10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

Linguistic experience and rhythm perception

Sumio Kobayashi[1], *Amalia Arvaniti*[2]

[1]Nihon University, [2]University of Kent
kobayashi.sumio@nihon-u.ac.jp, a.arvaniti@kent.ac.uk

Abstract

Many types of auditory perception are influenced by features of one's native language. The current work focuses on whether native language affects rhythm perception. Native English, Japanese and Russian speakers were asked to rate the rhythmic difference between a pair of sound files, a familiarization stimulus with either binary or non-binary rhythm and a *comparison* that included either a clash (succession of accented syllables) or a lapse (succession of unaccented syllables). Stimuli were of three types, linguistic, musical, and tonal; they all had the same rhythm structure but were tested in separate blocks. It was anticipated that Russian and English participants would be less sensitive to rhythm irregularities (clashes and lapses), as these are relatively rare in these languages, and thus they would be perceptually compensated. Since lapses occur frequently in Japanese due to the sparsity of accented syllables, participants were expected to be more familiar with and thus more sensitive to rhythm irregularities. The results confirmed that Japanese participants were better at detecting rhythm irregularities than English and Russian participants, between whom there were no differences; this applied to all three stimulus types. In conclusion, these group differences in rhythm perception appear to be influenced by linguistic experience.

Index Terms: rhythm perception, English, Japanese, Russian

1. Introduction

Many studies on the influence of native language on auditory perception demonstrate that one's native language influences sensitivity to phonemes and prosody (among others, [1], [2]). This has also been suggested for rhythm, which is here defined as a pattern involving the alternation of prominent and non-prominent prosodic elements (such as stressed or accented, and unstressed or unaccented syllables) [3].

There is no study examining whether when processing linguistic stimuli we perceive rhythm patterns that are frequent in our native language more accurately than infrequently used patterns. However, a related study on musical rhythm perception shows that listeners perceive familiar musical rhythms more accurately than non-familiar rhythms [4]. Based on [3] and [4], the main hypothesis tested in the current study is that listeners perceive speech rhythms familiar from their native language more accurately than non-native rhythms. This hypothesis was tested with speakers of three languages, British English, Japanese, and Russian. The languages were chosen because they have different rhythmic structures.

English rhythm tends to be a regular alternative pattern of prominent (stressed) and non-prominent (unstressed) syllables, with mechanisms constraining lapses (caused by successive unstressed syllables) and clashes (caused by successive stressed syllables); see [5] and [6]. Russian rhythm is similar to English, involving regularly alternating stressed and unstressed syllables [7], though other studies suggest that, unlike English, Russian does tolerate lapses [8]. Thus, in both English and Russian, linguistic rhythm is by and large binary in structure, though more regularly so in English than Russian. In Japanese, on the other hand, lapses are frequent due to the lack of a regular structure of prominences [9]. Thus, unlike in English and Russian, rhythm in Japanese is by and large non-binary. Considering that participants would accurately perceive rhythm used in their native language, it was hypothesized that Japanese speakers would more precisely perceive non-binary rhythm than English and Russian speakers due to their familiarity with it. It was also hypothesized that Japanese speakers would be more sensitive to lapses than English and Russian speakers, i.e. better able to detect them, due to the frequent occurrence of lapses in Japanese. Finally, the present study tested whether native language affects the perception of rhythm in modalities other than language, specifically in music and in stimuli involving tones. Testing all three types of stimuli using the same rhythm structure allows us to address the question of whether it is possible to extrapolate from studies with one of these types of stimuli to the perception of rhythm in other modalities.

2. Method

The experiments were based on the experiment paradigm used in [4]: participants listened to a *familiarization* stimulus and a *comparison* stimulus and had to rate the difference between them in terms of rhythm. Familiarization stimuli had either binary or non-binary rhythm (see Tables 1 and 2); comparisons differed by including either a clash (two adjacent beats) or a lapse (a missing beat).

