Differences in Metrical Structure Confound Tempo Judgments

Justin London, August 2009

Presented at the *Society for Music Perception and Cognition* biannual meeting
August 2009.

Abstract

Musical tempo is usually regarded as simply the rate of the *tactus* or beat, yet most rhythms involve multiple, concurrent periodicities. Two experiments were conducted to investigate relations between the absolute rate of the tactus versus a more global sense of speed via a tempo discrimination task involving typical rhythmic patterns. Stimulus rhythms in 4/4 meter were presented at two tempos (100 vs 120 BPM). Seven patterns using all combinations of the longest periodicity with one, two, or all three other components were employed. Trials consisted of a standard followed by a comparison, a 3AFC design (comparison = slower, same, or faster). In Exp1 participants simply judged relative speed. In Exp2 participants focused on the tactus rate. In both participants were told to refrain from tapping or making other synchronization movements. In both experiments when standard and comparison used the same pattern responses were highly accurate. This did not hold when standard and comparison involved different patterns. The presence of beat subdivisions (SDs) seems to be the key factor in influencing response accuracy. In both experiments, when Std and Comp both had SDs accuracy remained high. If SDs are lacking in either the standard or the comparison, performance was poor. In Exp. 2, focusing on the beat level did not result in a significant improvement in performance.

Research Background

Since the early nineteenth century, when the use of mechanical metronomes became common, composers and musicians have described tempo in terms of "beats per minute" (BPM) markings, whereby a certain note value (usually the quarter note) is yoked to a rate of the metronome. Prior to the invention of the metronome musicians used other means to designate the speed at which a piece should be performed. Time signatures and note values have carried tempo connotations from the earliest days of musical notation; mensural notation captured complex durational proportions and hence could be used to express tempo changes through a change in mensuration from section to section within a piece. From about 1600 onwards the now-familiar descriptive terms (usually in Italian) became widely used: Largo, Andante, Allegro, Presto, and so forth (see Fallows 2001). These terms of course are still in use today, an important supplement to the metronome markings.

BPM markings are both practical and convenient, and hence musicians have embraced their use for nearly two centuries. Given their ubiquity, BPM measurements are often regarded as a reasonably transparent measure of musical speed in studies of rhythm perception and cognition (Jones & Boltz 1989; Parncutt 1994; van Noorden & Moelants 1999; Quinn and Watt 2006; Moelants 2006). Likewise, tapping behaviors are often regarded as transparent indicators of tempo perception (Drake, Penel & Bigand 2000; Snyder & Krumhansl 2001; Martens 2005; McKinney & Moelants 2006).

But beat rate and/or tapping rate can be dissociated from perceived tempo (Epstein 1979, 1995; Drake, Gros & Penel 1999). While one can often identify a perceptually dominant pulse stream with a rhythmic sequence, "perceived tempo is not a property of the sequence itself, but the result of several complex, interacting perceptual and cognitive processes" (Drake, Gros & Penel, p 201). If one uses a tapping paradigm as a measure of pulse finding, Drake, Gros & Penel make note of the following factors:

- Tapping rates are constrained by neurological and kinematic factors, and tend toward an intermediate rate (≈600ms)
- Tapping rates are yoked to the stimulus at a 1:n ratios;
- Tapping studies do not "observe" the event density of stimulus (number of events per unit time), and event density is a highly salient cue for tempo.

Cautions regarding the oversimplification of tempo to simply tactus rate have also been made in the music theory community, especially by David Epstein:

> Tempo is a consequence of the sum of all factors within a piece--the overall sense of a work's themes, rhythms, articulations, "breathing," motion, harmonic progressions, tonal movement, contrapuntal activity. . . Tempo . . . is a reduction of this complex Gestalt into the element of speed per se, a speed that allows the overall, integrated bundle of musical elements to flow with a rightful sense (Epstein 1995, p. 99).

