1444

ELSEVIER

Available online at www.sciencedirect.com
ScienceDirect

Cognition 102 (2007) 434–454

COGNITION

www.elsevier.com/locate/COGNIT

Hearing a melody in different ways: Multistability of metrical interpretation, reflected in rate limits of sensorimotor synchronization ☆

Bruno H. Repp *

Haskins Laboratories, New Haven, CT, USA

Received 28 April 2005; revised 2 February 2006; accepted 7 February 2006

Abstract

Music commonly induces the feeling of a regular beat (i.e., a metrical structure) in listeners. However, musicians can also intentionally impose a beat (i.e., a metrical interpretation) on a metrically ambiguous passage. The present study aimed to provide objective evidence for this little-studied mental ability. Participants were prompted with musical notation to adopt different metrical interpretations of a cyclically repeated isochronous 12-note melody while tapping in synchrony with specified target tones in the melody. The target tones either coincided with the imposed beat (on-beat tapping) or did not (off-beat tapping). An adaptive staircase method was employed to determine the fastest tempo at which each synchronization task could be performed. For each metrical interpretation, a significant advantage for on-beat over off-beat tapping was obtained – except in a condition in which participants, instead of synchronizing, were in control of the target tones. By showing that a self-imposed beat can affect sensorimotor synchronization, the present results provide objective evidence for endogenous perceptual organization of metrical sequences. It is hypothesized that metrical interpretation rests upon covert rhythmic action.

☆ This manuscript was accepted under the editorship of Jacques Mehler.

* Tel.: +1 203 865 6163x236; fax: +1 203 865 8963.
E-mail address: repp@haskins.yale.edu

0010-0277/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2006.02.003

Keywords: Rhythm; Meter; Music perception; Metrical interpretation; Beat induction; Synchronization

1. Introduction

1.1. Multistability in perception

Ambiguous stimuli have long occupied an important place in the study of perception and cognition. It is only when the information impinging upon the sense organs is insufficient fully to determine the percept or motor response that the perceiver's contribution can be assessed. That contribution may derive from autonomous brain processes that give rise to biases over which the perceiver has no control, or it may result from a conscious decision or intention to impose a perceptual interpretation on the stimulus.

The textbook examples of ambiguous stimuli are reversible visual figures, such as Necker's cube or Rubin's vase-face stimulus, that lend themselves to two mutually exclusive interpretations, each of which is quite unambiguous. Such stimuli are said to be perceptually *multistable* (Attneave, 1971; Kelso, 1995). When such a figure is viewed for an extended time, the two percepts alternate spontaneously, changing as often as every few seconds. This alternation has been attributed to neural adaptation or satiation (Hebb, 1949; Köhler & Wallach, 1944). However, there is also evidence that the perceiver's knowledge and intentions play a role. When participants do not know the two alternatives in advance, they often see only one of them (Girgus, Rock, & Egatz, 1977; Rock, Hall, & Davis, 1994). When they are instructed to either accelerate or retard perceptual reversals, or to see one of the alternatives longer than the other, they can do so to some extent (Hochberg & Peterson, 1987; Liebert & Burk, 1985; Pelton & Solley, 1968; Toppino, 2003). When their attention is diverted, they tend to reverse their percepts less frequently (Reisberg & O'Shaughnessy, 1984; Rock et al., 1994). As long as the alternatives are known, however, reversal can generally not be prevented.

Most of these studies have relied exclusively on participants' subjective reports. However, it is important to prove that the reports really represent different percepts and not merely cognitive judgments of an ambiguous but constant percept. Hochberg and Peterson (1987) describe how the perceived orientation of a Necker cube can be revealed by judgments about the perceived direction of rotation of the cube. Other avenues have been opened up by the methods of neuroscience. Thus, Andrews, Schluppeck, Homfray, Matthews, and Blakemore (2002) found in an fMRI study that activation of the fusiform gyrus, a cortical area specifically involved in the perception of faces, was significantly increased whenever participants viewing Rubin's vase-face figure reported seeing the faces. Using a method of discontinuous stimulus presentation, Kornmeier and Bach (2004) found that event-related potentials (ERPs) following exogenous (i.e., physically unambiguous) and endogenous orientation reversals of the Necker cube were highly similar. Endogenous reversals affected ERPs as soon as 160 ms after stimulus onset, which suggested involvement of early visual processes.

Multistability is also known in auditory perception, where discontinuous presentation is the rule. There are a variety of stimuli that can give rise to different, mutually exclusive percepts. However, unlike reversible visual figures, repeatedly presented auditory stimuli do not usually cause the percepts to alternate. For example, a speech syllable drawn from the boundary region between two phonemic alternatives (e.g., /ba/-/da/) will upon repeated presentation be perceived as one or the other category, or perhaps as ambiguous between the two, but the two categorical percepts will not alternate in a regular fashion.[1] Repeated presentation of an ambiguous word can lead to verbal transformations (Warren, 1961, 1999), but these often involve multiple alternatives that follow an irregular progression. The tritone paradox (Deutsch, 1987) involves pairs of tones that can be perceived as either ascending or descending in pitch. Although such a pair may be perceived differently in different contexts (Repp, 1997), repeated presentation of the same pair generally does not lead to a change in percept. Although phenomena of neural adaptation do occur in audition, they seem to take place mainly at a more peripheral level than that of the categorical percept (see, e.g., Roberts & Summerfield, 1981).

Likewise, there is little evidence that participants' intentions can affect auditory perception. Knowledge of alternatives certainly plays a role (e.g., a speaker of a language that does not have a certain phoneme will not report hearing that phoneme in a speech perception test), and context effects abound, but the mere intention to hear one of two (or more) categorical alternatives generally has little effect on the perception of multistable auditory stimuli. As long as one percept occurs unambiguously (e.g., /ba/, or a rising pitch from one tone to another), it is difficult to make oneself hear anything else (e.g., /da/, or a falling pitch), at least according to the author's informal observations. There appears to be no scientific literature on reversible auditory figures comparable to that on reversible visual figures.

It is widely recognized that music is one of the most complex human achievements. Music encompasses a vast array of phenomena that pose unique perceptual, cognitive, and motoric challenges, and that provide excellent testing grounds for psychological theories. The present study is concerned with the perception of metrical structure in rhythmic sequences. Although metrical structure rarely changes spontaneously, it is both multistable and highly susceptible to effects of intention, as will be demonstrated. Thus, it is similar in some ways to reversible visual figures, but also different in significant ways.