2.1. Participants

Thirty British monolingual participants (17F, 13M), 18-27 years old (mean = 23.16, SD = 2.56) were recruited in South East England; they spoke Standard Southern British English. Thirty-one Japanese monolingual participants (20F, 11M), 18-31 years old (mean = 23.38, SD = 3.23) were recruited in Japan; they spoke standard Japanese and Kinki dialect. Thirty-three monolingual Russian participants (22F, 11M), 18-24 years old (mean = 22.84, SD = 1.29) were recruited in Izhevsk, Russia; they spoke standard Russian. They all reported they had normal hearing. None of the participants had extensive musical training or was a professional musician because, as shown by [4], musical experience influences rhythm perception. For this reason, musical education and activity were limited to obligatory classes at school. All participants were remunerated for their participation.

2.2. Stimuli

Three stimulus types (linguistic, musical, and tonal) were used to examine if rhythm is processed in a similar manner across

10.21437/SpeechProsody.2020-113

modalities. All three types of stimuli had the same rhythm structure. Two types of familiarization stimuli were used, stimuli with binary rhythm and stimuli with non-binary rhythm. In each trial, the familiarization stimulus was followed by one of seven types of test stimuli (*comparisons*): (1) control (a repetition of the familiarization stimulus); (2) structure-preserving, where the rhythmic structure remained the same between familiarization and comparison but the value of some elements in the latter changed; (3) clash1; (4) clash2; (5) lapse1; (6) lapse2; (7) lapse3; see Tables 1 and 2 for the rhythmic structures of the familiarization and stimuli with binary and non-binary rhythm respectively.

The linguistic stimuli consisted of repetitions of the syllable [ma] produced by a female native speaker of Greek. Greek was chosen because the quality of Greek [a] would be equally unfamiliar to all three groups of participants. The duration of the original syllable was adjusted to 200 ms, using PRAAT [10]. Prominent syllables had high falling pitch (from 252 Hz to 186 Hz), while non-prominent syllables had flat pitch (186 Hz). Both prominent and non-prominent syllables were 64 dB in average amplitude. Differentiating the syllables using pitch was chosen because it was most likely to be familiar to all three language groups and associated with prominence in all languages tested.

The musical stimuli were piano notes of 200 ms duration produced using Finale 2010 (https://www.finalemusic.com/). The musical stimuli consisted of C (262 Hz), E (330 Hz), and G (392 Hz) notes with accompaniment which consisted of C (131 Hz) and G (196 Hz) notes. Accented notes were expressed by the highest note (G) (392 Hz) underlined by the accompaniment.

Table 1: *List of stimuli (binary rhythm). "x" indicates a prominence which is higher or louder than "x" (non-promince); ":" indicates that the preceding element is twice as long as others. The stimuli are uploaded on:* http://sumiokobayashi.com/sumiokobayashiresearch.html

Binary Rhythm Familiarization Stimulus: x x x x x x x x xxxxxxxxxxxxxxxx
Binary Rhythm Test Stimuli 1: x x x x x x x x xxxxxxxxxxxxxxxx
Binary Rhythm Test Stimuli 2: x x x x x x x x x:xxxxxx x:xxxxx
Binary Rhythm Test Stimuli 3: xx x x xx x x xxxxxxxxxxxxx
Binary Rhythm Test Stimuli 4: xx xx xx xx xxxxxxxxxxx
Binary Rhythm Test Stimuli 5: x x x x x x x x xxxxxxxxxxxxxxxx
Binary Rhythm Test Stimuli 6: x x x x x x x x xxxxxxxxxxxxxxxxxx
Binary Rhythm Test Stimuli 7: x x x x x x x x xxxxxxxxxxxxxxxxxx

Tonal stimuli were created using the "Create sound as pure tone" function in PRAAT. The frequency of all tones was 440 Hz. Two types of sequences were produced, "long" sequences, in which tones were 150 ms long and separated by 50 ms of silence, and "short" sequences, in which the tones were 50 ms long and separated by 150 ms of silence. Thus, the overall duration of tone + silence was the same as that of [ma] in the linguistic stimuli and the notes in the musical stimuli. Accented tones were louder (77 dB) than unaccented tones (64 dB). Pitch and duration were not altered.

