Title:

Spectral consistency in sound sequence affects perceptual accuracy in discriminating subdivided rhythmic patterns

Running title:

Spectral consistency affects perceptual accuracy of subdivided rhythm

Author:

Jun Nitta[1], Sotaro Kondoh [1], Kazuo Okanoya[1,2**], Ryosuke O. Tachibana[1*]

Affiliation

[1] Graduate School of Arts and Sciences, the University of Tokyo, Tokyo, Japan

[2] Advanced Comprehensive Research Organization, Teikyo University, Tokyo, Japan

[**] co-corresponding author

E-mail: kazuookanoya@gmail.com

[*] corresponding author

E-mail: rtachi@gmail.com

Abstract:

Musical compositions are distinguished by their unique rhythmic patterns, determined by subtle differences in how regular beats are subdivided. Precise perception of these subdivisions is essential for discerning nuances in rhythmic patterns. While musical rhythm typically comprises sound elements with a variety of timbres or spectral cues, the impact of such spectral variation on our perception of rhythmic patterns remains unclear. Here, we show that consistency in spectral cue affects perceptual accuracy in discriminating subdivided rhythmic patterns. We conducted online experiments using rhythmic sound sequences consisted of band-passed noise bursts to measure discrimination accuracy. Participants were asked to discriminate between a swing-like rhythm sequence, characterized by a 2:1 in interval ratio, and its more or less exaggerated version. This task was also performed under two additional rhythm ratio conditions: an inversed-swing rhythm (1:2 ratio) and a regular subdivision (1:1 ratio). The center frequency of band noises was either held constant or alternated between two values. Our results revealed a significant decrease in discrimination accuracy when the center frequency was alternated, irrespective of the rhythm ratio condition. This suggests that rhythm perception is not only shaped by temporal structure but also affected by spectral properties.

Introduction

We detect temporal regularity and capture complex structures of sound events in almost all styles of music. Each perceived regular pulse is called a beat. Subdividing each beat interval provides various characteristics in a rhythmic pattern of the music. In other words, we perceptually organize various sound events into a rhythm structure according to their temporal properties by detecting and subdividing regular patterns in sound sequences [1]. For example, subdividing beats with simple integer ratios such as 1:1 and 2:1 is far more frequently observed in Western music [2]. In contrast, complex ratios such as 5:3 are very common in drum ensemble music in West Africa [3]. One of the famous subdivisions is swing. Swing is a long-short duration alternating pattern considered an essential factor in jazz music. The integer ratio of swing is generally considered to be 2:1, but it approaches 1:1 as the tempo increases. [4].

Various temporal factors influence our rhythm perception. We can easily perceive regular beats even though the actual timings of performed beats usually fluctuate and deviate from an exact regular pattern [5]. Despite such variation in beat durations, audiences perceive temporal regularity in the performed sound pattern and interpret the variation as intentional aesthetic expressions [6,7]. However, as beats are subdivided into finer or uneven intervals, rhythm perception becomes more variable. Equal subdivision causes the tempo to be perceived as slower, even though the actual tempo remains the same [8]. Rhythmic stimuli with the same pattern but different tempos are more likely to be perceived as different rhythmic patterns [9]. The complexity of rhythmic patterns also changes perceived rhythms. A study suggested that complex (syncopated) rhythms may be perceived by reinterpreting the phases of rhythms [10]. The perceptual accuracy of rhythm also changes. The just noticeable difference (JND) of monotonic isochronous

sound sequences is considered to be approximately 2.5 % when the inter-onset interval (IOI) is over 240 ms for the method of adjusting the displacement of only one note [4]. However, the JND varies widely depending on the measurement method and the time interval size. In the task of detecting phase shifts in equally spaced sound sequences, the detection threshold increases when the tempo is fast [11]. Moreover, many studies have indirectly investigated the perceptual accuracy of subdivision swing ratios using a tapping task. In the synchronization-continuation task, in which participants tap synchronously to a rhythmic pattern with a certain interval ratio and continued tapping even after the stimuli stopped, the tapping converged to a certain swing ratio [12-14]. Studies of behavioral experiments have also shown some findings: asymmetries have been reported in which subdivisions are perceived as equally spaced even though they are short-long patterns, and this does not happen in long-short patterns [15].

It is known that not only the temporal property but also sound timbre, or spectral property, influences how we organize the sound sequence structure. This issue has been intensively investigated in the research field of auditory stream segregation [16,17]. For example, a study using band noise suggested that differences in the center frequency of the band noise best predicted the degree to which stream segregation occurred [18]. Another study used a sound sequence alternating between pure tone and narrow-band noise regularly and delaying only the even-numbered sounds gradually [19]. The result showed that the more different in spectral aspect the two sound elements were, the less likely the subjects were to notice the delay of even-numbered sounds. The authors claimed that differences in timbre enhanced stream segregation, and the segregation made distinguishing the temporal relationship more difficult. This raises the possibility that variations in spectral pattern might influence the accuracy of rhythm perception. In real

music pieces, rhythmic patterns typically consist of sounds with different timbres, such as those produced by a drum set. Consequently, an experimental assessment of how varying spectral cues affects the accuracy of rhythmic pattern perception offers novel insights for understanding rhythm perception in practical musical contexts.

