SHORT REPORT

Certain non-isochronous sound trains are perceived as more isochronous when they start on beat

Emi Hasuo[1,2] · Hiroshi Arao[1]

Abstract

Perceiving the duration of neighboring time intervals is vital for rhythm perception. We discovered a phenomenon in which the perceived equality/inequality of neighboring time intervals in a sound sequence is changed by its metrical interpretation. The target sound sequence consisted of eight short sound bursts marking seven neighboring time intervals, which were repetitions of two durations (T1 and T2) presented in alternation (T1-T2-T1-T2 …). There were three tempos, which corresponded to T1 + T2 being 210, 420, and 630 ms. The physical difference between T1 and T2 (T1 – T2) was varied systematically for each tempo in the ranges of −100 to 100 ms (when T1 + T2 was 210 or 420 ms) or −150 to 150 ms (when T1 + T2 was 630 ms). Participants reported the level of perceived equality/inequality of these neighboring time intervals. For each target sequence, four isochronous lower-pitched preceding sounds were added at different phases so that the beginning of either T1 (Beat-on-T1 condition) or T2 (Beat-on-T2 condition) coincided with the beat induced by these preceding sounds. When T2 was longer than T1 by up to 60 ms, the neighboring time intervals of the same target sequence were perceived as more "equal" in the Beat-on-T1 condition compared with the Beat-on-T2 condition. Such a difference in the perceived equality/inequality appeared significantly only at the intermediate tempo of T1 + T2 = 420 ms. The difference in equality/inequality perception at limited temporal conditions could be accounted for by the occurrence of an illusion in time perception called time-shrinking.

Keywords Rhythm perception · Metrical interpretation · Time-shrinking

Introduction

Perceiving the relative durations of neighboring time intervals is the basis for rhythm perception and is of vital importance for auditory communication (e.g., Grondin et al., 2018; Handel, 1989). In the current work, we report a phenomenon in which perceived differences in the equality/inequality of the neighboring time interval durations arise from a physically identical sound sequence.

The phenomenon we report here is based on an illusion in time perception called time-shrinking (Nakajima et al., 1991). Time-shrinking typically occurs when there are three successive sounds marking two neighboring time intervals (T1 and

T2, in this order) and the second time interval is longer than the first time interval (T1 < T2). When time-shrinking occurs, the *subjective* duration of T2 is shortened remarkably. This phenomenon is very robust (e.g., Remijn et al., 1999) and is considered to be an example of *temporal assimilation* (e.g., Miyauchi & Nakajima, 2005; Nakajima et al., 2004). Time-shrinking has been studied mainly in the auditory modality, but has also been reported in visual (Arao et al., 2000) and tactile (Hasuo et al., 2014; van Erp & Spapé, 2008) modalities, suggesting that temporal assimilation is a universal phenomenon across modalities. In the auditory modality, time-shrinking occurs in a relatively limited time range: when T1 is 200 ms or shorter, and when the difference between T1 and T2 is below about 100 ms (e.g., Nakajima et al., 2004). Such short time intervals of a few hundred milliseconds appear frequently in music (e.g., Fraisse, 1982) and in speech (e.g., Greenberg & Arai, 2004).

Although many studies investigated the time-shrinking illusion and temporal assimilation (see ten Hoopen et al., 2008, for a review), most have focused on situations where the time-shrinking pattern was presented alone (without preceding or

✉ Emi Hasuo
emi8513@gmail.com

[1] Department of Human Sciences, Taisho University, 3-20-1
Nishi-Sugamo, Toshima-ku, Tokyo 170-8470, Japan

[2] Japan Society for the Promotion of Science, Tokyo, Japan

succeeding sounds). In a few studies, the number of preceding time intervals was increased (Remijn et al., 1999; Sasaki et al., 2002; ten Hoopen et al., 1995), and results showed that time-shrinking in the last time interval can take place even when the number of preceding time intervals increased (Remijn et al., 1999; ten Hoopen et al., 1995), although the duration combinations of these preceding time intervals seemed to influence the occurrence of time-shrinking (Sasaki et al., 2002). All of these studies focused only on the perceived duration of the last time interval in a sound sequence. In natural situations, however, it is also important to perceive the relative durations of neighboring time intervals (e.g., Grondin et al., 2018; Miyauchi & Nakajima, 2007). In addition, in real music, a rhythm pattern usually appears in a sequence of other sounds, which induce the feeling of beat and meter (e.g., Russo & Ammirante, 2018). Therefore, it seemed important to examine the perceived durations of adjacent time intervals within the context of other sounds.

As a casual first step, we made audio demonstrations of the time-shrinking pattern (e.g., T1 = 200 ms and T2 = 260 ms) presented repeatedly (T1-T2-T1-T2 …). When listening to such patterns, we noticed that the durations of the adjacent time intervals in a sound sequence sounded *almost equal* (isochronous) sometimes, while those in the *same* sequence sounded *clearly unequal* (non-isochronous) at other times. In other words, there seemed to be two possible rhythm percepts from the physically identical sound sequence. The occurrence of the two rhythm percepts appeared to be based on the metrical interpretation of the sound sequence, i.e., whether the sequence was perceived as cycles of "T1-T2" or as cycles of "T2-T1."

The purpose of the present study was to demonstrate the occurrence of the two equality/inequality percepts for a physically identical sound sequence, and to clarify the time conditions in which the two percepts occur. Because time-shrinking occurs in a limited time range, it was possible that the two equality/inequality percepts also occurred in a limited range. Thus, it was important to conduct a systematic experiment in which different tempos were employed and several steps of physical differences between the neighboring durations (T1 – T2 durations) were set for each tempo. Here, we employed three base tempos, corresponding to the total duration of T1 and T2 (T1 + T2) being 210, 420, and 630 ms. The faster two tempos (210 and 420 ms) were those in which time-shrinking was predicted to occur while the slowest tempo (630 ms) was not (see the *Stimuli and apparatus* section of Experiment 1 for details). To make it easier for the two rhythm percepts to occur naturally, we added lower-pitched preceding sounds, which had a regular beat, before the target rhythm sequence. By shifting the timing of the preceding sounds, the beat of those sounds would cause the same target sequence to be perceived as either cycles of "T1-T2" or as cycles of "T2-T1" (e.g., Repp et al.,

2008). Comparing the perceived equality/inequality of the neighboring time intervals' durations between the two preceding-sound-timings for the same target sequence would allow us to examine whether the two different equality/ inequality percepts occurred.