London (2001, 2004) has noted that the same rhythmic surface can give rise to alternative metrical construals on the part of the listener (c.f. the discussion of "metric malleability" London 2004, pp. 48-50). Thus two rhythmic patterns with the same number of onsets and same inter-onset intervals on the surface (i.e., equal event density), may nonetheless have different higher-level metrical organizations, with different component periodicities and nesting relationships.

If this is so, then such patterns will not be judged to be equivalent in tempo. More specifically, there is likely to be conflicts between the uniformity of surface event rates versus the difference between higher-level periodicities which will confound tempo judgements. The two experiments described below break down Epstein's "complex gestalt" and explore the influence of metrical structure on judgments of musical speed (more generally) and beat rate (in particular).

Differences in Metrical Structure Confound Tempo Judgments
Justin London, August 2009

Experiments

Two experiments were conducted to explore the following hypotheses:

- Differences in metrical structure will interfere with objective measures of tempo.
- The rate of surface activity (i.e. event density) will be more salient in judgments of tempo than the tactus rate.
- Focusing on the tactus level will improve the accuracy of tempo judgments.

Experimental Design and Method

A within-subjects, 3AFC design was employed in both experiments. In both a rhythmic pattern, exemplifying a basic metrical configuration was presented as a standard, followed by a comparison pattern. The experimental task was to judge if comparison is either slower, the same, or faster than the standard. There were four stimulus pairing possibilities (of which the participants were advised):

- Same Pattern @ Same Tempo (SPST)
- Same Pattern @ Different Tempi (SPDT)
- Diff. Patterns @ Same Tempo (DPST)
- Different Patterns @ Different Tempi (DPDT)

Appendix 1 lists the seven "master patterns" that were used as standards and comparisons; the musical notation here indicates the different wood-block sounds that were used to differentiate the metrical layers in the various stimuli. While there were pitch differences amongst the wood-block sounds, they were not "melodic" patterns in the usual sense. All stimuli were in 4/4 meter, and the tactus (beat) level was clear in all stimuli; pulse finding was not deemed to be a problem in any of these stimuli.

The stimuli were presented at either 100 and 120 BPM, which, dependent upon the stimulus pattern, involved component IOIs of 2400/1200/600/300 and 2000/1000/500/250ms, respectively. Participants were presented will all possible pairs in all possible orders. Stimuli were blocked & order counter-balanced among participants. Stimuli were sequenced using *Digital Performer* 5.13 and were presented over Beyerdynamic headphones (model DT 770) in quiet seminar room. Trials were self-paced: participants cued successive stimuli as they felt ready. Data were collected via a tally sheet, and then coded and entered into Excel/SPSS for analysis.

The two experiments differed only with respect to the instructions given to the participants. In experiment 1, participants were simply asked to make a judgment about the relative "speed" of the comparison in relation to the standard, while in experiment 2, they were instructed to focus on its beat rate in making their tempo comparison. In Experiment 1 if a participant asked if s/he were to "listen to or for the beat"(i.e., as sometimes happened if the participant had a high degree of musical training) they were enjoined not to count or attend to beats, but simply to make a judgment based on their overall impression of the speeds of the stimuli. Likewise, in experiment 2, subjects were asked if they understood what was meant by "the beat," and if unsure, a brief tutorial demonstration was given. In both experiments participants were enjoined not to tap along or employ other overt movements as they listened.

The subject pools for both experiments were as follows:

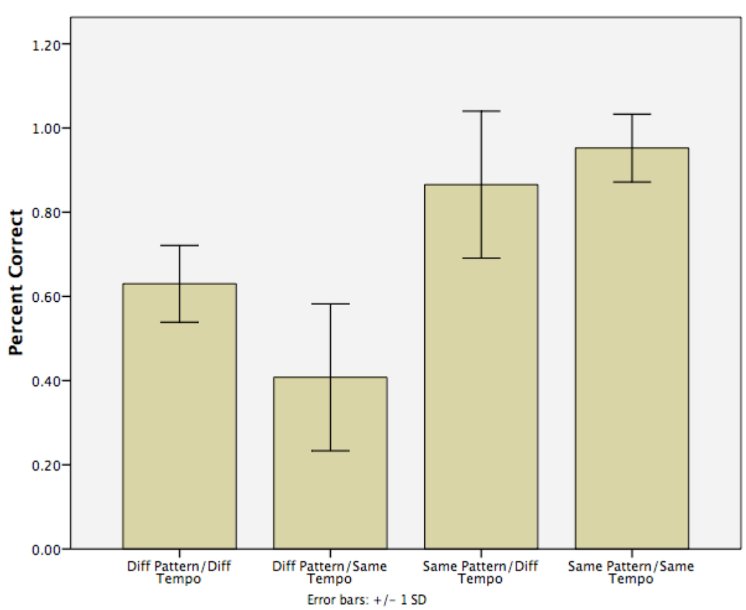
	Subjects	**M/F**	**Training***	**Avg. Age/STDev**
Exp 1	n = 24	10/14	6/7/11	20.2/1.2 yrs
Exp 2	n = 12	7/5	2/4/6	28.3/16.2 yrs

*Music Training Levels: <5yrs, 5-10yrs, >10yrs

Experimental Results

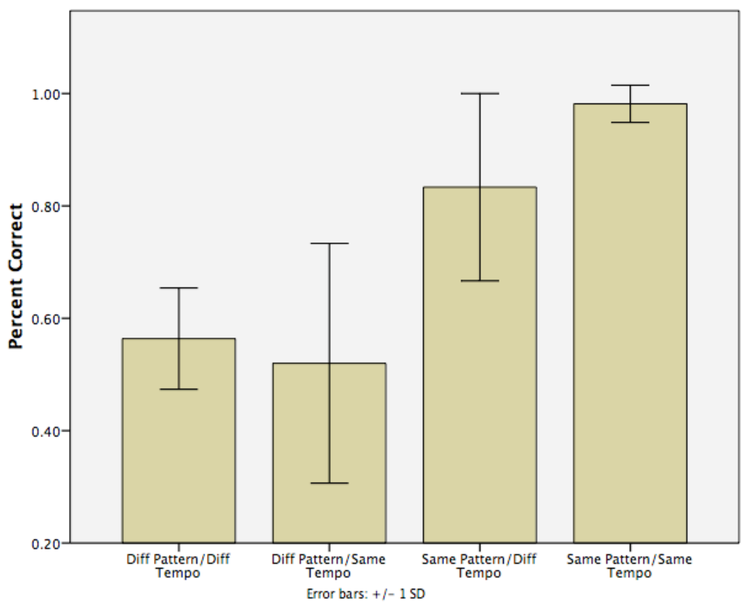
In both experiments no effects of order or gender were found; in experiment 1 some effect of training was apparent, but did not reach statistical significance. Hence the results were pooled for analysis.

In experiment 1 main effects were observed for **pattern**, $F(1,23) = 237.04$, $p < .001$, $\eta^2 = .912$, and **tempo**, $F(1,23) = 7.21$, $p = .013$, $\eta^2 = .239$. There was also a significant interaction between **pattern** x **tempo**, $F(1,23) = 37.72$, $p < .001$, $\eta^2 = .621$ (Greehouse-Geisser correction applied in all cases). Differences between trial conditions were significant in all cases (Wilcoxon signed ranks test). Note that here and below "Percent Correct" is operationalized relative to tactus rate (i.e., distinguishing between the 100 vs. 120 BPM relative to the quarter-note level of the stimulus).