In the present study, we aimed to examine the effect of different timbres on rhythm perception by psychophysically measuring the discrimination accuracy of multiple rhythmic patterns, including swing rhythms. We used a sound sequence consisting of band noise bursts, and manipulated the spectral cue by changing their center frequency. In order to recruit many participants, experiments were conducted online via a web browser. Three online experiments were conducted for the three rhythmic patterns: long-short rhythm (experiment 1), short-long rhythm (experiment 2), and straight rhythm (experiment 3). To confirm the reliability of online experiments, particularly in the aspect of data reproducibility across arbitrary sound listening environments, we additionally conducted an experiment with participants who performed their tasks in the local laboratory so that we could confirm the listening environment (Experiment 1').

Methods

Participants

We initially recruited a hundred participants for each of three online experiments via a crowd-sourcing service (CrowdWorks, Inc., Japan). None of them reported history of hearing problems. A subset of participants was excluded from the analysis (5, 6, and 14 participants in experiments 1, 2, and 3, respectively) due to incomplete data retrieval caused by malfunctions in the online experiment system. Then, we screened the dataset according to correct rates in screening trials and goodness of fit to the psychometric

function (see Analysis section). After the screening, data from 40 (20 males, 20 females; age: M ± SD = 39.2 ± 8.10 yo), 43 (21 males, 21 females, 1 other; age: 36.3 ± 8.66), and 30 (18 males, 12 females, age: 38.9 ± 8.96) participants for experiments 1, 2, and 3 were included in further analysis (**Table 1**). We allowed participants to join multiple experiments and additionally assessed intra-individual factors across different experiments. Several people participated in two experiments: after screening, 12, 10, and 12 participants overlapped in experiments 1-2, 2-3, and 3-1, respectively. Three participated in all experiments. We additionally conducted the experiment 1' on 16 people (8 males, 8 females; age: 27.5 ± 6.74) in our local laboratory without using the crowd-sourcing service so that we were able to confirm the acoustic environment of the participants.

We confirmed music expertise with a questionnaire. None of the participants was an expert musician, but almost half had experienced intensive music instrument practices as amateurs: 46, 47, and 46 participants in experiments 1, 2, and 3, respectively. Note that the intensive music experience refers to training at least two hours per day and may include experience in music classes and club activities but excludes experience in compulsory music classes. All participants were given informed consent prior to the experiment. All experimental procedures were approved by Ethics Review Committee on Experimental Research with Human Subjects of the Graduate School of Arts and Sciences, The University of Tokyo (No. 718-3).

Table 1. The number of participant information.

participants	**Exp 1**	**Exp 2**	**Exp 3**	**Exp 1'**
screened (recruited)	40 (95)	43 (94)	30 (86)	16 (16)

male	20 (53)	21 (51)	18 (50)	8 (8)
female	20 (42)	21 (42)	12 (36)	8 (8)
other	0 (0)	1 (1)	0 (0)	0 (0)

Sound stimuli

We employed sound sequences of brief band-passed noises as stimuli (**Figure 1**). First, the center frequency of the band noise was manipulated, and two frequency conditions were prepared: the different frequency condition (1000 Hz and 2250 Hz, equivalent to 14 semitones) and the same frequency condition (1500 Hz), the bandwidth at 1000 Hz. Then, each band noise was created from white noise by band-pass filtering (FIR, 1000 Hz: 582 taps, 1500 Hz: 1,020 taps, 2250 Hz: 2,038 taps, designed by the 'bandpass' function in MATLAB). The duration of each band noise was 30 ms, with a rise and a fall portion of 10 ms each.

The band noises were arranged chronologically; the inter-beat interval (IBI) was fixed at 300 ms, and stimuli were generated for four beats. The band noises were placed alternating between on-the-beat and off-the-beat. In the different frequency condition, there were two types of the order of band noise, depending on which was placed first. A previous study showed that subjective stream segregation occurred in the same IBI and center frequency separation [18]. Thus, we would expect to see auditory stream segregation in the different frequency condition of the present study.

The onset of off-the-beat was defined by interval ratio. The interval ratio *r* was defined as in the equation as follows:

$$\boldsymbol{r = 1 - (T/2 - d)/(T/2 + d)}$$

here, *T* represents IBI, which was fixed at 300 ms, and *d* represents onset displacement,

i.e., the time delay of the onset from 150 ms. By manipulating the value of *r*, we created standard and comparative stimuli. There were three types of standard stimuli; each value of *r* was 0.5, −0.5, 0, and the represented rhythm pattern was 2:1, 1:2, 1:1. For each standard stimulus, the comparison stimuli were prepared by varying the value of *r* in five small steps: $r = 0.2$, 0.4, 0.5, 0.6, and 0.8 (2:1); $r = -0.2$, −0.4, −0.5, −0.6, and −0.8 (1:2); $r = -0.3$, −0.15, 0, 0.15, and 0.3 (1:1).

All sound stimuli were digitally generated at 44.1 kHz, 16 bits sampling. Their amplitudes were adjusted to have the same root-mean-square level.

