RESEARCH ARTICLE

See what I hear? Beat perception in auditory and visual rhythms

Jessica A. Grahn

Received: 9 December 2011/Accepted: 27 April 2012/Published online: 24 May 2012 © Springer-Verlag 2012

Abstract Our perception of time is affected by the modality in which it is conveyed. Moreover, certain temporal phenomena appear to exist in only one modality. The perception of temporal regularity or structure (e.g., the 'beat') in rhythmic patterns is one such phenomenon: visual beat perception is rare. The modality-specificity for beat perception is puzzling, as the durations that comprise rhythmic patterns are much longer than the limits of visual temporal resolution. Moreover, the optimization that beat perception provides for memory of auditory sequences should be equally relevant to visual sequences. Why does beat perception appear to be modality specific? One possibility is that the nature of the visual stimulus plays a role. Previous studies have usually used brief stimuli (e.g., light flashes) to present visual rhythms. In the current study, a rotating line that appeared sequentially in different spatial orientations was used to present a visual rhythm. Discrimination accuracy for visual rhythms and auditory rhythms was compared for different types of rhythms. The rhythms either had a regular temporal structure that previously has been shown to induce beat perception in the auditory modality, or they had an irregular temporal structure without beat-inducing qualities. Overall, the visual rhythms were discriminated more poorly than the auditory rhythms. The beat-based structure, however, increased accuracy for visual as well as auditory rhythms. These results indicate that beat perception can occur in the visual modality and improve performance on a temporal discrimination task, when certain types of stimuli are used.

Keywords Music · Rhythm · Timing · Auditory perception · Visual perception · Cross-modal comparisons

Introduction

Accurate perception of time is crucial for normal perceptual, cognitive, and motor function. The perceptual accuracy of different temporal phenomena, however, is influenced by the modality in which the phenomenon is perceived. Auditory stimuli are perceived and discriminated more accurately than visual stimuli (Glenberg and Jona 1991; Guttman et al. 2005; Handel and Buffardi 1969; Repp and Penel 2002). The auditory system has a more fine-grained temporal resolution than the visual system (on the order of milliseconds, as opposed to tens of milliseconds (Holcombe 2009). However, modality differences in temporal resolution cannot explain why some temporal phenomena are elicited in certain modalities but not others. In particular, spontaneous perception of an underlying 'beat' [a perceived pulse that marks equally spaced points in time (Large and Palmer 2002; Nettl 2000)] in temporal sequences differs strongly between modalities. Perception of a beat occurs without effort in auditory sequences that have a regular temporal structure, such as periodically regular events occurring at particular points in time. For example, in "twinkle, twinkle, little star," one can tap regularly to every syllable, every other syllable, or every fourth syllable, and still be synchronized to the music (Drake et al. 2000; Parncutt 1994). Thus, there are several levels of periodicity present. Music or rhythm that has regular periodicities will generally induce beat perception. Beat perception, in turn, leads to better representation of the rhythm itself, as evidenced by higher accuracy in

Department of Psychology, Brain and Mind Institute, University of Western Ontario, London, ON N6A 5B7, Canada e-mail: jgrahn@uwo.ca



J. A. Grahn (\boxtimes)

rhythm discrimination or reproduction (Grahn and Brett 2007; Patel et al. 2005).

However, beat perception is a temporal phenomenon that has far greater affinity with the auditory modality than the visual modality (Glenberg and Jona 1991; Glenberg et al. 1989; Guttman et al. 2005; Handel and Buffardi 1969; Repp and Penel 2002). This is true even though beat perception occurs when periodicities are present at a rate of approximately 300-900 ms (van Noorden and Moelants 1999; Parncutt 1994), which is far from the limits of temporal resolution of the visual modality. The modalityspecificity of beat perception is even more puzzling when one considers the following: (1) Time itself is 'amodal', and conveyed in multiple sensory modalities, (2) The neural substrates of auditory and visual timing tasks overlap substantially (Schubotz et al. 2000; Wiener et al. 2010), and (3) Any optimization that beat perception offers to auditory timing accuracy should be equally beneficial to the visual modality. Thus, why do certain temporal phenomena, like beat perception, appear to be modality specific? Answering this question requires a more detailed consideration of the background literature.

Auditory versus visual performance in temporal tasks

Several studies have demonstrated that the auditory system is generally better than the visual system with temporal tasks (Glenberg and Jona 1991; Glenberg et al. 1989; Guttman et al. 2005; Handel and Buffardi 1969; Repp and Penel 2002). Much empirical evidence for this claim focuses on increased temporal sensitivity in the auditory modality compared to the visual modality. For example, short temporal intervals ($\sim 50-2,000$ ms) are discriminated and reproduced more accurately when the stimuli are auditory compared to when they are visual (Goldstone and Lhamon 1972; Grondin 1993; Stauffer et al. 2012). Not only is this true for timing of single intervals, but also for multiple intervals put together in a temporal sequence (Collier and Logan 2000; Gault and Goodfellow 1938; Handel and Buffardi 1969). In sequences, however, the auditory advantage may be limited to temporal patterns that involve interval durations that are related by simple integer ratios (Glenberg and Jona 1991; Glenberg et al. 1989; Schubotz et al. 2000), such as 1:2 (e.g., intervals of 200 and 400 ms). This would suggest that the source of the auditory advantage in temporal processing may be a modality difference in the 'beat-based' encoding of the rhythms. Beatbased encoding enables all intervals in the rhythm to be encoded as low-integer multiples of an underlying beat interval, rather than a series of unrelated durations. The idea of a beat-based advantage is supported by the fact that that rhythms that are very slow, and therefore outside the range of beat perception, do not show the same auditory performance advantage (Glenberg and Jona 1991).

Beyond the perceptual domain, studies that investigate tapping along to auditory or visual stimuli have also reported modality differences in synchronization performance (Chen et al. 2002; Patel et al. 2005; Repp and Penel 2002; Semjen and Ivry 2001). For example, Patel et al. (2005) used rhythmic patterns that varied from strongly metrical (having an easily detected beat-based structure) to weakly metrical (having a less easily detected beat-based structure) and found that visual rhythms elicited much worse synchronization performance than auditory rhythms. Moreover, accuracy in the visual condition did not appear to differ between the strong and weak beat-based structure (Patel et al. 2005), leading the authors suggest that beat perception and synchronization have a special affinity with the auditory system. Thus, studies involving both perception and production find little evidence for visual stimuli inducing a sense of the beat.

