Acquisition of L2 Rhythm

○Donna Erickson[1], Yu Tanaka[2],Keikichi Hirose[3], Nobuaki Minematsu[4]

[1]Showa University of Music, Kawasaki City, Japan,
[2]University of California, Los Angeles, U.S.A.
[3,4]University of Tokyo, Tokyo, Japan
[1]EricksonDonna2000@gmail.com, [2]yutanaka@ucla.edm, [3]hirose@gavo.t.u-tokyo.ac.jp,
[4]mine@gavo.t.u-tokyo.ac.jp

Abstract

Japanese middle-level learners of English recorded short English sentences. One class practiced using a metronome-based multimedia computer-presented format where each phrase of a sentence is associated with a visual image; another class practiced imitating the native English-speaking teacher, snapping fingers to indicate the rhythm of the sentences. After training, the students re-recorded their voices. Acoustic measurements of F0, intensity and formants were made for the recorded sentences and then compared with those of the recordings of the native speaker. In addition, perceptual judgments by native English speakers were done for the recordings. The results of these two methods of training of English pronunciation are discussed and compared.

1. Introduction

Language rhythm is difficult to teach. Babies learn it as part of learning their native language, but never is it learned formally, as are spelling, grammar, composition, etc.

The essential ingredients of rhythm might be said to be (a) prominence and (b) groupings of recurring patterns over time (Kohler, 2009). Recent ongoing work with spoken English examines rhythm from an articulatory point of view (Erickson, 2003, 2004, 2010a, 2010b). Patterns of jaw (mouth) displacement form the basic rhythm patterns of English. Simply speaking, English speakers open their mouths more for prominent syllables than for non-prominent syllables (e.g. Westbury and Fujimura, 1989;Jong, 1995; Erickson, 1998, 2002, 2003; Harrington, et al., 2000). These observations, among others, led Fujimura to propose the CD Model of phonetic implementation (Fujimura, 2000), in which is stated that the amount of jaw opening is directly related to syllable prominence. One interesting acoustic correlate of increased jaw opening is congruent changes in F1. It has been suggested that it is these patterning in jaw displacement/F1 that constitute the phonetic correlates of the metrical organization of spoken English utterances (e.g., Erickson et al., 2011)

This paper is a very preliminary essay into how these rhythmical patterns might be taught to Japanese learners of English, whose native language has a very different rhythmical structure than that of English.

2. Methods

Comparison of progress in acquisition of native-like rhythm was done in the following way. Two classes of middle-level English learners at a university in the Kanto area were taught by the same native English teacher, and both classes practiced with the same set of 14 utterances. For Class 2, the commercially available English teaching computer program, 脳学, was used (www.nowgaku.com). This involves (a) multimodal presentation of the English utterances (b) which are divided into three chunks (roughly corresponding to intermediate phrases in the ToBI transcription system), with (c) each chunk (phrase) associated with an easy-to-look at image related to the semantic content of the phrase, while (d) the students listen to the voice of a native (British female) speaker produce the utterance, at the same time (e) also listen to a metronome beating the "rhythm" of the utterance. After the multimodal presentation of each sentence, the students repeated the utterance several times, always associating a single image with a single verbal phrase, and keeping time to the metronome. Figure 1 below illustrates briefly this method. Notice that there are two metronome beats to each phrase-unit—one on the stressed word in the phrase, and the other at the end of the phrase.

x x x x x x
(Autumn is) (the best **sea**son) (to go to **Kyo**to)

Fig. 1. Images are presented synchronously with metronome and verbal phrases; x's indicate timing of metronome beat—one on the stressed syllable of the phrase and another at the phrase break; the utterance is written with parentheses indicating the phrases associated with each image above.

For Class 1, no computer program was used, only the native (American female) English language speaker. No images associated with each of the phrases were shown; however, the same sentence chunking was used, and instead of a metronome, the teacher kept the rhythm by snapping her fingers.

