependence of tonal and durational cues in the perception of rhythmic groups. *Phonetica*, 67(4), 219-242.