Potential explanations for lack of observed visual beat perception

Why might the visual system be less sensitive to beat-based structure? Surely the cognitive benefits of perceiving the structure (e.g., reducing working memory load by chunking with respect to an underlying beat) would be as useful in the visual domain as in the auditory domain. One possibility is that humans are not sensitive to the presence of beat structure per se, but rather we notice the structure if we are cued to its presence somehow, and these cues exist in the auditory domain but not in the visual domain. For example, although several studies have shown that when durations of intervals are related by integer ratios, they are reproduced more accurately than when the durations are related by noninteger ratios (e.g., 1:2.15, or 200 and 430 ms), other work shows that we do not always show an integer-ratio benefit (Grahn and Brett 2007). Grahn and Brett (2007) designed rhythms to test whether integer ratios were the key factor in beat-based rhythm performance, or if an additional cue, such as a 'temporal accent' was required for participants to be sensitive to the beat structure. Unlike volume or pitch accents (which are created by increasing the volume or changing the pitch of tones in a sequence) temporal accents arise solely from the pattern of durations in the sequence. For example, if two acoustically identical tones are sounded one after the other (close in time), the second tone is perceived as more salient than the first (Povel and Okkerman 1981). If 3 tones are sounded in a row, the first and third are perceived as more salient. Finally, a tone that occurs after a relative period of no tones will also be perceived as salient (Povel and Okkerman



1981). Thus, although there is no acoustic difference between tones that have a temporal accent and tones that do not, a difference in salience is perceived because of the temporal pattern. By aligning the occurrence of these temporal accents with periodicities (i.e., the beat) in the rhythm, the regularity in the sequence is highlighted to the listener, and they can then use the periodic beat to structure (or chunk) their perception of the rhythm. Grahn and Brett (2007) found that the best rhythm reproduction accuracy was achieved when temporal accents were aligned with the beat of integer-ratio rhythms. When temporal accents were not aligned, integer-ratio rhythms were not performed any better than noninteger-ratio rhythms. Thus, regular temporal accents appear to play an important role in whether a beat is perceived in integer-ratio rhythms. When temporal accents and integer ratios are present, participants are sensitive to the beat structure and can use beat-based encoding to remember the sequence, enhancing performance.

Accordingly, one reason that the visual system may not show better accuracy for beat-based rhythms is that we do not currently have evidence for perception of temporal accents in visual stimuli. For example, do two flashes close to each other in time convey a sense of 'accent' on the second flash? This and other forms of visual salience induced by temporal structure remain to be empirically studied, but lack of accents could be one reason that there is little evidence for beat structure improving performance of visual rhythms. In addition, the time-span of sensory memory appears to be shorter for visual information (0.3–1.0 s; Averbach and Coriell 1961; Eriksen and Collins 1967; Sakitt 1976; Sperling 1960) than for auditory information (1.5-4.0 s; Cowan 1984; Crowder 1982; Darwin et al. 1972). The more rapid decay of visual information may reduce the likelihood that higher-level temporal relationships (such as whether several successive durations are related by integer ratios) are perceived.

Alternatively, it may be that lack of evidence for visual beat perception results from the use of nonoptimal visual stimuli in previous studies. For example, musicians synchronize their playing to the visually perceived movements of a conductor. Empirically, the type of visual stimulus does affect synchronization of movement to visually presented sequences. Hove et al. (2010) compared tap synchronization performance to 300-600 ms intervals marked by flashes or different types of moving visual stimuli (e.g., a white bar moving up and down on the screen). Flashes elicited the worst synchronization performance of all tested stimuli. The best performance was elicited by visual stimuli that moved in a way that was compatible with the tapping motion of the finger (e.g., synchronizing the bottom of the tap trajectory with the bottom of the visual stimulus trajectory). However, motion per se was not enough to improve performance: synchronization with spatially incongruent motion (e.g., synchronizing the bottom of the tap trajectory to the top of the stimulus trajectory, which puts the tapping motion into anti-phase with the visual motion) elicited timing accuracy that was not significantly better than that for flashes. Orthogonal stimulus motion (synchronizing with a bar moving from side to side) elicited accuracy between congruent and incongruent motion. Thus, visual motion appears to improve tapping performance overall, although congruency of action motion and stimulus motion is clearly playing a role. However, whether motion will improve *perceptual* timing accuracy without action, or lead to visual beat perception, is unknown.

Another way to optimize visual stimuli for temporal processing may be to present stimuli at particular spatial locations. Schubotz et al. (2000) found similar error rates for auditory and visual modalities when participants monitored simple rhythmic sequences (composed of 3 intervals that were not consistently related by low integer ratios). Their visual stimuli did not have continuous motion, but rather 2 squares that appeared in different spatial positions for the length of each interval. Participants could therefore associate each spatial position with a time interval in the sequence. Although the deviants were somewhat large (transposing two intervals in a sequence that was only three intervals long), performance was not at ceiling ($\sim 10 \%$ error rate) and was not different between modalities. Thus, spatial association may also optimize temporal perception in visual rhythmic sequences, at least for simple sequences.

Additional support for the possibility of inducing visual beat perception (with optimized stimuli) comes from neuroimaging. Although synchronization to auditory and visual stimuli can elicit different patterns of brain activity (Jäncke et al. 2000; Jantzen et al. 2005), perceptual monitoring of visual and auditory rhythms activates relatively similar networks of (mainly motor) areas (Schubotz et al. 2000; Shih et al. 2009). Similarity in the neural correlates of temporal sequence perception across modalities implies that there may be potential to create similar perceptions. In addition, when visual rhythms have been preceded by auditory versions of the same rhythm (Grahn et al. 2011; McAuley and Henry 2010), a sense of beat can be induced for the visual rhythms. Moreover, activity in the putamen, an area that responds to beat perception in the auditory modality, increases during visual rhythms that are perceived after auditory rhythms (Grahn et al. 2011). It is therefore possible, under certain circumstances, to induce the feeling of a beat in visual rhythms, and the underlying neural substrates appear to be similar to those underlying beat perception in the auditory modality. The question remains, however, whether visual beat perception can be induced without concurrent or prior auditory exposure.



