Author(s): London, Justin; Burger, Birgitta; Thompson, Marc; Toiviainen, Petri

Title: Speed on the dance floor : auditory and visual cues for musical tempo

Year: 2016

Version:

Please cite the original version:

London, J., Burger, B., Thompson, M., & Toiviainen, P. (2016). Speed on the dance floor : auditory and visual cues for musical tempo. Acta Psychologica, 164, 70-80. https://doi.org/10.1016/j.actpsy.2015.12.005

Speed on the Dance Floor:
Auditory and Visual Cues for Musical Tempo

Justin London

Carleton College

Birgitta Burger, Marc Thompson, Petri Toiviainen

University of Jyväskylä

Author Note: This research was supported by a Finnish Core Fulbright Scholar grant to author JL, and by an Academy of Finland grant (project "Dynamics of Music Cognition," project numbers 272250, 274037) to authors PT, BB, and MT.

Correspondence regarding this article should be addressed to Justin London, Department of Music, Carleton College, Northfield, MN 55057 USA. jlondon@carleton.edu

Abstract

Musical tempo is most strongly associated with the rate of the beat or "tactus," which may be defined as the most prominent rhythmic periodicity present in the music, typically in a range of 1.67-2hz. However, other factors such as rhythmic density, mean rhythmic inter-onset interval, metrical (accentual) structure, and rhythmic complexity can affect perceived tempo (Drake, Gros, & Penel 1999; London 2011). Visual information can also give rise to a perceived beat/tempo (Iversen, et al. 2015), and auditory and visual temporal cues can interact and mutually influence each other (Soto-Faraco & Kingston 2004; Spence 2015). A five-part experiment was performed to assess the integration of auditory and visual information in judgments of musical tempo. Participants rated the speed of six classic R&B songs on a seven point scale while observing an animated figure dancing to them. Participants were presented with original and time-stretched (±5%) versions of each song in audio-only, audio+video (A+V), and video-only conditions. In some videos the animations were of spontaneous movements to the different time-stretched versions of each song, and in other videos the animations were of "vigorous" versus "relaxed" interpretations of the same auditory stimulus. Two main results were observed. First, in all conditions with audio, even though participants were able to correctly rank the original vs. time-stretched versions of each song, a song-specific tempo-anchoring effect was observed, such that sped-up versions of slower songs were judged to be faster than slowed-down versions of faster songs, even when their objective beat rates were the same. Second, when viewing a vigorous dancing figure in the A+V condition, participants gave faster tempo ratings than from the audio alone or when viewing the same audio with a relaxed dancing figure. The implications of this illusory tempo percept for cross-modal sensory integration and working memory are discussed, and an "energistic" account of tempo perception is proposed.

Keywords: music, rhythm, tempo, audio-visual feature binding, cross-modal perception

1. Introduction

The "BPM" (Beats Per Minute) measurement, used in contexts ranging from classical musicians playing piano sonatas to DJs in dance clubs, is usually regarded as a reliable index of musical speed. The "Beat" component of the BPM measure is a prominent rhythmic periodicity, typically in a range between 100-120 BPM (1.67-2hz). In musical scores it is represented by a particular notational value (e.g., a quarter note). Once established, other periodicities, both faster and slower, are understood relative to the beat, either as subdivisions of it, or as cycles of beats that form higher-level measures and hyper-measures. Researchers in rhythm perception (Jones & Boltz 1989; Parncutt 1994; van Noorden & Moelants 1999; Quinn & Watt 2006) and rhythmic synchronization, especially in tapping studies (Clynes & Walker 1986; Drake, Penel & Bigand 2000; Snyder & Krumhansl 2001; Martens 2005) have also treated BPM measures as reasonably transparent measures of musical speed (see London 2011).

However, cues for music's rhythmic and metric organization, including tempo, are many and complex. Drake, Gros, and Penel (1999) found that perceived tempo is an emergent property, one that is dependent upon how the listener perceptually organizes the musical sequence. In a tapping task in which participants were presented with stimuli at a wide range of BPM rates, they found tapping behavior to be influenced by (a) a tendency to tap at an intermediate rate around 600ms, (b) a tendency to tap at rates related to the BPM rate by integer ratios (e.g., twice or half as fast), (c) the number of events per unit of time, or "event density" of the rhythmic surface, and (d) the participant's musical background. Boltz (2011) found that register (high vs. low) and timbre (bright vs. dull) affected perceived tempo, and London (2011) found that rhythmic patterns with the same BPM rate but different event densities were often judged to be at different tempos in a standard vs. comparison task. London (2011) also found that the attentional focus of the listener affected tempo judgments. Moving along with the music also affects our perception of it: Manning and Schutz (2013) found that tapping along enhanced the detection of perturbed tones, and London and Cogsdill (2011) found that self-motion influenced perceived tempo for some listeners.

Temporal information may also be extracted from visual cues, though our ability to do so depends on the nature of the visual stimulus. It has repeatedly been shown that performance on rhythmic timing and synchronization tasks are much poorer when the cues are discrete visual stimuli (e.g., flashing lights) versus discrete auditory stimuli (e.g., clicks or brief tones--for a summary see Repp 2005; see also Patel, Iversen, Chen, & Repp 2005). Similarly, flashes do not give rise to a strong sense of beat (McAuley & Henry 2010), and different brain regions have been shown to be involved with discrete visual as opposed to discrete auditory stimuli (Grahn, Henry, & McAuley 2011; Hove, Fairhurst, Kotz, & Keller 2013). However, Hove, Iversen, Zhang, and Repp (2013) and Iversen, Patel, Nicodemus, & Emmorey (2015) have shown that when a continuous, colliding visual stimulus is used (i.e., a video animation of a bouncing ball) synchronization performance is nearly equivalent to that with discrete auditory tones. In another study of visual cues for beat and tempo Luck & Sloboda (2009) identified absolute acceleration as the most salient cue in synchronizing with a conductor's gesture. They found that changes in acceleration were related to the shape of the of the gesture, as changes of direction at any given velocity necessarily produced changes in acceleration: "In other words, perception of rhythmic elements of human movement (in this case, the beat in conductors' gestures) may be related not only to the kinematics of the movement, but also to the dynamics underlying that movement" (p. 472). Previously Brittin (1993) found that both musicians and non-musicians were able to detect tempo changes as indicated by a conductor's gestures, though musicians were better than non-musicians, and both musicians and non-musicians were more sensitive to tempo decreases than tempo increases.

There have been relatively few studies on the integration of auditory and visual information in specifically musical contexts, as most studies of audio-visual perception have employed combinations of discrete stimuli in each modality, such as words, pictures, or light flashes paired with individual tones or sounds (Shams, Kamitani, & Shimojo 2004; Soto-Faraco & Kingstone 2004; Spence 2015). In addition, the focus in many interaction studies has been on object detection and/or the recovery of semantic information from language-based stimuli. More recent studies have combined dynamic visual and auditory arrays, often probing the effect of auditory information on visual illusions. For example, Meyer & Wenger (2001) studied the effect of auditory direction cues on the perception of motion in random dot kinematograms. In trials where the kinematogram motion cue was ambiguous, the auditory cue would bias the response, but where visual motion was unambiguous, the auditory cues had little or no effect. In a set of recent studies that do engage a musical context, Schutz & Lipscomb (2007) and Schutz and Kubovy (2009) documented a visual-auditory illusion in which a percussionist's gestures altered the apparent duration of a marimba (or similar) tone. When the same marimba sound clip was paired with point-light displays of a percussionist striking the marimba with either an extended, relaxed gesture or a short, tense gesture, the former was heard as lasting longer than the latter. Another key aspect of their findings was that the durational illusion was dependent upon the pairing of the marimba sound (i.e., a tone produced by striking a resonating object) with the appropriate visual display (i.e., a striking motion temporally synchronized with the sound onset). When the point light displays were combined with other types of musical tones or temporally mis-aligned, they had no significant effect on perceived duration.