Table 2: *List of stimuli (non-binary rhythm); for a key, see caption of Table 1*

Non-Binary Rhythm Familiarization Stimulus: x x x x x x x x xxxxxxxxxxxxxxxxxx
Non-Binary Rhythm Test Stimuli 1: x x x x x x x x xxxxxxxxxxxxxxxxxx
Non-Binary Rhythm Test Stimuli 2: x x x x x x x x x:xxxxxxxx:xxxxxxx
Non-Binary Rhythm Test Stimuli 3: xx x x xx x x xxxxxxxxxxxxxxxxx
Non-Binary Rhythm Test Stimuli 4: xx xx xx xx xxxxxxxxxxxxxxx
Non-Binary Rhythm Test Stimuli 5: x x x x x x x x xxxxxxxxxxxxxxxxxxxxxx
Non-Binary Rhythm Test Stimuli 6: x x x x x x x x xxxxxxxxxxxxxxxxxxxxxxx
Non-Binary Rhythm Test Stimuli 7: x x x x x x x x xxxxxxxxxxxxxxxxxxxxxxx

2.3. Procedure

The experiment ran on OpenSesame (version 3.0.0a19) using a notebook, with the audio set at a comfortable listening level; stimuli were presented through headphones. The participants were seated in a sound-attenuated room separately from the experimenter. At most, two participants did the experiment at the same time in the same room.

In each trial, the participants were asked to judge the rhythmic difference between the familiarization stimulus and its comparison using a scale from 1 "very similar" to 6 "very different". Once a participant clicked the start button on the screen, Opensesame automatically played the familiarization stimulus and, after 1000 ms of silence, the comparison. Immediately after the comparison, the rating scale was displayed until the participant chose a rating. Ratings were chosen by left-clicking with a mouse on buttons with the numbers 1-6 displayed on the notebook monitor. After choosing their answer, participants clicked the start button for the next trial. Brightness and volume were kept constant.

The experiment started with 56 practice trials which had similar but different rhythm from the stimuli in the main experiment: the practice stimuli were based on 2 rhythm types [binary and non-binary] × 4 stimulus types [linguistic, musical, tonal long, and tonal short] × 7 violation types [control, structure-preserving, clash1, clash2, lapse1, lapse2, lapse3]. In the main experiment, each pair of familiarization and comparison was repeated 3 times, for a total of 42 trials per stimulus type (2 rhythm types × 7 violation types × 3 repetitions); this gave a total of 168 trials per experiment (42 trials × 4 stimulus types). The experiment was divided into four *sessions*, one per stimulus type; each session consisted of three blocks of 14 trials each (14 familiarization-comparison pairs). The order of trials within each block was randomized. Session order was counterbalanced across participants. A session lasted approximately ten minutes, for a total of

approximately 40 minutes for the entire experiment. Participants could take a short break between sessions.

2.4. Measurements & statistical analysis

Following [4], the ratings of the participants were converted into a measure of *accuracy*, the aim of which was to capture the extent to which participants understood the rhythm differences between the familiarization and test stimuli. Accuracy was calculated by separately subtracting from the mean rating of control stimuli (test stimulus 1 in Tables 1 and 2, which was identical to familiarization) the mean rating of each other stimulus type (test stimuli 2 to 7 in Tables 1 and 2); e.g., if a participant's average rating for binary controls was 1.5 and their average rating for binary clash1 was 4.7, the accuracy of clash1 was considered to be 3.2 (4.7 - 1.5).

Accuracy was statistically examined using linear mixed effects models (LMEMs) in R [11]. Accuracy was the dependent variable, with violation type (lapse, clash, structure-preserving, control), native language (English, Japanese, Russian), and rhythm type (binary, non-binary) as fixed factors, and participant as random factor. Responses for different types of lapses and clashes were pooled, as preliminary analysis did not show significant differences between different types of clashes or lapses. A p-value < 0.05 was considered significant. The three types of asterisks ("*" "**" "***") used in figures refer to "$p < 0.05$", "$p < 0.01$", "$p < 0.001$" respectively. Due to lack of space, the results reported here focus on two interactions of interest, the interaction between violation type and native language, and that between rhythm type and native language.