Issues addressed in the current study

To examine this issue, the current study used a novel visual stimulus presentation. Performance on a discrimination task was measured to examine whether a beat-based performance improvement could be observed for visual rhythms. The same rhythms as used in Grahn and Brett (2007) are used in the current study. There are three types of rhythmic sequences. Metric simple rhythms are the integer-ratio (1:2:3:4) rhythms that also have regular temporal accents aligned with the beat. Metric complex rhythms are rhythms in which the intervals are related by integer ratios, but the temporal accents are not regular and hence do not mark an underlying beat. Finally, nonmetric rhythms comprise intervals that are related by noninteger (1:1.4:3.5:4.5), or complex, ratios. In this case, no beat can be perceived as there is no underlying temporal regularity to the rhythm, nor any regular temporal accents.

To present visual rhythms, a line was used. The line "jumped" to different spatial positions, rotating around its midpoint, and stopping at each position for the duration of the interval in that particular rhythm (see Fig. 2). Participants compared the third presentation of a rhythm with the first two presentations, deciding whether the third presentation contained a temporal deviant. Participants also completed an auditory version of the task, so performance between the two modalities could be directly compared. On the basis of previous work, better discrimination performance was predicted in the auditory condition for metric simple rhythms compared to the metric complex and nonmetric rhythms (Grahn and Brett 2009). For the visual rhythms, it was predicted that if any difference in discrimination accuracy between conditions was observed, it would parallel that observed for auditory rhythms.

Methods

Materials

The rhythms were created using integer-ratio and noninteger-ratio related sets of intervals. The integer-ratio intervals were related by ratios of 1:2:3:4, and the noninteger-ratio intervals were related by 1:1.4:3.5:4.5. The shortest interval (i.e., 1) ranged from 220 to 270 ms, in 10 ms steps. The other intervals in the rhythm were multiples of the shortest interval. Three types of rhythms were created. Two types used integer-ratio intervals. For metric simple rhythms, the intervals were arranged in groups of four units (e.g., in the sequence 211314, an interval onset consistently occurs every four units). The patterns were constructed to induce a perceptual accent at the beginning of each group of four units (Povel and Okkerman 1981). The perceptual accents

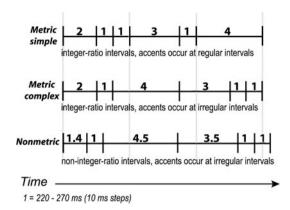


Fig. 1 Schematic representations of the intervals in three-sample rhythms. *Light gray lines* denote potential beat locations. In the metric simple condition, beat locations align with onsets in the stimulus, providing a clear beat structure. In the metric complex and nonmetric conditions, no regular beat locations regularly align with interval onsets

cued subjects to a regular beat structure, in which the beats coincided with the onset of each group (Essens 1995). In addition, if subjects choose a faster rate for the beat, the sequence can still be measured by that rate (e.g., measured in units of two rather than four). The intervals in the metric complex rhythms were arranged to not be reliably groupable into two, three, or four units (e.g., 341211), and therefore difficult to measure by any unit but the shortest unit. Since there is no regular grouping, and no regular accent occurrence, no beat is induced. Noninteger-ratio intervals were used to create nonmetric rhythms, using the same arrangement as the metric complex intervals but substituting the following noninteger-ratio interval lengths: 1.4 for 2, 3.5 for 3, and 4.5 for 4. Schematics of sample rhythms for each condition are shown in Fig. 1.

Participants had to compare the third presentation of a rhythm to the first two presentations. On half of the trials, the rhythm in the third presentation deviated from the first two presentations. The deviant sequences contained a transposition of intervals in the sequence. For example, the sequence 211413 can have as a possible deviant sequence 211431, in which the 3 interval and the 1 interval have been transposed. Only deviant sequences that were in the same category as the standard sequences were allowed. That is, a metric simple standard sequence could not have a metric complex deviant sequence, and a metric complex standard sequence could not have a metric simple deviant sequence. Thus, 43122 would not have 43212 as a possible deviant sequence, because the onsets would no longer be grouped in units of four and would violate the regular accent structure of the sequence. If these types of deviants were present, subjects could accomplish the task by detecting that onsets in deviant sequences either no longer



Table 1 Table of standard and deviant rhythm sequences for each condition

Metric simple		Metric complex		Nonmetric	
Standard	Deviant	Standard	Deviant	Standard	Deviant
5 Intervals					
22413	22431	33141	31341	1 1 3.5 4.5 3.5	1 3.5 1 4.5 3.5
31413	31431	41133	14133	1 3.5 1.4 4.5 1.4	3.5 1 1.4 4.5 1.4
31422	13422	41232	14232	1.4 3.5 1.4 4.5 1	1.4 3.5 1.4 1 4.5
41331	43131	11343	13143	3.5 3.5 1 4.5 1	3.5 3.5 4.5 1 1
43113	41313	13242	31242	4.5 1 1 3.5 3.5	1 4.5 1 3.5 3.5
43122	41322	23241	23214	4.5 1 1.4 3.5 1.4	1 4.5 1.4 3.5 1.4
6 Intervals					
112314	112134	121233	121323	1 3.5 1.4 3.5 1.4 1	1 1.4 3.5 3.5 1.4 1
112422	211422	122142	122412	1.4 1 4.5 1.4 1.4 1	1.4 4.5 1 1.4 1.4 1
211134	211314	124113	124131	1.4 1 4.5 3.5 1 1	1.4 4.5 1 3.5 1 1
222114	221124	221241	221214	3.5 1.4 3.5 1.4 1 1	3.5 1.4 1.4 3.5 1 1
223113	223131	231123	213123	4.5 1 1.4 1.4 1 1.4	1 4.5 1.4 1.4 1 1.4
311322	313122	321411	324111	4.5 1.4 1 3.5 1 1	4.5 1.4 1 1 3.5 1
211224	112224	132321	312321	1 1.4 1 1.4 3.5 3.5	1 1.4 1 3.5 1.4 3.5
211413	211431	214221	241221	1 1.4 1.4 1 4.5 1.4	1 1.4 1.4 4.5 1 1.4
221331	221313	214311	241311	1 1.4 4.5 1 1 3.5	1 1.4 4.5 1 3.5 1
312213	312231	323211	322311	1.4 1.4 1 1.4 4.5 1	1.4 1.4 1 4.5 1.4 1
411231	411213	412212	142212	1.4 3.5 1 1 1.4 3.5	1.4 3.5 1 1 3.5 1.4
422112	422211	421311	241311	3.5 1.4 1 4.5 1 1	3.5 1.4 1 1 4.5 1
7 Intervals					
1123122	1121322	1112412	1112142	1 1 1 1.4 4.5 1 1.4	1 1 1 4.5 1.4 1 1.4
2112231	2112213	1132131	1132311	1 1 3.5 1.4 1 3.5 1	1 1 3.5 1.4 3.5 1 1
2113113	2113131	1132212	1123212	1 1 3.5 1.4 1.4 1 1.4	1 1 1.4 3.5 1.4 1 1.
3141111	1341111	2141211	2411211	1.4 1 4.5 1 1.4 1 1	1.4 4.5 1 1 1.4 1 1
4111131	4111113	2331111	2313111	1.4 3.5 3.5 1 1 1 1	1.4 3.5 1 3.5 1 1 1
4221111	4211211	3113121	3131121	3.5 1 1 3.5 1 1.4 1	3.5 1 1 1 3.5 1.4 1
1111431	1111413	1314111	1311411	1 3.5 1 4.5 1 1 1	3.5 1 1 4.5 1 1 1
1122114	1121124	1411311	1141311	1 4.5 1 1 3.5 1 1	1 4.5 1 1 1 3.5 1
1123113	1123131	2123211	2132211	1.4 1 1.4 3.5 1.4 1 1	1.4 1 1.4 1.4 3.5 1
2211114	2112114	3114111	1314111	3.5 1 1 4.5 1 1 1	3.5 1 1 1 4.5 1 1
3121113	3121131	3221112	2321112	3.5 1.4 1.4 1 1 1 1.4	1.4 3.5 1.4 1 1 1 1.
3122112	1322112	4111221	1411221	4.5 1 1 1 1.4 1.4 1	1 4.5 1 1 1.4 1.4 1

