Effects of pitch of short pure tones on onomatopoeic expressions:
A cross-linguistic study of Japanese and Chinese

Megumi Matsui*

Graduate School of Science and Technology, Sophia University,
7–1, Kioi-cho, Chiyoda-ku, Tokyo, 102–8554 Japan

(Received 10 May 2018, Accepted for publication 19 September 2019)

Abstract: The purpose of the present study was to examine the effects of the pitch of short-duration pure tones on onomatopoeic expressions, as well as the common relationships between frequency and onomatopoeic expressions. The sound stimuli were 85 pure tones (duration: 240 ms) that spanned 7 octaves from 62.5 Hz to 8 kHz in 1/12-octave steps. The participants were randomly presented with each material twice and were told to write or choose what they heard using an onomatopoeic expression. The results indicated that the participants tend to use /u/ or /o/ for low frequencies and /i/ for high frequencies, and the distribution of vowels tends to be similar for speakers of different languages. Therefore, in vowels, there are common relationships between the pitch of pure tones and onomatopoeic expressions.

Keywords: Pure tones, Frequency, Onomatopoeic expressions

PACS number: 43.66.Hg, 43.70.Kv, 43.70.Mn [doi:10.1250/ast.41.472]

1. BACKGROUND

1.1. Sound Symbolism

Sound symbolism (phonetic symbolism) is the direct relationship between a specific sound and a specific meaning [1]. It is of interest because language is considered to be arbitrary [2]. However, in fact, past studies have suggested relationships between sound and shape or meaning, and a relationship between a mono mora or a phoneme and the impression evoked from it has been revealed in multiple languages [e.g., 3–8]. On the other hand, there are also many studies dealing with the phenomenon as a cross-modality, cross-sensory, or cross-modal association [e.g., 9–11]. In addition, there are studies about the relationships between brain functions and the cross-sensory phenomenon [e.g., 12–14]. Furthermore, this sound symbolism and the cross-sensory phenomenon have also been reported to occur in children [e.g., 15–20]. From the above, the phenomena seem strongly related and universal.

1.2. Relationship between Physical Attributes of Sound and Onomatopoeic Expressions

In addition, the relationship between physical attributes of sound and language has been examined [21–36]. Studies include those related to the relationships between the frequency in the physical attribute of sounds and the expressions upon hearing the sounds.

Hiyane et al. [24] used short (duration: 200 ms) gammatones manipulated at the center frequency. In addition, on the basis of the work of Tanaka et al. [25], 21 expressions (e.g., “kan,” “shan,” and “rin”) were prepared as answer options. The results indicated the following tendencies: the occurrence rate of voiced consonants and /o/ increases when the tone is less than 1 kHz; the occurrence rate of /a/ increases from 1 to 2 kHz, that of /i/ increases at more than 2 kHz, and those of /k/ and /chi/ increases at approximately 3 kHz. Yamauchi et al. [26] manipulated the number of repetitions of materials, duration, attack time, and release time of the materials to study the relationship between the onomatopoeic expression for sine waves and sound images of these sounds. The results of the experiment showed the tendency of evoking emergency warning impression when the center frequency is low, and the occurrence rate of voiced consonants (for example, “bu”) increases in the low frequency band. In Yamauchi and Iwamiya [27] order to study functional images, Yamauchi and Iwamiya [27] used the frequency modulation tones. The results of the experiment showed, for example, that repeated expressions including long vowels (e.g., “/pi:po:pi:po:/)” were used for stimuli with modulation tones less than 5 kHz. Furthermore, the tones...
were associated with emergency images. Takada et al. [28] studied the relationships between onomatopoeic expressions for environmental sounds and impressions. There have also been studies on expressions for tinnitus. Ouchi et al. [29] indicated the following expression tendencies of Japanese listeners: “bu” for 125 Hz and “pi” or “ki” for 8 kHz in the case of normal hearing listeners. In addition, there is a fundamental and important series of studies by Oishi and co-workers [30–36]. In particular, as work that underlies the series of studies, Oishi et al. [30], in their Experiment 1, instructed participants to describe three pure tones with different frequencies using onomatopoeic expressions. Results indicated that the following onomatopoeic expressions were used the most frequently in the responses for the following pure tones: “bo” for 63 Hz, “pu” for 500 Hz, and “pi” for 4 kHz. In Experiment 2 of the same study [30], participants chose the frequency ranges they felt were appropriate for the three expressions “bo,” “pu,” and “pi” in Experiment 1 from among 19 pure tones that spanned 6 octaves from 63 Hz to 4 kHz in 1/3-octave steps. Results indicated the following boundaries between each expression: 195 Hz between “bo” and “pu” and 875 Hz between “pu” and “pi” (Fig. 1).