In the present paper, we report two experiments. Experiment 1 was our main experiment in which the occurrence of the two equality/inequality percepts was examined as described above, with a four-point rating task. Experiment 2 was designed as a follow-up experiment to relate our task and data to previous time-shrinking/temporal assimilation studies (e.g., Miyauchi & Nakajima, 2007) in which a different task was employed, and to provide data of a control condition with no preceding sounds.

Experiment 1

Method

Participants

Twelve participants (five females and seven males with a mean age of 24 years; range: 20–35 years) took part in the experiment. None had hearing deficits. All provided written informed consent, and all were allowed to drop out of the experiment at any time. The experiment was approved by the Ethics Committee of Taisho University.

Stimuli and apparatus

Each presentation consisted of four preceding sounds and eight target sounds (Fig. 1). All sounds were 20 ms in duration, consisting of rise and fall times of 10 ms that were raised-cosine shaped. The preceding sounds were pure-tone bursts of 500 Hz and the target sounds were pure-tone bursts of 1,000 Hz. Lower-pitched sounds were employed as the preceding sounds to make the higher-pitched target sounds clearly distinguishable from the preceding sounds (e.g., Moore, 2003). The peak amplitudes of the preceding sounds and the target sounds were kept constant, and the 1,000-Hz target sounds were presented at 80 dB, measured as the level of a continuous tone of the same amplitude.

The three time intervals between the four preceding sounds were equal in duration. These preceding sounds were employed to induce a regular beat. There were three tempos, which corresponded to the time intervals between the four preceding sounds being 210, 420, or 630 ms. Note that by *time interval* in the present study, we refer to the temporal distance between the onsets of two successive sounds (i.e., inter-onset interval). For the eight target sounds, there were two time intervals (T1 and T2, in this order), which were presented in alternation (T1-T2-T1-T2 …). The total duration
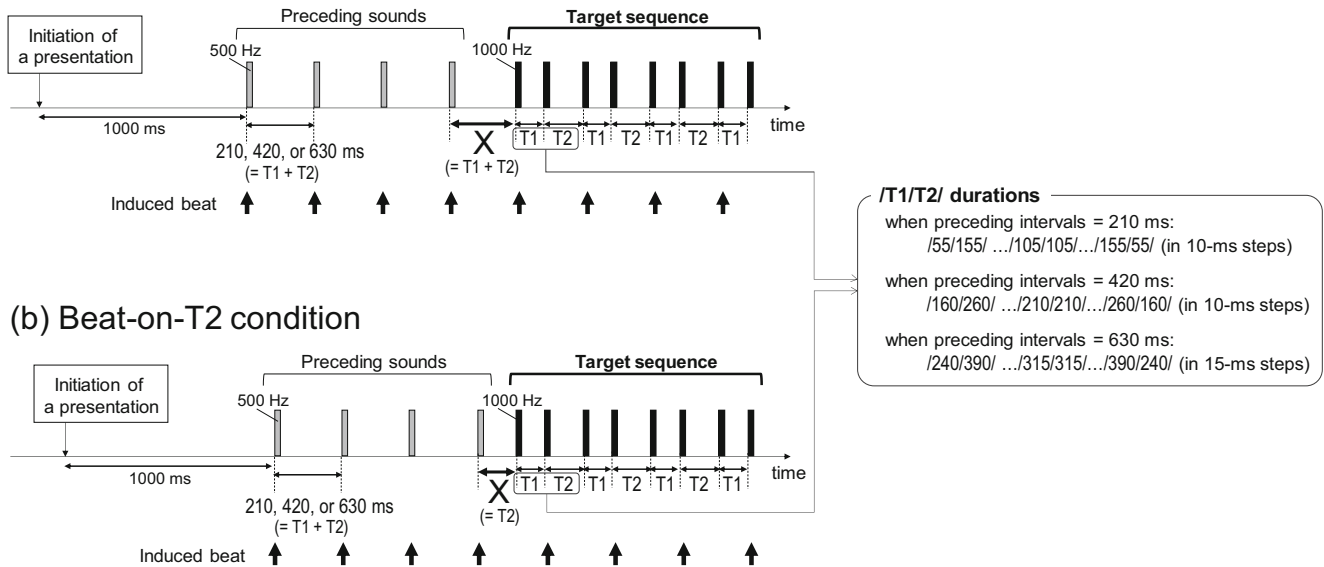
(a) Beat-on-T1 condition

Initiation of a presentation

Preceding sounds

Target sequence

500 Hz

1000 Hz

1000 ms

210, 420, or 630 ms
(= T1 + T2)

X
(= T1 + T2)

T1 T2 T1 T2 T1 T2 T1

time

Induced beat

(b) Beat-on-T2 condition

Initiation of a presentation

Preceding sounds

Target sequence

500 Hz

1000 Hz

1000 ms

210, 420, or 630 ms
(= T1 + T2)

X
(= T2)

T1 T2 T1 T2 T1 T2 T1

time

Induced beat

/T1/T2/ durations
when preceding intervals = 210 ms:
/55/155/ .../105/105/.../155/55/ (in 10-ms steps)

when preceding intervals = 420 ms:
/160/260/ .../210/210/.../260/160/ (in 10-ms steps)

when preceding intervals = 630 ms:
/240/390/ .../315/315/.../390/240/ (in 15-ms steps)

Fig. 1 Time charts of stimulus patterns in Experiment 1. The participant initiated each presentation by pressing a button on the computer screen. The stimulus sounds in the Beat-on-T1 condition (**a**) and the Beat-on-T2 condition (**b**) were identical except for the duration of the time interval (X) between the preceding sounds and the target sequence. In the Beat-on-T1 condition, X was equal to the time intervals between the preceding sounds; thus, the beginning of T1 in the target sequence coincided with the beat suggested by the preceding sounds. In the Beat-on-T2 condition, X was equal to the duration of T2; thus, the beginning of T2 in the target sequence coincided with the beat suggested by the preceding sounds. The task for the participants was to judge how equal/unequal the durations of the neighboring time intervals in the target sequences were.

of T1 and T2 (T1 + T2) was made equal to the time interval between the four preceding sounds (i.e., 210, 420, or 630 ms). Hereafter, we refer to temporal patterns of T1 and T2 as /T1/T2/ (with numbers representing the duration of each time interval in milliseconds). When the preceding time intervals were 210 ms, T1 and T2 were varied from /55/155/ to /155/55/ in steps of 10 ms (i.e., /55/155/, /65/145/, 75/135/, /85/125/, /95/115/, /105/105/, /115/95/, /125/85/, /135/75/, /145/65/, /155/55/). When the preceding intervals were 420 ms, T1 and T2 were varied from /160/260/ to /260/160/ in 10-ms steps. When the preceding intervals were 630 ms, T1 and T2 were varied from /240/390/ to /390/240/ in 15-ms steps. Note that lengthening T1 automatically meant that T2 was shortened to keep T1 + T2 constant (i.e., 210, 420, or 630 ms), and that the above /T1/T2/ patterns corresponded to the difference between T1 and T2 (T1 – T2) being −100 to 100 ms in 20-ms steps (when T1 + T2 = 210 or 420 ms) or −150 to 150 ms in 30-ms steps (when T1 + T2 = 630 ms). There were 11 /T1/T2/ patterns for each of the three tempos. The faster two tempos (i.e., T1 + T2 = 210 and 420 ms) were the tempos in which time-shrinking would occur because the shorter of the two time intervals (T1 or T2) was 200 ms or shorter at these tempos (e.g., Nakajima et al., 2004).

