EFFECT OF REPEATING RHYTHMIC BEATS OF SHORT SENTENCES ON L2 PRONUNCIATION OF JAPANESE LEARNERS OF ENGLISH

Kaori Sugiura & Tomoko Hori

Ritsumeikan University & Juntendo University
k-sugi51@fc.ritsumei.ac.jp

ABSTRACT

The role of rhythm is essential in language acquisition. This study examined sentences produced by 20 Japanese adolescents currently learning English to observe the effects of repeating musical rhythms on aspects of pronunciation, such as sentence duration, ratio of unstressed to stressed vowel (i.e., duration ratio), and pitch range. To gauge a pronunciation baseline, participants repeated sentences without the prime. They also repeated the rhythmical prime and sentences with matched and mismatched rhythm. The results revealed a significant improvement in sentence duration, duration ratio (in both the elementary and intermediate proficiency groups) and revealed a tendency towards an improvement in pitch range (only in the intermediate proficiency groups) in matching conditions. These findings suggest that rhythmic priming can enhance the pronunciation of L2 learners of English with non-stress-timed language backgrounds, which not only occurs in durational features but also in other prosodic aspects.

Keywords: L2 pronunciation, rhythmic priming, sentence duration, duration ratio, pitch range

1. INTRODUCTION

Language has rhythm, that is, a certain temporal regularity. Stress-timed languages like English have a temporal pattern of sequences comprising stressed and unstressed events. First language (L1) acquisition studies have shown that the ability to perceive metrical structures plays a vital role in language development, including syntactic, semantic, and phonological processing and production. Recent priming studies have revealed that musical rhythm has the potential to aid phonological language processing [4] and production [5]. Cason, Hidalgo, Isoard, Roman, and Schön [5] demonstrated that hearing-impaired children with cochlear implants improved the accuracy of their reproduction of words, syllables, vowels, and consonants by listening to and repeating a model sentence immediately after the rhythmic prime that matched the metrical structure of the model sentence (rhythmic priming). They proposed that the metrical predictability of a heard sentence can promote its perception, thereby facilitating its reproduction.

In L2 learning, beats are widely used in classroom settings for pronunciation training. For instance, upbeat chants using jazz rhythms enhance the natural stress and intonation patterns of English (e.g., [7]). Although several studies have reported the ways rhythm contributes to improvements in the development of L2 pronunciation, the empirical literature on the effects of training with beats is scant.

Gluhareva and Prieto [6] studied the effects of beat gesture training on accentedness ratings in university students who were Catalan-dominant as well as Catalan/Spanish bilinguals. They found that the pronunciation of words regarded as difficult by the learners improved. This finding suggests that observing a beat gesture, which highlights rhythmic information, can contribute to enhancing perceived pronunciation.

Baills, Zhang, and Prieto [1] examined the possible benefits of hand-clapping in language trainings. They studied a group of Chinese adolescents clapping to the rhythm of newly learned French words. The effects of this training were assessed by accentedness ratings of native speakers and an acoustical analysis of the final syllable duration. The results demonstrated that clapping to highlight the metrical word structure during short-term trainings can improve students' pronunciation. This study also revealed that students' working memory contributes significantly to their pronunciation.

Nakano and Natsume [12] found that adolescent Japanese learners of English who chanted sentences with beat sounds also benefitted from reproducing the target rhythm. It should be added that this training visually enhanced the target sentences by indicating the alternation of stressed and unstressed syllables using black and white circles, respectively.

These studies provide crucial evidence that beats distinguish essential information in stream speech, which helps listeners direct their attention to the beats and leads to improvements in L2 pronunciation. However, these experiments involved factors that were ultimately supplementary to the rhythmic beats—i.e. gestures, hand-clapping, or visually enhanced unstressed and stressed syllables [1, 6, 12]—which might have obscured the independent role played by the beats themselves. Furthermore, if rhythmic training reduces the cognitive demand of learning the rhythmic property of words or sentences, then it may allow learners more cognitive resources to attend to and internalize phonetic information; this could enhance their pronunciation of other segmental and suprasegmental sounds beyond rhythm itself. However, this issue has not yet been explored. In addition, regarding the possible impacts of listening ability and working memory [1] on phonological improvement from rhythmic beat training, there is still room for further research. Therefore, the present study, addresses the following research questions:

1) Does the repetition of musical rhythm help improve Japanese English-learners' pronunciation (rhythmic priming effect), duration of sentences (sentence duration, i.e., speech rate[1]), duration ratio of unstressed vowels to stressed vowels (duration ratio), and ratio of minimum to maximum in pitch range at the sentence level (pitch range)?
2) If so, how do factors such as one's level of English proficiency and working memory capacity affect the rhythmic priming effect?