aligned with the beat (if the shift is from simple to complex) or now aligned with the beat (if the shift is from complex to simple). The complete list of sequences is shown in Table 1. Other than interval order, all characteristics of the deviant sequences were the same as the standard sequences.

In the auditory condition, sine tones (rise/fall times of 8 ms) sounded for the duration of each interval, ending 40 ms before the specified interval length to create a silent gap that demarcated the intervals. The sequences used filled intervals, as piloting indicated performance was similar for empty and filled interval sequences, and filled intervals were more analogous to the visual presentation (presence of a stimulus for the duration of the interval, as opposed to

a brief presentation of the stimulus marking interval onset). We also added an additional tone, the length of the 1 interval, to the end of each sequence. This was done for all sequences, so that the stimuli were directly analogous to that used in a previous reproduction experiment (Grahn and Brett 2007). In that study, the onset of each reproduced interval was indicated by the participant's tap, and the reproduced lengths of each interval were measured by the intertap time. An additional tone was therefore needed at the end of each sequence. Otherwise, without this final onset for participants to tap, the last reproduced interval's length would not have been measured.

The visual stimulus was a line with a small cross-bar in the middle. The line appeared sequentially at different



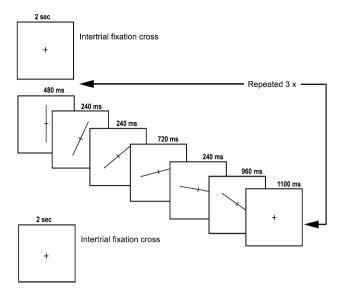


Fig. 2 Trial structure for visually presented rhythms. A fixation cross is presented for 2 s, followed by presentation of the rhythm using a *line* that jumps from one position to the next. The fixation cross is present for 1,100 ms between rhythm presentations. The cycle is repeated three times. In the third presentation, two time intervals may be transposed to create a deviant (not shown). The times shown are for a six-interval 211314 rhythm created with 1 = 240 ms

positions around a central axis of rotation, with the positions separated by 25° increments. The positions were vertical (0°) , 25° , 50° , 75° , 100° , 125° , and 150° . The five-and six-interval sequences jumped to only the first five or six positions, respectively. The amount of time spent at each position was equivalent to the duration of that interval in the rhythmic sequence. After each presentation of the sequence, a fixation cross was displayed and the new presentation started at the vertical position. A schematic of the visual presentation is shown in Fig. 2.

Procedure

Seventeen participants completed the experiment. Participants were not pre-selected for music or dance training. Twelve participants had two or more years of musical training (2–5 years, apart from one participant with 17 years), the remaining five participants had 6 months or less training (one with 6 months of training, the rest with none). Five participants had one or more years of dance training (1–6 years, apart from one participant with 14 years), the remaining 12 had less than 6 months of training (one with 5 months, the rest with none).

Auditory rhythms were presented in one half of the experiment and visual rhythms in the other half. The modality order was chosen at random for each participant. In the auditory condition, rhythms were presented binaurally over headphones. Subjects completed four example trials of a pseudorandom selection of rhythms in one

modality before starting the half of the experiment in that modality. Each half of the experiment contained three blocks with 30 trials each. Each trial consisted of three presentations of a rhythm, separated by 1,100 ms. There were 1,100 ms of silence in the auditory condition, and 1,100 ms of fixation cross-presentation in the visual condition. The intertrial interval was 2 s of silence in the auditory condition, and 2 s of a fixation cross in the visual condition. Half the trials contained a deviant in the third repetition. Subjects indicated whether the third rhythm presentation was the same as or different from the first two presentations by pressing one key for 'same' and another key for 'different' on a computer keyboard. For the visual condition, a red box outlined the edge of the display during the first two presentations, to indicate to participants that these were the standard rhythms. During the third presentation, the box turned green, indicating that this rhythm was to be compared to the previous two. During auditory presentation, the words 'First time', 'Second time', and 'Same or different?' were displayed on the screen during the first, second, and third rhythm presentations, respectively. For each condition, d' scores and percent correct were calculated.

Results

Preliminary ANOVAs to assess the effects of music or dance training were conducted. A $2 \times 2 \times 3$ mixed ANOVA with the between-subjects factor Music group (2+ years musical training, 6 months or less musical training), and the within-subjects factors Modality (auditory, visual) and Rhythm type (metric simple, metric complex, nonmetric) was conducted on d' scores and percent correct. Music group had a significant main effect on d' scores (d':F(1, 15) = 5.08, p = .04) and a marginally significant effect on percent correct (F(1, 15) = 3.98,p = .065). Thus, on average, participants with two or more years of musical training did better than those with 6 months or less (mean d': .95 vs. .47; mean percent correct: 66 vs. 59 %). However, Music group did not interact with any of the other factors (all Fs < 1). To assess the effects of dance training, a $2 \times 2 \times 3$ mixed ANOVA with the between-subjects factor Dance group (1+ year of dance training, <6 months musical training), and the withinsubjects factors Modality (auditory, visual) and Rhythm type (metric simple, metric complex, nonmetric) was conducted on d' scores and percent correct. No main effects of or interactions with Dance group were observed (all Fs < 1). Therefore, all subsequent analyses are collapsed across participants.