There were two main conditions regarding the timing of the beginning of eight target sounds: the Beat-on-T1 condition and the Beat-on-T2 condition. In the Beat-on-T1 condition, the beginning of T1 coincided with the beat suggested by the preceding sounds. In other words, the time interval between the preceding sounds and the target sounds (indicated as "X" in Fig. 1a) was equal to the duration of the preceding time intervals. In the Beat-on-T2 condition, the beginning of T2 coincided with the preceding beat. In other words, the time interval between the preceding sounds and the target sounds (indicated as "X" in Fig. 1b) was equal to the duration of T2. In total, there were 66 experimental conditions [11 (/T1/T2/ patterns) × 3 (tempos) × 2 (Beat-on-T1/Beat-on-T2 conditions)].

Sound signals were generated digitally (16 bits; sampling frequency of 44,100 Hz) on a computer (Panasonic, Let's note, CF-RZ6), and were presented diotically via headphones (Beyerdynamic DT1770 Pro) in a quiet room. Sound levels were measured with a sound-level meter (Custom, SL-1320).

Procedure

In each trial, the participant initiated the presentation of stimuli by clicking the "play" button on the computer screen. The task for the participants was to judge how equal (or unequal) the durations of the neighboring time intervals in each target sequence were, and to choose a response from four response choices: (1) The neighboring time intervals sounded clearly equal and could not be heard as unequal ("Clearly Equal"), (2) The neighboring time intervals sounded almost equal but could also be heard as slightly unequal ("Almost Equal"), (3) The neighboring time intervals sounded a little unequal but

could also be heard as close to equal ("A Little Unequal"), and (4) The neighboring time intervals sounded clearly unequal and could not be heard as equal ("Clearly Unequal"). Although the task could be carried out without listening to the preceding sounds (and focusing only on the target sounds), participants were instructed not to intentionally ignore the preceding sounds. Participants were generally allowed to listen to each stimulus pattern only once before choosing a response, but were allowed to click the "play" button and listen to the sounds again when they could not hear the sounds for any incidental reason (e.g., coughing). After choosing a response, the participant clicked the "next" button on the computer screen and moved on to the next trial.

The four-point rating method used in this study provided two additional response choices compared to the two-alternative forced choice task used in previous studies on temporal assimilation (e.g., Miyauchi & Nakajima, 2007). The use of the four-point rating was suitable for the present experiment because it can capture subtle differences in perceived equality/inequality. Its correspondence to the two-alternative forced choice method as used in the previous studies was examined in Experiment 2.

There were three experimental blocks, one block for each tempo (i.e., T1 + T2 = 210, 420, and 630 ms). Each block consisted of three sub-blocks: one practice sub-block (six trials) and two experimental sub-blocks (22 trials each). The practice sub-block was employed to familiarize the participants with the task, and six of the 22 [11 (/T1/T2/ patterns) × 2 (Beat-on-T1/Beat-on-T2 conditions)] conditions for that tempo were chosen pseudo-randomly[1] for each participant. In each experimental sub-block, the stimuli of the 22 conditions were presented in random order. Only the responses obtained in the experimental sub-blocks were used for analysis. Participants were free to take breaks between blocks, and each participant completed the experiment (all three blocks) in about 30 min. The order of the three blocks was counterbalanced between participants (there were six possible orders).

The responses from participants were converted to numerical values as "Clearly Equal" = 0, "Almost Equal" = 1, "A Little Unequal" = 2, and "Clearly Unequal" = 3, and these values were used for further analysis. For each stimulus condition, two responses were obtained from each participant (one from each of the two experimental sub-blocks), and these two response values were averaged to determine the response value for each participant.

Results and discussion

For each of the 66 stimulus conditions [11 (/T1/T2/ patterns) × 3 (T1 + T2 durations/tempos) × 2 (Beat-on-T1/Beat-on-T2 conditions)], 12 response values were obtained from the 12 participants. Figure 2 shows the average of these response values. Generally, the response values were larger (meaning a greater perceived inequality) when the physical difference between T1 and T2 was large, as was expected naturally. When focusing on the difference between the Beat-on-T1 condition (black bars) and the Beat-on-T2 condition (gray bars), the response values tended to be smaller in the Beat-on-T1 condition (than the Beat-on-T2 condition) when T1 was shorter than T2 (thus, T1 − T2 being a negative value; left half of the graphs in Fig. 2). An opposite tendency, i.e., smaller response values for the Beat-on-T2 condition (than the Beat-on-T1 condition), was seen when T1 was longer than T2 (thus, T1 − T2 being a positive value; right half of the graphs in Fig. 2). Such tendencies were clearest when T1 + T2 = 420 ms (Fig. 2b).

A two-way (/T1/T2/ patterns × beat conditions) ANOVA using the response values was conducted separately for each tempo (T1 + T2 durations). Because it was natural that the perceived equality/inequality (thus, the response values) changed as the /T1/T2/ patterns (i.e., physical difference between T1 and T2 durations) changed, we will not detail the main effects for this factor (its main effect was significant in all conditions, $p < .001$). For T1 + T2 = 210 ms, the main effect of beat conditions and the interaction were both non-significant, $F(1, 11) = .576$, $p = .464$, $\eta_p^2 = .050$, and $F(10, 110) = 1.355$, $p = .211$, $\eta_p^2 = .110$, respectively. For T1 + T2 = 420 ms, the main effect of beat conditions was not significant, $F(1, 11) = 4.162$, $p = .066$, $\eta_p^2 = .275$, but the interaction between the /T1/T2/ patterns and beat conditions was significant,[2] $F(10, 110) = 3.831$, $p < .001$, $\eta_p^2 = .258$. A follow-up simple main-effect test revealed a significant effect of beat conditions when the /T1/T2/ pattern was /180/240/ (i.e., T1–T2 = −60 ms; $p = .021$), /190/230/ (i.e., T1–T2 = −40 ms; $p < .001$), and /200/220/ (i.e., T1–T2 = −20 ms; $p = .034$), and a marginally significant trend when /240/180/ (i.e., T1–T2 = 60 ms; $p = .054$). For T1 + T2 = 630 ms, neither the main effect of beat conditions nor the interaction were significant, $F(1, 11) = .989$, $p = .341$, $\eta_p^2 = .082$, and $F(10, 110) = 1.497$, $p = .150$, $\eta_p^2 = .120$, respectively.