The suprasegmental features investigated in the present study play an important role in accent rating and the comprehensibility of native English speakers (e.g., [8, 13, 15]). L2 learners tend to speak at slower rates than native speakers [11]. Duration ratio and pitch range also tend to be poorly acquired, due to L1 and L2 phonological differences. Because Japanese is a mora-timed language [16], for Japanese speakers, the duration ratio of unstressed to stressed vowels in English words tends to be higher than that of the average native English speaker (around .45) [10]. Intonations produced by Japanese learners generally tend to be narrower in pitch range when compared with the dynamic realization of native speakers [14]. These features are worth investigating.

2. METHOD

2.1. Participants

Twenty Japanese students of English (3 females and 17 males) with an average age of 19.65 years participated in this study. All were native Japanese speakers who studied English twice a week at a university in Japan but were not majoring in English.

2.2. Materials

For the rhythmic prime, two different rhythmic patterns were prepared: an xXxX structure (Prime A) and an XxxX structure (Prime B). When the rhythmic prime had an XxxX structure, sentences following the prime were either structured as XxxX (metrically matching), xXxX (metrically mismatching), or vice versa. The musical primes consisted of stressed sounds (X) and unstressed sounds (x). The stressed sound was provided by a snare drum sound with a duration of 375 ms and a mean energy intensity of 80 dB. The unstressed percussion sound was provided by a closed high-hat sound with a duration of 155 ms and a mean energy intensity of 65dB. The primes lasted on average 1250 ms. The stimuli were created using Apple's GarageBand, music production software. Sentences consisting of four monosyllabic words matching the metrical structure of the A prime (e.g., "He sings a song.") or B prime (e.g., "Bring me a pen.") were prepared (n = 20 each, 40[2] in total) and then recorded by a male native speaker of American English. The mean duration of the sentences was 1233 ms, and the durations of stressed and unstressed syllables were 394 ms and 173 ms, respectively. The mean durations of the sentences and stressed and unstressed syllables were comparable to those of the rhythmic primes. Pitch range in the sentences averaged 191Hz for max. and 87 Hz for min.

2.3. Procedures

The participants conducted two tasks: (a) a rhythmic priming experiment and (b) Digit Span Test (UCLA).

(a) The rhythmic priming experiment consisted of baseline and experimental phases. The same sentences were presented in baseline and experimental sessions. For the baseline phase, the participants were asked to repeat 20 target sentences (10 sentences from each rhythmic type), which they had each heard once. Participants then listened to and repeated a musical rhythmic prime (xXxX or XxxX) using the sound "ta." Immediately afterwards, they listened to and repeated the 20 target sentences with an xXxX structure or XxxX structure. Each target sentence was presented twice: once for the matching condition, where the rhythmic prime and rhythmic structure of the sentences matched, and again for the mismatching condition. The presentation of the stimuli was self-paced using SuperLab (ISI=Max. 5 sec).

(b) A digit span test was conducted to assess the participants' capacity for working memory [3], i.e., processing and storing numbers in the brain. The participants listened to a series of numbers and repeated them back.

The session included both tasks and lasted approximately 40 minutes.

2.4. Data analysis

Two experts acoustically analyzed sentence duration, duration ratio, and pitch range of each reproduced sentence using Praat [2].