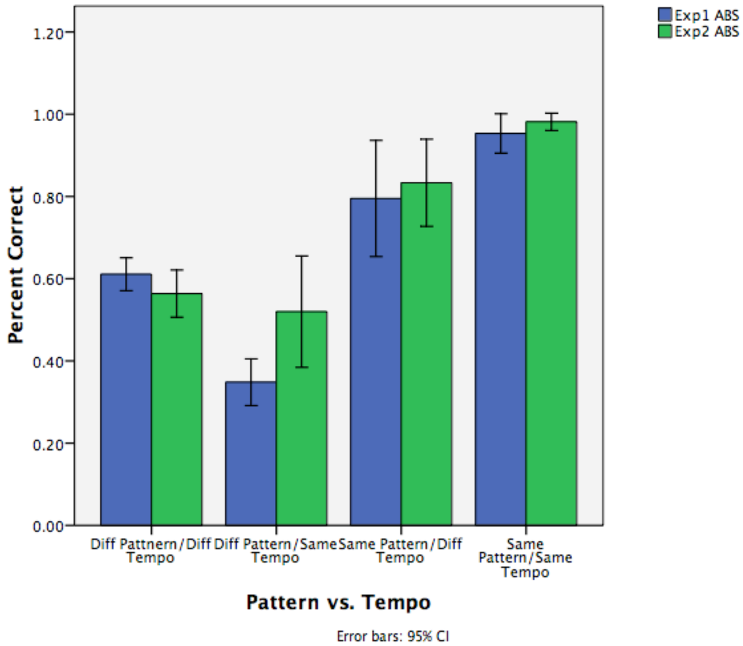


When the patterns are the same, judgments are near-perfect if the tempo is also the same, and very good if the tempo is different. When the patterns are different performance is substantially worse, though still above chance (N.B. recall this is a 3AFC design). Interestingly, the hardest task is when different patterns are presented at the same tempo.

In experiment two, in considering the absolute tempo (i.e., overall pattern length, as well as tactus rate), a significant main effect was observed for **pattern**, $F(1, 11) = 130.74$, $p < .001$, $\eta^2 = .922$, and there was a significant interaction between **pattern** x **tempo**, $F(1, 11) = 5.922$, $p < .033$, $\eta^2 = .350$; (Greehouse-Geisser correction applied in all cases). However, there was not a significant main effect for **tempo**. Differences between trial conditions were not significant between DPDT/DPST, but were significant in all other cases (Wilcoxon signed ranks test)



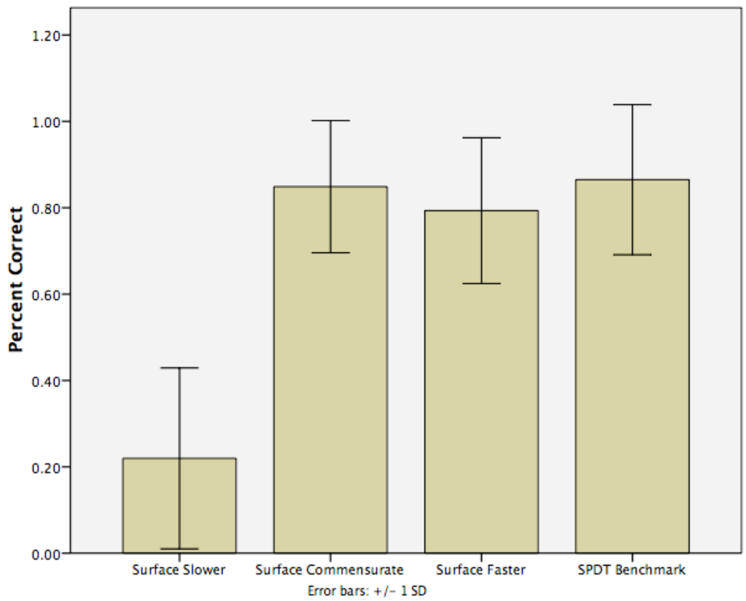
Recall that the difference between experiments 1 and 2 was a difference in participant focus: in the first experiment participants simply judged the "speed" of each stimulus, while in the second they were told to do so while focusing on the beat rate. However, no significant differences in performance were found between the two trial conditions (Mann-Whitney U Test):