For both classes, after training with each utterance, there was "substitution practice". That is, for example, for the utterance "Autumn is/ the best season/ to go to Kyoto" (where / indicate phrase breaks), students suggested various substitions for "autumn" (e.g., summer, spring, winter) or for "Kyoto" (e.g., Italy, France, etc) and these were practiced as a group, and also individually.

Recording of the utterances by students was done using skype-type headmounted microphone sets and a windows-based PC, with the free acoustic analysis software WAVESURFER. Students recorded their voices before training and after training (two 90-minute class periods). This paper reports on one of the sentences, "Autumn is the best season to go to Kyoto" recorded before and after training.

Acoustic measurements were made of the maximum intensity and F0 and F1 at the time of maximum intensity within each vowel of each of the syllables of the utterance. The acoustic software PRAAT was used. In this paper, the focus is on the F1 measurements.

Perception tests were done with the pre-training/post-training utterances paired for each student using a Runtime Revolution software. For each class, 8 students successfully recorded pre- and post-training utterances. These were randomized 3 times and presented to three native teachers of English (two British, one American), who listened to each of the paired utterances and judged how much better one utterance was compared to another. They used a slide bar on the computer screen from -4 to +4, with the positive numbers indicating improvement. The presentation of the utterance pairs also randomly varied in terms of whether the pre- or post-trained utterance was first. Before the test, there was one practice paired utterance (from actually the 9th speaker in the 2nd class).

3. Results

3.1. Rhythmical analyes of native English speakers (F1 patternings)

Rhythmical analyses of native English speakers in terms of F1 patternings are shown in the figure below.

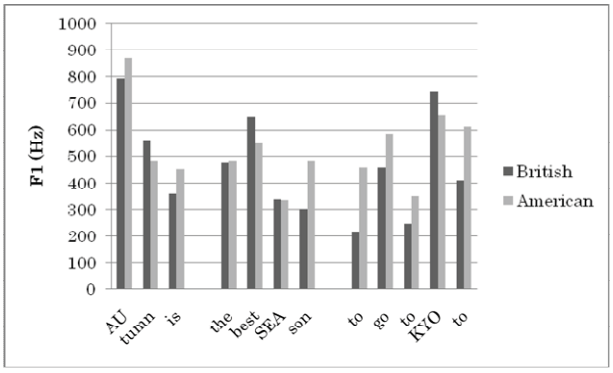

Fig. 2. F1 values for each syllable are shown for the British (dark) and American (light) teacher. The three phrases are indicated by breaks in the bar graph. The stressed syllables in each of the phrases are indicated by capital letters.

Roughly speaking "strong syllables" have higher F1 (because the jaw is more open). The pattern of F1 for both speakers shows alternations of strong (S) and weak (W) syllables. Notice that (1) for the most part there is considerable similarity in the F1 patterns for the two native English speakers: the first phrase shows a pattern of S-W-W, and the last phrase, W-S-W-S-W; (2) the second phrase, is either W-S-W-W (British speaker) or W-S-W-S (American speaker); and (3) the stressed syllable in each phrase always has the highest F1, except of course, when the stress is on the first syllable of "seasons", which contains the high vowel (/i/), and it is well-known that jaw displacement is less for high vowels than low vowels. Future work along these lines will need to incorporate not only stress

but also vowel height (as well as position in the phrase/utterance) in order to derive a more accurate assessment of prominence (S-W) patternings. F0 is not shown here; however, previous studies have shown that peaks in F0 tend to indicate where the stress is, rather than the rhythmical patterning (e.g., Erickson, 2004; Erickson et al, 2011 (Montreal)

3.2. Results of perception tests with native English teacher listeners

The results suggest that there is more improvement in performance with the group of students who practiced with the multimedia computer-based language teaching method, compared to those who practiced only with the native English speaker. The "improvement score" for the former group was 1.7, and 1.05 for the later group. The range of improvement for the multimedia computer-based language teaching method was from 0.3 to 2.2; that of the other class was -1.2 to 2.8. It is interesting that the best individual improvement was seen, however, with one of the students in the non-multimedia computer-based language teaching method.