Thus, in the series of studies by Oishi and co-workers [30–36], typical expressions for three sounds, their boundary values, and factors affecting the typical relationships were examined. However, the previous study [30] may be difficult to reproduce because the following points were not examined. First, the experimental environment (soundproof room/others) and output device (headphones/loudspeaker) were not standardized among the experiments. Furthermore, samples were collected from a wide range of age groups. This may cause the perception and expressions for each sound to be different, and large variation may occur among the participants. Second, in Oishi et al.’s Experiment 2 [30], the hearing frequency of each sound and ways of hearing were not standardized among participants. Because of this, the selection accuracy may differ between participants with high listening frequency and low listening frequency, which may have influenced the boundary distribution. Finally, the number of stimuli used for each frequency was limited. For example, none of the above three expressions may have been appropriate near the boundary values 195 and 875 Hz. That is, there is the possibility that expressions suitable for each audible frequency range were insufficient.

In our present study, we aim to address the above points to obtain more reliable distributions of onomatopoeic expressions. Specifically, in order to examine the language response of young adults to the frequency range of 4–8 kHz and frequencies other than those of the 19 tones (62.5 Hz to 4 kHz) that Oishi’s group [30–36] did not examine, we adopted the following three points and precisely examined the relationship between the frequencies of pure tones and onomatopoeic expressions: 1) control of the experimental method (experimental environment, equipment, and the number of repetitions were standardized among the participants); 2) limit the age of participants (in their 20s and younger), and 3) expand and subdivide the target frequency band (pure tones that spanned 7 octaves from 62.5 Hz to 8 kHz in 1/12-octave steps).

2. PURPOSE

The purpose of the present study was to examine the following three points. 1) In Experiments 1 and 2, typical onomatopoeic expressions appropriate for each frequency band. 2) In which frequency range (Hz) each onomatopoeic expression changes to another expression (Experiment 2). 3) In Experiment 3, compare speakers of Japanese and Chinese to examine the relationships between the pitch of sounds and onomatopoeic expressions. In addition, although Japanese and Chinese belong to the East Asian languages and both languages have onomatopoeic expressions based on environmental sounds, their pronunciations differ (e.g., consonants). To examine whether differences in such linguistic characteristics affect the verbalization of pure tones or whether there are any common tendencies between native speakers of the two languages beyond the differences, those two languages were selected in the present study.

In addition, a slightly different procedure from that of Oishi et al. [30] (refer to Appendix 1) was used to acquire more precise results.

3. METHODS

3.1. Design

The independent variable was the frequency of the pure tone stimuli and the dependent variables were the occurrence rate of onomatopoeic expressions (consonant and vowel in the first mora: “CV”) in a free-response task [Experiments 1 and 3] and the selection rate of onomatopoeic expressions in a forced-choice task [Experiment 2].

3.2. Participants

The participants in Experiment 1 were 34 students (male: 9, female: 25, mean age: 18.35, SD: 0.64), who were
native Japanese speakers (JNs). In addition, the areas where the participants grew up were as follows (regions with the longest living period from 0 to 18 years of age): Tohoku region (northeast Japan), 3 people; Kanto region, 27; Chubu region (central Japan), 2; Kansai region, 2.

The participants in Experiment 2 were 34 JN students (male: 12, female: 22, mean age: 19.91, SD: 1.17). In addition, the areas where the subjects of analysis grew up were as follows: Tohoku region, 2 people; Kanto region, 28; Chubu region, 2; Chugoku region, 1; Kyushu-Okinawa region, 1.

The participants in Experiment 3 were 20 students (male: 2, female: 18, mean age: 24.55, SD: 1.50), who were native Chinese speakers (CNs) and also learning the Japanese language as second or multi-language learners. In addition, the areas where these subjects grew up were as follows: Shanghai, 2 people; Beijing, 1; Hebei, 1; Hubei, 2; Chongqing, 3; Guizhou, 3; Liaoning, 2; Henan, 2; Jilin, 1; Zhejiang, 1; Guangxi Zhuang, 2.