Data Subset Analyses

The overall results presented above employ an operational definition of tempo as indicated by the tactus/beat rate of each stimulus—that is, as determined by the 600ms or 500ms component of the pattern, whether or not that periodicity was explicitly articulated in the pattern, and whether or not "faster" periodicities were present in the standard or comparison. But consider the following stimulus pair: the Standard has 2400ms and 300ms components (i.e., Master Pattern 8.0.0.1 in the appendix); the comparison articulates the 2000ms and 500ms periodicities (Pattern 8.0.2.0—N.B. see website for sound demo). Which one is truly faster? In the standard the 600ms tactus level is implicit (and according to research such as Parncutt 1994, maximally salient). So if tempo is simply a matter of tactus judgment, it should be a simple task (with large inter-subject agreement) that the comparison is faster, and task performance accuracy should be high. But if speed is cued more by surface activity/event density, then the standard will (a) more often be judged as faster, but also (b) accuracy and inter-subject agreement should decline, as there is a confound between tactus rates and surface speeds.

To gain a better sense of this surface vs. tactus interaction, a subset of the data from each experiment was examined which used different patterns presented at different tempos, specifically when the standard was slower than the comparison. In this context the "correct" answer should in all cases be "Faster." However, here are the results from experiments 1:

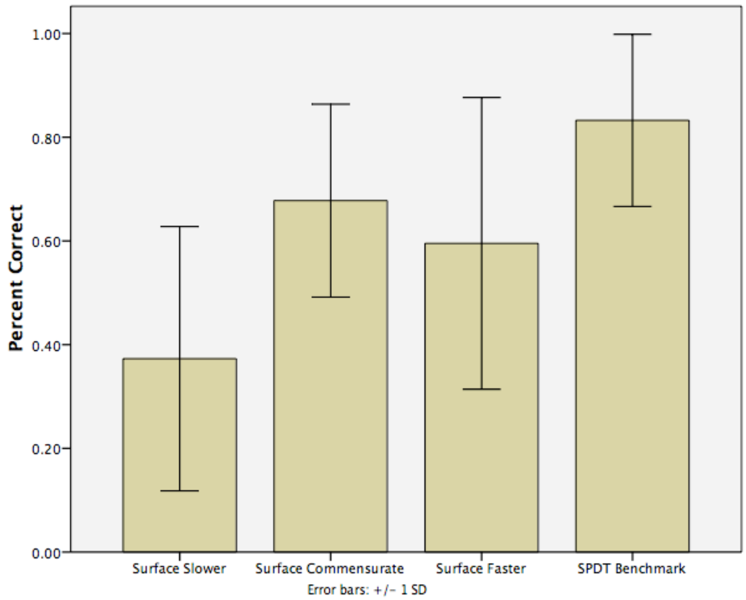


The labels on the ordinate refer to the surface activity of the comparison (the target) versus the standard. So, "surface Slower" is a case like that given in the audio demonstration--e.g., a 2400/300ms standard followed by a 2000/500ms comparison. "Surface Commensurate" refers to trials where the standard and comparison have analogous cues—e.g., a 2400/300ms standard followed by a 2000/250ms comparison. The "Surface Faster" refers to those where the comparison was faster and involved non-analogous rhythms—e.g., a 2400/600ms standard followed by a 2000/250ms comparison. Lastly, the SPDT results are given as benchmark. You can see that in "Surface Slower" cases, most participants got the "wrong" answer, while in other cases they mostly got the "right" answer. The SPDT benchmark was that of comparisons between the same pattern at SF and FS tempo conditions. The following table gives the Z-score results of Wilcoxon Signed Ranks Test comparisons between the various conditions:

	Slow/Cmns	Slow/Fast	Cmns/Fast	Slow/SPDT	Cmns/SPDT	Fast/SPDT
Z	-4.294	-4.162	-1.181	-4.291	-.187	-1.545
Sig.	.000	.000	.238	.000	.852	.122

*significance is two-tailed

Next is the analogous data comparison from Experiment 2. Notice when focusing on the beat rate, the differences between Slower, Same, and Faster are less significant (and in the case of Slower vs. Fast, non-significant), and performance relative to the benchmark degrades.