1.2. Metrical structure and grouping

Most forms of music are rhythmic, which means they consist of events whose onsets are separated by intervals that form simple ratios. Rhythmic sequences often elicit in listeners the feeling of a regular beat, which in turn engages (or, more likely, reflects the engagement of) their motor system and facilitates rhythmic movement,

---

[1] This statement is based on the author's experience with such stimuli many years ago. There does not seem to be a study in the literature that addresses this issue directly.

such as dance – a facilitation that seems to be specific to humans. The purpose of research on *beat induction* is to determine how various structural properties of a musical sequence (event timing, event intensities, pitch contour, repetition, etc.) determine the period and phase of the experienced beat. There are many studies in that area (for recent examples, see Hannon, Snyder, Eerola, & Krumhansl, 2004; Snyder & Krumhansl, 2001), and various beat-finding algorithms have been proposed and tested (e.g., Desain & Honing, 1999; Large & Kolen, 1994; Toiviainen & Snyder, 2003). Listeners in these studies are usually required to indicate the beat that fits the music best, often by tapping along. In principle, there are many ways of tapping along with a musical passage, but usually only one way seems "right" to a given individual at a given time. If there is strong physical support for a beat in the music (e.g., if heavy accents recur regularly), there will be general agreement among participants. However, it is quite common to observe individual differences in the period and/or relative phase of the tapped beat. Such differences reflect different *metrical interpretations* of the same rhythmic sequence. Two different metrical interpretations are mutually exclusive unless their beats are in phase and their frequencies exhibit a simple ratio, in which case one beat is a subdivision of the other, so that the two beats are hierarchically related. Indeed, metrical structures are generally hierarchical, and beats can be perceived at multiple levels, albeit with different degrees of salience (Drake, Penel, & Bigand, 2000; Parncutt, 1994).

Musicians frequently encounter situations in which a given musical passage must be interpreted according to a notated meter or according to preceding musical context, regardless of whether or not the interpretation provides a best fit to the structural characteristics of the music. Especially when prompted by notation, such an interpretation is not induced by the sequence (exogenous) but rather *imposed* intentionally by the listener (endogenous) and results in hearing the passage in a particular way. Subjectively, the same passage sounds very different when it is interpreted to be in a different meter, or in the same meter but with the beat shifted in relative phase.[2] However, a passage may resist imposition of a metrical interpretation that is strongly incompatible with its structure.

Musicians' ability to impose a beat on music while listening has not been studied much, and the evidence for it to date is mainly informal and subjective.[3] It is closely related, however, to the phenomenon of *subjective grouping*, which was first investigated by Bolton (1894). Subjective grouping into twos, threes, or fours often occurs spontaneously when an isochronous sequence of identical tones is presented at a moderately fast tempo. Such grouping implies a beat that (for Western listeners, at least) coincides with the group-initial tones. Conversely, a given musical meter (notated as a "time signature") implies a particular default grouping of the notes

---

[2] Sloboda (1983, p. 383) reports a nice example. He generated two melodies from a single pitch sequence by phase-shifting the meter in the musical notation. Pianists requested to play the notated melodies in the course of an experiment remained unaware that the pitch sequences were identical, as shown by their surprise when it was pointed out to them afterwards.

[3] There are some studies of metrical interpretation in music performance (Sloboda, 1983, 1985) and in rhythm production (Repp, London, & Keller, 2005; Repp & Saltzman, 2002).

of an isochronous sequence (reflected by beams connecting note stems in standard musical notation). Although grouping can in principle be dissociated from meter, this is not attempted in the present study, so that metrical interpretation is synonymous with subjective grouping. Thus, triple and quadruple meters (time signatures such as 3/4 and 4/4) imply grouping of isochronous events into threes and fours, respectively, and vice versa; and a phase shift of the beat in either of these meters is accompanied by a phase shift of the subjective groups.

An isochronous sequence of identical tones lends itself to numerous possible metrical interpretations, although listeners will generally prefer the simplest one(s). If the sequence is more structured, for example, if it has different pitches that form a melodic contour or if the event inter-onset intervals are of different durations, the sequence may strongly encourage certain metrical interpretations and resist others. For a study of subjective beat imposition, it is desirable to reduce such strong biases but at the same time to provide some structure that makes it possible to distinguish different metrical interpretations. An isochronous monotone sequence sounds exactly the same when a beat of a given period is phase-shifted, but an isochronous melody changes radically because its pitches are being regrouped by the phase shift. The present study employed isochronous melodies because, unlike rhythmic structure, pitch structure is a relatively weak cue to metrical structure (Hannon et al., 2004).

When dealing with a purely subjective phenomenon, such as metrical interpretation, it is desirable to find some objective measure that reflects the experience of the listener in an indirect but obligatory fashion. This approach is analogous to that of using perceived rotation as an index of the perceived orientation of a Necker cube (Hochberg & Peterson, 1987). In the case of musical stimuli, synchronized tapping is an obvious choice. It would not be sufficient, however, to ask participants simply to tap along with their subjective beat because the tapping itself may induce the feeling of a beat (Repp, 2005b), leading to circularity. Rather, it is necessary to dissociate the motor activity from the subjective beat and vary both orthogonally. The present experiments took advantage of the fact that off-beat tapping is generally more difficult than on-beat tapping, especially when the sequence tempo is fast (e.g., Fraisse & Ehrlich, 1955; Repp, 2005a). Accordingly, it was hypothesized that tapping on a self-imposed beat should be easier than tapping off that beat, even when (unlike the typical off-beat tapping paradigm) both on-beat and off-beat locations are marked by sequence events. Because different metrical interpretations were combined factorially with different tapping targets in the sequence, the hypothesis can also be stated as follows: For any given set of tapping targets, it should be easier to synchronize when the targets coincide with the subjective beat (on-beat tapping) than when they do not (off-beat tapping). The measure of ease of synchronization was the fastest sequence tempo at which synchronization could be maintained.

A previous study (Repp, 2005b, Experiment 3) used the same approach, but with very limited success. The predicted result was shown by only one of eight musically trained participants – the author himself. Although this alone was sufficient to demonstrate that metrical interpretation *can* affect synchronization, the negative results for the other participants were disconcerting. Many of them, however, had difficulty maintaining the requested metrical interpretations, which suggested that the materials (rhythmic patterns with a temporal structure) were not ideally suited for the purpose. The beat implied by the rhythmic patterns (see Povel & Essens, 1985) may have been too strong to be overcome reliably by a conflicting metrical interpretation. The study provided ample evidence that metrical interpretation can change spontaneously, either by reverting to the one favored by the stimulus structure or (more frequently) by aligning itself with the finger taps. A subsequent experiment (Repp, 2005b, Experiment 4) indicated that the variability of asynchronies during on-beat and off-beat synchronization with moderately paced sequences is not a sensitive measure of metrical interpretation. This is because the instability at fast tempi depends not only on the variability of taps but also on their relative phase. Failure to synchronize at fast tempi is usually caused by phase drift, not just increased variability (see Repp, 2005a). Therefore, the present study kept using the fastest possible tempo of synchronization (the *synchronization threshold*; Repp, 2003) as the dependent variable. The synchronization threshold is substantially higher for off-beat than for on-beat tapping in the standard case where the off-beat taps fall between sequence events (Repp, 2005b). A smaller difference was expected in the present paradigm, where the taps coincided with sequence events in both cases and were on-beat or off-beat only with respect to the participants' internal beat.