Furthermore, all participants fulfilled the following conditions, 1) they agreed to participate in each experiment, 2) had no hearing disorders, 3) had no knowledge about this study, and 4) did not have absolute pitch. In addition, all participants were assigned to only one experiment.

3.3. Materials

The materials were 85 pure tones (duration: 240 ms, onset: 20 ms, offset: 20 ms) that spanned 7 octaves from 62.5 Hz to 8 kHz in 1/12-octave steps (Fig. 2). These were made by Audacity 2.1.1. (free software) [37]. Furthermore, amplitude was normalized so that the sound pressure levels agree with the A weighting function. In measurement, the sound level meter was fixed to the stereo headphones. Furthermore, short-duration materials were used in the present study because the purpose was not to analyze word-final consonants (e.g., “bu” “chip”) but to analyze CVs.

3.4. Apparatus

A PC (ARROWS Tab, FARQ02012, Fujitsu), attached keyboard, and Superlab 4.5 (Cedrus) were used to present and control the materials, and record the responses from the participants. In addition, stereo headphones (CONTROL-HD270, made by Sennheiser) were used to present the sound materials to the participants. Furthermore, a sound level meter (NL-42, made by Rion) was used to measure the loudness level of each sound.

3.5. Procedure

All experiments were conducted individually in a soundproof room at Hosei University. The materials were played on the PC and presented through headphones to the participants.

First, participants heard eight pure tones (frequency: 62.5 Hz, 125 Hz, 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz, 8 kHz, duration: 240 ms, ISI: 500 ms) as the sample sounds. In Experiments 1 and 3, participants were told to write what they heard using an onomatopoeic expression and move on to the next trial by pressing the space key. JNs responded using hiragana or katakana (Japanese syllabary, both of which are components of the Japanese writing system). In addition, participants were told the following. 1) There is no correct answer. 2) Listen to each sound carefully. 3) Respond intuitively without thinking deeply. 4) Do not respond with a musical note name. 5) Do not compare with other sounds. 6) The same onomatopoeic expression can be used multiple times.

In Experiment 2, participants were told to choose an appropriate onomatopoeic expression for the presented sound from a fixed set of alternatives (chosen on the basis of the results of Experiment 1). Participants were told the following. 1) There is no correct answer. 2) Listen to each sound carefully. 3) Respond intuitively without thinking deeply. 4) Do not compare with other sounds. 5) The same onomatopoeic expression can be used multiple times. 6) Push the key carefully to avoid key-press errors.

Furthermore, CNs responded using pinyin and Chinese tones (first tone: Tone 1, high and smooth tone; second tone: Tone 2, rising tone; third tone: Tone 3, falling and rising tone; fourth tone: Tone 4, falling tone; light tone: natural tone [38]).

In addition, in all experiments, participants were randomly presented with each material twice (Total: 170 trials). Eighty-five tones in random order made up one block. The time required for each experiment was approximately 10 min. Each experiment was conducted separately.

4. RESULTS

4.1. Experiment 1: Free-response Task for JN

In Experiment 1, the participants responded with a total of 41 CVs as onomatopoeic expressions in all of the frequency bands. The CVs included some musical scale names in Japanese (e.g., “do” (C), “re” (D), “mi” (E)). However, all responses were considered in the analysis.
because of the large number of participants who used musical scale names even though the participants had been instructed specifically not to do so and were told to write what they heard using an onomatopoeic expression by the experimenter. Therefore, we concluded that participants were used to using musical scale expressions as onomatopoeic expressions.

To examine the tendencies of the expressions that were associated with each frequency bands the CVs with the most and second most number of occurrences in the responses given by the participants were chosen for analysis. Figure 3 shows the mean occurrence rate and standard deviation of each CV for each 1/3-octave band. It shows that “pi” and “ki” tended to appear frequently in the high frequency bands (HF), “pu” and “po” tended to appear frequently in the medium frequency bands, and “bu” and “bo” tended to appear frequently in the low frequency bands (LF).