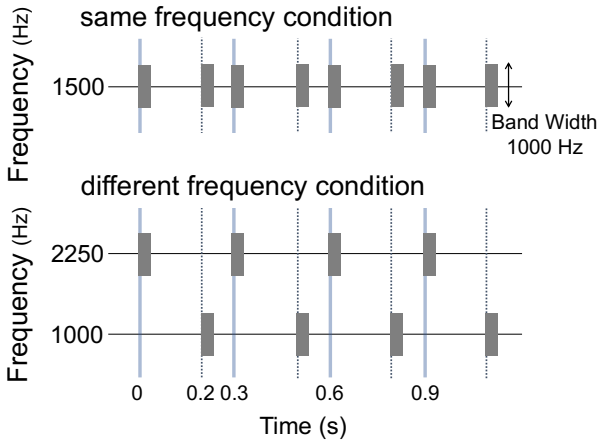

Figure 1. Schematic drawing of stimuli design for *r* = 0.5. Upper indicates a stimulus used in the same frequency condition. Lower shows a stimulus used in the different frequency condition. The thick solid line (blue) indicates beats, and the vertical dotted line (black) indicates subdivisions.

We performed three different experiments that contained different stimuli as standards to be discriminated; long-short rhythm (2:1), short-long rhythm (1:2), and straight rhythm (1:1). Participants in experiments 1 and 1' were assigned to long-short rhythm,

experiment 2 to short-long rhythm, and experiment 3 to straight rhythm.

Task

We employed the two-alternative forced-choice task. There were two types of tasks: for experiments 1, 1’ and 2, and for experiment 3. In the experiments 1, 1’ and 2, a standard stimulus and a comparison stimulus were presented in this order continuously. The interstimulus interval was fixed at 300 ms, corresponding to one beat of the stimuli. Participants were asked to choose "the bouncier one," with the larger absolute value of the interval ratio, between the two stimuli. In the latter task, the standard and comparison stimuli were presented continuously in randomized order. participants were asked to choose "the one equally spaced." In both tasks, participants responded by mouse click within 2.5 s. After participants responded, feedback for correct or incorrect answers was presented by ○ (correct) or × (incorrect). The above process was defined as one trial.

To confirm whether participants appropriately focussed on the task, we inserted screening trials in experimental trial sequences. Interval ratios of comparative stimuli in screen trials were selected to be as distant from the standard stimuli as possible (r = 0.2, -0.2, and 0.3 in experiments 1, 2, and 3 respectively), so that participants easily chose the correct response if they kept paying attention to the presented stimuli. The screening trials were added at the beginning of the first block and the end of each block.

Procedure

All experiments were programmed using a JavaScript-based experiment bulider (lab.js [20]) and conducted online via Open Lab (Open Lab Online UG, Konstanz, Germany). Participants were required to participate from a PC and wear headphones or earphones

when participating in the experiments.

Experiments 1, 2, and 3 consisted of 240 trials with two frequency difference conditions × five onset displacement conditions × 24 repetitions. The experiment was conducted in blocks of 40 trials × 6, with breaks between blocks. In Experiment 1', the number of repetitions was increased to 40 to prioritize the number of trials over the time required. The experiment was conducted in blocks of 40 trials × 6, with breaks between blocks. The order of high and low frequencies was counterbalanced among participants. The frequency conditions and onset displacement conditions were randomized within blocks.

Before the main experiment, we conducted a headphone screening task, confirming whether participants were wearing headphones and instruction/practice of the experiment. After the main experiment, a questionnaire about the music experience was administered to investigate the effects of musical experience on discrimination accuracy additionally. The entire experiment took about 30 minutes. In the headphone screening task, participants were asked to select the sound with the lowest loudness among three different sound stimuli, based on a previous report [21]. The three sound stimuli consisted of (1) a 200 Hz pure tone, (2) a 200 Hz pure tone with a sound pressure 6 dB lower than (1), and (3) an inverse phase stereo presentation of (1).

The practice session consisted of 8 trials, one for each of the eight conditions except for the one with an onset displacement of 0 ms. In the practice session, correct/incorrect feedback was presented after each response, and if the response time was exceeded, the participants were told that their time was up.

The questionnaire survey included the following information on musical experience (experience with musical instruments or not, hours of music listening per day,

presence or absence of musical instrument experience and type of instrument/most proficient instrument, duration of training/training hours per day, starting age) as well as favorite musical genres and songs, dance and rhythm game experience (presence or absence of experience, starting age and starting age and years), as well as their favorite music genres and songs, and experience with dance and rhythm games (presence/absence, starting age and years of experience). The last item included a question to check whether the respondent had read the questionnaire correctly.

Analysis

The data were averaged within-subject response rates by frequency condition. In experiments 1 and 2, The obtained response rate was approximated by a normal cumulative distribution function (sigmoid), defined by the 'cdf' function in MATLAB. In the sigmoid function, μ gives the gradient of the curve maximum, which corresponds to the point of subjective equality (PSE), and σ corresponds to the extent of the gradient of the curve, which is equivalent to the discrimination accuracy. We calculated average response rates for all subjects and compared them with frequency conditions.

Data with one or more incorrect answers in the screening trials were excluded. as not having appropriately participated. Furthermore, data with an adjusted coefficient of determination (R^2) smaller than 0.5, the absolute value of PSE ($|\mu|$) larger than one or smaller than 0, or discrimination accuracy (σ) smaller than 0, in both same frequency conditions and different frequency condition, were excluded as the approximation did not fit.

Homogeneity of variance was examined using μ and σ by the two-sample F test, which failed to guarantee the homogeneity in μ in experiment 1 (F (39, 39) = 2.32, p =

0.010), σ in experiment 1 (F (39, 39) = 7.88, $p < 0.001$), σ in experiment 2 (F (42, 42) = 3.02, $p < 0.001$), and σ in experiment 3 (negative: F (17, 17) = 7.47, $p < 0.001$; positive: F (17, 17) = 3.22, $p = 0.021$). Therefore, we employed Wilcoxon signed rank tests, one of the paired nonparametric tests. The significance level α was set at 0.05. Post hoc analysis confirmed that power ($1-\beta$) was more significant than 0.8 in all comparisons.