Schutz's work combines the dynamic visual array of a single action sequence with the presentation of a unitary tone. And though it has great ecological validity, as it involves the very cues that would be involved in one's experience of a real musical performance, it is difficult to generalize to most musical contexts, as music involves complex sequences of successive tones that form rhythms and melodies. For such sequences, the perception of tempo is analogous to the perception of duration for isolated tones or inter-stimulus intervals. Thus to probe the interaction between auditory and visual cues for musical tempo, one needs auditory and visual sequences that each individually convey a sense of tempo. At the same time, one must be mindful that cross-modal sensory integration crucially depends on the ecological "relevance" of both the auditory and visual cues, as Schutz & Kubovy have shown. This relevance is fine-grained, for it is not just the combination of any musical sound" with ant musical performance gesture, but co-presentation of the particular sounds and gestures that occur together in real-world musical contexts.

To explore the effect of visual information on the perception of tempo, our experiment uses a carefully chosen set of sound clips from classic American rhythm and blues (R&B) songs, along with visual stimuli that are directly related to to the auditory signal: point-light displays produced from motion capture recordings of people dancing to them. We are thus able to present our participants with auditory stimuli with robust and unambiguous tempo cues paired with natural and continuous movement sequences. The challenge, of course, in using real as opposed to artificial auditory and visual stimuli is that they may introduce uncontrolled confounds or cues. We acknowledge this challenge, and as detailed below, have taken care in the selection of our stimuli and the design of our experiment to minimize these potential problems. Our research hypotheses are as follows:

1. That participants will be able to discriminate and properly rank the tempos of original and temporally manipulated unimodal auditory stimuli. This is essentially a baseline condition, as our ability to make tempo discriminations amongst artificial and real musical stimuli is already well established (Miller & McAuley 2005; Honing 2006).
2. That stable and matched combinations of musical and visual cues would yield more precise tempo judgments than in unimodal auditory or visual contexts. That is, the presence of more information/redundant temporal information will reduce the variability of participant responses.
3. That systematically varied visual cues that are ecologically relevant will affect the perception of concurrently presented music. In plain terms, changing the dance interpretations will change the perception of the music's tempo.

In addition, we want to determine if participants will be able to make veridical tempo rakings from the video stimuli alone, and if their ability to do so is affected by the character of the movement(s) they observe.

2. Motion Capture Experiment

2.1 Method

The video stimuli used in our experiment are derived from data obtained in a companion experiment which explored tempo-driven (i.e., musically "forced") versus volitional (i.e., musically "unforced") changes in spontaneous movements to music. That is, while speeding up or slowing down a song should lead to changes in movement characteristics, we also wanted to establish that dancers could make analogous changes when prompted to change their interpretive framework, even when the musical tempo remained constant. We give a brief report of the motion capture experiment here, as it will be helpful in understanding the stimuli used in main experiment reported below.

2.1.1 Participants

Thirty participants (15 female) were recruited from the Jyväskylä University community (average age: 28.2, SD: 4.4, range: 21-36 years). Four participants had received professional music education. Twenty-two participants had undergone music education as children or adults, of which 13 were still actively playing and instrument or singing. Fourteen participants had taken dance lessons of various styles. Participants were given a movie ticket (value ≈10€) for their participation in the experiment.

2.1.2 Audio Stimuli

For this experiment to be successful, auditory stimuli were needed that would reliably induce movement in our participants. Classic Motown R&B songs, known for their danceability/high "grooviness" ratings (Janata, Tomic, & Haberman 2012), were chosen as auditory stimuli. "Groove" has been operationally defined as the extent to which a piece of music gives rise to spontaneous movement and/or the desire to move. It was recognized that high groove ratings would also be desirable when the same stimuli would be used in the companion tempo rating experiment, as stronger kinematic reactions would presumably give rise to correspondingly stronger impressions of tempo. The core audio stimuli consisted of the first 30-35 seconds of six Motown/R&B songs released between 1964-1970 (see Table 1).

Table 1

Musical stimuli used in the experiments.

Artist	Title	Original BPM	R&B Chart Ranking
Temptations	Get Ready	134.5	#1 (1966)
Supremes	Where Did Our Love Go?	133	#1 (1964)
Supremes	Stop, In the Name of Love	117	#2 (1964)
Wilson Pickett	The Midnight Hour	113	#1 (1965)
Stevie Wonder	Signed, Sealed, Delivered	105.5	#1 (1970)
Temptations	My Girl	103	#1 (1964)

These songs were chosen according to the following criteria:

- Having an objective beat rate at/near 105, 115, or 130 BPM.[1]
- Having the same metrical structure; all had four beats per measure with light to moderate amounts of swing, meaning that the typical binary divisions of the beat were played somewhat unevenly but without overt triplet divisions of the beat.
- Similar rhythmic surface characteristics for each pair of songs at each BPM level.
- Homogeneity of musical style, as all songs were from the same genre and historical era (1964-1970).
- Ubiquity in popular music culture, as all songs have achieved the status of R&B classics.

Objective BPM measurements were determined by averaging the results of two independent raters who tapped along to each song using a beat-finding metronome, and were also checked using the Matlab-based MIRtoolbox “mirtempo” function[2]. The original versions were first time-stretched so their tactus rates aligned precisely at 105, 115, or 130 BPM using Audacity (ver. 2.0.5), an open-source sound editor (audacity.sourceforge.net). The stimuli were then time-stretched a second time to produce tactus rates that were ±5% of these three baseline rates (an example set of original and time-stretched audio stimuli is given in the Audio Appendix). These core tempos and time-stretch amounts were intentionally chosen both to yield stimuli in the 100-135 BPM range, a range in which we are maximally sensitive to temporal distinctions (Fraisse

[1] Note that here and below, "BPM" will be used in conjunction with different levels of musical speed (with respect to at least one level of rhythmic periodicity) in the *stimuli*, and "tempo" will be used to refer to participants' *judgments* of musical speed. Likewise "core BPM" refers to the original (or slightly time-stretched) versions of each song at 105, 115, or 130 BPM, and "time stretched" refers to the BPM-altered versions of each song, as is explained below.

[2] MIRtoolbox ver 1.5, www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox; for details see http://link.springer.com/chapter/10.1007/978-3-540-78246-9_31

1984; Penel, Rivenez, & Drake 2001; Drake & Bertrand 2003), and to yield BPM overlaps between stimulus groups. These modest amounts of time stretching created readily perceivable differences in tempo, while preserving pitch and timbre without introducing any significant audio artifacts.