Two experiments were conducted. In Experiment 1, the different metrical interpretations concerned the phase of the self-imposed beat (with meter and group size held constant), whereas in Experiment 2, they concerned the period of the beat (i.e., meter and group size). Experiment 1 also included a condition in which participants, instead of synchronizing with certain target tones in a computer-controlled sequence, controlled those tones themselves by means of their taps. The results of that condition proved informative with regard to the temporal references used in synchronization with metrical sequences.

2. Experiment 1

As just stated, Experiment 1 manipulated the phase of the imposed beat while keeping meter and group size constant, and it also compared synchronization with production of (i.e., control over) target tones. These two conditions were originally just thought of as variants, but as will be seen, they yielded quite different results, which could be given a meaningful interpretation.

2.1. Methods

2.1.1. Participants

Seven paid volunteers (6 women, 1 man; ages 19–24) and the author (age 59) participated. All had considerable musical training (7–16 years, ranging from advanced amateur to professional level of instrument playing) and were regular participants in synchronization experiments.

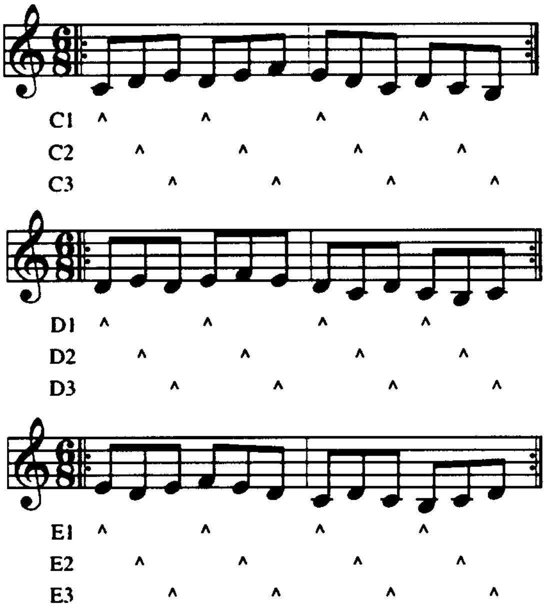

Fig. 1. Materials and tasks of Experiment 1. Wedges indicate target tones for tapping.

2.1.2. *Materials and tasks*

The materials are shown in Fig. 1 in musical notation. The figure shows three 12-note melodies notated in 6/8 meter, which implies a grouping of tones into threes (indicated by the beams connecting note stems) and a group-initial beat. The vertical bar line in the center is not important for present purposes. The melodies could have been notated just as well in 3/8 meter (three bar lines) or 12/8 meter (no bar line). In other words, the beat of interest here is the one that occurs at the metrical level just above that of the individual events and is aligned with the initial events of groups comprising three events each. The double vertical lines with colons at the beginning and end of each melody signify that the melody is to be repeated over and over. A closer look reveals that the three melodies are composed of the same notes: The first melody begins on C; the second melody begins on D (the second note of the first melody) while the C appears at the end; and the third melody begins on E (the third note of the first melody) while the C and D appear at the end. Thus, when each melody is repeated continuously, the same pitch sequence results; only the starting point (C, D, or E) is different. In other words, each melody constitutes a different metrical interpretation of the same pitch sequence, differing only in the phase of the beat.

The wedges below the melodies indicate the target tones in the tapping tasks. For each of the three metrical interpretations, participants were required to tap on the first, second, or third tone of each group. Thus, there were nine experimental tasks. Tasks C1, D1, and E1 were on-beat tapping tasks in which participants tapped on the initial events of the groups; the others were off-beat tapping tasks in which the taps fell on the second or third events of groups. Note that in tasks C1, D3, and

E2, participants were tapping on exactly the same target tones (C-D-E-D), but with different metrical interpretations in mind. The same holds for tasks C2, D1, and E3 (targets D-E-D-C), and for tasks C3, D2, and E1 (targets E-F-C-B).

2.1.3. Procedure

Participants came for four sessions, typically one week apart. Sessions 1 and 2 constituted the *synchronization* condition, whereas Sessions 3 and 4 constituted the *production* condition. The tasks were blocked by metrical interpretation. In Sessions 1 and 3, participants performed the tasks in the order shown in Fig. 1; in Sessions 2 and 4, they did them in the reverse order.

In the synchronization condition, the pitch sequence was played under computer control on a digital piano (Roland RD-250s) in *legato* style: Each tone ended shortly after the next tone began. The tones all had the same nominal intensity (MIDI velocity). In the production condition, the pitch sequence started in the way just described in each trial, but as soon as the first tap occurred, the computer relinquished control over (i.e., omitted) the target tones, which were now controlled by the participant's taps. In other words, the participant now "produced" the target tones during gaps in the melody, trying to place them as precisely as possible between the rigidly timed non-target tones. Consequently, the timing of the target tones was variable, and there were no longer any asynchronies between taps and target tones – a situation known as "pseudo-synchronization" when participants are unaware that they are in control (Flach, 2005; Fraisse & Voillaume, 1971). In the present case participants always knew they were in control of the target tones because the production condition was explained to them in advance. The computer still controlled all properties of the target tones other than timing: The tones were unaffected by tapping force and duration of finger-pad contact, and the *legato* style of the melody was preserved.

The different tasks were explained to the participants before the experiment began. Before each task, participants were given a prompt card showing the musical notation for the appropriate metrical interpretation, with the target tones indicated (as in Fig. 1, but on a separate card for each task). Participants were free to view the card throughout the run. They were instructed to "hear" the melody in accord with the notation while tapping in synchrony with the target tones. It was stressed that maintenance of the required metrical interpretation was crucial. If they found themselves involuntarily switching to another interpretation, they were to stop after the trial, take a short break, report the occurrence on a note pad, and then continue with the original interpretation in mind.

An adaptive staircase procedure (implemented in MAX 3.0.9 software running on an iMac G4 computer) was used to estimate the fastest tempo at which each task could be carried out. Each task consisted of a run of trials that took 5–10 min. Participants sat in front of the computer, wore Sennheiser HD540 II earphones, and held an electronic percussion pad (Roland SPD-6) on their lap. They started a trial by pressing the space bar of the computer keyboard. The sequence started playing and participants began tapping with the index finger of their preferred hand on the pad whenever they felt ready (usually in the first group of the second cycle of the melody). In the production task, participants had to start tapping in the first group, or else they heard incorrect pitches. A successful trial consisted of 40 taps, each of which had to be within IOI/2 of a target position, where IOI is the inter-onset interval of the (non-target) sequence tones. The target positions were marked by computer-controlled tones in the synchronization condition but constituted the midpoints of long IOIs (i.e., virtual targets) in the production condition. In the first trial of a run, the IOI was 200 ms, implying a beat period and tapping period of 600 ms. Following a successful trial, the IOI was decreased by x ms, with $x = 10$ initially. If any tap did not meet the accuracy criterion, the sequence stopped immediately, the IOI for the next trial was increased by x ms, and x was decreased by 2. A run ended when x reached zero (i.e., after five unsuccessful trials). The final value of the IOI was the estimate of the rate limit or *synchronization threshold* for that task.