In experiment 3, we performed a different analysis from experiments 1 and 2 since the fitting function was later found to be incorrect. The fitting function was defined in the equation as follows, a modified version of the normal distribution equation with μ fixed at 0 and the maximum value fixed at 0.5 (symmetric fitting):

$$f(x) = \frac{1}{2}\, exp(-\frac{x^2}{2\sigma})$$

In addition, we also performed a fitting with a similar function where σ differs according to the sign of the interval ratio in order to investigate asymmetry of different accuracy (asymmetric fitting).

Data with one or more incorrect answers in the screening trials were excluded. as not having appropriately participated, as in experiments 1 and 2. Furthermore, Data with R^2 of symmetric fitting smaller than 0.5 or σ smaller than 0, in both the same and different frequency conditions, were excluded as the approximation did not fit. We employed Wilcoxon signed rank tests as in experiments 1 and 2, and there were some data (n = 3) in which σ was very close to zero. Therefore, we conducted the test in each case when they were excluded and not excluded.

Result

Validity and reliability of the online experiment

To investigate the influence of a sound-listening environment, we performed a sigmoidal approximation for experiments 1 and 1' using the same technique, plotting the spacing ratio on the horizontal axis and the response rate on the vertical axis (**Figure 2a**). Visual comparison of the sigmoid approximation curves between the participants in experiments 1 and 1' revealed commonality under the same frequency condition. This confirmed the validity and reliability of the online experiment.

Exp 1: Long-short rhythm (2:1)

To clarify the effect of timbre on rhythmic perception for long-short rhythm, we measured discrimination accuracy by calculating σ of the sigmoid approximation curve. We screened data according to the exclusion by screening trials and fitting. The remaining 40 data in experiment 1 were used in the analysis (**Figure 2a, 2b, 2c**). We calculated the response rate in each interval ratio, compared by frequency conditions, and found that the gradient in the different frequency condition was smaller than that in the same condition (**Figure 2a**). Concerning PSE (μ) (**Figure 2b**) and discrimination accuracy (σ) (**Figure 2c**), Wilcoxon signed rank tests for μ and σ showed that μ was not significantly different between two conditions ($Z = 1.83$, $p = 0.068$), and σ was found to be significantly larger ($Z = 4.73$, $p < 0.001$) in the different frequency condition than the same frequency condition. This indicates that discrimination accuracy was lower when the center frequency of narrow bands was different.

Exp 2: Short-long rhythm (1:2)

To examine the discrimination accuracy of short-long rhythm, the remaining 43 data in experiment 2 were used in the analysis (**Figure 2d, 2e, 2f**). We calculated the response rate in each interval ratio and compared by frequency conditions, and found that the gradient in different frequency condition were smaller than that in the same frequency condition, as well as in long-short rhythm (**Figure 2d**). Concerning μ (**Figure 2e**) and σ (**Figure 2f**), Wilcoxon signed rank tests showed that μ was significantly different between conditions ($Z = 3.90$, $p < 0.001$), and σ was significantly larger in the different frequency condition ($Z = 3.75$, $p < 0.001$). This indicates that discrimination accuracy was lower when the center frequencies differed, as with the long-short rhythm.

Exp 3: Straight Rhythm (1:1)

To examine the discrimination accuracy of straight rhythm, the remaining 30 data in experiment 3 were used in the analysis (**Figure 3**). We calculated the response rate in each interval ratio, compared by frequency conditions, and found that the absolute value of gradient in different frequency conditions was smaller than that in the same frequency condition (**Figure 3a**). Concerning σ of symmetric fitting (**Figure 3b**), Wilcoxon signed rank tests showed that σ was significantly larger in the condition with frequency difference ($Z = 4.78$, $p < 0.001$). This indicates that discrimination accuracy was lower when the center frequencies differed, as with the long-short and short-long rhythms.

Concerning σ of asymmetric fitting (**Figure 3c**), Wilcoxon signed rank tests showed that σ was significantly larger when the interval ratio was negative than positive in the same frequency condition (same frequency condition: $Z = 4.22$, $p < 0.001$; different frequency condition: $Z = 3.18$, $p = 0.001$). After excluding data with σ less than 0.01 ($n =$

27), σ was significantly larger in both frequency conditions (same frequency condition: $Z = 4.22$, $p < 0.001$; different frequency condition: $Z = 3.32$, $p < 0.001$). This suggests that discrimination accuracy is higher in long-short rhythms than in short-long rhythms when the rhythmic pattern is close to the straight rhythm.