Knowing that event density, in addition to BPM, can also affect tempo judgments (and thus potentially one's movement response--Drake, Gros, & Penel 1999), a pair of rhythmically contrastive songs was chosen at each core BPM level. A score-based analysis of each song was performed which indexed the number of notes at the 8th note level of the meter (i.e., binary subdivisions of the beat) in each bar of their vocal melody, bass, and percussion parts. From these measurements, an aggregate rhythmic density score was calculated for each song. As a corresponding measure, the low-frequency spectral flux for each song was calculated by choosing an octave-wide frequency range between 100 and 200 Hz and calculating the sub-band flux (MIRtoolbox function “mirflux”) by taking the Euclidean distances of the spectra for each two consecutive frames of the signal (Alluri & Toiviainen 2010), using a frame length of 25 ms and an overlap of 50% between successive frames and then averaging the resulting time series of flux values. Low frequency flux has been shown to be related to rhythmic density features in music (Burger, Ahokas, Keipi, & Toiviainen 2013).

2.1.3 Apparatus and Procedure

Dancer movements were recorded using an eight camera Qualisys Oqus 5+ motion capture system with a capture rate of 120 frames per second (fps); 28 reflective markers were attached to each participant. The musical stimuli were played back via a pair of Genelec 8030A loudspeakers using a Max patch running on an Apple computer. The direct (line-in) audio signals of the playback and the synchronization pulse transmitted by the Qualisys cameras when recording, were recorded using ProTools software in order to synchronize the motion capture data with the musical stimulus afterwards. Additionally, a video camera was used to record the sessions for reference purposes.

Participants were recorded individually while being asked to imagine being in a social setting such as a club or disco, that is, to dance to the music as "naturally" as possible, under the circumstances. The six Motown songs were presented in random order for each participant, in blocks including all of the versions of each particular song. Each block began with one of the two time-stretched versions of the stimulus, and then two presentations of the baseline tempo version of the stimulus. When presented with the time-stretched versions, participants were asked to dance freely. When presented with the baseline tempi versions, they were instructed to either move in a “fast/vigorous” or “slow/relaxed” way. Participants were advised that both presentations of the song in the instructed condition would be at the same tempo. The order of the uninstructed condition pair (-5% vs. +5%) was counterbalanced among participants, as was the order within the instructed condition pair (relaxed vs. vigorous). Participants were further advised to remain synchronized to the music and stay in the capture area marked on the floor (approximately 3 x 4 m) during all trials. Participants were free to rest whenever they wished during the experiment; the experiment took an average of 45 minutes per participant.

2.2 Results and Discussion

Recall that our participants gave spontaneous/non-choreographed dance responses, and thus there was no common movement pattern present in all dancers, nor were consistent patterns of movement present within most trials (see the Video Appendix for a sample block of one participant's responses in all four conditions). Given the dynamic and fluid nature of the responses, our analysis is necessarily restricted to more global measures of movement characteristics. As a summary measure of the participants' responses to the various experimental conditions (time-stretched and instructed), we will report on our acceleration data for the center of mass (CoM) marker in the conditions of interest, (i.e., Time-Stretched -5% vs. +5%, and Relaxed versus Vigorous interpretation). Acceleration, rather than speed or gestural shape, has been shown to be a more salient cue for timing information (Luck & Sloboda, 2009). Likewise, the CoM has been shown to be a useful index of global movement characteristics, as the movement of the CoM strongly influences the movement of more distal markers/parts of the body in the kinematic chain (Toiviainen, Luck & Thompson, 2010; Burger et al., 2014). Our measure of acceleration is the mean of the magnitude of the acceleration over the course of a given trial, expressed in mm/sec^2.

A 3x2 repeated measures ANOVA (3 BPM levels x 2 Time-Stretch levels) found a main effect for BPM (105 vs. 115 vs. 130 core levels), $F(1.972, 116.319) = 4.835$, $p = .010$, $\eta_p^2 = .076$ and a main effect for Time-Stretch, $F(1,59) = 77.288$, $p < .001$, $\eta_p^2 = .567$ (Greenhouse-Geisser correction applied in all cases). In addition, a BPM x Time-Stretch interaction was also found, $F(1.983, 117.021) = 3.48$, $p = .034$, $\eta_p^2 = .056$, though this is largely due to the flatness of the BPM levels relative to the more substantial difference in the two Time-Stretch conditions (note effect sizes). The grand mean of all -5% time-stretched data was 2237.99 mm/sec^2, while for the +5% data it was 2700.47 mm/sec^2. A 3x2 repeated measures ANOVA (3 BPM levels x 2 Instruction levels) found a main effect for BPM, $F(1.793, 105.788) = 3.124$, $p = .048$, $\eta_p^2 = .050$ and a main effect for Instruction (Relaxed vs. Vigorous Interpretation), $F(1,59) = 147.328$, $p < .001$, $\eta_p^2 = .714$. There was no interaction between BPM and Instruction ($F(1.939, 114.382) = 1.213$, $p = .301$). The grand mean for all relaxed trials was 1414.57 mm/sec^2, and for all vigorous trials was 3773.45 mm/sec^2.

In summary, in the Time-Stretched conditions, overall increase in core BPM rates had a very small but significant effect on the acceleration of the CoM, while time stretching produced a far greater effect size and significance. Similarly, in the Instructed conditions, core BPM rates again had a very small but significant effect, while the dancer's volitional interpretation had an even greater effect than time-stretching; note the grand averages reported above. These data show that while spontaneous dancer acceleration is reliably correlated with BPM rate, dancers were able to create dance interpretations with demonstrably different movement characteristics even when the speed of the music is held constant. Thus both conditions would be able to provide video materials suitable for creating the stimuli used in the main experiment.

3. Main Experiment

3.1 Method

In using familiar and/or highly memorable musical stimuli for a tempo judgment task, a problem arises, in that listeners can quickly learn to associate a particular tempo with a particular stimulus. Thus another motivation for using original and time-stretched versions of each song in the motion-capture experiment was to forestall this association when we re-used the same audio stimuli in the current experiment. Participants were informed that they would be presented with original and time-stretched versions of the stimuli; they thus realized that each time they heard a particular song they would need to make a fresh tempo judgment. To explore the second hypothesis--that stable and matched combinations of musical and visual cues would yield more precise tempo judgments than from audio alone--the time-stretched versions of each song were each paired with a single video. To explore the third experimental hypothesis--that appropriate visual cues can affect the perceived tempo of concurrently heard music--recordings of songs at their original tempos were paired with two different videos: one of a dancer giving the "relaxed" interpretation of the song, and the other with the same dancer giving the "vigorous" interpretation of the same song. In all three conditions--audio only, video only, and audio+video (A+V)--the participants' task was the same: to rate the speed of the music using a seven-point Likert scale.