	Slow/Cmns	Slow/Fast	Cmns/Fast	Slow/SPDT	Cmns/SPDT	Fast/SPDT
Z	-2.197	-1.413	-1.020	-2.981	-2.591	-2.040
Sig.	.028	.158	.308	.003	.010	.041

*significance is two-tailed

Discussion

The main effects in both experiments show that differences in metrical structure do interfere with judgments of tempo—or more precisely, with tactus-based judgements of tempo and speed. Were tempo judgments a simple matter of beat-rate comparison, we should not have seen such large differences between same pattern and different pattern trial conditions. Perhaps most telling is that the worst performance on the comparison task occurs when the beat rate is constant but the metrical structure differs between standard and comparison (DPST condition), and, tellingly, performance worsens in this context when participants focus on the tactus level. Likewise, even when the rhythmic surfaces of the standard and comparison were commensurate (DPDT Slow-fast condition), participant accuracy declined relative to SPDT performance in experiment 2.

While the first hypothesis was supported by both experiments, the third was not. Focusing on the beat or tactus level does not improve the accuracy of tempo judgments, as evidenced by the similar results in Exp 1 vs. Exp 2. The DPDT analysis also seems to indicate that rate of surface activity (i.e., event density) is the more salient cue for tempo, in line the claims of Drake, Gros, and Penel (1999). The unanimity of response in the DPDT slow-fast condition where the standard's musical surface moved faster than the comparison's surface (even though their beat rates were slow -> fast) demonstrates the relative salience of these cues, even in experiment 2 where participants were focusing on the beat rate.

More broadly, the interaction between the tactus, metrical structure and the rhythmic surface is complex. Further analysis and additional experiments will investigate whether differences in metrical structure alone (i.e., when event density and tactus rate are held constant, but the number and periodicities of metrical levels differs) can give rise to similar confounds. Moreover, the studies described here involve only a minimal pitch component. Musical surfaces which involve clear and cumulative pitch transitions, as is the case with real melodies, may add yet another cue for musical speed and motion, as these transitions can be heard as relative distances in "musical space" as well as patterns in musical time.

Works Cited

Drake, C., L. Gros, et al. (1999). How fast is that music? The relation between physical and perceived tempo. In *Music, Mind, and Science,* ed. S. W. Yi. Seoul National University: 190-203.

Drake, C., A. Penel, et al. (2000). Tapping in time with mechanically and expressively performed music. *Music Perception* **18**(1): 1-23.

Epstein, D. (1979). *Beyond Orpheus*. Cambridge, MIT Press.

Epstein, D. (1995). *Shaping Time: Music, the Brain, and Performance*. New York, Schirmer.

Fallows, D. (2001). Tempo and Expression Marks. In *The New Grove Dictionary of Music and Musicians*, rev. edition, J. Tyrell & S. Sadie eds. Oxford Music Online.

Jones, M. R. and M. Boltz (1989). Dynamic attending and responses to time. *Psychological Review* **96**(3): 459-91.

London, J. (2004). *Hearing in Time.* New York, Oxford University Press.

London, J. (2001). Rhythm. In *The New Grove Dictionary of Music and Musicians*, rev. edition, J. Tyrell & S. Sadie eds. Oxford Music Online.

Martens, P. A. (2005). *Beat-Finding, Listener Strategies, and Musical Meter*. PhD. Diss, University of Chicago.

McKinney, M. F. and D. Moelants (2006). Ambiguity in tempo perception: What draws listeners to different metrical levels? *Music Perception* **24**(2): 155-165.

Moelants, D. (2006). "Perception and performance of aksak metres." Musicae Scientiae **10**(2): 147-172.

Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception* **11**(4): 409-464.

Quinn, S. and R. Watt (2006). The perception of tempo in music." *Perception* **35**(2): 267-280.

Snyder, J. and C. L. Krumhansl (2001). Tapping to ragtime: cues to pulse-finding. Music *Perception* **18**(4): 455-89.

van Noorden, L. and D. Moelants (1999). Resonance in the perception of musical pulse. *Journal of New Music Research* **28**(1): 43-66.

Differences in Metrical Structure Confound Tempo Judgments
Justin London, August 2009

Appendix: Rhythmic Stimuli