3. Results

3.1. Linguistic stimuli

The model showed a significant interaction between language and rhythm type ($F = 8.04$, df = 2, 848; $p < 0.001$). However, there was no interaction between language and violation type ($F = 1.24$, df = 4, 848; n.s.). Post hoc tests (Tukey's honestly significant difference [HSD] post hoc test) were done using the emmeans function in R for the interaction.

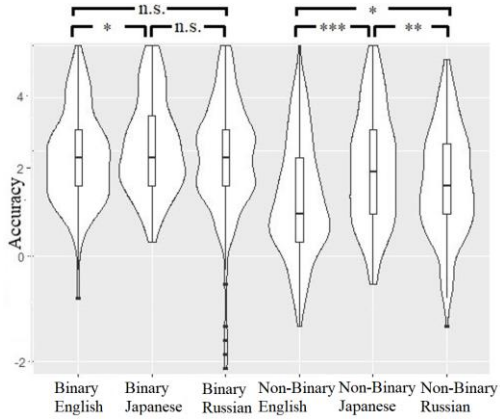

Figure 1: *Violin plots of accuracy (showing distribution shape, median, and interquartile range) for the interaction between language and rhythm type in linguistic stimuli.*

As shown in Figure 1, for binary rhythm, there was no significant difference between English, and Russian speakers ($t = -0.18$, df = 136; n.s.). The Japanese speakers' accuracy was higher than that of English speakers ($t = -1.98$, df = 136; $p = 0.05$), though not different from that of Russian speakers ($t = 1.77$, df = 136; n.s.) [English mean = 2.1, SE = 0.074; Russian mean = 2.13, SE = 0.1; Japanese mean = 2.44, SE = 0.082].

Regarding non-binary rhythm, Japanese accuracy was higher than that of both English and Russian speakers ($t = -5.33$, df = 136; $p < 0.0001$, and $t = 2.96$, df = 136; $p = 0.004$, respectively). Accuracy was also higher for Russian than English speakers ($t = -2.32$, df = 136; $p = 0.022$) [English mean = 0.96, SE = 0.093; Russian mean = 1.35, SE = 0.092; Japanese mean = 1.86, SE = 0.103].

3.2. Musical stimuli

The model revealed an interaction between language and rhythm type ($F = 11.46$, df = 2, 848; $p < 0.001$). However, there was no interaction between language and violation type ($F = 0.4$, df = 4, 848; n.s.).

Looking at Figure 2, it is clear that accuracy is higher for binary than non-binary rhythm regardless of the native language. As is also apparent from Figure 2, there was no difference between English and Russian speakers both for binary ($t = -1.43$, df = 120.1; n.s.) and non-binary rhythm ($t = 1.44$, df = 120.1; n.s.). On the other hand, Japanese accuracies were higher than English both for binary ($t = -2.85$, df = 120.1; $p = 0.005$) and non-binary rhythm ($t = -3.03$, df = 120.1; $p = 0.003$). Finally, Japanese accuracy in binary rhythm was statistically not different from that of Russian speakers ($t = 1.39$, df = 1201; n.s.), but Japanese accuracy for non-binary rhythm was higher than that of Russian speakers ($t = 4.41$, df = 120.1; $p < 0.0001$) [binary rhythm: English mean = 2.618, SE = 0.09; Russian mean = 2.883, SE = 0.103; Japanese mean = 3.133, SE = 0.091; non-binary rhythm: English mean = 1.688, SE = 0.104; Russian mean = 1.447, SE = 0.116; Japanese mean = 2.236, SE = 0.11.]