3.1.1 Participants

Twenty-seven participants (15 female) were recruited from the Jyväskylä University community. Mean age of the participants was 29.3 years (SD 7.0 yrs; max = 49, min = 21). Five participants had no musical training and were unable to read music, 10 participants had 1-10 years of musical training, and the remaining 12 had >10 years of training. Seven did not actively engage in music making (either singing or playing an instrument) and 13 claimed to be actively making music at least 3-4 times per week. Twenty-three participants reported that, in general, they had a low familiarity with the Motown/1960s R&B style of the music used in the experiment, and 5 participants had never heard any of the songs used in the experiment. Nonetheless 17 participants were familiar with at least 3 of the 6 songs used as stimuli. 12 participants were Finnish, and the remaining 15 were from 13 other countries. All were fluent in English and able to follow the experimenter’s directions as well as the on-screen prompts in the experimental set-up. Participants were given a movie ticket (value ≈10€) for their participation in the experiment.

3.1.2 Stimuli

Participants were presented with audio, audio+video (AV), and video-only versions of the stimuli. The audio stimuli were the same six song excerpts and their time-stretched versions were the same as described above in the motion capture experiment. In preparing the video stimuli, data from responsive and reasonably skilled dancers were required in order to have effective video stimuli. In the motion capture experiment there was a wide range of dance interpretations: some participants exhibited little movement in any condition, while others moved enthusiastically but were not well synchronized to the beat, and some may have been moved well in one trial, but not in others. To avoid experimenter bias in the selection of the video stimuli, the level of dance skill of all of the 30 participants in the motion capture experiment was first informally rated by 15 observers who viewed two randomly chosen 10-second clips of each dancer. Clips were presented to the observers as a group, and ratings were made on a seven-point scale. Animations from the three top rated male and three top rated female dancers were then selected for further consideration.

To avoid confounds due to different dancers being presented in a pair of target stimuli, the same dancer was used in each pair (i.e., either the -5% vs. +5% time-stretched versions, or the "relaxed" vs. "vigorous" versions of each song). Dancers were used in different songs in different blocks of the experiment (i.e., time-stretched versus instructed conditions), counter-balanced by BPM rate. Once the dancers were selected, point-light animations at a downsampled frame rate of 30 fps were produced from the original motion capture data using the MoCap Toolbox (Burger & Toiviainen 2013), trimmed and synchronized to the audio stimuli. For the animations, the original configuration of 28 markers was reduced to a 20-point stick figure. A periodicity analysis of the movement data using the MoCap Toolbox (Burger & Toiviainen 2013) confirmed that all of motion capture data used to produce the video stimuli used in this experiment exhibited movements that were period-locked in the hip, feet, and/or head markers to the BPM rate of the accompanying music within a 5% range.

3.1.3 Apparatus and Procedure

Each experimental session lasted approximately 40 minutes, and consisted of an introduction and pretest followed by five experimental blocks (Block 1 = audio-only, Block 2 time-stretched A+V; Block 3 instructed A+V; Block 4 time-stretched video-only; Block 5 instructed video-only; see Table 2). Stimuli were presented in a random order for each participant within each block, with the randomization constrained so that different versions of the same stimulus were not presented consecutively. A decision was made not to randomize the order of blocks, so that the effect of sequential block design could be tracked for all participants (see discussion). As an alternative to counterbalancing the ordering of Blocks 2 & 3 and 4 & 5, six foils taken from the "opposite" block type (e.g., instructed A+V stimuli included in the the time-stretched A+V block) were included in each block to provide BPM and visual variety. The two blocks of video-only stimuli were simply the A+V stimuli with the audio muted and presented in a re-randomized order for each participant.

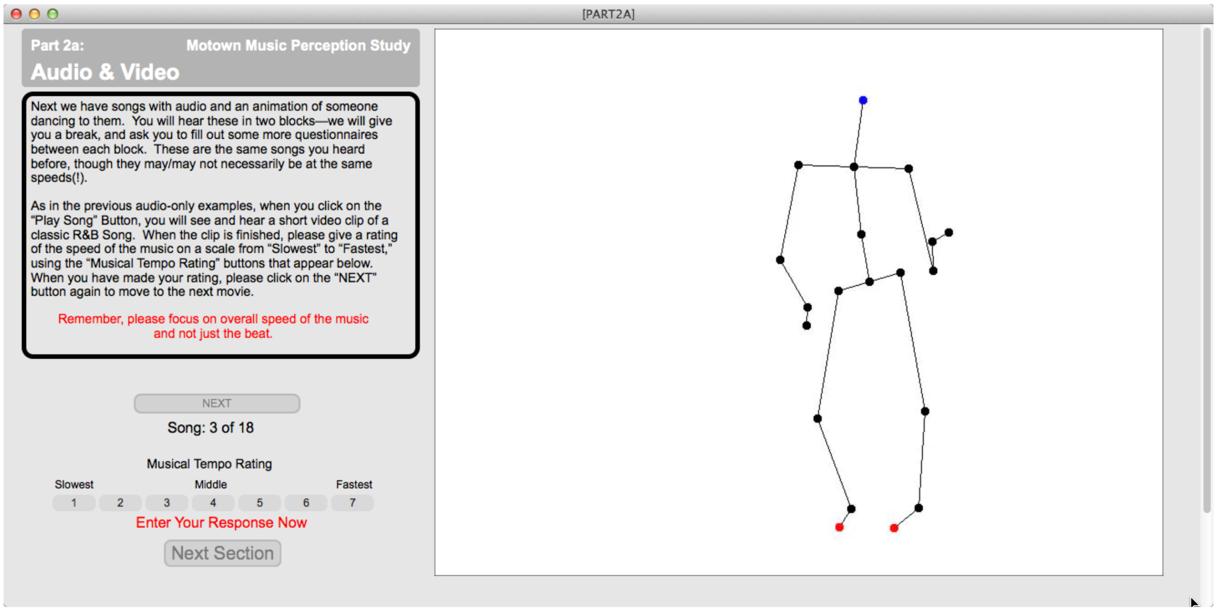
Table 2

Summary of Block Design and Stimuli Used in the Experiment.

	Block 1	**Block 2**	**Block 3**	**Block 4**	**Block 5**
Modality	Audio	A+V	A+V	Video	Video
Auditory Variables	3 Core BPM x 3 Time-Stretch (-5/0/+5)	3 Core BPM x 2 Time-Stretch (-5/+5)	3 Core BPM	3 Core BPM x 2 Time-Stretch (-5/+5)	3 Core BPM
Visual Variables	(none)	Single, free interpretation	2 Interpretations (vigor/relaxed)	Single, free interpretation	2 Interpretations (vigor/relaxed)

The introduction and pretest consisted of demonstration songs at the low and high end of the tempo range, as well as time-stretched versions of a sample song to familiarize participants with the range of stimuli used in the experiment (demo songs were not used in the experiment). The pretest then presented participants with a simple rock drumming pattern to precisely indicate the range of tempos used in the experiment (100-135 BPM) as well as to familiarize them with the response interface and tempo rating procedure, using a seven-point scale (1 = slowest to 7 = fastest). All participants were able to successfully rank order the rock drumming patterns, indicating that they were able to make tempo discriminations within the time-scale used in the experiment.

All stimuli (audio and video) were 10 seconds in duration. Each began on the first significant downbeat following the introductory portion of each song. Figure 1 gives a screenshot of the stimulus presentation in an A+V trial. Participants were able to provide a response only after the entire stimulus had been presented.