3.3. Acoustic analyses of students productions

The two bar graphs below show pre- and post-production patterns for the most-improved speaker in each class. Data from the Class 1 most improved speaker (speaker 105) are shown in Figures 3 and 4.

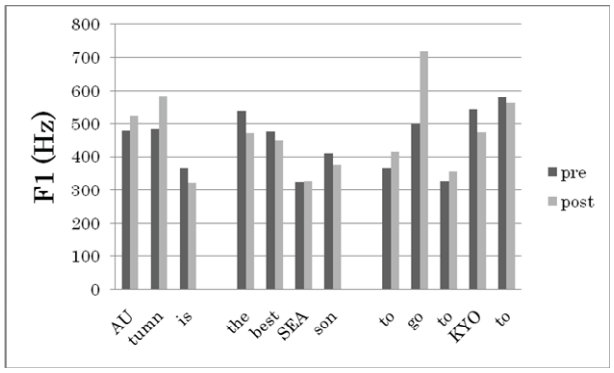

Fig. 3. F1 patterns of pre and post-training utterances for most improved student in Class 1.

Comparison of the F1 pattern for Class 1 most-improved speaker (Figure 3) with that of the native English speakers (Figure 2) does not show a native-like rhythmic S-W F1 pattern, either for the pre or post training sentences.

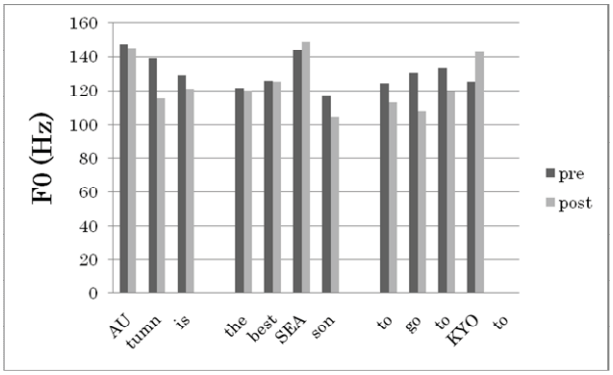

Fig.4. F0 patterns of pre and post-training utterances for most improved student in Class 1.

However, if we look at F0 for the Class 1 speaker, it seems maybe there is an improvement in that now the stressed syllable in each phrase has the highest F0, and maybe this speaker was focusing on F0 changes, rather than other correlates of rhythm. For the final syllable, "to" (of kyoto"), there was no measurable F0 value for either pre- or post-training utterances.

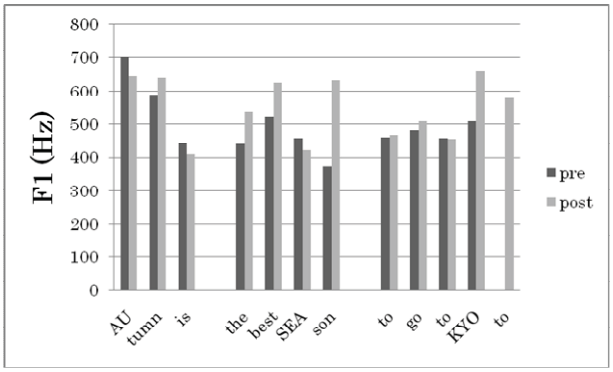

Fig. 5. F1 patterns of pre and post-training utterances for most improved student in Class 2.

Comparison of the F1 pattern for the most improved Class 2 speaker (Speaker 203) with that of the native speakers suggests that this speaker actually did improve her rhythm after training with the multimedia metronome computer-based language teaching method to be more native-like. It is interesting that this most improved speaker was one of the poorer (more shy, less confident) speakers in the class.