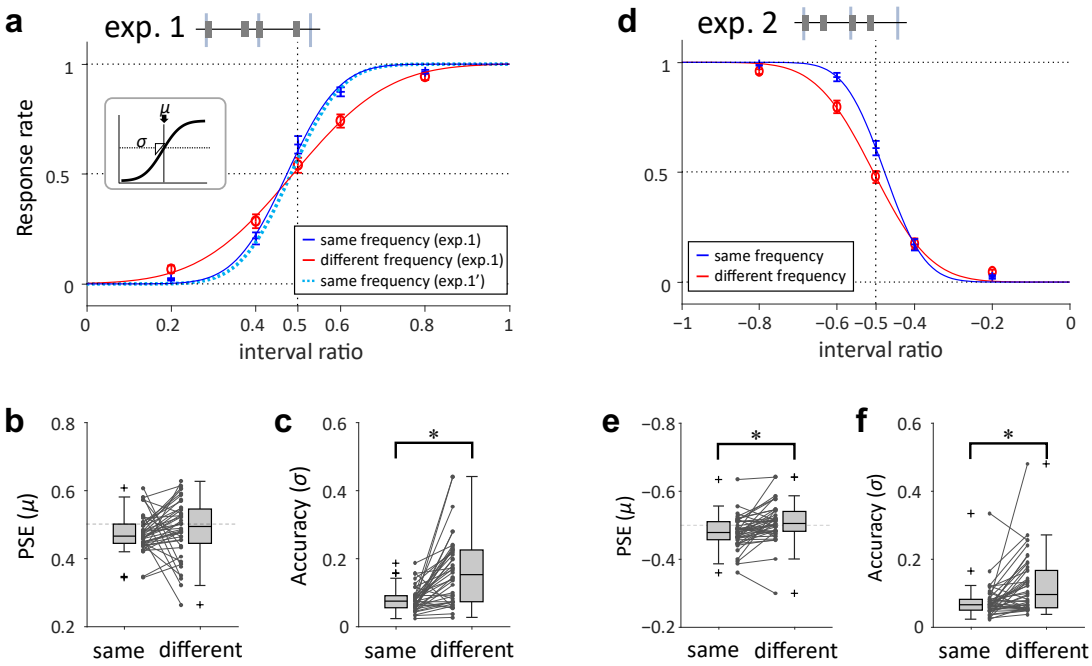

Figure 2. PSE and discrimination accuracy in long-short (Exp 1 and 1', a-c) and short-long rhythm (Exp 2, d-f). (a) Response rates that comparative stimuli were "more bouncing" for each interval ratio. The red line indicates the different frequency condition in experiment 1, and the blue line indicates the same frequency condition in experiment 1. The light-blue dotted line indicates the same frequency condition in experiment 1'. Error bars are standard errors. Curves are sigmoidal approximations. (b) Distribution of PSE (μ) for different and same frequency conditions. (c) Distribution of discrimination accuracy (σ) for different and same frequency conditions. (d) Response rates that the comparative stimuli were "more bouncing" for each interval ratio. The red line indicates the different frequency condition, and the blue line indicates the same frequency condition. Error bars are standard errors. Curves are sigmoidal approximations. (e) Distribution of PSE (μ) for different and same frequency conditions. (f) Distribution of discrimination accuracy (σ) for different and same frequency conditions. * indicates $p < 0.001$.

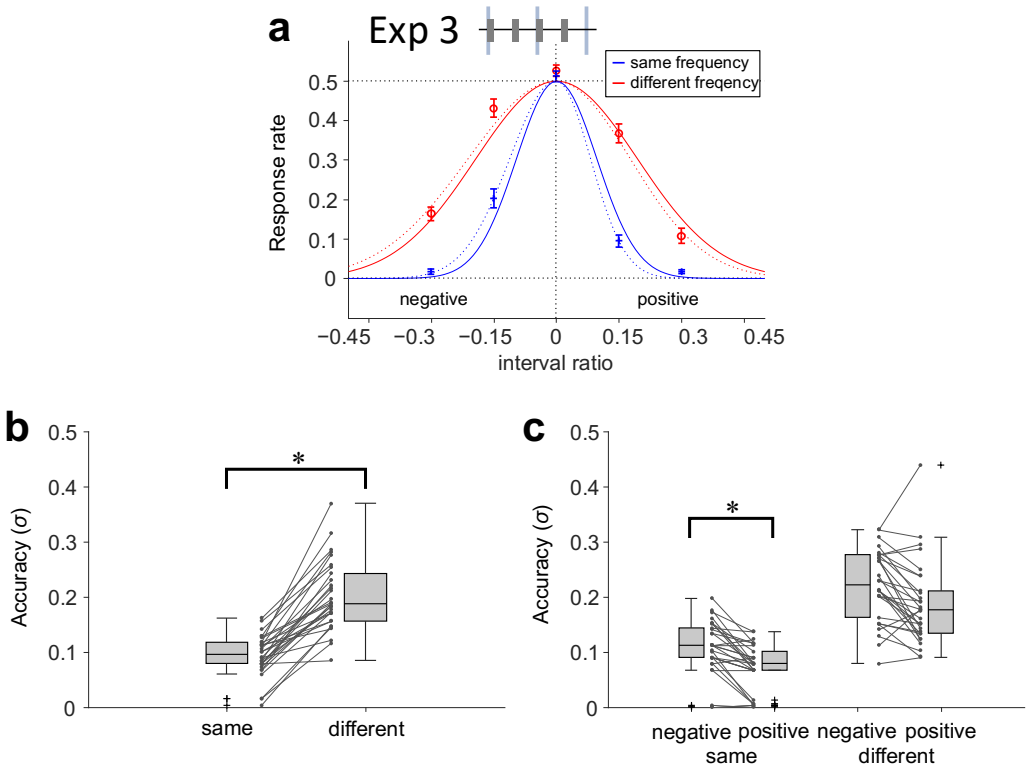

Figure 3. Discrimination accuracy in straight rhythm (Exp 3). (a) Response rates that the comparative stimuli were "more bouncing" for each interval ratio. The blue line indicates the condition with a frequency difference, and the red line indicates the condition without a frequency difference. Error bars are standard errors. Curves drawn by a solid line indicate symmetric fitting. Curves drawn by dotted line indicates asymmetric fitting. (b) Distribution of PSE ($|\mu|$) for different and same frequency conditions. (c) Distribution of accuracy (σ) for different and same frequency conditions.