2.2. Results

It seemed easier to establish and maintain different metrical interpretations with the present isochronous melodies than with the monotone rhythmic sequences in Repp's (2005b) study. Nevertheless, participants did encounter some difficulties, usually at fast rates, when they approached the synchronization threshold. Spontaneous switches to another interpretation were reported in 72 runs (25% of all runs). As a run typically contained between 10 and 20 trials, this amounts to about 2% of the trials (each of which, if successful, contained at least 11 cycles of the melody). In most cases, participants seemed to be able to recover the intended metrical interpretation after taking a short break. Because there were large individual differences in the frequency of switches, these data were not analyzed statistically. However, switches were less frequent in the last session (8) than in the first three sessions (18, 26, and 20); they were more frequent in the melody beginning on D (36) than in the melodies beginning on C or E (18, 18); and they tended to be less frequent in on-beat tapping (18) than in off-beat tapping (27, 27). Information about the direction of switches (i.e., to which of the two alternative interpretations) had not been requested and was rarely provided by the participants. Importantly, failures to maintain metrical interpretations worked in favor of the null hypothesis in this study.

The synchronization thresholds for the nine tasks in the two conditions are shown in Fig. 2. In the synchronization condition (upper panel), the on-beat tapping tasks (C1, D1, and E1) were easier (i.e., had lower thresholds) than the off-beat tapping tasks, as predicted.[4] In the production condition (lower panel), however, all tasks were about equally difficult.

The results for each condition were subjected to a $2 \times 3 \times 3$ repeated-measures ANOVA with the variables of session, metrical interpretation, and target position (1st, 2nd, or 3rd within tone groups). In the synchronization condition, the main effect of target position was significant, $F(2,14) = 19.76$, $p < .001$, $\varepsilon = .75$ (Green-

---

[4] The large range of thresholds in the C3, D3, and E3 conditions was largely due to a single individual, a percussionist, who generally found it as easy to tap on the third tone of a group (the "pickup" to the beat) as on the first tone.

house-Geisser correction), $\eta_p^2 = .74$.[5] The only other significant effect was a main effect of session, $F(1,7) = 9.06$, $p < .02$, $\eta_p^2 = .56$, which reflects lower thresholds in the second session – a practice effect. In the production condition, the main effect of target position was far from significance. The only significant effect here was the Session × Metrical Interpretation interaction, $F(2,14) = 5.87$, $p < .04$, $\varepsilon = .66$, $\eta_p^2 = .46$, which was due to improvement of performance in the course of each session, though not across sessions. Because the order of the metrical interpretations was reversed in the later session, the within-session improvement resulted in an interaction.

A combined repeated-measures ANOVA of the synchronization and production conditions, with the additional variable of condition, showed the Condition × Target Position interaction to be significant, $F(2,14) = 10.86$, $p < .005$, $\varepsilon = .77$, $\eta_p^2 = .61$, which confirms that the effect of target position was different in the two conditions. Several additional effects, largely predictable from the results of the separate ANOVAs, were also significant: the main effects of session, condition, and target position, as well as the Condition × Session × Metrical Interpretation interaction.

One additional analysis concerned the direction of deviation of the final tap of unsuccessful trials (the tap that violated the ±IOI/2 accuracy criterion) from its target position. It was more common for participants to tap too late (62%) than too early (38%). An ANOVA on the percentages of late taps showed no differences between the two conditions. The only significant effect was the Session × Metrical Interpretation interaction, $F(2,14) = 6.59$, $p < .02$, $\varepsilon = .91$, $\eta_p^2 = .48$, which reflected a decrease in the tendency to lag behind as a session progressed. This decrease resulted in an interaction because metrical interpretation was blocked and reversed in order between the two sessions.

2.3. Discussion

The results of the synchronization condition provide objective evidence for the subjective phenomenon of metrical interpretation: It is easier to tap on a self-induced beat than to tap off that beat. A more detailed discussion of these findings will follow Experiment 2. Here, an explanation of the results of the production condition will be given.

In contrast to the synchronization condition, the production condition did not show an advantage for on-beat over off-beat tapping. In both on-beat and off-beat tapping, the on-beat tones probably served as primary temporal references, and the off-beat tones as secondary ones. Temporal references are necessary for phase error correction (see Repp, 2005c). When the on-beat tones were under the participant's control, they exhibited temporal variability and thus could no longer serve as stable temporal references, which made on-beat tapping more difficult. By contrast,

---

[5] The effect was still significant after removing the author's data, $F(2,12) = 14.47$, $p < .002$, $\eta_p^2 = .71$. Alternatively, the data could have been analyzed with target pitches (C-D-E-D, D-E-D-C, and E-F-C-B) as a variable, instead of target position. In that case, the Metrical Interpretation × Target Pitches interaction would have been the significant effect.

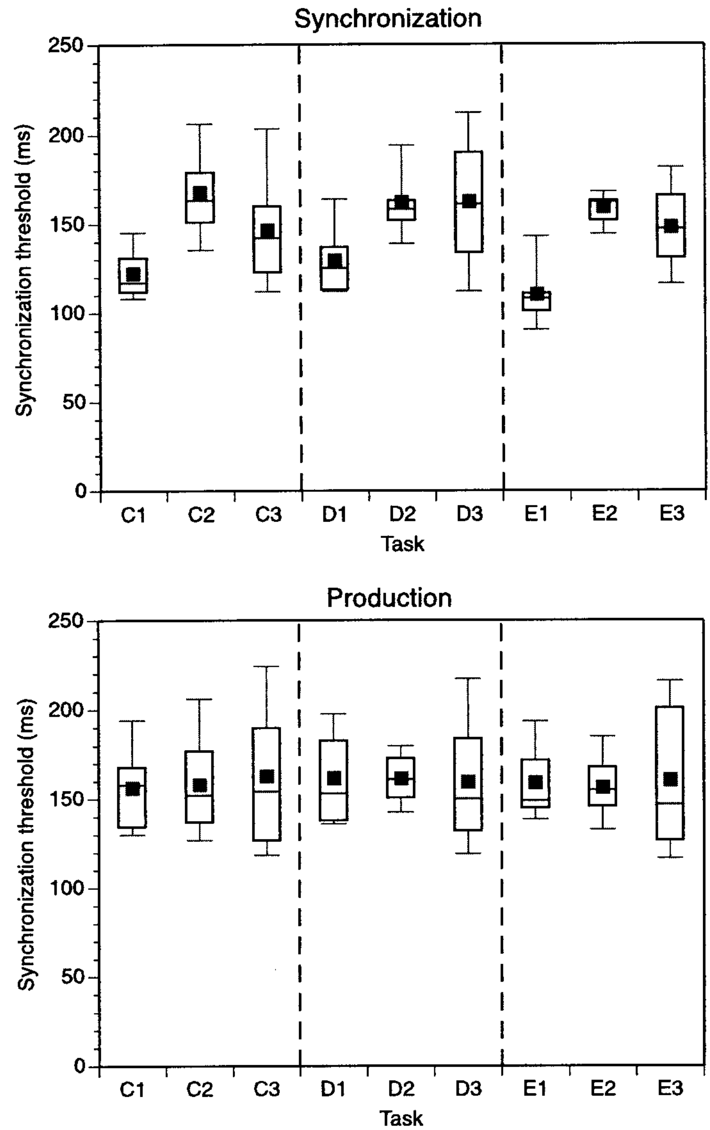

Fig. 2. Results of Experiment 1. Box plots show means (squares), medians, quartiles, and 1st and 9th deciles for 16 data points (8 participants × 2 sessions). The on-beat tapping conditions are C1, D1, and E1.