In addition, to verify what consonants were associated with each frequency band, the CVs with occurrence rates of 5% or more were extracted. The occurrence rates of consonants in the CVs are shown in Fig. 4. The figure indicated the boundary value of 157 Hz between /b/ and /p/. As with consonants, vowels were also extracted. The occurrence rates of vowels in the CVs are shown in Fig. 5. The figure indicated the boundary values of approximately 500 Hz between /o/ and /u/ and 1 kHz between /u/ and /i/.

4.2. Experiment 2: Forced-choice Task for JN

On the basis of results of Experiment 1, response alternatives for Experiment 2 were selected using the following criteria. The CVs with the most and second most number of occurrences in any of the octaves in Experiment 1 were selected as the response alternatives for Experiment 2. As a result, the following seven expressions were selected: “bo,” “bu,” “po,” “pu,” “pa,” “pi,” and “ki” (Fig. 3). These were used as representative syllables (RSs) in Experiment 2. The mean selection rate of RSs in 1/3-octave bands are indicated in Fig. 6. This figure shows that the following RSs tend to dominate each frequency band: “bo” between 63 and 187 Hz; “pu” between 198 and 375 Hz; “po” between 397 and 1,189 Hz; “pa” between 1,260 and 1,498 Hz; “pi” between 1,587 and 5,040 Hz; “ki” between 5,040 and 8,000 Hz.

In addition, the degree of agreement between the two responses to the same material by each participant is shown in Fig. 7. This figure shows that: while responses within individuals do not match at LF, /i/ selections within individuals match at HF.

Furthermore, Fig. 8 shows a 6th-order polynomial regression model (1) fitted to each selection rate of RSs at each frequency ($R^2 = 0.98$ for “bo” and “bu”; 0.92 for “po”; 0.87 for “pu”; 0.86 for “pa”; 0.94 for “pi”; 0.97 for “ki”).

$$y_i = b_0 + b_1x_i + b_2x_i^2 + \cdots + b_6x_i^6 + e_i$$  (1)
Figure 8 indicates the following boundary values between expressions: 66 Hz between “bo” and “bu,” 177 Hz between “bu” and “pu,” 315 Hz between “pu” and “po,” 1,260 Hz between “po” and “pi,” and 5,657 Hz between “pi” and “ki.” Furthermore, it also shows the following maximum selection rates for each expression: “bo” at 74 Hz, “bu” at 88 Hz, “pu” at 315 Hz, “po” at 530 Hz, “pa” at 1,888 Hz, “pi” at 3,776 Hz, and “ki” at 8,000 Hz.

4.3. Experiment 3: Free-response Task for CN

From the results of Experiment 3, although CNs were instructed to answer using Pinyin, they also answered with expressions that were not Pinyin (e.g., “puo,” “bou,” “dù”). Therefore, the expressions were categorized by first mora (CV). Among them, four CVs (“bu,” “du,” “bi,” “di”) satisfied the previous criterion, that is, their occurrence rates were the highest or second highest in one of the octave bands. The mean occurrence rates of these CVs in each 1/3-octave band are indicated in Fig. 11. This figure shows that the following CVs tend to dominate in the respective frequency bands: “bu” between 63 and 1,498 Hz, and “bi” and “di” appear to compete with each other between 1,587 and 8,000 Hz.

Figure 12 shows the tendency of each consonant appearing in each of the frequency bands. The consonants were included in 5% or more in any octave band. According to the figure, /b/ and /d/ also appeared to be competing with each other, whereas /p/ (aspirated consonant) and /b/ (un aspirated consonant) did not appear to be in conflict with each other (Fig. 13). Furthermore, as with consonants, vowels were similarly extracted.

Figure 7 Match rate within individuals (N = 34). 1: bo, 2: bu, 3: po, 4: pu, 5: pa, 6: pi, 7: ki.

Fig. 7 Match rate within individuals (N = 34). 1: bo, 2: bu, 3: po, 4: pu, 5: pa, 6: pi, 7: ki.

Fig. 8 Fitting of polynomial regression model (N = 34).

Fig. 8 Fitting of polynomial regression model (N = 34).

Fig. 9 Distribution of consonants (N = 34).

Fig. 9 Distribution of consonants (N = 34).

Fig. 10 Distribution of vowels (N = 34).

Fig. 10 Distribution of vowels (N = 34).