Discussion

Rhythm perception and its accuracy vary depending on the temporal properties represented by beat subdivision. Timbre also influences how the sound sequence is perceptually organized according to stream segregation studies. However, the effect of differences in timbre on the accuracy of rhythm perception remained unclear. Here, we showed that discrimination accuracies (σ) decreased when center frequencies differed, regardless of the rhythmic pattern (Figures **2c, 2f, and 3b**). This result indicates that the spectral structure of sound sequences influences rhythm perception no matter what ratio the beats are subdivided into.

The effect of timbre on perceptual accuracy

Our analysis revealed that the discrimination accuracy got lower when the center frequency of band noise sequences was different (**Figure 2, 3**). The result may arise because the band noise sequences were perceived as two separate streams when the center frequency differed. This is consistent with the previous studies [18,19]. However, we did not ask the participants to self-report whether or not they perceived sound streams as integrated or segregated. This is one of our limitations. We showed that the discrimination accuracy decreased with different center frequencies, even if the standard rhythmic pattern was not straight (**Figure 2**). The result suggests that timbre affects discrimination accuracy regardless of the rhythmic pattern. The discrimination threshold of swing ratio in straight rhythm is around 10 ms according to the previous study [4,22], which is roughly consistent with the present study. However, discrimination accuracy in rhythmic patterns of various subdivision remained unclear. This is the first study to reveal a decrease in discrimination accuracy regardless of various subdivisions. The findings of the present study are significant in that they link rhythmic perception research, which is

an aspect of the temporal structure of perceptual structuring, and stream segregation research, which is an aspect of spectral timbre.

Asymmetry in perceptual accuracy

We measured the point of subjective equality (PSE) of the interval ratio in experiments 1, 1', and 2. In experiment 3, PSE was not measured because it was assumed to be fixed as 0 in interval ratio. The result showed that PSE ($|\mu|$) shifted closer to 0 in the same frequency condition (**Figure 2b, 2e**). This suggests an asymmetry in perceptual accuracy depending on whether the direction of displacement is positive or negative. This result is consistent with the previous study [14]. The authors showed that subjects detected a negative shift of interval ratio more correctly than a positive shift in 1:2 rhythm, although the inter-beat interval (IBI) in their study was 1080 ms, which was different from our study (300 ms). Interestingly, the shift of PSE occurred only in the same frequency condition. This may be because the encoded rhythmic pattern shifted toward a straight rhythm in the retention phase of short-term memory, resulting in an underestimation of the interval ratio of the initial rhythmic stimulus. When the center frequency is different, retaining the relationship of the time interval may be more difficult. However, the shift of PSE may also be due to the order effect of fixing the standard stimulus to the initial one. Thus, it would be essential to look into the causality in the future.

In experiment 3, there was a significant difference of σ between positive and negative discrimination accuracy (**Figure 3c**). This result may be explained by time-shrinking [23]. Time-shrinking is an auditory illusion when there are three consecutive tones representing two adjacent time intervals (T1, T2 in that order) and the second time interval is longer than the first (T1 < T2). When time-shrinking occurs, T2 is perceived

as shorter robustly. Although the effect was weaker in the repeated presentation condition, there was a tendency for the sound to be perceived as more straight [15], which was consistent with the asymmetry that discrimination accuracy decreases when the direction of displacement is negative, as observed in experiment 3. This suggests that the discrimination accuracy varies depending on the direction of displacement, regardless of the center frequency condition.

The effect of musical experience

To examine the effect of instrumental experience additionally, participants in experiment 1 were divided into two groups (with experiences: $n = 23$; without experiences: $n = 20$) by reporting whether they had instrumental experience, and σ was calculated for each center frequency condition. The results showed a significant difference in σ between the center frequency conditions in both groups (with experiences: $Z = -3.47$, $p < 0.001$; without experiences: $Z = -3.54$, $p < 0.001$). This suggests that discrimination accuracy decreased under center frequency conditions, regardless of musical experience. This may be because no experts were recruited in this study, which would not result in visible differences. However, several studies suggest that musical experience increases the accuracy with which rhythms are perceived. For instance, one study showed that the detection threshold for tempo changes in a group with high musical experience was smaller than that of a group with low musical experience [24]. Furthermore, detection thresholds for phase shifts and tempo changes were also lower in the group of musical experts [25].

A previous study found a relationship between the perceptual accuracy of interval ratio and nationality [26]. For example, the accuracy of American participants

was high only in the 2:1 ratio, while that of Turkish was high in 2:1 and 3:2, which were familiar swing ratios for them. This suggests perceptual accuracy may vary depending on culture-specific listening experiences and acquired musical knowledge. Therefore, it will be essential to investigate in future studies how the properties of rhythm perception, observed in the present study, vary with musical experience and belonging to culture to understand how we acquire rhythm perception.