when participants controlled off-beat tones, participants could still use the computer-controlled on-beat tones as stable references, and so performance was similar to that in the off-beat synchronization tasks.

In on-beat tapping, participants could have adopted one of two strategies: Either they kept using the on-beat tones as primary references, despite their variability, or they relied instead on the computer-controlled off-beat tones as references. In fact, there were considerable individual differences in the production condition: Some participants still showed a small advantage for on-beat tapping, whereas others showed a disadvantage for on-beat tapping. This suggests that participants adopted different strategies of coping with variable on-beat tones, with the first strategy probably being the less beneficial of the two.

3. Experiment 2

The purpose of Experiment 2 was to replicate the basic finding of the synchronization condition of Experiment 1, but with a different manipulation of metrical interpretation that concerned the period of the imposed beat (i.e., the meter itself and, consequently, group size), rather than its phase. Because replication of the effect of metrical interpretation on the synchronization threshold was the main concern at this point, a production condition was not included in Experiment 2.

3.1. Methods

3.1.1. Participants

The same participants returned for Experiment 2.

3.1.2. Materials and tasks

The materials and tasks are shown in Fig. 3. One of the three melodies of Experiment 1, the one starting on C, was used again but it was notated in two different meters: 12/8 and 3/2. (These meters were chosen because they avoid bar lines between groups, but alternative notations with bar lines – 6/8 or 3/8 for 12/8, and 4/8 for 3/2 – would have served just as well for the present purpose.) As indicated by the beaming of notes, the 12/8 meter implies four groups of three notes each (i.e., 4 × 3), whereas the 3/2 meter implies three groups of four notes each (i.e., 3 × 4), with the beat being group-initial in each case. The wedges below the notation indicate the target tones in the tapping tasks. P1 and P2 refer to the position of the first tap in the first group (coinciding with the first or second tone, respectively), whereas T3 and T4 refer to the number of taps made in each cycle of the melody (three or four). In the 4 × 3 condition, P1T4 is an on-beat tapping task and P2T4 is an off-beat tapping task (both are replications of conditions of Experiment 1), whereas P1T3 and P2T3 are *polymetric* tapping tasks in the sense that the taps suggest a 3 × 4 grouping but are to be executed within the 4 × 3 mental framework. Conversely, in the 3 × 4 condition, P1T3 is an on-beat tapping task and P2T3 is an off-beat tapping task, whereas P1T4 and P2T4 are polymetric in that the taps encourage a 4 × 3 grouping but are to be executed within the 3 × 4 metrical interpretation. Note that the target tones are exactly the same in tasks that have the same name; only the metrical interpretation is different. On-beat tapping was predicted to be easier than off-beat tapping, and the polymetric tasks were expected to be most difficult.

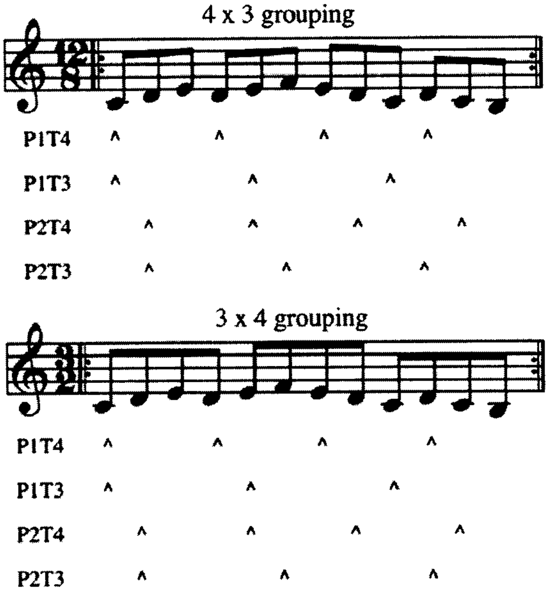

Fig. 3. Materials and tasks of Experiment 2. Wedges indicate target tones for tapping.

3.1.3. Procedure

The procedure was the same as in the synchronization condition of Experiment 1. That is, in Session 1 the eight tasks were performed in the order given in Fig. 3, and in Session 2 they were done in the reverse order. Because of the expected difficulty of the polymetric tasks, participants were encouraged to practice each of them for a few minutes before starting a run in Session 1. They were also given the option of starting at an IOI of 250 ms instead of 200 ms, if necessary.

3.2. Results

Difficulties in maintaining the designated metrical interpretation were encountered about as often as in Experiment 1. Spontaneous switches were reported in 29 out of 128 runs (23%). There were no striking differences in their incidence between Session 1 (17) and Session 2 (12), between the two metrical interpretations (13, 16), between the polymetric conditions (17) and the other conditions (12), and between on-beat and off-beat tapping (4, 8). Again, any failures to maintain the intended metrical interpretations worked in favor of the null hypothesis.

The synchronization thresholds are shown in Fig. 4. It is evident that the on-beat tapping tasks (P1T4 in the 4 × 3 condition and P1T3 in the 3 × 4 condition) were easier than the other tasks, as predicted. Surprisingly, the polymetric tapping tasks (P1T3 and P2T3 in the 4 × 3 condition, and P1T4 and P2T4 in the 3 × 4 condition) were no more difficult than the off-beat tapping tasks (P2T4 in the 4 × 3 condition and P2T3 in the 3 × 4 condition), but they did show a greater range of synchroniza-