4. Summary & future work

Previous studies have shown that jaw displacement/F1 patterns may well reflect a

hierarchical metrical (rhythmic) structure of spoken English. However, how to apply this knowledge to practical applications for teaching rhythm is quite challenging. This study is an initial foray into using research-based findings for applications to teaching English rhythm to Japanese learners of English. The data presented here are preliminary. However, the tendencies we see here suggest that a multimedia metronome computer-based language teaching method may be a good approach for accomplishing this goal. In fact, the pilot studies reported here show that there is more improvement in students using the multimedia metronome computer-based language teaching method than being taught by a native-English speaking teacher. Moreover, depending on the teaching method there may be a difference in terms of which acoustic aspects of the speech signal improve. Very preliminary results suggest that the students taught by more traditional methods (a native-English speaking teacher) may make increase F0 on the stressed syllable, but may not necessarily change the basic underlying rhythm of the utterance to be more native-like. Students taught by the multimedia metronome computer-based language teaching method may actually improve their rhythmic patterns to be more native-speaker like.

One possible explanation for this might be related to the video image the students are looking at. The images are extremely well-selected, and easily draw the students' attention. The process of learning the language becomes in a way more natural, in that there is a meaning associated with an image, thus allowing the students to associate visual and verbal meanings. At the same time, in the background is the metronome which helps give rhythmical structure to the learning process of visual-verbal association.

These speculations are mentioned here to encourage further thinking along these lines. Before anything conclusive can be said, more substantive research along these lines is needed

Acknowledgements

I thank Albert Rilliard for his help with making the perception test, and the English teachers who took the test. I also thank the Nowgaku staff who has allowed us to use their tool, and reprint their images. This work was supported in part by the Japanese Ministry of Education, Science, Sport, & Culture, Grant-in-Aid for Scientific Research (C)#22520412.

References

Erickson, D. (1998) Effects of contrastive emphasis on jaw opening. *Phonetica*, *55*, 147-169.

Erickson, D. (2002) Articulation of extreme formant patterns for emphasized vowels. *Phonetica, 59*, 134-149.

Erickson, D. (2003a) The jaw as a prominence articulator in American English, *Acoust Soc Jn, Fall Meeting,* 311-2.

Erickson, D. (2003b) Some effects of prosody on articulation in American English. In *A New Century of Phonology and Phonological Theory, A Festschrift for Prof. Haraguchi*, T. Honma, M. Okazaki, T. Tabata & S. Tanaka (eds.) Tokyo: Kaitakusha, pp. 473-491.

Erickson, D. (2004) On phrasal organization and jaw opening. *Proceedings of From Sound to Sense, June 13, MIT,*p. 24, CDRom publication.

Erickson, D. (2010a). An articulatory account of rhythm, prominence, and articulatory organization, *Speech* Prosody 2010.

Erickson, D. (2010b) More about jaw, rhythm and metrical structure. Acoustical Society of Japan, fall meeting, p. 103.

Erickson, D., Shibuya, S., and Suemitsu, A. (2011). Rhythm and Emphasis in American English : Comparison of native and non-native speakers' productions. *Proceedings of International Seminar of Speech Production, Montreal, Canada*, pp.345-352.

Fujimura, Osamu (2000) The C/D Model and Prosodic Control of Articulatory Behavior, *Phonetica* 57,128-138.

Harrington, J., Fletcher, J., & Beckman, M. (2000) Manner and place conflicts in the articulation of accent in Australian English, *Papers in Laboratory Phonology* (edited by M. Broe and J. Pierrehumbert), Cambridge: Cambridge University Press, 40-51.

Jong, K.,de (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *J. Acoust. Soc. Am.* **97**, 71-96.

Kohler, K. (2009). Whither speech rhythm research? *Phonetica*, **66**, 5-14.

Westbury, J. and Fujimura, O. (1989) An articulatory characterization of contrastive emphasis. *J. Acoust. Soc. Am*.**85**, 1782-1791.