3. RESULTS

3.1. Overall

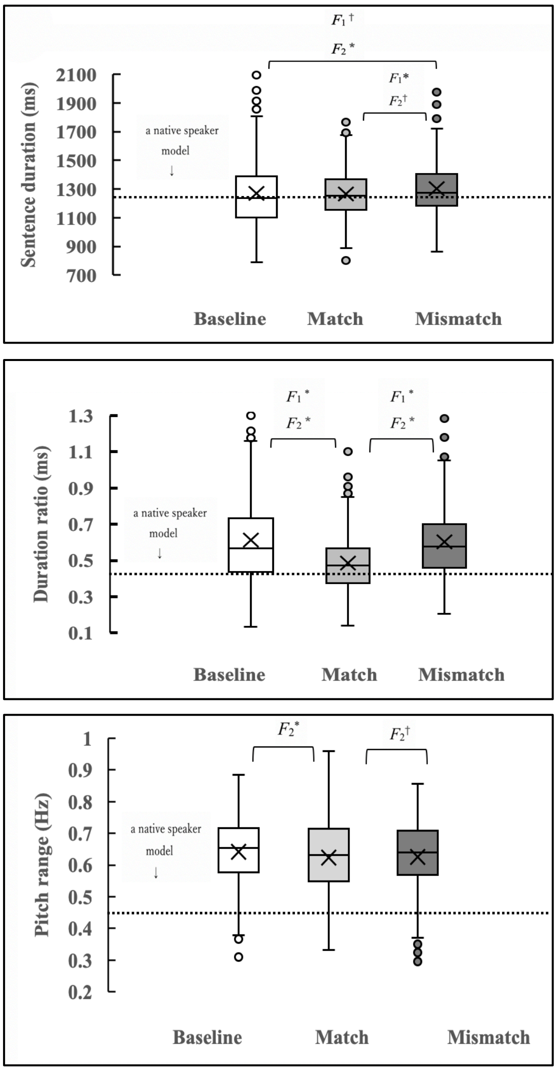
Repeated-measure analyses of variance tests were run on the data separately for sentence duration, duration ratio, and pitch range. Post hoc analyses were then conducted using Fisher's tests to examine the differences among the three conditions (baseline, match, mismatch) (see Fig.1)

For sentence duration, the main effect of the condition was: F_1 (2, 38) = 2.96, p = .08; F_2 (2, 76) = 2.42, p < .05. Post hoc analyses revealed a significant difference in matching vs. mismatching, t_1 (19) = 3.01, p < .01; t_2 (38) = -1.98, p = .06, and t_1 (19) = 2.47, p = .08; t_2 (38) = 2.4, p < .05 in baseline vs. mismatch. There was no significant difference in baseline vs. matching conditions, t_1 (19) = 0.29, p = .77; t_2 (38) = 0.91, p = .36. That is, sentence duration in the baseline and matching conditions was significantly shorter than that of mismatching, respectively.

Regarding duration ratio, the main effect of the condition was: F_1 (2, 38) = 31.17, p <.001; F_2 (2, 76) = 47.56, p <.001. Post hoc analyses revealed a significant difference in baseline vs. matching, t_1 (19) = 5.87, p < .001; t_2 (38) = -7.13, p < .001 and matching vs. mismatching, t_1 (19) = 8.19, p < .001; t_2 (38) = -9.10, p < .001. No significant difference was confirmed in baseline and mismatching, t_1 (19) = -0.45, p = .65; t_2 (38) = -1.18, p = .24. Duration ratio in the matching condition was significantly smaller than the others and was the closest to the native ratio.

As for pitch range, the subject analysis did not reveal a main effect, F_1 (2, 38) = 0.48, p =.60, but the item analysis showed a marginal main effect, F_2 (2, 76) = 3.22, p = .07. Post hoc analyses showed a significant difference in baseline vs. matching, t_2 (38) = 2.18, p < .05 and a marginally significant difference in matching vs. mismatching, t_2 (38) = -1.97, p = .06. No significant difference was observed in baseline vs. mismatching, t_2 (38) = -1.32, p = .19. Pitch range in the matching condition was smaller than the other conditions and was the closest to the native norm.

Figure 1: Main effect of each condition.