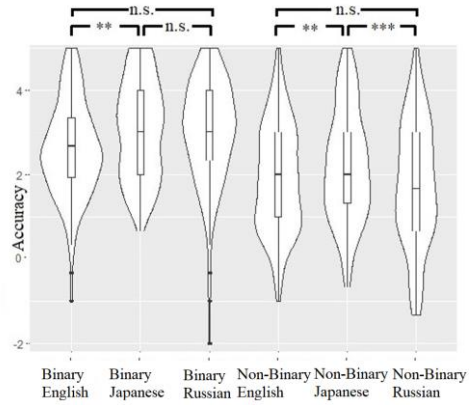

Figure 2: *Violin plots of accuracy for the interaction between language and rhythm type in musical stimuli.*

3.3. Tonal stimuli

The two types of tonal stimuli (long and short tonal stimuli) were pooled in the statistical analysis to increase statistical power because there was no significant difference in accuracy between short and long tonal stimuli ($t = 0.875$, df = 1823; n.s.). The model revealed an interaction between language and rhythm type ($F = 3.84$, df = 2, 1798; $p = 0.0215$), but no interaction among language and violation type ($F = 1.84$, df = 4, 1790; n.s.).

Figure 3 shows that there were no differences between English and Russian speakers both for binary (t = -1.07, df = 112.7; n.s.) and non-binary rhythm (t = -0.1, df = 112.7; n.s.). While Japanese accuracy of binary rhythm was not different from that of Russian speakers (t =1.32, df = 112.7; n.s.), Japanese accuracy of non-binary rhythm was higher (t = 3.05, df = 112.7; p = 0.003). Japanese accuracies were higher than those of English speakers as well, both for binary (t = -241, df = 112.7; p = 0.018) and non-binary rhythm (t = -3.19, df = 112.7; p = 0.002) [binary rhythm: English mean = 2.481, SE = 0.041; Russian mean = 2.676, SE = 0.065; Japanese mean = 2.881, SE = 0.062; non-binary rhythm: English mean = 1.546, SE = 0.078; Russian mean = 1.559, SE = 0.086; Japanese mean = 2.075, SE = 0.077]. These results with tonal stimuli are identical to the results with musical stimuli.

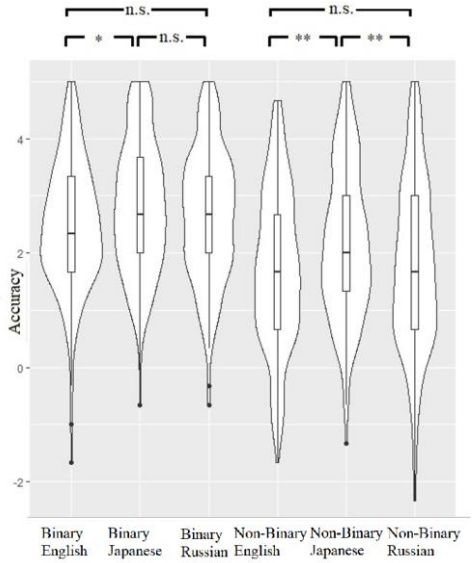

Figure 3: *Violin plots of accuracy for the interaction between language and rhythm type in tonal stimuli.*

4. Discussion and Conclusions

It was hypothesized that Japanese speakers who are familiar with lapses in their native language would be more sensitive to lapses than English and Russian speakers. However, all the results show that there was no interaction between native language and violation type. With respect to the accuracy of binary rhythm stimuli, results were identical for all stimulus types (linguistic, musical, and pure-tone stimuli): Japanese accuracies were higher than those of English speakers, while there were no statistical differences between Japanese and Russian speakers, on the one hand, and English and Russian speakers, on the other.

Looking at the accuracies of stimuli with non-binary rhythm, Japanese speakers' accuracies for all stimulus types were higher than those of English and Russian speakers. Although English and Russian speakers' accuracies for non-binary rhythm were not statistically different from each other's with respect to musical and tonal stimuli, English speakers' accuracy was lower than that of Russian speakers with respect to linguistic stimuli. As was expected, Japanese participants more accurately perceived non-binary rhythm than English and Russian speakers. We attribute this to the Japanese speakers' greater familiarity with non-binary rhythms due to the rhythmic structure of their language. On the other hand, there were largely no statistical differences between English and Russian speakers, a result that can be attributed to the similar rhythm structure of these languages. As mentioned earlier, the sole difference between English and Russian was in the results regarding non-binary rhythm accuracy with linguistic stimuli. A possible explanation for this result may be the tolerance of lapses in Russian [8], which renders Russian speakers more familiar with non-binary rhythm than English speakers.