Summarizing the results of the statistical tests above, perceived equality/inequality of the target sequence was significantly influenced by the beat of the preceding sounds when T1

---

[1] In the six practice trials, three trials were chosen from the Beat-on-T1 condition, and three from the Beat-on-T2 condition.

[2] The interaction was also significant in the Greenhouse–Geisser and Huynh–Feldt results, $p = .014$ and $p = .004$, respectively, which were consulted because sphericity could not be assumed for the interaction when T1 + T2 = 420 ms.
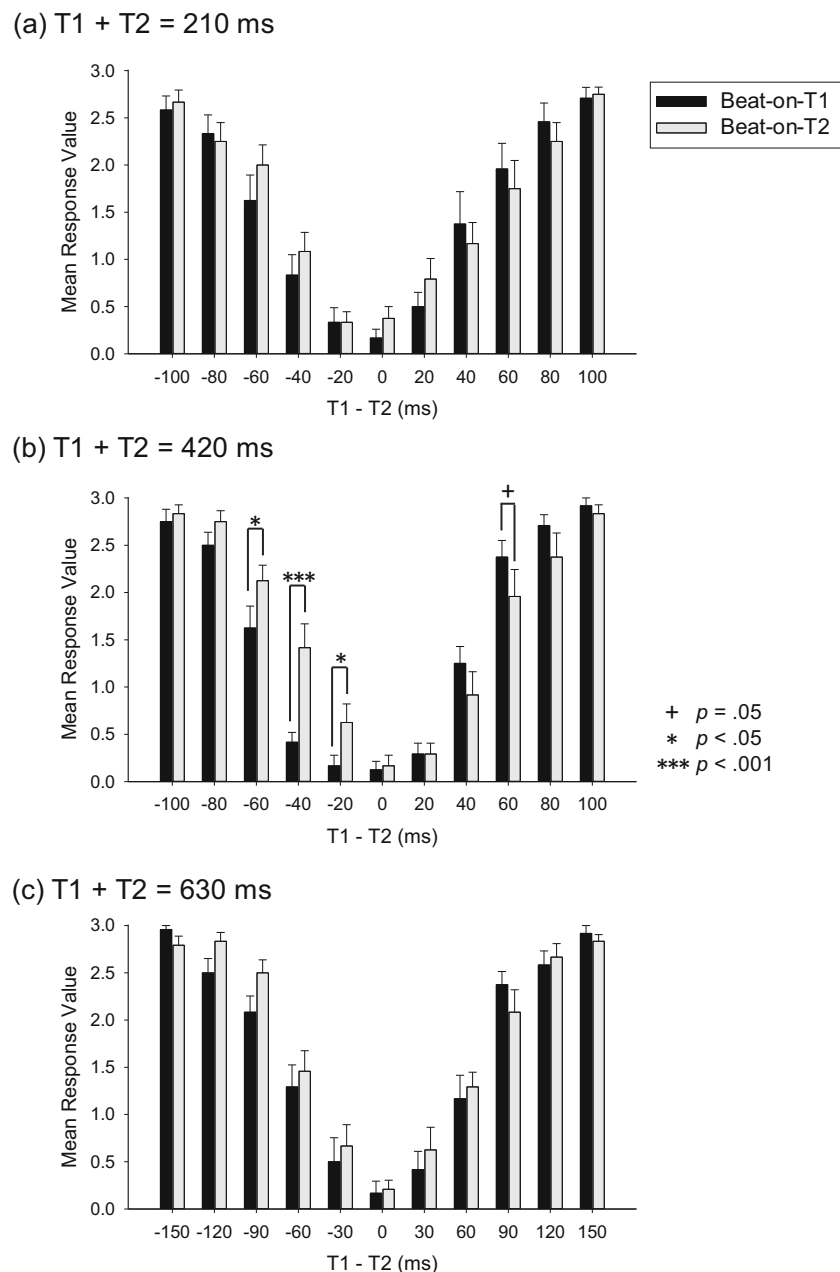
Fig. 2 Mean response (converted to numerical values) for each stimulus pattern in Experiment 1. Larger response values indicate greater perceptual inequality. Different /T1/T2/ patterns are expressed along the horizontal axes as the physical difference between T1 and T2 (T1 – T2).

Note that the range of values of the horizontal axis are different for T1 + T2 = 630 ms (**c**) from the other two tempos (**a** and **b**). Error bars represent standard errors. The results of the simple main-effect test, following the significant interaction of the ANOVA, are indicated with asterisks

+ T2 = 420 ms. At this tempo, when T2 was a little longer than T1 (by up to 60 ms), the same target sequence was perceived as more "equal" when T1 started on beat to the preceding sounds compared with when T2 started on beat. This effect was clearest when T2 was 40 ms longer than T1 (T1 – T2 = −40 ms). The results clearly demonstrated that the metrical interpretation of a sound sequence can influence the perception of equality/inequality of neighboring time intervals.

Experiment 2

Experiment 1 differed from previous studies on temporal assimilation (e.g., Miyauchi & Nakajima, 2007) in a few ways: (1) Four-point rating was used instead of a two-alternative forced-choice task to measure the perceived equality/inequality of the time intervals, (2) the T1-T2 pattern was presented repeatedly instead of only once, and (3) preceding sounds were employed

to induce the feeling of beat. To examine the effects of these discrepancies, Experiment 2 was conducted with two main conditions: Single-presentation condition and Repetitive-presentation condition. In the Single-presentation condition, the T1-T2 pattern was presented only once, as in typical studies on time-shrinking and temporal assimilation (e.g., Nakajima et al., 2004; Miyauchi & Nakajima, 2007). In the Repetitive-presentation condition, the same target sequence as in Experiment 1 (eight sounds marking T1 and T2 in alternation) was employed. In both conditions, there were no preceding sounds, and the four-point rating was used to measure the perceived equality/inequality of the time intervals, as in Experiment 1. Thus, these conditions may be considered as a kind of control condition for Experiment 1. By comparing the Single-presentation with the previous studies on temporal assimilation (e.g., Miyauchi & Nakajima, 2007), the validity of the four-point rating in capturing the characteristics of temporal assimilation can be examined (corresponding to (1) above). By comparing the Single-presentation condition and the Repetitive-presentation condition, the effects of simply repeating the T1-T2 pattern can be examined (corresponding to (2) above). Finally, by comparing the Repetitive-presentation condition and the Beat-on-T1 condition and the Beat-on-T2 condition in Experiment 1, the effect of adding the preceding sounds can be examined (corresponding to (3) above).