Within-subjects factors

To investigate whether discrimination accuracy varies within individuals depending on their rhythmic patterns, we compared μ and σ for the duplicate participants in experiments 1 and 2 who had one or fewer incorrect responses in the screening trial (n = 23). For participants who answered all questions correctly in the screening trials, $|\mu|$ was smaller in the same frequency condition (**Figure 4a**), and σ was larger in the differential frequency condition (**Figure 4b**). A Wilcoxon signed rank test showed significant differences between μ and σ in experiment 1(μ: $Z = -2.55$, $p = 0.011$; σ: $Z = -3.22$, $p = 0.007$) and μ in experiment 2 (μ: $Z = -2.68$, $p = 0.007$; σ: $Z = -1.43$, $p = 0.153$). Thus, we can say that for those who performed well in the screening trial, there was a consistent trend in PSE and discrimination accuracy, even when the same subjects participated in different experiments.

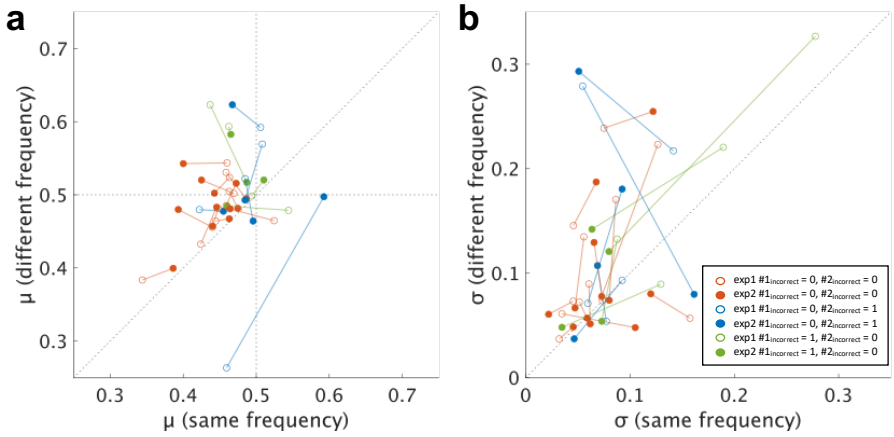

Figure 4. Comparison of μ and σ for overlap participants in Exp 1 (a) and 2 (b). Scatterplots of the absolute value of PSE ($|\mu|$) in the different frequency condition for each $|\mu|$ in the same frequency condition. Open circles indicate the values in experiment 1, and filled circles indicate the values in experiment 2. Red indicates all correct answers to the screening trials, blue indicates one incorrect answer to the screening trials in experiment 1, and green indicates one incorrect answer to the screening trials in experiment 2. Solid lines connect the same participants.

Reliability regarding online experiments

In order to ensure data reliability, we screened experimental data according to their performance on screening trials and R^2 to the sigmoid approximation curve. A closer look at the excluded data because of the R^2 criterion revealed that participants did not respond correctly even at the maximum or minimum interval ratios, which was the most straightforward trial. This suggests that there is a relationship between the criterion of screening trials and the goodness of fit to the approximation curve because we applied the maximum or minimum interval ratio of the same frequency in the screen trials. Failure to respond correctly on the most straightforward trials means that data validity for the

entire trial is unacceptable. Therefore, the criterion applied in this study was appropriate. A possible reason for the failure of correct response in the most straightforward trials could be the limit of rhythm perception and concentration capability.

Still, there were several concerns. First, the acoustic environment could differ for each participant in online experiments. We tested that all participants wore headphones during the experiment. We also minimized the influence of frequency characteristics by reducing the differences in stimulus frequencies. However, it is possible that each participant heard differently by audio devices or background noises, resulting in a random effect on the discrimination accuracy. Such data could have been excluded by screening, but we cannot reveal if they were omitted because of the person's listening ability or the environment. Second, the screening trials were relatively small compared to the overall number. We had to make this choice to obtain as much data as possible in a limited experimental time. Also, the position of the screening trial was fixed at the end of the block. This might influence the performance of screening trials due to a lack of concentration. Finally, the attribution and the performance of the participants might have changed depending on the date and time of the experiment. At least, since experiments 1 and 2 were conducted at the same time, the populations would be relatively similar. Nevertheless, we found a close agreement in the sigmoid approximation curves between experiments 1 and 1'. This suggests that the data from the online experiments are consistent with those from the in-laboratory experiments. Thus, we can say that the essential reliability of the online experiment has been ensured.

Implications for cultural aspects of music

Our study also provides implications for the cultural aspects of music. The findings

suggest that a decrease in perceptual accuracy may also occur in the musical performance since rhythmic patterns usually consist of multiple musical instruments. For example, drummers often use hi-hats and ride cymbals with the same center frequency in jazz music, where the swing ratio is considered an essential factor. The audience may feel more swing by composing rhythmic patterns with such instruments. However, many other parameters are involved in the actual rhythmic pattern, such as intensity, accents, and duration. For more accurate implications, it would be essential to investigate the effects of these factors on the accuracy of rhythm perception.

Our paradigm can also be used to examine the effect of cultural differences on perceptual accuracy. For example, prior probability distributions over integer ratio on musical rhythm differ by cultural experience [27]. This may be due to the varying accuracy of discrimination depending on the cultural experience. A similar online experiment recruiting participants from different cultural backgrounds may provide mechanistic insights into the cultural evolution of music.