Next we have songs with audio and an animation of someone dancing to them. You will hear these in two blocks—we will give you a break, and ask you to fill out some more questionnaires between each block. These are the same songs you heard before, though they may/may not necessarily be at the same speeds(!).

As in the previous audio-only examples, when you click on the "Play Song" Button, you will see and hear a short video clip of a classic R&B Song. When the clip is finished, please give a rating of the speed of the music on a scale from "Slowest" to "Fastest," using the "Musical Tempo Rating" buttons that appear below. When you have made your rating, please click on the "NEXT" button again to move to the next movie.

Remember, please focus on overall speed of the music and not just the beat.

Figure 1. Screenshot of Max/MSP environment used for stimulus presentation and data collection in each block of the experiment; this screenshot is from Block 3 (A+V Condition)

After making their response, participants then cued the next stimulus. In the audio and A+V conditions the next stimulus was presented after a variable 4-5 second delay, to minimize carry-over effects of auditory beat entrainment, given the BPM rates used in the experiment (Van Noorden & Moelants 1999; London 2012); no added delay was deemed necessary in the video-only condition (Grahn, Henry, & McAuley 2012). In every trial participants were reminded to focus on the speed of the music and not just the beat rate and/or the speed of the dancing figure. In the video only conditions, participants were told they were watching dancers at a club through a window, and to imagine the speed of the music to which the observed figures were dancing.

Stimuli were presented to participants in a quiet room on an iMac Desktop Computer (20-inch screen, 2.16 Ghz Intel Core2 Duo, with 3 or 4GB Ram, running OS 10.7.5 or 10.8.5) via a Max 6 patch (www.cycling74.com) for presentation of both audio and video stimuli and collection of participant responses. Participants listened via Sennheiser HD25 headphones, which provided additional attenuation of ambient noise, with the headphone volume adjusted to a comfortable listening level. Once the pre-test was complete, the experimenter left the room to avoid any biasing of the participant's responses.

3.2 Results

3.2.1 Individual Block Results

Audio-only condition (Block 1). For the audio-only condition (block 1), a 3x3 (BPM x Time Stretch) repeated-measures ANOVA showed a main effect for BPM category (105 vs. 115 vs. 130 core BPM), $F(1.879, 99.599) = 65.219$, $p < .001$, $\eta_p^2 = .552$ and a main effect for time-stretching within each category, $F(1.892, 100.291) = 162.005$, $p < .001$, $\eta_p^2 = .753$ (Greenhouse-Geisser corrections applied here and in all other ANOVAs).[3] The interaction between BPM category and time-stretching was not significant. Post-hoc pairwise comparisons showed the differences between all core BPM levels were statistically significant, though the difference between the 105 and 115 BPM levels was small, with average tempo ratings of 3.54 and 3.83, respectively ($F(1, 26) = 9.63$, $p = .005$, $\eta_p^2 = .270$). As is evident from Figure 2, participants were readily able to discern and correctly rank the original and time-stretched versions of each song.

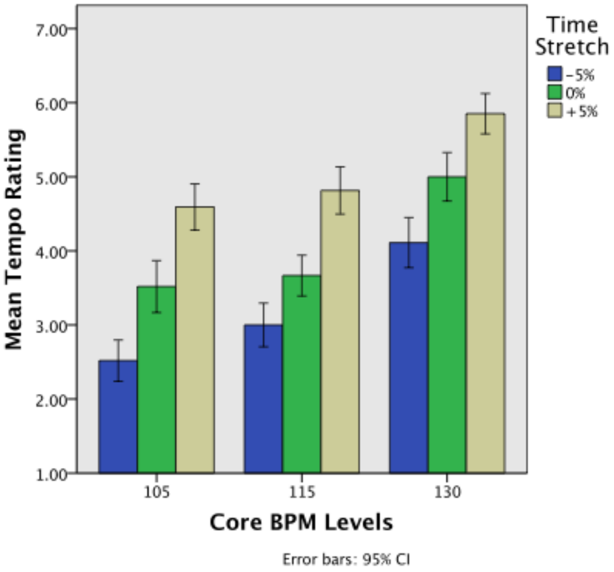


Figure 2. Average participant tempo ratings (y axis) for stimuli in audio-only condition (Block 1).

The -5% time stretched versions at the two slowest BPM levels were given significantly different ratings ($t(53) = 3.08$, $p = .003$, $d = .419$). However, not only were the original and +5% versions of the 105 and 115 BPM songs given similar ratings (see Figure 2), the +5% versions of the 105 BPM songs (now at 110 BPM) were given a higher average rating (4.59) than the

[3] The proper statistical analysis of Likert and Likert-type data has been a matter of recent debate (see esp. Jamieson 2004). As Norman (2010) points out, however, in many contexts ordinal data closely approximate true interval data and hence are suitable for parametric analyses such as ANOVA, including the analysis of both main effects and interactions. In our case, because (a) our data are averaged from several trials, grounded in their relation to an interval value scale (the BPM measures), and continuous, and (b) ANOVA has been shown to be very robust, we believe ANOVA methods are appropriate for our analysis. Moreover, almost all of our results are extremely significant (most *p* values for our response data analysis are ≤ .005), which means that we are likely to meet the more stringent criteria as would be used in non-parametric tests.

original versions of the songs at 115 BPM (3.67; $t(53) = -5.12$, $p < .001$, $d = -.699$). Had the participants tempo ratings been veridical, Figure 3 would have shown a monotonic increase in tempo ratings from left to right. Thus there is an apparent confusion between the absolute BPM level of each different stimulus and the relative ratings of the time-stretched versus original versions of each song (see discussion of the "anchoring effect" below).

Finally, there were two small effects of gender and musical background. Male participants tended to rate stimuli faster than female participants ($F(1, 15) = 8.39$, $p = .004$, $\eta_p^2 = .018$) and musically experienced participants tended to use a reduced range of the scale, as they were less apt to use the highest tempo ratings: $F(1, 15) = 8.53$, $p = .004$, $\eta_p^2 = .018$). There were no significant interactions between gender or musical background and tempo rating.

Time-stretched stimuli, A+V condition (Block 2). For the A+V condition there were no participant effects of musical background or stimulus familiarity, though the effect of gender was nearly significant ($p = .055$), as again males tended to rate songs slightly faster than females. A 3x2 (Core BPM x Time Stretch) repeated measures ANOVA found main effects for BPM category (F(1.929, 109.974) = 62.757, $p < .001$, $\eta_p^2 = .524$) and (unsurprisingly) for time stretching (F(1, 57) = 176.484, $p < .001$, $\eta_p^2 = .756$); there was no significant interaction. Figure 3 shows the participant ratings for each stimulus category in the A+V condition. Pairwise comparisons found non-significant differences only between the 105(-5%)/115(-5%) and 105(+5%)/115(+5%) pairs; all other differences were significant ($p < .001$ for all, except 115(+5%)/130(-5%), $p = .026$, and 105(+5%)/130(-5%), $p = .009$).