Method

Participants

Six participants (two females and four males, all of whom had participated in Experiment 1) took part in Experiment 2. All provided written informed consent, and all were allowed to drop out of the experiment at any time. The experiment was approved by the Ethics Committee of Taisho University.

Stimuli and apparatus

Each target sound was the same as those in Experiment 1. In the Single-presentation condition, three target sounds marked T1 and T2 only once in this order. In the Repetitive-presentation condition, eight target sounds marked T1 and T2 in alternation (T1-T2-T1-T2…), as in Experiment 1. No preceding sounds were presented.

The durations of T1 and T2 were the same as in Experiment 1. Thus, there were 66 conditions [11 (/T1/T2/ patterns) × 3 (tempos) × 2 (Single-presentation/Repetitive-presentation)]. The apparatus was the same as that in Experiment 1.

Procedure

The procedure was the same as in Experiment 1. There were three experimental blocks for each of the Single-presentation and Repetitive-presentation conditions, one block for each tempo (i.e., T1 + T2 = 210, 420, and 630 ms). Each block consisted of three sub-blocks: one practice sub-block (six trials) and two experimental sub-blocks (11 trials each). The practice trials were chosen randomly for each participant from the 11 /T1/T2/ patterns for that tempo. In each experimental sub-block, the stimuli of the 11 conditions were presented in random order. Only the responses obtained in the experimental sub-blocks were used for analysis. Each participant completed the experiment (all six blocks) in about 30 min. The order of the Single-presentation and Repetitive-presentation conditions, as well as the order of the three blocks within each condition, were counterbalanced between participants.

The responses were converted to numerical values in the same way as in Experiment 1.

Results and discussion

For each of the 66 stimulus conditions [11 (/T1/T2/ patterns) × 3 (T1 + T2 durations/tempos) × 2 (Single-presentation/Repetitive-presentation)], six response values were obtained from the six participants. Figure 3a, c, and e show the average of these response values.

The results for the Single-presentation condition showed a clearly asymmetric tendency, i.e., T1 and T2 were perceived to be more "equal" when T1 < T2 (see the dark gray bars in the left half of Fig. 3a, c, and e) compared to when T1 > T2 (see the same bars in the right half of Fig. 3a, c, and e). This asymmetry reflects the occurrence of time-shrinking (e.g., Hasuo et al., 2011; Miyauchi & Nakajima, 2007). The asymmetric tendency was clearer when T1 + T2 = 210 and 420 ms compared to when T1 + T2 = 630 ms. These results were consistent with Miyauchi and Nakajima (2007), indicating that the four-point rating method is capable of capturing the characteristics of temporal assimilation.

As for the effects of repeating the T1-T2 pattern, comparison between the Single-presentation condition and the Repetitive-presentation condition showed that simply repeating the /T1-T2 pattern diminished the asymmetric temporal assimilation (i.e., T1 and T2 were generally correctly perceived to be more "unequal" when T1 ≠ T2, and the asymmetric tendency in the Single-presentation condition was not clear in the Repetitive presentation condition; Fig. 3a, c, and e).[3]

A two-way (/T1/T2/ patterns × repetitions) ANOVA using the response values was conducted separately for each tempo (T1 + T2 durations). As in Experiment 1, we will not detail the main effects for /T1/T2/ patterns (its main effect was

---

[3] In the Repetitive-presentation condition, there were a few reports from the participants that the perceived equality/inequality changed within the eight target sounds in a trial. Such change in equality/inequality perception seemed to have occurred mostly in the T1 + T2 = 420 ms tempo.
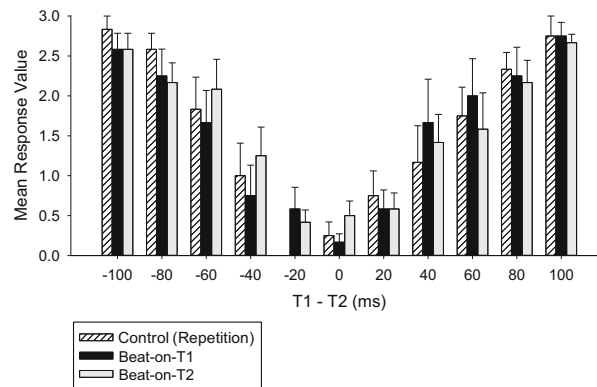
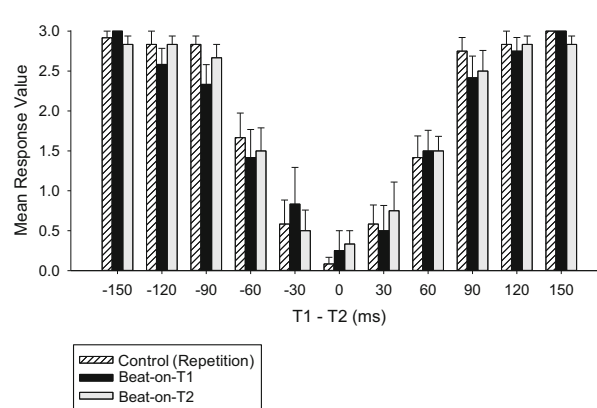# T1 + T2 = 210 ms
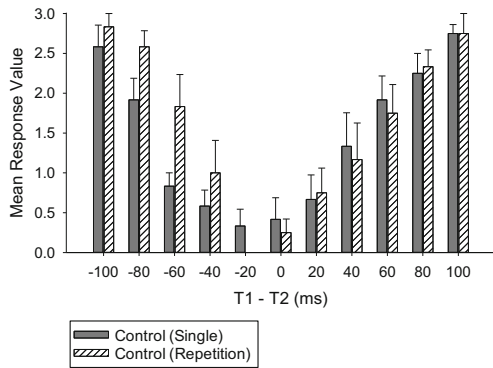
T1 + T2 = 420 ms
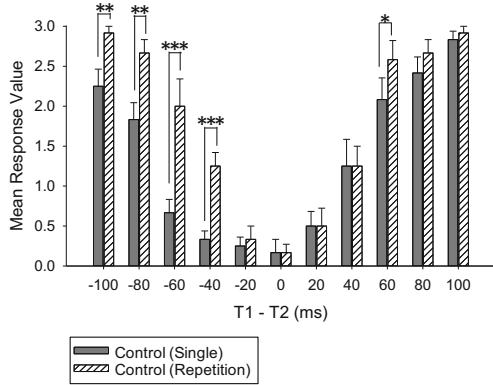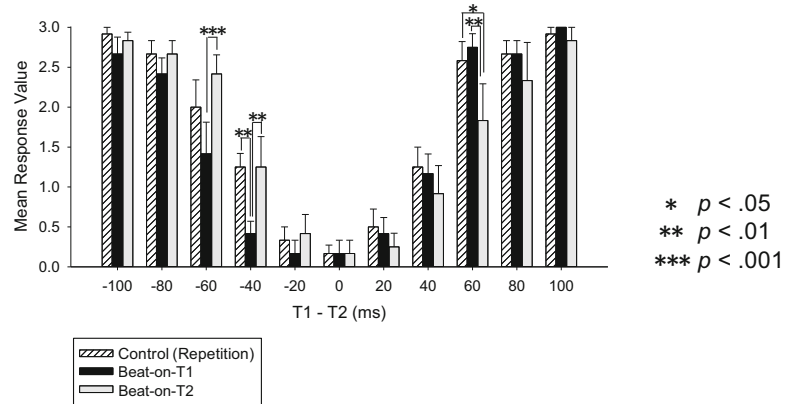
T1 + T2 = 630 ms
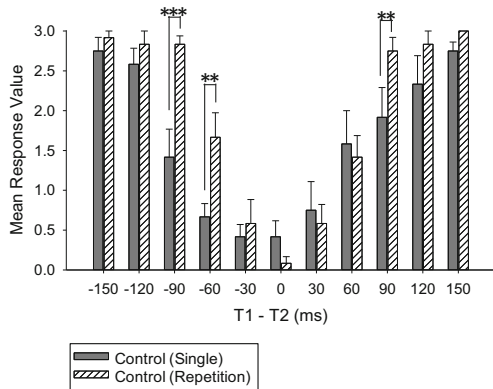
Fig. 3 Mean response (converted to numerical values) for each stimulus pattern in Experiment 2 are shown in **a**, **c**, and **e**. In **b**, **d**, and **f**, the data for the Repetitive-presentation condition are plotted with the data obtained in Experiment 1. Note that the data for the Beat-on-T1 and Beat-on-T2 conditions in **b**, **d**, and **f** are different from those in Fig. 2, because the data of only the six participants that took part in Experiment 2 are shown in the present figure. Error bars represent standard errors. The results of the simple main-effect test are indicated in **c** and **e**, and those of the Ryan's post hoc test are indicated in **d** with asterisks