3.2. Predictive factors

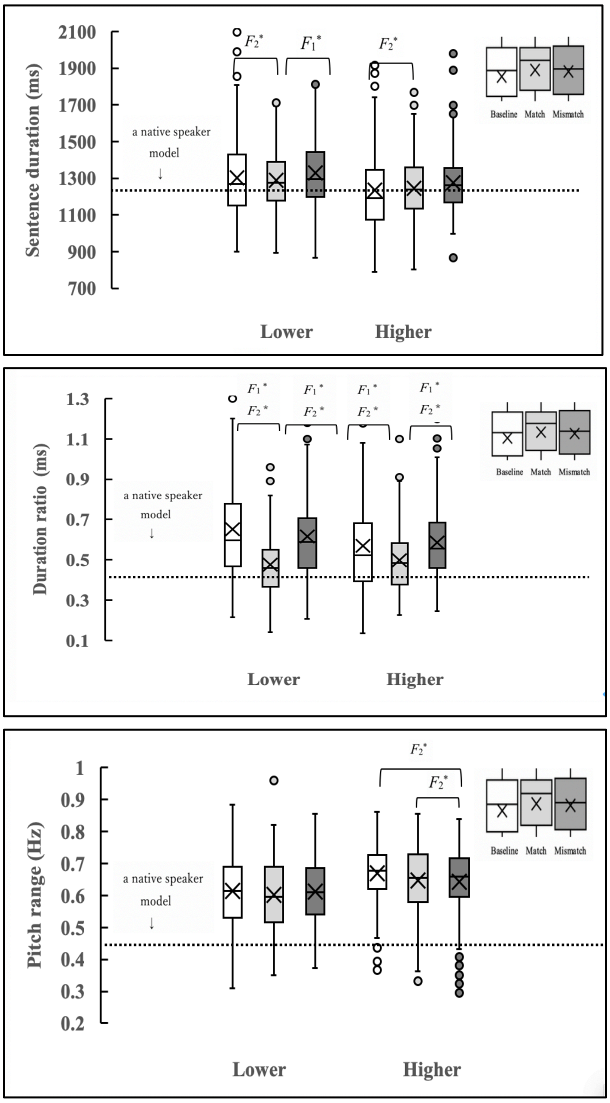
Two multiple regression analyses were run to identify any predictive factors on experimental effects: (1) for the matching and baseline conditions (matching minus baseline data) and (2) for the matching and mismatching conditions (matching minus mismatching data), with the scores of the TOEIC® Listening & Reading Test (English proficiency) and Digit Span Test serving as regression factors. These analyses were conducted on each pronunciation item. [3] English proficiency was found to be a predictive factor.

Friedman tests were then separately performed on two English proficiency groups[4]: the Lower Group (n = 11, Avg. 425, SD =79; CEFR: A2) and the Higher Group (n = 9, Avg. 651, SD = 68, CEFR: B1) (see Fig.2). Both the Lower and Higher Groups showed significant effects on sentence duration and duration ratio, p < .001 (F_1, F_2) and the Higher Group also demonstrated a significant effect on pitch range (F_2),

p < .001. Post hoc comparisons (Wilcoxon's test) demonstrated the following results: For the Lower Group, a significant difference emerged in sentence duration between matching vs. mismatching, $p < .01$ (F_1), and baseline vs. matching, $p<.05$ (F_2). Regarding duration ratio, a significant difference emerged between baseline vs. matching, $p < .005$ (F_1) and $p < .001$ (F_2) and matching and mismatching, $p < .005$ (F_1) and $p < .001$ (F_2).

For the Higher Group, the tests revealed a significant difference in sentence duration between baseline vs. matching, $p < .005$ (F_2). As with duration ratio, a significant difference was evident between baseline vs. matching, $p < .05$ (F_1) and $p < .001$ (F_2) and matching vs. mismatching, $p < .05$ (F_1) and $p < .001$(F_2). As for pitch range, a significant difference was revealed between matching and mismatching, $p < .005$ (F_2) and baseline vs. mismatching, $p < .005$ (F_2).

Figure 2: Performance in the three pronunciation features across the three conditions shown separately for the Lower and Higher Groups.



4. DISCUSSION AND CONCLUSION

This study investigated whether Japanese learners of English can improve their pronunciation when the metrical prime matches the prosodic structure of a subsequent heard sentence. The present study highlighted that the rhythmic predictability of a sentence can enhance L2 pronunciation in terms of sentence duration and the duration ratio and might improve the pitch range for learners whose L1 rhythm is non-stress-timed. These results suggest that even a short session of repeating beats that match the following metrical structures has the potential to facilitate participants' language processing [4, 5]—thereby reducing the cognitive demand required in learning rhythmic aspects to the extent that participants are able to allocate more cognitive resources to processing and storing different prosodic properties, such as pitch.

The participants' proficiency seemed to influence the effect of rhythmic repetition. Regarding sentence duration and duration ratio, the Higher Group performed better than the Lower Group in the baseline condition, and was closer to the native norm, but this was not the case for pitch range. Interestingly, however, only the Higher Group benefited from the anticipation of rhythm in the pronunciation of the pitch range. [5] Therefore, it is speculated that the Higher Group's smooth production of durational features might have enabled the allocation of more cognitive resources for attending to and storing pitch information, which led to an improvement in pitch.