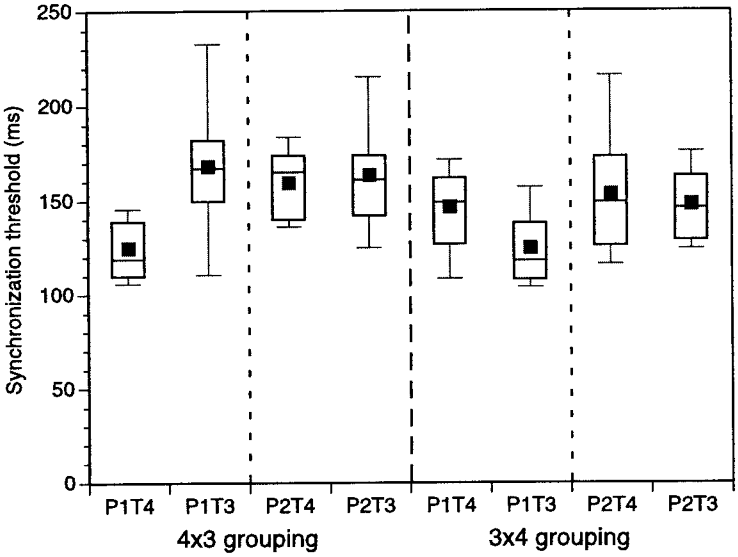

Fig. 4. Results of Experiment 2. Box plots show means (squares), medians, quartiles, and 1st and 9th deciles for 16 data points (8 participants × 2 sessions). The on-beat tapping conditions are P1T4 for 4 × 3 grouping and P1T3 for 3 × 4 grouping.

tion thresholds. This was mainly due to one participant (the one with the least musical training) who had particular difficulty with these tasks.

A 2 × 2 × 2 × 2 repeated-measures ANOVA with the variables of session, metrical interpretation, position of first tap in the first group, and number of taps per cycle revealed a significant Metrical Interpretation × Position × Number of Taps interaction, $F(1,7) = 15.33$, $p < .006$, $\eta_p^2 = .69$, which reflects the lower synchronization thresholds in the two on-beat tapping tasks than in the six other tasks and is the most important result.[6] Other significant effects included the main effect of metrical interpretation (slightly better performance with 3 × 4 than with 4 × 3), the main effect of position, and the Metrical Interpretation × Number of Taps interaction, the latter two effects being offshoots of the triple interaction.

To unpack the triple interaction, four 2 × 2 × 2 ANOVAs were conducted on subsets of the data. The first ANOVA compared the on-beat and off-beat tapping tasks (P1T4 and P2T4 in 4 × 3 grouping; P1T3 and P2T3 in 3 × 4 grouping) and revealed only a reliable main effect of position, $F(1,7) = 23.53$, $p < .002$, $\eta_p^2 = .77$, which reflects the superiority of on-beat over off-beat tapping. The second ANOVA compared the four polymetric tapping tasks and yielded a main effect of metrical interpretation, $F(1,7) = 19.55, p < .003$, $\eta_p^2 = .74$, reflecting better performance with 3 × 4 than with 4 × 3 grouping, and a marginally significant effect of session, a practice effect. A third ANOVA considered all P1 tasks – the two on-beat tapping tasks

---

[6] The interaction was still significant after removing the author's data, $F(1,6) = 13.19, p < .02$, $\eta_p^2 = .69$.

and two polymetric tasks. Here there was a significant Metrical Interpretation × Number of Taps interaction, $F(1,7) = 11.59$, $p < .02$, $\eta_p^2 = .62$, which reflects the advantage of on-beat over polymetric tapping within each metrical interpretation. The main effect of number of taps also reached significance (making three taps was a little easier than making four taps). The fourth ANOVA considered all P2 tasks – the two off-beat tapping tasks and two polymetric tasks. Here the significant effects were the main effects of session (a practice effect) and of metrical interpretation (the 3×4 tasks being a little easier than the 4×3 tasks).

An analysis of the final taps of unsuccessful trials showed that the tendency to lag behind (72%) was even stronger than in Experiment 1. An ANOVA on these data revealed a significant effect of number of taps, $F(1,7) = 13.92$, $p < .007$, $\eta_p^2 = .34$: The tendency to lag behind was stronger when four taps were made per cycle than when only three taps were made.

4. Discussion

4.1. Metrical interpretation and musical expertise

The results of the synchronization tasks of both experiments demonstrate that the difficulty of a particular tapping task depends not on the identity of the target tones (a stimulus property) but on the position these tones occupy within a particular metrical interpretation (a mental framework). Regardless of whether the metrical interpretation requires maintenance of a particular phase (Experiment 1) or period (Experiment 2) of the beat, synchronization is facilitated when the taps coincide with the beat. The mean synchronization thresholds for on-beat tapping resemble those obtained previously for tapping with every 2nd, 3rd, or 4th tone in a monotone isochronous sequence (Repp, 2003, 2005d).

The advantage of on-beat over off-beat and polymetric synchronization was obtained despite the fact that participants had occasional difficulties in maintaining the required metrical interpretation. Such difficulties work against the observed advantage because they reduce the differences among tasks. On the whole, however, participants were quite successful in manipulating and maintaining their metrical interpretations, as one should expect given their extensive musical training.

It might be asked whether the present findings can be generalized to individuals with little musical training. Most likely, non-musicians would not be able to succeed in the present tasks, which were quite challenging even for musicians. One notable feature of the present methods is that metrical interpretations were induced by means of musical notation. This would not be possible with musically illiterate or less sophisticated participants, but different preceding metrical contexts (e.g., a simple metronome, or the melody with the beat tones accentuated), as well as different starting points, could be used instead. There is little doubt that different metrical interpretations could be induced in that fashion even in musically untrained individuals (see, e.g., Desain & Honing, 2003). However, non-musicians probably could not be relied upon to maintain and carefully monitor their metrical interpretation throughout a run, and to stop and try to recover a lapsed interpretation. They would be more likely to yield to influences of stimulus structure and revert to a preferred metrical interpretation in the course of a run, perhaps without noticing it. Furthermore, their synchronization thresholds would be much higher, and they might not be able to carry out certain tasks at all, such as the polymetric tapping required in Experiment 2, and even off-beat tapping. Therefore, the present results should be regarded as reflecting musical expertise.

4.2. The role of musical structure

In Experiment 1, different metrical interpretations were induced not only by means of musical notation but also by different starting pitches of the melodic sequence. Because listeners tend to perceive the beat on the first event of a sequence (e.g., Toiviainen & Snyder, 2003), the starting pitch helped launch the required metrical interpretation. However, it hardly can have played a role later in the sequence, where maintenance of the induced beat was up to the participant. In Experiment 2, starting pitch was not a factor because it was held constant.

It was not the purpose of the present study to determine which metrical interpretation listeners prefer, given the structure of the stimulus sequence. In that respect, it contrasts with studies of beat induction. Indeed, the literature on beat induction generally gives the impression that listeners are entirely at the mercy of stimulus properties. By contrast, the emphasis of the present study is on the relative liberty listeners have to impose their own beat on a metrical sequence. To be sure, stimulus factors are important in metrical interpretation, but situations in which a required metrical interpretation overrides stimulus structure abound in music. The required metrical interpretation is usually both specified by notation and induced by preceding context. Given these constraints, composers and performers are free to introduce all kinds of counterevidence to the prevailing beat (e.g., syncopations, off-beat accents, and various melodic patterns), assuming all along that listeners will be able to hold on to the beat. This is a common strategy for introducing rhythmic tension that subsequently can be released.