It may be counter-intuitive that Japanese speakers, whose native language rhythm is non-binary, perceived binary rhythm more accurately than English and Russian speakers, who should be more familiar with binary rhythms. However, there is a possible explanation: as shown in Table 1, test stimuli 3 to 7 for binary rhythm, by virtue of including clashes and lapses, became rhythmical but non-binary. Due to the method of accuracy calculation, the use of such non-binary rhythms for comparison with binary familiarization could have increased the accuracy scores of Japanese speakers. In addition, a possible reason for English and Russian speakers' relatively low accuracies with binary rhythm is that they could misperceive the non-binary rhythms created by clashes and lapses. English speakers tend to perceive that non-binary rhythm is avoided by stress shift even when they hear a phrase with a stress clash (i.e. when no stress shift is present) [12]. This would suggest that English speakers at least did not perceive clash and lapse violations as such, but compensated for their presence, as predicted. This type of auditory illusion was found by [4] too with respect to music. Not enough is known about Russian rhythm, but it is likely that similar mechanisms are in place and can account for the results.

The difference in musical experience between participants might affect the results too. However, looking at data provided by the British Government's Department for Education (https://www.gov.uk/government/organisations/department-for-education) and Russian Employee Social Network of Education (Социальная сеть работников образования; https://nsportal.ru/), the musical backgrounds of English and Russian participants are clearly different from each other's, while they have similar tendencies in their responses to the present experiment. For instance, there is a minimum of 30 hours of mandatory music teaching in UK schools, while for Russia the minimum is 72 hours. Considering the similarity between English and Russian linguistic rhythm, the difference in music education between the UK and Russia, and the finding that the accuracies of English and Russian speakers were similar to each other, linguistic experience seems to affect rhythm perception more than music training.

The influence of language is also reflected in the fact that the results were by and large the same independently of stimulus type. This indicates that the role of language is primary and that it is possible to extrapolate, with some caution, from studies with one of these types of stimuli to the perception of rhythm in other modalities. In conclusion, the present results support the idea that linguistic experience shapes our perception of rhythm.

5. Acknowledgments

The financial support of The Kao Foundation for Arts and Sciences (2016-2017) and the Kawai Foundation for Sound Technology & Music (2018) to the first author is hereby gratefully acknowledged.

6. References

[1] Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds ‘l’ and ‘r’. *Neuropsychologia, 9*: 317-323.

[2] Cutler, A. (2000). Listening to a second language through the ears of a first. *Interpreting, 5*(1): 1-23.

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

[4] Hannon, E. E., & Trehub, S. E. (2005). Tuning in to musical rhythms: Infants learn more readily than adults. *Proceedings of the National Academy of Sciences, 102*(35), 12639-126.

[5] Prince, A. (1983). Relating to the grid. *Linguistic Inquiry, 14*(1), 19-100.

[6] Hayes, B. (1984). The phonology of rhythm in English. *Linguistic Inquiry, 15*, 33-74.

[7] Mills, M. H. (1988). Perceived stress and the rhythmical organization of the utterance in Colloquial Russian. *Russian Language Journal/Русский язык, 42*(141), 51-65.

[8] Gouskova, M., & Roon, K. (2013). Gradient clash, faithfulness, and sonority sequencing effects in Russian compound stress. *Laboratory phonology*, *4*(2), 383-434.

[9] Tanaka, S., & Kubozono, H. (1999). Introduction to Japanese pronunciation: theory and practice. Tokyo: Kuroshio.

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

[11] Core, R. T., & Team, R. (2014). R: A language and Environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2013. *URL http://www. R-project. org*.

[12] Grabe, E., & Warren, P. (1995). Stress shift: do speakers do it or do listeners hear it. *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, ed. by B. Connell and A. Arvaniti. Cambridge: Cambridge University Press, 95-110.