For the main analysis, a 2×3 repeated measures ANOVA with Modality and Rhythm type was conducted



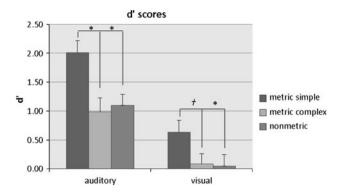


Fig. 3 The sensitivity to the presence of a deviant (as indexed by d'scores) for each rhythm type in each modality. Performance in the auditory modality is significantly better than the visual modality, and metric simple performance is significantly better than metric complex performance and nonmetric performance in both modalities. *p < .05, 2-tailed test. †p < .05 1-tailed test. *Error bars* show standard error of the mean

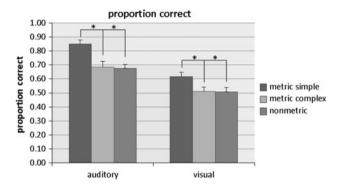


Fig. 4 The proportion of correct judgments of the third repetition for each rhythm type in each modality. Performance in the auditory modality is significantly better than the visual modality, and metric simple performance is significantly better than metric complex performance and nonmetric performance in both modalities. *p < .05, 2-tailed test. *Error bars* show standard error of the mean

on d' scores and percent correct. A significant main effect of Modality was present for d' scores and percent correct (d': F(1, 16) = 60.07, p < .001; percent correct: F(1, 16) = 65.96, p < .001). The auditory condition was performed more accurately than the visual condition (d': 1.37 vs. .25, percent correct: 74 vs. 55 %). D' scores and percent correct are shown for each condition in each modality in Figs. 3 and 4.

A main effect of Rhythm type was found in both measures (d': F(2, 32) = 9.90, p < .001; percent correct: F(2, 32) = 15.70, p < .001). Planned contrasts showed that the metric simple condition was performed better than the metric complex condition and nonmetric conditions (d': 1.32, .54, and 0.57, percent correct: 74, 60, and 59 %). The metric simple condition was significantly different from the metric complex and nonmetric conditions, which did not differ from

each other (metric simple vs. metric complex d': F(1, 16) = 10.48, p = .001; percent correct: F(1, 16) = 16.98, p = .001; metric simple vs. nonmetric d': F(1, 16) = 13.19, p = .002; percent correct: F(1, 16) = 30.96, p < .001; metric complex vs. nonmetric d': F < 1; percent correct: F < 1).

There was no significant interaction between Modality and Rhythm Type, indicating that the relative discrimination accuracy pattern across rhythm conditions in the visual condition did not significantly differ from the auditory condition (d': F(2, 32) = 1.07, p = .35, percent correct: F < 1). Planned contrasts indicated that in the auditory condition the metric simple condition was significantly different from the metric complex and nonmetric conditions, which did not significantly differ from each other (metric simple vs. metric complex d': F(1, 16) = 19.90, p < .001; percent correct: F(1, 16) = 19.08, p < .001; metric simple vs. nonmetric d': F(1, 16) = 15.51, p = .001; percent correct: F(1, 16) = .00116) = 47.67, p < .001; metric complex vs. nonmetric d': F < 1; percent correct: F < 1). In the visual condition the pattern was similar, although the difference between metric simple and metric complex did not reach 2-tailed significance in the d' measure (metric simple vs. metric complex d': F(1,16) = 3.43, p = .08; percent correct: F(1, 16) = 5.67, p = .03; metric simple vs. nonmetric d': F(1, 16) = 4.97, p = .04; percent correct: F(1, 16) = 7.77, p = .01; metric complex vs. nonmetric d': F < 1; percent correct: F < 1). However, on the whole, performance in the visual condition was much worse than in the auditory condition. In one-sample tests for each condition relative to chance performance (0 for d', 50 % for percent correct) indicated that performance was significantly above chance in all auditory conditions (metric simple: d' t(1, 16) = 10.3, p < 001; percent correct t(1, 16) = 10.316) = 13.6, p < .001; metric complex: d' t(1, 16) = 4.27, p = .001; percent correct t(1, 16) = 4.67, p < .001; nonmetric: d' t(1, 16) = 6.09, p < .001; percent correct t(1, 16) = 0.0916) = 5.87, p < .001), but in the visual modality, performance was only significant above chance in the metric simple condition (metric simple: d' t(1, 16) = 3.07, p = .007; percent correct t(1, 16) = 3.91, p = .001; metric complex: d' t(1, 16) = 0.5, p = .62; percent correct t(1, 16) = 0.44, p = .66; nonmetric: d' t(1, 16) = 0.22, p = .83; percent correct t(1, 16) = .21, p = .84).

To determine whether the overall length of rhythms or the number of intervals in a rhythm significantly affected performance, particularly in the visual condition (as visual sensory memory) decays more rapidly than auditory sensory memory), supplementary analyses were conducted that compared performance on the different rates (220–270 ms) and the different number of intervals (5, 6, or 7). D' scores could not be reliably calculated because many participants did not make a hit, miss, false alarm, and correct rejection in every possible condition when broken down this way, so



only percent correct scores were analyzed. A $2 \times 3 \times 3$ repeated measures ANOVA with Modality, Rhythm type, and Number of Intervals (5, 6, or 7) revealed no significant main effects of or interactions with Number of Intervals (main effect F < 1; Modality × Number of Intervals F < 1; Rhythm type × Number of Intervals F(4,64) = 1.5, p = .22; Modality × Rhythm type × Number of Intervals F(4, 64) = 1.18, p = .33). A separate 3×3 ANOVA with Rhythm Type and Number of Intervals on only the visual condition data still revealed no significant main effects of or interactions with Number of Intervals (main effect F < 1, Number of Intervals \times Rhythm type F(4, 64) = 2.23, p = .12). Although the interaction is marginally significant, it is because the 6-interval rhythms were discriminated accurately more often than the 5- and 7-interval rhythms (58 vs. 52 and 52 %, respectively), so no consistent effect of increasing the numbers of intervals can be inferred.