Fig. 11 Distribution of CVs (N = 20).

Fig. 11 Distribution of CVs (N = 20).
Figure 14 indicates the boundary value of about 1 kHz between the vowels /u/ and /i/.

In addition, to examine the effect of differences in the pitch of pure tones on the occurrence rate of “Tones,” the data were analyzed using a two-factor 7 × 6 (7 oct. bands × 6 tones: Tone 1–Tone 4, the light tone, and no tone) ANOVA design (Appendix 2). From the results of the ANOVA analysis, there were significant effects in “Tones” (F(5, 95) = 17.56, p < .01, η² = 0.40) and interaction between the two factors (F(30, 570) = 3.33, p < .01, η² = 0.03). From the results of a multiple comparison (Bonferroni test), the mean occurrence rate of Tone 1 was found to be higher than those of Tones 2, 3, 4 and the light tone in all frequency bands (Figs. 15 and 16). Also the mean occurrence rate of Tone 1 was higher in the 125 Hz band than in the 63 Hz band, and that of Tone 1 was higher in the 500 Hz band than in the 63 Hz band. In addition, in each octave band, the mean occurrence rate of Tone 1 tended to be higher than those of the other tones.

Furthermore, to compare these expressions between the two groups of native speakers, the occurrence rates of CVs were analyzed for vowels (in both cases the highest and second highest occurrence rates of CVs in any of the octave bands were extracted). The results of the analysis showed that the occurrence rate of vowels in both groups of native speakers was predominantly /u/ in the LF regions and /i/ in the HF region, with the boundary between them being around 1 kHz (Fig. 17). On the other hand, a comparison of Figs. 4, 12, and 13 revealed that there was no common tendency, similarly to consonants.

5. DISCUSSION

The present study had the following three purposes. 1) In Experiments 1 and 2, determination of typical onoma-
topoecic expressions appropriate for each frequency band. 2) Determine the frequency range (Hz) in which each onomatopoeic expression changes (Experiment 2). 3) In Experiment 3, comparison between speakers of Japanese and Chinese to examine the relationships between the pitch of sounds and onomatopoeic expressions that may exist across languages.

From the results of conducting experiments to fulfill purposes 1 and 2, we observed the following distributions: “bo” between 63 and 66 Hz, “bu” between 66 and 177 Hz, “pu” between 177 and 315 Hz, “po” between 315 and 1,260 Hz, “pi” between 1,260 and 5,657 Hz, and “ki” between 5,657 and 8 kHz. Compared with the results of a previous study [30], the present results indicate a certain common tendency among Japanese listeners, for example, their perception of the two consonants /b/ and /p/ changes at similar values in LF regions, and they have an increased tendency to perceive pure tones as /u/ or /o/ in LF ranges and /i/ in HF ranges. The result also indicated differences, such as a different boundary value between vowels.

On the other hand, the present study indicated the following new finding. With a greater number of stimuli in a larger range of target frequencies, and a controlled experimental environment, for instance, the following tendency was observed: there is competition between /bo/ and /bu/ in the LF, /pu/ and /po/ in the middle frequency bands (approximately 500 Hz), and /pi/ and /ki/ in HF, and also, there are more /ki/ expressions in HF (approximately 5 kHz over). The above tendencies were also seen in past studies but were underestimated. In actuality, the tendency may be important. There may be more variation in the onomatopoeic expressions associated with different pitches of pure tones.