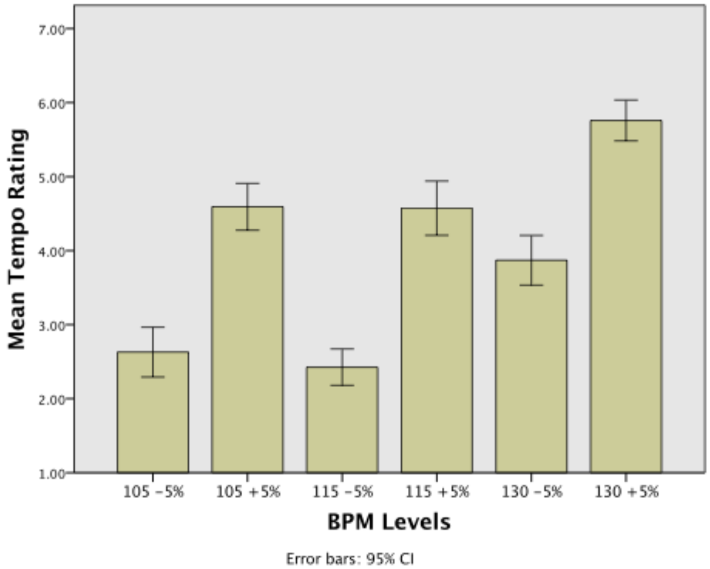


Figure 3. Average participant tempo ratings (y axis) for time-stretched stimuli, A+V condition (Block 2).

As can be seen, participant tempo ratings for the time-stretched stimuli in the A+V condition were similar to those for the time-stretched stimuli in the audio-only condition, the difference being that the presence of the video information eliminated the tempo rating distinction between the 105(-5%) and 115(-5%) BPM levels that occurred in Block 1.

Core BPM stimuli with instructed video, A+V condition (Block 3). In this part of the experiment the core BPM audio stimuli were paired with videos either having a slow, relaxed interpretation or a fast, vigorous dance interpretation. Thus there were three BPM levels (105, 115, and 130 BPM) and two contrasting Video Conditions (Relaxed vs. Vigorous). A 3x2 repeated-measures ANOVA found main effects for both BPM ($F(1.978, 104.811) = 38.041$, $p < .001$, $\eta_p{}^2 = .418$) and Video Condition ($F(1, 53) = 43.321$, $p < .001$, $\eta_p{}^2 = .449$); the interaction was not significant. Figure 4 shows the participant ratings for the Relaxed vs. Vigorous video conditions, grouped by BPM level. Stimuli with relaxed videos were consistently rated slower than vigorous videos, but post-hoc pairwise comparisons between the corresponding Relaxed and Vigorous A+V stimuli at the 105 and 115 BPM levels were also not statistically significant (though the Relaxed pair was near significance, $t(53) = -1.88$, $p = .066$, two tailed).

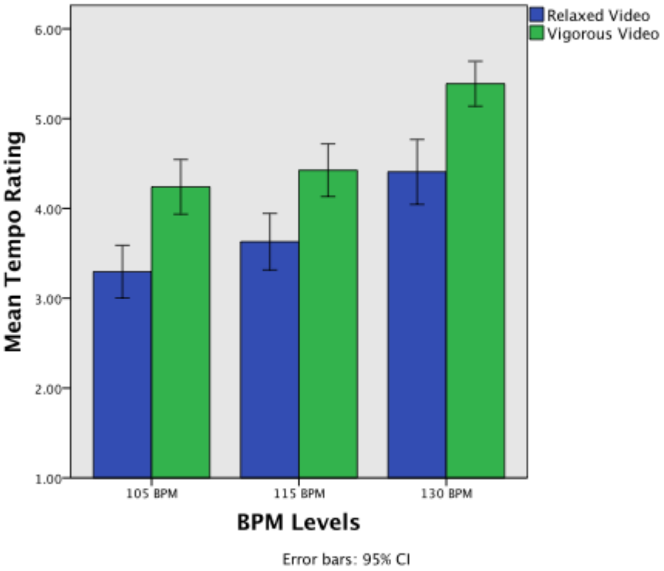


Figure 4. Average participant tempo ratings (y axis) for A+V condition (Block 3), Relaxed vs. Vigorous video stimuli.

There were no effects of gender or musical training on participant ratings in Block 3, but a small effect of familiarity (i.e., operationally defined as whether a participant had previously heard three or more of the songs used in the experiment) was found. A 2x2 (Familiarity X Video Condition) Independent Measures ANOVA found a very small main effect for familiarity ($F(1, 482) = 4.61$, $p = .032$, $\eta_p{}^2 = .009$), and another small interaction between Familiarity x Video Condition ($F(1, 482) = 11.33$, $p = .001$, $\eta_p{}^2 = .023$), as those with greater familiarity were slightly less affected by the video stimuli.

Time-Stretched stimuli, Video-Only condition (Block 4). Block 4 is the video-only analogue to Block 2; participant tempo ratings are summarized in Figure 5. A 3x2 repeated measures ANOVA (Core BPM x Time Stretch) found main effects for BPM ($F(1.867, 98.962) = 45.923$, $p < .001$, $\eta^2{}_p = .464$) and for Time Stretch ($F(1, 53) = 66.277$, $p < .001$, $\eta^2{}_p = .556$). A post-hoc pairwise comparison again showed that the difference between the core 105 and 115 BPM levels was not significant, but the differences between 130 BPM and both 105 and 115 BPM levels were highly significant ($p < .001$). Differences between all time-stretched pairs were highly significant ($p \leq .001$). There were no significant differences in tempo rating between

105(+5%) and 115(+5%) BPM, as well as between 115(+5%) and 130(-5%) BPM.

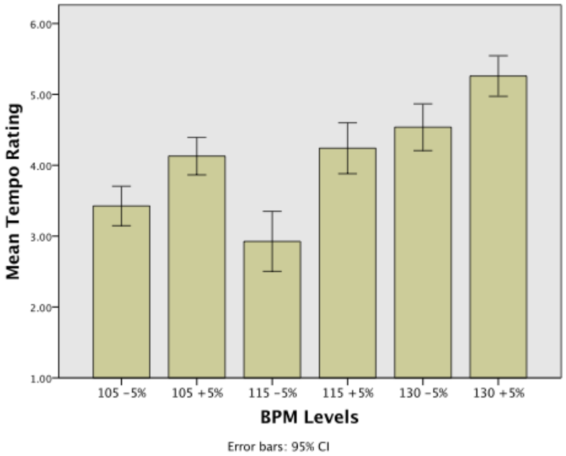


Figure 5. Average participant tempo ratings (y axis) for Time-Stretched stimuli, video-only condition (Block 4).

Core BPM stimuli with instructed video, Video-Only condition (Block 5). Block 5 is the video-only analogue to Block 3; participant tempo rankings are summarized in Figure 6. Thus as with block 3, there were three BPM levels (105, 115, and 130 BPM) and two Video Conditions (Relaxed vs. Vigorous). A 3x2 repeated-measures ANOVA found main effects for both variables, BPM (*F*(1.389, 73.603) = 20.295, $p < .001$, $\eta_p^2 = .277$), and Video Condition (*F*(1, 53) = 197.231, $p < .001$, $\eta_p^2 = .788$). Here the interaction between BPM and Video Condition was highly significant (*F*(1.748, 92.654) = 31.724, $p < .001$, $\eta_p^2 = .374$); as can be seen from Figure 6, this interaction was largely due to the extreme contrast between the relaxed and vigorous conditions at the 130 BPM level. A separate repeated measures ANOVA comparing the six individual stimulus categories (*F*(3.16, 167.29) = 54.53, $p < .001$, $\eta_p^2 = .507$) found that differences between many individual BPM/Video categories were no longer significant (105 BPM Relaxed and 130 BPM Relaxed, as well as between 105 BPM Vigorous, 115 BPM Relaxed, and 115 BPM Vigorous). All other differences were highly significant ($p < .001$). Three distinct tempo rating categories are thus evident, but they are not applied to the stimuli in any consistent manner. More precisely, a trend of linearly increasing tempo ratings relative to increased BPM is evident in the Vigorous condition (though the difference between the 105 and 115 BPM levels is n.s.). This trend is absent in the Relaxed condition.