To analyze the effects of overall length of the rhythms, the six rates that were used were binned into three groups: Fast rhythms (220–230 ms base interval), Medium rhythms (240–250 ms base interval), and Slow rhythms (260-270 ms base interval). The Fast rhythms are the shortest ($\sim 2,860-2,990$ ms), the Medium rhythms are middle in length ($\sim 3,120-3,250$ ms), and the Slow rhythms are the longest ($\sim 3,380-3,510$ ms). A 2 \times 3 \times 3 repeated measures ANOVA with Modality, Rhythm type, and Rate (Fast, Medium, or Slow) revealed no significant main effects of or interactions with Rate (all Fs < 1). A separate 3 × 3 ANOVA with Rhythm Type and Rate on only the visual condition data still revealed no significant main effects of or interactions with Rate (Rate: F(2), 32) = 1.23, p = .31; Rate × Rhythm Type F < 1). Numerically, the fastest rhythms were more accurately discriminated than the medium and slowest rhythms (65, 65, and 62 % in the auditory condition; 54 50, and 48 % in the visual condition), thus the lack of significant effect may be the result of the restricted tempo range. However, overall, neither longer rhythms nor rhythms with greater number of intervals were related to a significant performance decrement.

Finally, because other studies have found that prior auditory exposure enhances beat perception in visual rhythms, we compared visual discrimination performance when the visual condition was first (no prior auditory exposure) versus second (prior auditory exposure). A 2 (Order) \times 3 (Rhythm type) mixed measures ANOVA was conducted on d' and percent correct scores in the visual condition. No significant main effects of Order nor the interactions between Order and Rhythm type were found (all F's < 1). Thus, prior auditory exposure did not affect visual rhythm discrimination.



Discrimination accuracy across the auditory conditions was better overall than across the visual conditions. This is in line with other research (Collier and Logan 2000; Ferrandez et al. 2003; Grondin et al. 1999; Patel et al. 2005) that finds an auditory advantage for temporal discrimination tasks. In addition, the metric simple condition was discriminated more accurately than the metric complex and nonmetric conditions, in both the auditory and visual modalities. For the auditory condition, this is consistent with previous work measuring rhythm reproduction accuracy for these rhythms. However, the observation of the same pattern for the visual condition is novel and is in contrast to previous studies using different rhythmic stimuli (Collier and Logan 2000; Glenberg and Jona 1991; Glenberg et al. 1989; Patel et al. 2005). This suggests that the spatially based visual presentation was effective at conveying the beat-based structure of the rhythms. In addition, only the rhythms with this structure were discriminated above chance in the visual modality, a finding that will be discussed in more detail later.

Examining the temporal factors that improve performance shows that the presence of low integer-ratio relationships between the durations in a rhythmic sequence does not necessarily lead to enhanced discrimination performance. The metric complex rhythms are discriminated less accurately than the metric simple rhythms, even though both are composed of intervals related by integer ratios. This is consistent with previous work with these rhythms in the auditory domain that measured rhythm reproduction accuracy (Grahn 2009; Grahn and Brett 2007). Moreover, the metric complex and nonmetric rhythms are discriminated similarly, even though the former are composed of integer-ratio intervals and the latter of noninteger-ratio intervals. The important structural element that improves performance therefore appears to be the additional grouping of the onsets into consistent 4-unit groups, or chunks, creating a higherlevel periodicity. The onset of nearly every chunk in the metric simple sequences was also accented by the temporal pattern of the intervals (Povel and Essens 1985; Povel and Okkerman 1981). However, these accenting rules have not been shown to apply in the visual modality, so it is unknown whether the grouping itself or if the presence of temporal accents that emphasize grouping is most important. One point to note is that with filled durations, some of the temporal accents in the stimuli will also coincide with long notes, thus long durations may sometimes provide an additional salience cue that could be used to help detect the beat-based structure in metric simple visual stimuli.



Whether it is the grouping or grouping plus temporal accents that convey the beat-based structure, it appears that visual rhythm discrimination is enhanced by its presence. The pattern of discrimination accuracy between different visual conditions was similar to the auditory conditions: The metric simple condition was discriminated more accurately than the metric complex and nonmetric rhythms, and the metric complex and nonmetric did not differ from each other. However, overall, visual rhythms were only discriminated with above chance accuracy in the metric simple condition. Visual rhythms therefore do benefit from a beat-based structure. This finding contrasts with previous studies that do not find beat-based encoding for visual rhythms (McAuley and Henry 2010; Patel et al. 2005). In Patel et al.'s study, participants tapped a steady beat along with the rhythm (rather than discriminating changes in rhythms). Two of the rhythm conditions in Patel et al. are analogous to the metric simple and metric complex rhythms in the current study, although they were concatenated to make a single long rhythm (~ 48 s). Even though the longer stimulus may have given participants more opportunities to find the beat, there was no difference in tapping accuracy and variability between the two conditions when presented visually, suggesting that the temporal structure did not improve performance in the visual modality. However, in those studies, brief visual stimuli were used: the stimulus was composed of LED flashes. The use of a line that moves from position to position in the current study appears to be more successful at communicating temporal structure. It may be that the perception of relationships between the time intervals in the rhythm was strengthened by the perception of apparent motion of a single object, instead of separate flashes of stimuli in the same space. Or perhaps the association of each duration with a particular spatial position of the line enabled participants to keep track of the durations, increasing the chances of perceiving a relationship between successive durations. Perception of the relationship between durations could enable perception of the grouping of intervals. Then, as in the auditory modality, the intervals can be encoded as groups of intervals that relate to the underlying beat, instead of as separate arbitrary durations (as is required for the nonmetric stimuli).

An alternate interpretation is that the visual rhythms were recoded into an auditory or other (perhaps supramodal) representation and that beat perception was only present after recoding. The fact that stimuli were repeated three times would allow time for recoding to occur and the beat-based behavioral benefit to be observed during the discrimination of the third presentation. According to this interpretation, in the Patel et al. (2005) study, the lack of repeated presentations may have prevented any recoding benefit from being observed, therefore no difference

between conditions was found. In addition, auditory recoding could explain how visual discrimination performance was improved from near-chance (in the metric complex and nonmetric conditions) to significantly above chance (in the metric simple condition). If beat-based rhythms are more easily remembered than nonbeat-based rhythms in the auditory modality, recoding of visual rhythms into an auditory representation would only improve beat-based rhythm performance. The auditory representation for the metric simple condition would result in improved visual discrimination performance.