Furthermore, we compared the results from speakers of different languages to examine the common relationships between the frequency of pure tones and onomatopoeic expressions (purpose 3). The results indicated that there are similar distributions of each vowel in both groups of native speakers. Japanese listeners used /o/ in LF and /i/ in HF. Chinese listeners used /u/ in LF and /i/ in HF. The results showed the possibility that there are strong relationships between the second formant frequency (F2) of vowels and the frequency of presented pure tones. This is because F2 of /o/ and /u/ have the lowest frequencies in the vowels of both languages [39,40], and F2 of /i/ has the highest frequency in the vowels in both languages. Therefore, in terms of vowels, there are common relationships between the pitch of pure tones and onomatopoeic expressions, which may be a perceptual characteristic common to humans. On the other hand, when considering consonants, there are the following two problems. Firstly, the way consonants are produced is different in the two languages (“voiced consonants” and “unvoiced consonants” in Japanese, “aspirated consonants” and “unaspirated consonants” in Chinese) [41], and therefore, a simple comparison is not valid. In addition, the reason why the same tendency of consonant occurrence in Japanese (Fig. 4) cannot be seen in Chinese (Fig. 13) may be as follows. The Chinese aspirated/unaspirated consonants (for example, /p/ and /b/) are acoustically similar to each other compared with in Japanese, as indicated by the phonetic symbols [ph] and [p] [41]. On the other hand, the reason Japanese speakers used voiced consonants for low-frequency tones may be related to the fact that the voiced consonants have lower F0 of subsequent vowels than do unvoiced consonants. The reason may also be that the voiced sounds appear acoustically with low frequency energy. However, Haryu and Zhao [9] suggested a difference among Japanese speakers, Chinese speakers, and Japanese-learning Chinese speakers. Haryu and Zhao [9] revealed that when a native speaker of Chinese is a Japanese language learner, the impressions of unvoiced sounds are closer to those of a native Japanese speaker than a non-Japanese learner. On the other hand, the results of our present study suggested that Chinese speakers did not use voiced/unvoiced consonants to the difference in frequency even though all participants were Japanese-learning Chinese speakers. However, the experimental procedures were not the same in the two studies; Haryu and Zhao [9] used a forced-choice task but we used a free-response task. An additional cause may be that we instructed that Chinese (pinyin) be used. Therefore, the difference in the results may be indicative of these differences. Secondly, the onset of the materials was 20 ms, but if the onset changes, the perceived consonant may also change. In addition, post analysis was conducted on the effects of differences in the pitch of pure tones on Chinese tones because each tone is uttered at a different frequency, but there was no tendency of the tone associated with the pure tone being changed according to the pitch of pure tones, and in all frequency bands, the average
occurrence rate of Tone 1 was higher than those of other tones. However, there is a possibility that this result may reflect physical attributes (such as sound pressure changes) other than the frequency of pure tones. Furthermore, the results should also be analyzed by birthplace because the number of tones differs for different dialects of Chinese, and even if the number of tones matches, speaking involves different frequencies.

Finally, here are four important future tasks: 1) continue to conduct experiments on Chinese speakers to match the number of samples and accurately compare the results for Chinese and Japanese; 2) investigate the relationship between the pitch of pure tones and F2 of vowels, and the reason for the correspondence (e.g., shape of oral cavity at utterance). Furthermore, we need to investigate the reason why the match rate in individuals differs between LF and HF (e.g., auditory peripheral system); 3) compare results of native speakers of other languages; 4) conduct experiments that manipulate other physical attributes of sound (e.g., duration, amplitude).

6. CONCLUSION

In this study, by subdividing the target frequency range, we clarified onomatopoeic expressions for indicating a target frequency range and the boundary values of each expression (“bo” between 63 and 66 Hz, “bu” between 66 and 177 Hz, “pu” between 177 and 315 Hz, “po” between 315 and 1,260 Hz, “pi” between 1,260 and 5,657 Hz, and “ki” between 5,657 and 8 kHz) (Experiments 1 and 2). The results of this study partially supported the results obtained by Oishi and co-workers [30–36] and provided more details of the relationship between the frequency of pure tones and onomatopoeic expressions.

Furthermore, by targeting not only Japanese native speakers but also Chinese native speakers (Experiment 3), it was possible to investigate the common points and differences between the two languages.

However, since the number of participants and the number of languages were still small, the universality of the relationship between the frequency of pure tones and onomatopoeic expressions could not be definitively established; this is a task for the future.

ACKNOWLEDGMENTS

This paper is a partial revision of my master’s thesis submitted to Hosei University. In carrying out this study, Professor Keiichi Tajima of Hosei University has given me meaningful suggestions and advice. I also thank the two anonymous reviewers for their careful reading of this paper and their many valuable comments.

REFERENCES

[1] L. Hinton, J. Nichols and J. J. Ohala, (Eds.), Sound Symbolism (Cambridge University Press, Cambridge, 2006).
[2] F. D. Saussure, Course in General Linguistics (W. Baskin, Trans.) (Philosophical Library, New York, 1959).
[3] W. Köhler, Gestalt Psychology: An Introduction to New