Conclusion

We found that differences in center frequency in band noise lower the perceptual accuracy of subdivided rhythm regardless of the rhythmic pattern. This indicates that rhythm perception is affected not only by temporal structure but also by spectral parameters. Our findings bridge the gap between rhythm perception and auditory stream segmentation studies. Our paradigm also has the potential to examine the mechanism of cultural differences in music.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number JP20K21804. We thank Dr. Kazutoshi Kudo for his helpful comments on early manuscripts. Furthermore, we want to express our gratitude to the colleagues from Okanoya Laboratory who provided insight and expertise that greatly assisted the research.

Reference

1. Large EW, Palmer C. Perceiving temporal regularity in music. Cognitive Science. 2002;26(1):1–37.

2. Lerdahl F, Jackendoff RS. A Generative Theory of Tonal Music, reissue, with a new preface. MIT Press; 1983. 390 p.

3. Polak R. Rhythmic Feel as Meter: Non-Isochronous Beat Subdivision in Jembe Music from Mali. Music Theory Online [Internet]. 2010 Dec 1 [cited 2022 Aug 9];16(4). Available from: https://mtosmt.org/issues/mto.10.16.4/mto.10.16.4.polak.html

4. Friberg A, Sundströöm A. Swing Ratios and Ensemble Timing in Jazz Performance: Evidence for a Common Rhythmic Pattern. Music Perception. 2002 Mar 1;19(3):333–49.

5. Repp BH. Patterns of expressive timing in performances of a Beethoven minuet by nineteen famous pianists. The Journal of the Acoustical Society of America. 1990 Aug;88(2):622–41.

6. Palmer C. On the Assignment of Structure in Music Performance. Music Perception: An Interdisciplinary Journal. 1996;14(1):23–56.

7. Drake C, Penel A, Bigand E. Tapping in Time with Mechanically and Expressively Performed Music. Music Perception: An Interdisciplinary Journal. 2000;18(1):1–23.

8. Repp BH. Metrical Subdivision Results in Subjective Slowing of the Beat. Music Perception. 2008 Sep 1;26(1):19–39.

9. Handel S. The effect of tempo and tone duration on rhythm discrimination. Perception & Psychophysics. 1993 May 1;54(3):370–82.

10. Fitch WT, Rosenfeld AJ. Perception and Production of Syncopated Rhythms. Music Perception. 2007 Sep 1;25(1):43–58.

11. Ehrlé N, Samson S. Auditory discrimination of anisochrony: Influence of the tempo and musical backgrounds of listeners. Brain and Cognition. 2005 Jun;58(1):133–47.

12. Povel DJ. Internal representation of simple temporal patterns. Journal of Experimental Psychology: Human Perception and Performance. 1981 Feb 1;7(1):3–18.

13. Repp BH. Sensorimotor synchronization: a review of the tapping literature. Psychonomic Bulletin & Review. 2005 Dec 1;12(6):969–92.

14. Repp BH, London J, Keller PE. Perception–production relationships and phase correction in synchronization with two-interval rhythms. Psychological Research-psychologische Forschung. 2011 May 1;75(3):227–42.

15. Hasuo E, Arao H. Certain non-isochronous sound trains are perceived as more isochronous when they start on beat. Atten Percept Psychophys. 2020 May;82(4):1548–57.

16. Van Noorden LPAS. Temporal coherence in the perception of tone sequences. PhD thesis, Eindhoven University of Technology [Internet]. 1975 [cited 2022 Jan 6]; Available from: https://ci.nii.ac.jp/naid/10012929296/

17. Bregman AS. Auditory Scene Analysis: The Perceptual Organization of Sound. MIT Press; 1990. 800 p.

18. Bregman AS, Ahad PA, Van Loon C. Stream segregation of narrow-band noise bursts. Perception & Psychophysics. 2001 Jul 1;63(5):790–7.

19. Cusack R, Roberts B. Effects of differences in timbre on sequential grouping. Perception & Psychophysics. 2000 Jul 1;62(5):1112–20.

20. Henninger F, Shevchenko Y, Mertens UK, Kieslich PJ, Hilbig BE. lab.js: A free, open, online study builder. Behav Res. 2022 Apr;54(2):556–73.

21. Woods KJP, Siegel MH, Traer J, McDermott JH. Headphone screening to facilitate web-based auditory experiments. Atten Percept Psychophys. 2017 Oct 1;79(7):2064–72.

22. Frane AV, Shams L. Effects of tempo, swing density, and listener's drumming experience, on swing detection thresholds for drum rhythms. The Journal of the Acoustical Society of America.

2017 Jun;141(6):4200–8.

23. Nakajima Y, ten Hoopen G, Sasaki T, Yamamoto K, Kadota M, Simons M, et al. Time-Shrinking: The Process of Unilateral Temporal Assimilation. Perception. 2004 Sep;33(9):1061–79.

24. Ellis MC. Research Note. Thresholds for Detecting Tempo Change. Psychology of Music. 1991 Oct 1;19(2):164–9.

25. Repp BH. Sensorimotor synchronization and perception of timing: effects of music training and task experience. Human Movement Science. 2010 Apr 1;29(2):200–13.

26. Hannon EE, Soley G, Ullal S. Familiarity overrides complexity in rhythm perception: A cross-cultural comparison of American and Turkish listeners. Journal of Experimental Psychology: Human Perception and Performance. 2012;38(3):543–8.

27. Jacoby N, McDermott JH. Integer Ratio Priors on Musical Rhythm Revealed Cross-culturally by Iterated Reproduction. Current Biology. 2017 Feb 6;27(3):359–70.