significant in all conditions, $p < .001$). For T1 + T2 = 210 ms, the main effect of repetitions and the interaction were both non-significant, $F (1, 5) = 2.841$, $p = .153$, $\eta_p^2 = .362$, and $F (10, 50) = 1.753$, $p = .095$, $\eta_p^2 = .260$, respectively. For T1 + T2 = 420 ms, the main effect of repetitions and the interaction were both significant, $F (1, 5) = 36.636$, $p = .002$, $\eta_p^2 = .880$, and $F (10, 50) = 3.631$, $p = .001$, $\eta_p^2 = .421$, respectively. A follow-up simple main-effect test revealed a significant effect of repetitions when the /T1/T2/ pattern was /160/260/ (i.e., T1–T2 = −100 ms; $p = .008$), /170/250/ (i.e., T1–T2 = −80 ms; $p = .001$), /180/240/ (i.e., T1–T2 = −60 ms; $p < .001$), /190/230/ (i.e., T1–T2 = −40 ms; $p < .001$), and /240/180/ (i.e., T1–T2 = 60 ms; $p = .042$). For T1 + T2 = 630 ms, the main effect of repetitions and the interaction were both significant, $F (1, 5) = 7.004$, $p = .046$, $\eta_p^2 = .583$, and $F (10, 50) = 3.963$, $p < .001$, $\eta_p^2 = .442$, respectively. A follow-up simple main-effect test revealed a significant effect of repetitions when the /T1/T2/ pattern was /270/360/ (i.e., T1–T2 = −90 ms; $p < .001$), /285/345/ (i.e., T1–T2 = −60 ms; $p = .001$), and /360/270/ (i.e., T1–T2 = 90 ms; $p = .006$).

The increased sensitivity for the Repetitive-presentation condition compared to the Single-presentation condition would be consistent with the Multiple-look model (e.g., Drake & Botte, 1993; Miller & McAuley, 2005) in that increasing the number of time intervals improved temporal sensitivity, although we dealt with perceived durations of neighboring time intervals whereas the studies on the Multiple-look model dealt with perceived tempo of isochronous sound sequences.

To examine the effects of adding the preceding sounds, we plotted the data of the Repetitive-presentation condition with the two conditions in Experiment 1 (Fig. 3b, d, and f). When T1 + T2 = 420 ms, adding the preceding sounds caused T1 and T2 to be perceived as more "equal" at some /T1/T2/ conditions (Fig. 3d): in the Beat-on-T1 condition when T1 – T2 = −40 ms, and in the Beat-on-T2 condition when T1 – T2 = 60 ms. This result was supported by a two-way (/T1/T2/ patterns × beat conditions) ANOVA using the response values, conducted separately for each tempo (T1 + T2 durations). For T1 + T2 = 210 ms, the main effect of beat conditions and the interaction were both non-significant, $F (2, 10) = .011$, $p = .990$, $\eta_p^2 = .002$, and $F (20, 100) = .907$, $p = .579$, $\eta_p^2 = .154$, respectively. For T1 + T2 = 420 ms, the main effect of beat conditions was not significant, $F (2, 10) = 1.091$, $p = .373$, $\eta_p^2 = .179$, but the interaction was significant, $F (20, 100) = 1.962$, $p = .016$, $\eta_p^2 = .282$. A follow-up simple main-effect test revealed a significant effect of beat conditions when the /T1/T2/ pattern was /180/240/ (i.e., T1–T2 = −60 ms; $p = .004$), /190/230/ (i.e., T1–T2 = −40 ms; $p = .006$), and /240/180/ (i.e., T1–T2 = 60 ms; $p = .005$). For these significant simple main effects, Ryan's post hoc test revealed significant difference between the Repetitive-presentation condition and the Beat-on-T1 condition when /190/230/ (i.e., T1–T2 = −40 ms), significant difference between the Repetitive-presentation

condition and the Beat-on-T2 condition when /240/180/ (i.e., T1–T2 = 60 ms), and significant differences between the Beat-on-T1 condition and the Beat-on-T2 condition when /180/240/ (i.e., T1–T2 = −60 ms), /190/230/ (i.e., T1–T2 = −40 ms), and /240/180/ (i.e., T1–T2 = 60 ms) ($p < .05$). For T1 + T2 = 630 ms, the main effect of beat conditions and the interaction were both non-significant, $F (2, 10) = .188$, $p = .831$, $\eta_p^2 = .036$, and $F (20, 100) = .822$, $p = .682$, $\eta_p^2 = .141$, respectively. The interpretation of these results is discussed in the *General discussion*.