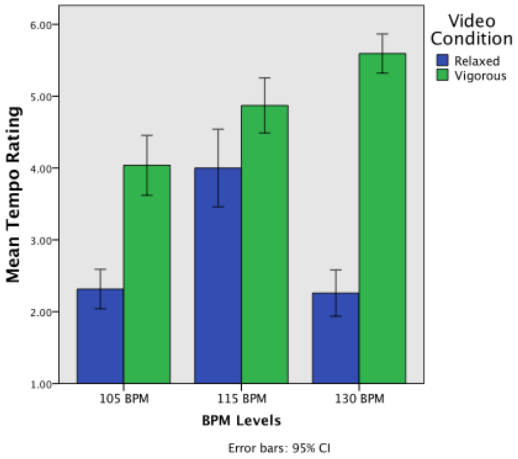


Figure 6. Average participant tempo ratings (y axis), Relaxed vs. Vigorous dance interpretation, Video-Only condition (Block 5).

3.2.2 Inter-Block Comparisons

BPM and tempo judgments: Time-Stretched stimuli (Blocks 1, 2, and 4). As can be seen in Figure 7, tempo ratings across these three blocks are remarkably consistent; a 3x6 repeated-measures ANOVA (Blocks x BPM levels) showed no main effect for Block, but a statistically significant interaction between Block and Tempo (*F*(7.92, 419.93) = 8.00, $p < .001$, $\eta_p^2 = .131$). A within-subjects contrast showed that the only significant difference between Blocks 1 and 2 occurs at 115-5% (*F*(1, 53) = 8.04, $p = .006$, $\eta_p^2 = .132$), whereas all differences in tempo ratings between Blocks 2 and 4 are significant ($p \leq .001$). Paired samples *t*-tests showed no significant differences between blocks in terms of their grand means or their standard deviations. A narrowing of the range of responses in the video-only condition (Block 4) is evident, as is a tendency toward increasing variance in tempo judgments in both the A+V and video-only conditions, contradicting research hypothesis #2.

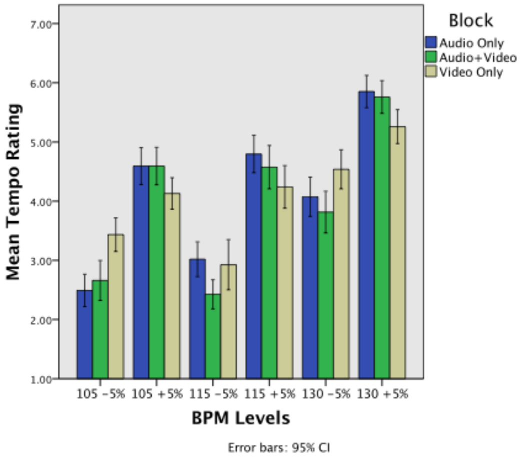


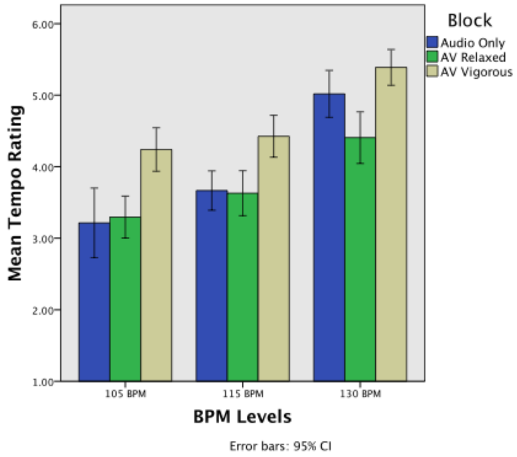
Figure 7. Comparison of average participant tempo ratings (y axis) in Blocks 1, 2, and 4.

BPM and tempo judgments: Core BPM stimuli (Blocks 1, 3, and 5). The relationships between Blocks 1, 3, and 5 are quite different from Blocks 1, 2, and 4, as one would expect. Given the nature of the results, and for clarity, we will discuss the the audio-only compared to the A+V condition (Blocks 1 & 3) and the audio-only compared to the video-only condition (Blocks 1 & 5) separately.

As can be seen in Figure 8 panel (a), the results between the audio-only and A+V blocks are quite similar. A 3x3 repeated measures ANOVA (BPM x Presentation Condition) found an effect of stimulus presentation (*F*(1.88, 100.09) = 22.39, $p < .001$, η_p^2= .297), BPM (*F*(1.97, 104.19) = 53.12, $p < .001$, η_p^2= .501), but no significant interaction between blocks. Pairwise comparisons found no significant difference between the audio-only and relaxed A+V interpretation at the 105 and 115 BPM levels, but a highly significant difference between the vigorous A+V presentation and the other two conditions at all BPM Levels (Bonferroni correction applied for multiple comparisons). Thus, at the two slowest core BPM levels, only the vigorous interpretation was able to modulate participant tempo ratings, whereas at the fastest BPM level, both relaxed and vigorous videos affected participant tempo ratings.

As can be seen in Figure 8 panel (b), the effects in the video-only condition are strikingly different in comparison with the audio-only condition. A similar 3x3 repeated measures ANOVA found significant effects for Condition (*F*(1.77, 93.85) = 97.20, $p < .001$, η_p^2 = .647), BPM (*F*(1.44, 76.14) = 25.35, $p < .001$, η_p^2 = .324), and a significant interaction between Condition and BPM (*F*(3.54, 187.453) = 27.97, $p < .001$, η_p^2 = .345).

(a)



(b)

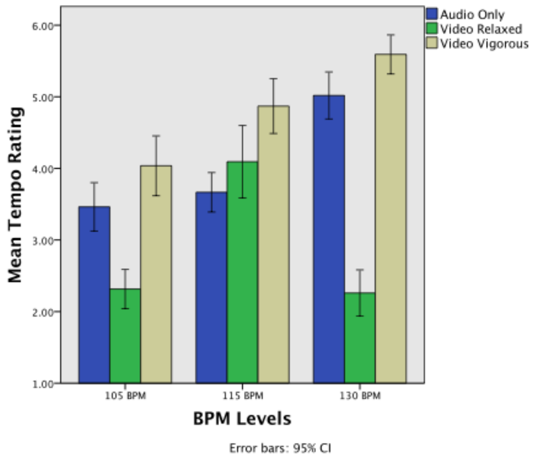


Figure 8. Comparison of average participant tempo ratings (y axis) in Blocks 1, 3, and 5. Panel (a): Audio-Only vs. A+V conditions; Panel (b): Audio-Only vs. Video-Only conditions.