Support for the general idea of recoding visual rhythms into auditory representations is conflicting. Some studies do suggest that visual rhythms receive obligatory automatic auditory encoding (Guttman et al. 2005). For example, Guttman et al. (2005) conducted a series of studies to examine how concurrent auditory information affected visual rhythm discrimination performance. Visual rhythms were presented using sequentially appearing Gabor patches that could change in spatial frequency or orientation to indicate an 'event' or interval onset. The concurrent presentation of incongruent auditory rhythms significantly impaired visual rhythm discrimination performance. In addition, the authors suggest that irrelevant auditory information decreased accuracy more than irrelevant visual information, in support of the idea that auditory recoding of the visual rhythms was occurring. However, in their study, the irrelevant visual information was not analogous to the irrelevant auditory information. The irrelevant visual condition involved presenting a visual rhythm using one type of visual change (e.g., in spatial frequency) to signify interval onsets, then discriminating that rhythm when presented using another type of visual change (e.g., in spatial orientation). Participants indicated whether the timing of the changes was the same between the two rhythm presentations, regardless of what visual change was used to indicate the rhythm. The fact that the visual change that indicated the rhythm differed from the first to the second presentation constituted the irrelevant visual information. The irrelevant auditory information, however, consisted of tones presented at the same time as the visual rhythms, but with the timing of the tones not congruent with the timing of the visual rhythm. Therefore, it is not clear that the performance was worse because incongruent auditory information was more disruptive than incongruent visual information, or whether the presence of concurrent irrelevant information was more disruptive than subsequent irrelevant information, because there was no presentation of concurrent visual information to compare to. Thus, the support for specific auditory recoding is not clear cut.

Other work suggests that auditory encoding of visual rhythms is neither obligatory nor automatic (McAuley and Henry 2010). McAuley and Henry's study asked



participants to complete 'speeding up or slowing down' judgments on visual and auditory rhythms. The judgment could be made one of two ways: The first was with respect to the individual intervals in the rhythm: Does the final tone make the last interval shorter (speeding up) or longer (slowing down) than the earlier intervals in the rhythm? Alternatively, the judgment could also be made with respect to an implied, regular beat: Does the final tone come earlier than the implied beat (speeding up) or later (slowing down)? They found that many auditory rhythms, but almost no visual rhythms, were judged in relation to the beat (indicated by using the second strategy). Visual judgment strategies could be changed, however, by exposing listeners to the auditory rhythms prior to the visual rhythms. After auditory exposure, a greater number of visual judgments were made using in relation to the beat. Thus, the prior auditory exposure seemed to encourage the auditory encoding of visual rhythms, leading to the judgments based on the beat. Importantly, auditory recoding did not spontaneously occur for the visual rhythms without prior auditory exposure, indicating that the recoding was not automatic.

In the current study, visual rhythm discrimination did not benefit from prior auditory exposure. A key difference between this study and previous work by McAuley and Henry (2010) may explain how this is possible. In McAuley and Henry (2010), the speeding up or slowing down judgment was made about the final interval in the rhythmic pattern. The rhythmic patterns therefore only varied in this to-be-judged final interval, and the rest of the intervals were one of two possible rhythmic patterns repeated throughout the experiment. The two possible patterns could easily be learned after a few exposures. Pattern repetition would maximize the chances that any beat perception induced by hearing the auditory rhythms could easily be carried over to the temporally identical visual rhythms, even though the visual rhythms were not intermixed with auditory rhythms in the same block. In the current study, all rhythms were novel, presented in only one trial per modality, and were highly unlikely to be remembered from one half of the experiment to the other. Therefore, the auditory exposure could not benefit any individual visual rhythm.

On the whole, the current results would be consistent with an auditory or supramodal recoding account, with visual rhythms benefitting from this recoding when the rhythms have temporal structure that is conducive to beat perception. The fact that participants knew the rhythms would be repeated and did not have to track a continuously unfolding rhythm (as in Patel et al. (2005)) may have facilitated the use of this strategy. This strategy may have been implicit, as there were no clear indications that explicitly reported strategies of auditory recoding

consistently led to better performance. Commonly reported strategies included counting the 'beats' or the 'sounds', trying to visualize the auditory rhythms (sometimes closing the eyes), and trying to 'hear' the visual rhythms. For example, in the visual condition one participant said they tried "locating a note to each movement and rehearsing the 'song' in my head." However, only one of the top four performers in the visual condition reported any explicit strategy for the visual condition: "I tried to make the visual rhythm another in my head," and in general the reporting of strategies did not appear to be consistently associated with better performance.

Why was performance generally low in the visual modality? The metric complex and nonmetric conditions were discriminated at chance levels. One possibility is that visual sensory memory is temporally too short (0.3–1.0 s; Averbach and Coriell 1961; Eriksen and Collins 1967; Sakitt 1976; Sperling 1960) relative to auditory sensory memory (1.5-4.0 s; Cowan 1984; Crowder 1982; Darwin et al. 1972), such that the rhythms (approximately 3 s) were too long to store. In the auditory condition, the metric simple rhythm is encoded more accurately by virtue of the beat-based structure, as shown by several previous studies. However, the auditory metric complex and nonmetric discrimination judgments likely rely on the raw sensory memory trace. In the visual condition, this sensory trace has likely already deteriorated even by the end of the rhythm. Thus, unless a beat-based structure is found to enable beat-based encoding, there is little to fall back on to make the visual discrimination judgment. Future work may be able to further optimize visual rhythm perception with continuous motion stimuli, perhaps taking advantage of visual prediction and motion processing mechanisms to highlight the visual sequence structure.

These experiments show that results of timing experiments that test one or two different interval lengths cannot necessarily be generalized to timing of three or four different interval lengths. Several studies find that simple ratios of intervals are reproduced more accurately than complex ratios (Collier and Wright 1995; Essens and Povel 1985), but the current experiment shows that, in stimuli that use a greater variety of interval lengths, accurate discrimination of simple ratios can differ, depending on the interval ordering.

In conclusion, the current study provides evidence for a beat-based benefit in discrimination performance for visual rhythms that parallels that found for auditory rhythms. The benefit for visual rhythms is evident regardless of prior auditory exposure to the rhythms, and consequently may indicate an auditory or supramodal recoding strategy. In addition, the presence of integer-ratio relationships between interval durations is not enough to induce a performance advantage—higher-level grouping of the time



intervals is also necessary. Future research may focus on how to further optimize the visual stimulus presentation, perhaps a 'bouncing ball' approach (Schutz and Kubovy 2009) or the use of biological motion.