Summarizing Experiment 2, the results suggested that (1) the four-point rating method can capture the characteristics of temporal assimilation (reflecting time-shrinking) as in the previous studies, (2) simply repeating the T1-T2 pattern increases sensitivity to inequality of neighboring time intervals, and (3) adding the preceding sounds causes the neighboring time intervals to be perceived as more "equal" in some cases when T1 + T2 = 420 ms.

General discussion

The results of Experiment 1 showed that, especially at the tempo of T1 + T2 = 420 ms, the preceding beat influenced the perceived equality/inequality of the target sequence. Specifically, when T2 > T1 and the difference was within about 60 ms, the same target sequence was perceived as more "equal" in the Beat-on-T1 condition compared with the Beat-on-T2 condition. A possible explanation for this is that in the Beat-on-T2 condition, the preceding sounds induced a regular beat, which coincided with the beginning of T2; thus, the T1 and T2 repetition in the target sequence was more likely to be perceived as cycles of "T2-T1" (i.e., the target sequence was perceived as anacrustic), while in the Beat-on-T1 condition, the beat coincided with the beginning of T1, and the target sequence was more likely to be perceived as cycles of "T1-T2" (e.g., Repp et al., 2008; Povel, 1984). In the latter case, the time-shrinking illusion may have occurred because the /T1/T2/ patterns were within the time condition range in which time-shrinking occurs (T1 < 200 ms, and T2 was longer than T1 by up to about 100 ms, e.g., Nakajima et al., 2004). The occurrence of time-shrinking shortens the subjective duration of T2, thus making the *physically* longer T2 *perceptually* more similar to T1, consequently increasing the "equal" responses. In the Beat-on-T2 condition, the perceived "T2-T1" rhythm cycle would *not* cause time-shrinking because the first time interval in the (perceptual) rhythm cycle (i.e., T2) was longer than the second one (i.e., T1). This is not the time condition in which time-shrinking occurs (e.g., Nakajima et al., 2004; see also Miyauchi & Nakajima, 2005, 2007); thus, the physically longer T2 must have also been perceived as longer than T1, consequently inducing more "unequal" responses.

If this was the case, the same logic should apply for patterns in which T1 > T2: in this case, time-shrinking should have occurred in the Beat-on-T2 condition because the target sequence would have been perceived as cycles of "T2-T1," which is the time condition in which time-shrinking occurs (T2 < 200 ms and T2 < T1 in the "T2-T1" cycle; e.g., Nakajima et al., 2004). Conversely, time-shrinking would not have occurred in the Beat-on-T1 condition because the "T1-T2" cycle in this condition is not the pattern in which time-shrinking occurs (the first time interval in the rhythm cycle would have been longer than the second, T1 > T2). Indeed, such tendencies did appear in the results (see the right half of Fig. 2b). The effect seemed weaker than when T1 < T2 (left half of Fig. 2b), which suggests that time-shrinking is more likely to occur when the shorter time interval (i.e., T1 when T1 < T2) appears at the beginning of the target sound sequence.

It was slightly surprising that the difference between the Beat-on-T1 condition and the Beat-on-T2 condition did not appear significantly when T1 + T2 = 210 ms, because time-shrinking was expected to also occur at this tempo (e.g., Miyauchi & Nakajima, 2007; Nakajima et al., 2004). The reason for the absence of the beat condition effects at this tempo is currently unclear, but one explanation may be that participants had difficulty maintaining the beat positions (metrical interpretation) suggested by the preceding sounds through the target sequence presentation when T1 and T2 were as short as 55–155 ms as in the T1 + T2 = 210 ms tempo (e.g., Repp, 2006). This explanation is in line with the phase shifting (or phase adaptation) reported in Barnes and Jones (2000) in terms of the Dynamic Attending Theory (Jones & Boltz, 1989). Their dynamic attending interpretation posits that an adaptive oscillator entrains with the external sound sequence, and that this oscillator can modulate its phase when there is a sound at an unexpected timing (a perturbation). It is possible that when the sounds of the target sequence were close to each other in time (i.e., the T1 and T2 durations were short) as in the T1 + T2 = 210 ms tempo, one of the sounds in the target sequence acted as the unexpected sound (perturbation) and caused the phase of the oscillator to adapt to this sound, regardless of the beat suggested by the preceding sounds. In such a case, it is natural that there were not many effects of preceding beat conditions on the perception of the target sequence.

The non-significant effects of beat conditions at the slowest tempo of T1 + T2 = 630 ms was probably because the T1 and T2 durations employed at this tempo (i.e., 240–390 ms) all exceeded 200 ms. Time-shrinking diminishes when the preceding time interval becomes longer than 200 ms (e.g., Nakajima et al., 2004), thus it is likely that time-shrinking did not appear much at this slowest tempo. The diminishing of time-shrinking must have decreased the difference between the Beat-on-T1 and the Beat-on-T2 conditions.

Comparison of Experiment 1 and the Repetitive-presentation condition in Experiment 2 showed that, for the same target sequence, adding the preceding sounds sometimes caused T1 and T2 to be perceived as more "equal" compared to when there were no preceding sounds at the tempo of T1 + T2 = 420 ms. Such an effect occurred in the Beat–on–T1 condition when T1 − T2 = −40 ms, and in the Beat-on-T2 condition when T1 − T2 = 60 ms. Since these were the conditions in which time-shrinking would occur, as stated in the first two paragraphs of this *General discussion*, it can be considered that the beat induced by the preceding sounds promoted the occurrence of time-shrinking. In other words, it is possible that the asymmetrical temporal assimilation (caused by time-shrinking), which typically appears when the T1-T2 pattern is presented only once (as in the Single-presentation condition in Experiment 2), diminishes when the T1-T2 pattern is repeated (as in the Repetitive-presentation condition in Experiment 2), but appears again when the preceding sounds are added at the right timing.