Some indirect information about participants' preferred metrical interpretation(s) is provided by the switch frequencies in Experiment 1. The switch frequency was twice as high for the melody starting on D than for the melodies starting on C or E, which suggests that the D melody was less stable metrically. It is known that tones preceding a point of change of direction ("turn") in a pitch contour (i.e., local pitch maxima and minima) tend to be perceived as accented (see, e.g., Drake & Palmer, 1993; Jones & Pfordresher, 1997). A look at Fig. 1 shows that in the E melody all four beats (i.e., the initial tones of groups) coincide with turns in the pitch contour, in the C melody only the second and fourth beats do, and in the D melody none does. Another factor that affects grouping, in particular, is parallelism – the repetition of pitch contours (Temperley & Bartlette, 2002). In both the C and D melodies the pitch contour of a group is repeated (transposed upward or downward) before it changes, whereas in the E melody each group has a unique pitch contour. Together, the effects of turns and repetitions can explain why the C and E melodies were more stable than the D melody.

In addition to reverting to metrical interpretations favored by the sequence structure, participants may also have had a tendency to align the beats with their taps (cf. Repp, 2005b). Indeed, switches were more frequent in off-beat than in on-beat tapping. However, the fact that switches occurred at all during on-beat tapping indicates that the structure of the sequence also played a role. Structural factors are strong in non-isochronous rhythmic sequences (Povel & Essens, 1985) and are the likely reason for the previous failure to find reliable effects of metrical interpretation on tapping with such sequences (Repp, 2005b). The present sequences were less strongly biased in that regard because, as mentioned earlier, pitch contour is a relatively weak cue to metrical structure (Hannon et al., 2004). The tendency for the subjective beat to move into congruence with the taps also seems to have been weaker than in the earlier monotone materials, probably because the pitch variation provided better anchors for intended metrical interpretations.

4.3. The multistability of musical meter

The present results demonstrate that musical sequences are essentially (more or less) ambiguous figures that can be organized perceptually in multiple ways, not only by different individuals or in different contexts but also by the same individual at different times, and at will – as long as the sequence structure does not provide very strong support for a particular metrical interpretation. With respect to perceptual multistability, melodic sequences are comparable to visual reversible figures, although there is a larger number of categorical alternatives in metrical interpretation, and there may be stronger a priori asymmetries among those alternatives. Two striking differences between visual chimeras and melodies, however, are in the switching rate and in the extent to which the switching is under intentional control. Visual figures tend to reverse every few seconds, and the viewer's intentions have only relatively mild effects on the rate of switching. By contrast, metrical interpretations are stable for minutes before they switch, if they switch at all. If they switch, it seems unlikely that they will ever switch back spontaneously to the original interpretation. To return them to their original state, a considerable mental effort is required from the participant, who often succeeds. Thus, although there are multiple categorical attractors, as it were, they do not adapt rapidly, as they do in the visual case. Rather, there is a prolonged competition between attractors in which the strongest (the one most favored by bottom-up factors) may win eventually, if the listener does not intervene. However, it is listener intervention that establishes and maintains particular metrical interpretations and ensures their relative endurance. Persistence in metrical interpretations is generally required and desirable in music, in contrast to visual figures, where the demand characteristics of experiments favor changing percepts (Rock et al., 1994). If participants were instructed to shift metrical interpretations every few seconds, they might be able to do so, and indeed there is music (e.g., by Stravinsky) in which frequent changes of meter seem to require this facility. However, most familiar music has a constant meter, and listeners' natural tendency is to maintain a metrical interpretation. Therefore, metrical interpretations are much more resistant to change than are ambiguous visual figures.

What exactly goes on in listeners' minds and brains when they impose a beat and thereby make themselves hear a sequence as a particular metrical structure? Repp (2005b) suggested three theoretical possibilities, which are not mutually exclusive: Voluntarily heightened attention at the temporal beat positions (cf. Barnes & Jones, 2000; Large & Jones, 1999); generation of a covert action that coincides with the intended beat position; and internal entrainment of a complete rhythmic pattern in which the intended beats are emphasized somehow.[7] Heightened attention may go along with a readiness to act at a particular time, even when motoric activity is suppressed, and hence may also go along with internal (imaginary) action, but it is also conceivable that modulation of attention alone is sufficient to undergird metrical interpretation. The latter two hypotheses have in common the necessary involvement of the action system but differ in the extent of the involvement: Only on beats, or on all events but more strongly on beats.

Any explanation must be able to account for the striking phenomenology of metrical interpretation: An identical sequence interpreted as a different metrical structure simply does not sound the same, even though the constancy of the stimulus may be recognized when it is pointed out, just as a Necker cube does not look the same when viewed in different orientations and yet remains recognizably the same stimulus. Action, and even the mere planning of action, can affect the perception of visual stimuli (e.g., Schubö, Prinz, & Aschersleben, 2004; Wohlschläger, 2000), and so it should not be too surprising that imaginary action (if it underlies metrical interpretation) can also change the perception of auditory stimuli. However, it should not be concluded that the synchronization advantage for on-beat tapping is mediated by the perceptual change brought about by metrical interpretation, although that remains a theoretical possibility. Rather, it is more straightforward to assume that the on-beat advantage results from the in-phase coordination of an overt movement (finger taps) with a covert one (internal beats).

Different hypotheses about the processes underlying metrical interpretation could be addressed with the methods of neuroscience available now. Brochard, Abecasis, Potter, Ragot, and Drake (2003) have found evidence in ERPs for a more vigorous brain response to attenuated odd than to attenuated even events in an isochronous monotone sequence (IOI = 600 ms), which led them to infer that listeners automatically impose a binary metrical structure on such a sequence, with the beat occurring on the odd elements. However, they did not attempt to manipulate the metrical interpretation to see whether the beat and the enhanced brain potentials could be shifted intentionally to the even sequence elements. A recent magnetoencephalographic study using the materials of Repp (2005b) has found preliminary evidence for different brain activation patterns corresponding to differ-

---

[7] If required, internal beats could easily be "acted out" with another body part while carrying out the synchronization task with the preferred hand. However, that was not permitted in the present experiments and, as far as could be ascertained, did not occur. Such a strategy would facilitate off-beat tapping and thus would work in favor of the null hypothesis.

ent endogenous metrical interpretations, possibly involving premotor areas (Iversen, Repp, & Patel, 2005). Future studies along these lines should yield more specific information about the brain areas involved in the generation of an internal beat or rhythm, and to reveal the relative contributions of attentional and motor processes to metrical interpretation.

Acknowledgments

This research was supported by NIH Grant MH-51230, as well as by NIH Grants HD-01994 (Carol Fowler, P.I.) and DC-03663 (Elliot Saltzman, P.I.). Peter Keller and three anonymous reviewers made helpful comments on the manuscript.