In summary, the present study showed a possible impact of prosodic expectation on L2 phonological development of durational properties and other phonetic features, such as pitch range. To verify the present finding, however, further studies are warranted: an investigation into the extent to which differences in music-speech rhythms in a mismatching condition may interfere with phonological processing and production (cf. [4]) as well as whether intensive training with musical rhythm would be more effective in facilitating L2 phonological improvement in pronunciation. Also, the order in which the three conditions (baseline, matching, mismatching) are presented should be more carefully managed. Overall, however, the results presented here have suggested that an implicitly induced rhythmic expectation has significant potential for both L2 phonological instruction and phonological acquisition research.

5. ACKNOWLEDGEMENTS

This work was supported by Grant-in-Aid for Scientific Research (C) Grant Numbers JP18K 00765.

6. REFERENCES

[1] Baills, F., Zhang, Y., Prieto, P. (2018). Hand-clapping to the rhythm of newly learned words improves L2 pronunciation: Evidence from Catalan and Chinese learners of French. *Proc. 9th International Conference on Speech Prosody 2018*, 853–857.

[2] Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International* 5:9/10, 341–345.

[3] Cambridge Brain Sciences. (n.d.). https://www.cambridgebrainsciences.com/science/tasks/digit-span

[4] Cason, N., Astésano, C., Schön, D. 2015. Bridging music and speech rhythm: Rhythmic priming and audio–motor training affect speech perception. *Acta Psychologica,* 155, 43–50.

[5] Cason, N., Hidalgo, C., Isoard, F., Roman, S., Schön, D. 2015. Rhythmic priming enhances speech production abilities: Evidence from prelingually deaf children. *Neuropsychology,* 29(1), 102–107.

[6] Gluhareva, D., Prieto, P. 2017. Training with rhythmic beat gestures benefits L2 pronunciation in discourse-demanding situations. *Language Teaching Research,* 21(5), 609–631.

[7] Graham, C. 1979. Jazz Chants; Small Talks. OUP

[8] Kang, O. 2010. Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System,* 38(2), 301–315.

[9] Kormos, J., Dénes, M. 2004. Exploring measures and perceptions of fluency in the speech of second language learners. *System,* 32(2), 145–164.

[10] Lee, B., Guion, S. G., Harada, T. 2006. Acoustic analysis of the production of unstressed English vowels by early and late Korean and Japanese bilinguals. *Studies in Second Language Acquisition,* 28(3), 487–513.

[11] Munro, M.J., Derwing, T.M., 1995. Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech* 38, 289– 306.

[12] Nakanno, N., Natsume, K. 2011. English rhythm learning and changes in EEG using RIM with beat. *Computer and Education* 31, 88–93.

[13] Tajima, K., Dalby, J., Port, R. 1996. Foreign-accented rhythm and prosody in reiterant speech. *The Journal of the Acoustical Society of America,* 99(4), 2493–2500.

[14] Taniguchi,M., Abberton, E. 1999. Effect of interactive visual feedback on the improvement of English intonation of Japanese EFL learners. *Speech, Hearing and Language: Work in Progress,* 11. Department of Phonetics and Linguistics, University College London. 76–89.

[15] Trofimovich, P., Baker, W. 2006. Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, 28(1), 1–30.

[16] Warner, N., Arai, T. 2001. Japanese mora-timing: A review. *Phonetica,* 58(1-2), 1–25.

---

[1] Sentence duration is considered almost equivalent to speech rate in this study because speech rate—for instance, syllable per second—is a measure of the total number of syllables produced in a given sentence divided by the amount of total time (ms) required to produce the target sentence (including pause time) [9]. All the sentences analyzed in this study had four syllables and included pause time.

[2] Although 40 sentences were originally prepared as the study material, due to a setting mistake, 39 total sentences were actually presented to the participants.

[3] There was a difference between the matching and baseline conditions in sentence duration ($p <. 05$) and duration ratio ($p <.001$), as well as between the matching and mismatching conditions in pitch range ($p < .001$).

[4] The two groups' average scores were significantly different, $t\,(18) = 6.73$, $p < .001$.

[5] A significant difference was observed between matching vs. mismatching and baseline vs. mismatching, but not between baseline vs. matching. However, a numerical difference was evident between the last pair; overall, it can be said that the Higher Group benefited from rhythmic predictability in the feature of pitch range.