The potential effects of contexts such as beat and measure on perceived rhythm have been identified in some empirical studies (e.g., Desain & Honing, 2003; Iversen et al., 2009; Povel & Essens, 1985; Repp, 2005), and the effects of metrical interpretation may also appear in one's perception of real music (e.g., Stobart & Cross, 2000). In the present study, it was clearly demonstrated that the perceived durations of neighboring time intervals (which is the basis for perceiving rhythm) were influenced by the preceding beat: at a certain tempo, shifting the phase of the preceding beat changed the perceived equality/inequality of the neighboring time intervals of *physically identical* simple rhythm patterns. In other words, the same rhythm was perceived as different rhythms when organized within a different metrical framework. It is possible that this phenomenon is a type of bistable perception in rhythm perception.

In conclusion, the present study was the first to report different equality/inequality perceptions of neighboring time intervals in a physically identical sound sequence. This phenomenon can provide a physically controlled approach for future neuroscientific investigations that explore underlying mechanisms and awareness of rhythm perception.

Open Practices Statement None of the data or materials for the experiments reported here is available at present, and none of the experiments were preregistered.

References

Arao, H., Suetomi, D., & Nakajima, Y. (2000). Does time-shrinking take place in visual temporal patterns? *Perception* 29, 819–830. doi: https://doi.org/10.1068/p2853

Barnes, R., & Jones, M. R. (2000). Expectancy, attention, and time, *Cognitive Psychology*, *41*, 254-311.

Desain, P., & Honing, H. (2003). The formation of rhythmic categories and metric priming. *Perception*, *32*, 341-365. doi:https://doi.org/10.1068/p3370

Drake, C., & Botte, M.-C. (1993). Tempo sensitivity in auditory sequences: Evidence for a multiple-look model. *Perception & Psychophysics*, *54*, 277-286.

Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *The psychology of music* (pp. 149–180). New York: Academic Press.

Greenberg, S., & Arai, T. (2004). What are the essential cues for understanding spoken language? *IEICE Transactions on Information and Systems*, *E87-D*, 1059–1070.

Grondin, S., Hasuo, E., Kuroda, T., & Nakajima, Y. (2018). "Auditory time perception," in *Springer Handbook of Systematic Musicology*, ed. R. Bader (Cham: Springer), 407-424.

Handel, S. (1989). *Listening: An introduction to the perception of auditory events*. Cambridge, MA: MIT Press.

Hasuo, E., Kuroda, T., & Grondin, S. (2014). About the time-shrinking illusion in the tactile modality. *Acta Psychologica, 147*,122–126.

Hasuo, E., Nakajima, Y., & Hirose, Y. (2011). Effects of sound marker durations on rhythm perception. *Perception*, *40*, 220–242.

Iversen, J. R., Repp, B. H., & Patel, A. D. (2009). Top-down control of rhythm perception modulates early auditory responses. *Annals of the New York Academy of Sciences*, *1169*, 58-73. doi: https://doi.org/10.1111/j.1749-6632.2009.04579.x.

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, *96*, 459-491. doi:https://doi.org/10.1037/0033-295X.96.3.459

Miller, N., & McAuley, J. D. (2005). Tempo sensitivity in isochronous tone sequences: The multiple-look model revisited. *Perception & Psychophysics*, *67*, 1150-1160.

Miyauchi, R., & Nakajima, Y. (2005). Bilateral assimilation of two neighboring empty time intervals. *Music Perception*, *22*, 411-424. doi: https://doi.org/10.1525/mp.2005.22.3.411.

Miyauchi, R., & Nakajima, Y. (2007). The category of 1:1 ratio caused by assimilation of two neighboring empty time intervals. *Human Movement Science*, *26*, 717–727. doi:https://doi.org/10.1016/j.humov.2007.07.008.

Moore, B. C. J. (2003). An introduction to the psychology of hearing: Fifth edition, San Diego, CA: Academic Press.

Nakajima, Y., ten Hoopen, G., Sasaki, T., Yamamoto, K., Kadota, M., Simons, M., & Suetomi, D. (2004). Time-shrinking: The process of unilateral temporal assimilation. *Perception*, *33*, 1061–1079. doi: https://doi.org/10.1068/p506.

Nakajima, Y., ten Hoopen, G., & van der Wilk, R. (1991). A new illusion of time perception. *Music perception*, *8*, 431-448.

Povel, D. J. (1984). A theoretical framework for rhythm perception. *Psychological Research*, *45*, 315-337.

Povel, D. J., & Essens, P. (1985). Perception of temporal patterns. *Music Perception*, *2*, 411-440.

Remijn, G., Meulen, G. V. D., ten Hoopen, G., Nakajima, Y., Komori, Y., & Sasaki, T. (1999). On the robustness of time-shrinking. *Journal of the Acoustical Society of Japan (E)* 20, 365–373.

Repp, B. H. (2005). Rate limits of on-beat and off-beat tapping with simple auditory rhythms: 2. The roles of different kinds of accent. *Music Perception*, *23*, 165-187.

Repp, B. H. (2006). Rate limits of sensorimotor synchronization. *Advances in Cognitive Psychology*, *2*, 163-181.

Repp, B. H., Iversen, J. R., & Patel, A. D. (2008). Tracking an imposed beat within a metrical grid. *Music Perception*, *26*, 1-18.

Russo, F. A., & Ammirante, P. (2018). Music Perception. In J.T. Wixted & J. Serences (Eds.) *Volume 2 of Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience: Sensation, Perception, and Attention* (pp. 317-342). New York, NY: John Wiley & Sons.

Sasaki, T., Suetomi, D., Nakajima, Y., & ten Hoopen, G. (2002). Time-shrinking, its propagation, and Gestalt principles. *Perception & Psychophysics*, *64*, 919–931. doi:https://doi.org/10.3758/BF03196796.

Stobart, H., & Cross, I. (2000) The Andean anacrusis? Rhythmic structure and perception in Easter songs of Northern Potosí, Bolivia, *British Journal of Ethnomusicology*, *9:2*, 63-92, doi: https://doi.org/10.1080/09681220008567301.

ten Hoopen, G., Hartsuiker R., Sasaki, T., Nakajima, Y., Tanaka, M., and Tsumura, T. (1995). Auditory isochrony: time shrinking and temporal patterns, *Perception*, *24*, 577-593.

ten Hoopen, G., Miyauchi, R., & Nakajima, Y. (2008). Time-based illusions in the auditory mode. In S. Grondin (Ed.), *Psychology of time* (pp. 139–188). Bingley: Emerald Group Publishing.

Van Erp, J. B. F., & Spapé, M. M. A. (2008). Time-shrinking and the design of tactons. In M. Ferre (Ed.), *Haptics: Perception, devices and scenarios*, *Vol. 5024*. (pp. 289–294). Berlin Heidelberg: Springer.