References

Andrews, T. J., Schluppeck, D., Homfray, D., Matthews, P., & Blakemore, C. (2002). Activity in the fusiform gyrus predicts conscious perception of Rubin's vase-face illusion. *NeuroImage, 17*, 890–901.

Attneave, F. (1971). Multistability in perception. *Scientific American, 225*, 62–71.

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

Bolton, T. L. (1894). Rhythm. *American Journal of Psychology, 6*, 145–238.

Brochard, R., Abecasis, D., Potter, D., Ragot, R., & Drake, C. (2003). The "ticktock" of our internal clock: Direct brain evidence of subjective accents in isochronous sequences. *Psychological Science, 14*, 362–366.

Desain, P., & Honing, H. (1999). Computational models of beat induction: The rule-based approach. *Journal of New Music Research, 28*, 29–42.

Desain, P., & Honing, H. (2003). The formation of rhythmic categories and metric priming. *Perception, 32*, 341–365.

Deutsch, D. (1987). The tritone paradox: Effects of spectral variables. *Perception & Psychophysics, 41*, 563–575.

Drake, C., & Palmer, C. (1993). Accent structures in performance. *Music Perception, 10*, 343–378.

Drake, C., Penel, A., & Bigand, E. (2000). Tapping in time with mechanically and expressively performed music. *Music Perception, 18*, 1–24.

Flach, R. (2005). The transition from synchronization to continuation tapping. *Human Movement Science, 24*, 465–483.

Fraisse, P., & Ehrlich, S. (1955). Note sur la possibilité de syncoper en fonction du tempo d'une cadence [Note on the possibility of syncopation as a function of sequence tempo]. *L'Année Psychologique, 55*, 61–65.

Fraisse, P., & Voillaume, C. (1971). Les repères du sujet dans la synchronization et dans la pseudo-synchronisation [The subject's references during synchronization and during pseudo-synchronization]. *L'Année Psychologique, 71*, 359–369.

Girgus, J. J., Rock, I., & Egatz, R. (1977). The effect of knowledge of reversibility on the reversibility of ambiguous figures. *Perception & Psychophysics, 22*, 550–556.

Hannon, E. E., Snyder, J. S., Eerola, T., & Krumhansl, C. L. (2004). The role of melodic and temporal cues in perceiving musical meter. *Journal of Experimental Psychology: Human Perception and Performance, 30*, 956–974.

Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.

Hochberg, J., & Peterson, M. A. (1987). Piecemean organization and cognitive components in object perception: Perceptually coupled responses to moving objects. *Journal of Experimental Psychology: General, 116*, 370–380.

Iversen, J. R., Repp, B. H., & Patel, A. D. (2005). *Modulation of brain responses to rhythmic sequences by subjective metrical interpretation*. Poster presented at the *Neurosciences and Music II* conference, Leipzig, Germany, May.

Jones, M. R., & Pfordresher, P. Q. (1997). Tracking musical patterns using joint accent structure. *Canadian Journal of Experimental Psychology, 51*, 271–290.

Kelso, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge, MA: MIT Press.

Köhler, W., & Wallach, H. (1944). Figural after-effects, an investigation of visual processes. *Proceedings of the American Philosophical Society, 88*, 269–357.

Kornmeier, J., & Bach, M. (2004). Early neural activity in Necker-cube reversal: Evidence for low-level processing of a gestalt phenomenon. *Psychophysiology, 41*, 1–8.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How we track time varying events. *Psychological Review, 106*, 119–159.

Large, E. W., & Kolen, J. F. (1994). Resonance and the perception of musical meter. *Connection Science, 6*, 177–208.

Liebert, R. M., & Burk, B. (1985). Voluntary control of reversible figures. *Perceptual & Motor Skills, 61*, 1307–1310.

Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception, 11*, 409–464.

Pelton, L. H., & Solley, C. M. (1968). Acceleration of reversals of a Necker cube. *American Journal of Psychology, 81*, 585–588.

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

Reisberg, D., & O'Shaughnessy, M. (1984). Diverting subjects' concentration slows figure reversals. *Perception, 13*, 461–468.

Repp, B. H. (1997). Spectral envelope and context effects in the tritone paradox. *Perception, 26*, 645–665.

Repp, B. H. (2003). Rate limits in sensorimotor synchronization with auditory and visual sequences: The synchronization threshold and the benefits and costs of interval subdivision. *Journal of Motor Behavior, 35*, 355–370.

Repp, B. H. (2005a). Rate limits of on-beat and off-beat tapping with simple auditory rhythms: 1. Qualitative observations. *Music Perception, 22*, 479–496.

Repp, B. H. (2005b). 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. (2005c). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review, 12*, 969–992.

Repp, B. H. (2005d). *Counting auditory events at rapid rates: Emergence of metrical structure from automatic subdivision of repeated counts*. Manuscript submitted for publication.

Repp, B. H., London, J., & Keller, P. E. (2005). Production and synchronization of uneven rhythms at fast tempi. *Music Perception, 23*, 61–78.

Repp, B.H., & Saltzman, E.L. (2002). *Influences of grouping and metrical structure on the kinematics of rhythmic finger tapping*. Unpublished manuscript (http://www.haskins.yale.edu/haskins/STAFF/repp.html).

Roberts, M., & Summerfield, Q. (1981). Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. *Perception & Psychophysics, 30*, 309–314.

Rock, I., Hall, S., & Davis, J. (1994). Why do ambiguous figures reverse?. *Acta Psychologica 87*, 33–59.

Schubö, A., Prinz, W., & Aschersleben, G. (2004). Perceiving while acting: Action affects perception. *Psychological Research, 68*, 208–215.

Sloboda, J. A. (1983). The communication of musical metre in piano performance. *Quarterly Journal of Experimental Psychology, 35A*, 377–396.

Sloboda, J. A. (1985). Expressive skill in two pianists: Metrical communication in real and simulated performances. *Canadian Journal of Psychology, 39*, 273–293.

Snyder, J., & Krumhansl, C. L. (2001). Tapping to ragtime: Cues to pulse finding. *Music Perception, 18*, 455–489.

Temperley, D., & Bartlette, C. (2002). Parallelism as a factor in metrical analysis. *Music Perception, 20*, 117–150.

Toiviainen, P., & Snyder, J. S. (2003). Tapping to Bach: Resonance-based modeling of pulse. *Music Perception, 21*, 43–80.

Toppino, T. C. (2003). Reversible-figure perception: Mechanisms of intentional control. *Perception & Psychophysics, 65*, 1285–1295.

Warren, R. M. (1961). Illusory changes of distinct speech upon repetition—the verbal transformation effect. *British Journal of Psychology, 52*, 249–258.

Warren, R. M. (1999). *Auditory perception: A new analysis and synthesis.* Cambridge, UK: Cambridge University Press.

Wohlschläger, A. (2000). Visual motion priming by invisible actions. *Vision Research, 40*, 925–930.