References

- Averbach E, Coriell E (1961) Short-term memory in vision. Bell Syst Tech J 40:309–328
- Chen Y, Repp BH, Patel AD (2002) Spectral decomposition of variability in synchronization and continuation tapping: comparisons between auditory and visual pacing and feedback conditions. Hum Movement Sci 21:515–532
- Collier GL, Logan G (2000) Modality differences in short-term memory for rhythms. Mem Cognition 28:529–538
- Collier GL, Wright CE (1995) Temporal rescaling of simple and complex ratios in rhythmic tapping. J Exp Psychol Human 21:602–627
- Cowan N (1984) On short and long auditory stores. Psychol Bull 96:341-370
- Crowder RG (1982) Decay of auditory memory in vowel discrimination. J Exp Psychol Learn Mem Cogn 8:153–162
- Darwin CJ, Turvey MT, Crowder RG (1972) An auditory analogue of the Sperling partial report procedure: evidence for brief auditory storage. Cognit Psychol 3:255–267
- Drake C, Penel A, Bigand E (2000) Tapping in time with mechanically and expressively performed music. Music Percept 18:1–24
- Eriksen CW, Collins JF (1967) Some temporal characteristics of visual pattern perception. J Exp Psychol 74:476–484
- Essens PJ (1995) Structuring temporal sequences: comparison of models and factors of complexity. Percept Psychophys 57:519–532
- Essens PJ, Povel DJ (1985) Metrical and nonmetrical representations of temporal patterns. Percept Psychophys 37:1–7
- Ferrandez AM, Hugueville L, Lehéricy S et al (2003) Basal ganglia and supplementary motor area subtend duration perception: an fMRI study. NeuroImage 19:1532–1544
- Gault RH, Goodfellow LD (1938) An empirical comparison of audition, vision, and touch in the discrimination of temporal patterns and ability to reproduce them. J Gen Psychol 18:41–47
- Glenberg AM, Jona M (1991) Temporal coding in rhythm tasks revealed by modality effects. Mem Cogn 19:514–522
- Glenberg AM, Mann S, Altman L, Forman T (1989) Modality effects in the coding and reproduction of rhythms. Mem Cogn 17:373–383
- Goldstone S, Lhamon WT (1972) Auditory-visual differences in human temporal judgment. Percept Motor Skills 34:623–633
- Grahn JA (2009) The role of the basal ganglia in beat perception: neuroimaging and neuropsychological investigations. Ann NY Acad Sci 1169:35–45
- Grahn JA, Brett M (2007) Rhythm perception in motor areas of the brain. J Cogn Neurosci 19:893–906
- Grahn JA, Brett M (2009) Impairment of beat-based rhythm discrimination in Parkinson's disease. Cortex 45:54–61
- Grahn JA, Henry MJ, McAuley JD (2011) FMRI investigation of cross-modal interactions in beat perception: audition primes vision, but not vice versa. NeuroImage 54:1231–1243
- Grondin S (1993) Duration discrimination of empty and filled intervals marked by auditory and visual signals. Percept Psychophys 54:383–394

- Grondin S, Meilleur-Wells G, Ouellette C, Macar F (1999) Sensory effects on judgements of short time-intervals. Psychol Res 61:261–268
- Guttman SE, Gilroy LA, Blake R (2005) Hearing what the eyes see: auditory encoding of visual temporal sequences. Psychol Sci 16:228–265
- Handel S, Buffardi L (1969) Using several modalities to perceive one temporal pattern. Q J Exp Psychol 21:37–41
- Holcombe AO (2009) Seeing slow and seeing fast: two limits on perception. Trends Cogn Sci 13:216–221
- Hove MJ, Spivey MJ, Krumhansl CL (2010) Compatibility of motion facilitates visuomotor synchronization. J Exp Psychol Human 36:1525–1534
- Jäncke L, Loose R, Lutz K, Specht K, Shah NJ (2000) Cortical activations during paced finger-tapping applying visual and auditory pacing stimuli. Cog Brain Res 10:51–66
- Jantzen KJ, Steinberg FL, Kelso JAS (2005) Functional MRI reveals the existence of modality and coordination-dependent timing networks. NeuroImage 25:1031–1042
- Large EW, Palmer C (2002) Perceiving temporal regularity in music. Cogn Sci 26:1–37
- McAuley JD, Henry MJ (2010) Modality effects in rhythm processing: auditory encoding of visual rhythms is neither obligatory nor automatic. Atten Percept Psycho 72:1377–1389
- Nettl B (2000) An ethnomusicologist contemplates universals in musical sound and musical culture. In: Wallin NL, Merker BH, Brown S (eds) The origins of music. MIT Press, Cambridge, MA, pp 463–472
- Parncutt R (1994) A perceptual model of pulse salience and metrical accent in musical rhythms. Music Percept 11:409–464
- Patel AD, Iversen JR, Chen Y, Repp BH (2005) The influence of metricality and modality on synchronization with a beat. Exp Brain Res 163:226–238
- Povel DJ, Essens PJ (1985) Perception of temporal patterns. Music Percept 2:411–440
- Povel DJ, Okkerman H (1981) Accents in equitone sequences. Percept Psychophys 30:565–572
- Repp BH, Penel A (2002) Auditory dominance in temporal processing: new evidence from synchronization with simultaneous visual and auditory sequences. J Exp Psychol Human 28:1085
- Sakitt B (1976) Iconic memory. Psychol Rev 83:257-276
- Schubotz RI, Friederici AD, von Cramon DY (2000) Time perception and motor timing: a common cortical and subcortical basis revealed by fMRI. NeuroImage 11:1–12
- Schutz M, Kubovy MJ (2009) Causality and cross-modal integration. J Exp Psychol Hum Percept Perform 35:1791–1810
- Semjen A, Ivry RB (2001) The coupled oscillator model of betweenhand coordination in alternate-hand tapping: a reappraisal. J Exp Psychol Human 27:251–265
- Shih LYL, Kuo W-J, Yeh T-C et al (2009) Common neural mechanisms for explicit timing in the sub-second range. NeuroReport 20:897–901
- Sperling G (1960) The information available in brief visual presentation. Psychol Monogr 74:1–29
- Stauffer CC, Haldemann J, Troche SJ, Rammsayer TH (2012) Auditory and visual temporal sensitivity: evidence for a hierarchical structure of modality-specific and modality-independent levels of temporal information processing. Psychol Res 76:20–31
- van Noorden L, Moelants D (1999) Resonance in the perception of musical pulse. J New Music Res 28:43-66
- Wiener M, Turkeltaub P, Coslett HB (2010) The image of time: a voxel-wise meta-analysis. NeuroImage 49:1728–1740