In the absence of audio, relaxed versus vigorous interpretations of the same song are generally distinguished, but there are marked differences between the vigorous versus relaxed videos. The vigorous videos show a monotonic increase in rating that corresponds to increasing core BPM levels, analogous to the A+V condition (compare Figure 8a). The relaxed videos, however, are inconsistent: at 105 BPM the video-only ratings are slower than in the audio condition, at 115 BPM they are faster, and at 130 BPM they are again slower.

3.2.3 Results Summary

Our first hypothesis--that participants will be able to discriminate and properly rank the tempos of original and temporally manipulated unimodal auditory stimuli--was (surprisingly) only partially confirmed. While the three BPM levels were discerned and ranked properly, a more fine-grained assessment shows a confusion between the core BPM levels and the time-stretched versions of particular songs, producing tempo ratings that were incommensurate with the actual BPM rates for many of the stimuli. Our second hypothesis--that stable and matched combinations of musical and visual cues would yield more precise tempo judgments than in the unimodal auditory context--was refuted, as there was no decrease in variability in participant ratings in Block 2 as compared to Block 1. Our third research hypothesis--that systematically varied visual cues that are ecologically relevant will affect the perception of concurrently presented music--was confirmed. Vigorous dance interpretations paired with audio (Block 3) produced faster tempo ratings in comparison to the audio alone (Block 1), though relaxed dance interpretations had no effect. In addition, we found that tempo judgments could be reliably extracted from the video information alone, at least in some contexts (Block 4, Block 5 vigorous interpretation), although the pattern of participant ratings suggests some influence of the previous experimental blocks (see discussion below).

4. Discussion.

In this experiment participants were presented with audio-only (Block 1), A+V (Blocks 2 & 3) and video-only (Blocks 4 & 5) stimuli and given a tempo-rating task. Audio stimuli consisted of original and time-stretched excerpts from classic Motown/R&B songs, and video stimuli were point-light displays created from motion capture of dancers moving to the audio stimuli. Audio and video stimuli were paired either (a) to exhibit a stable match between the objective beat rate in the audio and the movements of the dancers (Blocks 2 & 4), or (b) to present dancers moving in both a "relaxed" and "vigorous" fashion to the same audio stimulus (Blocks 3 & 5). The results in each of the stimulus contexts (audio-only, A+V, and video-only) are discussed in turn below.

4.1 Tempo judgments in the audio-only context

The pattern of results in the audio-only condition (Block 1) was unexpected. Our participants were able to distinguish the three core BPM categories used in the experiment, though the difference between the two slowest categories (105 vs. 115 core BPM), while statistically significant, was quite small (3.54 vs. 3.83). Participants were also able to make fine-grained distinctions amongst faster versus slower versions of particular songs, and put these song-clusters in a rank order that corresponded to their BPM rates. Honing (2007) found that listeners could distinguish original from time-stretched versions of music in a familiar style or genre, and he posited that this may be due to links between expressive timing nuances and particular BPM rates. That is, the timing ratios between successive notes music that has a degree of rhythmic "swing" (as is the case with the R&B songs used here) are yoked to particular BPM rates, such that when these recordings are manipulated via time-stretching, the resulting ratios are slightly "off." The goodness-of-fit (or lack thereof) between the original vs. time-stretched timing nuances many have served as a cue for distinguishing original versus time-stretched versions of songs in Block 1. At the same time, Franěk & Fabiánová (2003) found that listeners have good short-term memories for BPM rates in the context of reasonably complex musical sequences, as opposed to simply retaining a metronome rate, which also may have also played a role here.

However, as noted above, participants failed to correctly rank the sped-up versions of the songs at one BPM level relative to the slowed-down or original tempo songs at the next-highest level. There were significant overlaps between the stimulus subcategories in adjacent BPM levels, though the very slowest songs (105-5% BPM) were rated slower than any other stimuli. Fast versions of slow songs were judged to be both (a) faster than slow versions of faster songs (e.g., 105+5% vs. 115-5%), even if they were at the same objective BPM rate, and (b) faster than the original version of a faster song (e.g., 105+5% vs. 115 at the original BPM), even though the time-stretched songs are objectively slower than the unstretched songs. Thus it seems unlikely that participants made their tempo judgments on the basis of comparing each stimulus to some internal/absolute tempo standard.

This confusion of tempo ratings may be due to the limitations of the rating scale, as a range of seven values may have been too small to individuate all of the tempos presented here. More generally, participants using Likert-type scales may tend to avoid the rating extremes, which could have contributed to the overlaps. However, a song-specific "tempo anchoring effect" also seems plausible. Given the multiple presentations of each song, over the course of Block 1 each participant was able to build both a sense of the tempo of "the song" and a sense of the tempo for a particular "version" of a song (recall that in the pre-test participants were informed that they would be presented with versions of the same song at different tempos, and given a demonstration of the range of time-stretching they would encounter). Their tempo ratings would thus involve an absolute sense of the tempo of a particular song, versus the relative sense of the tempos of different versions of a song. The relative closeness of the core BPM levels used in the experiment versus the need to sort out the tempos of different versions led to overlaps in tempo judgment. Listeners often have to mediate between absolute versus relative senses of tempo, and both of these may be influenced by prior encounters with the same or similar music. The distinction between absolute and relative tempo judgments requires stimuli that are musically distinguishable, such that one can speak of faster versus slower versions of the same piece, or music of the same type or genre. To produce the distinguishability required to create these nested tempo distinctions requires sufficient musical richness and complexity. These distinctions will not arise in the context of impoverished stimuli such as metronome clicks or even fairly complex stimuli that are musically generic such as standard rock drumming patterns. Levitin & Cook (1996) demonstrated that long-term memory for the tempo of particular pieces can be quite precise, and this long-term memory could anchor relative judgments of tempo for particular pieces. This is also why it makes sense when musicians speak of a "slow March" or a "fast Sarabande," as slow Marches are not really slow, nor are fast Sarabandes really fast. Rather, there is an interplay between the normative tempos associated with particular genres (absolute) versus the tempo of a particular piece in comparison to generic norms (relative), so "slow march" really means "slow (for a march)."

4.2 Tempo judgments in the A+V context

In the A+V contexts where there was a fixed relationship between the audio-video stimuli (Block 2), participant tempo ratings were consistent with the ratings made in the audio-only condition. Toiviainen, Luck & Thompson (2010) found that bodily movements often embody the multiple periodicities present in the music. Burger, Thompson, Saarikallio, Luck & Toiviainen (2014) found movement characteristics to be related to clear pulse structures at the beat level. We hypothesized that the additional temporal information in the video would reduce the variability of participant responses, but this did not occur. One reason may be that there is a ceiling effect from the audio-only condition: the auditory modality contains all the information needed to perform the experimental task and is highly salient, thus little temporal information was added by the video. Another explanation may be that while the videos do add some tempo information, they also add some amount of rhythmic "noise," and those two aspects more or less cancel out, yielding the same degree of precision as in the audio-only context. A final explanation may be, following conventional wisdom, that for temporal judgment tasks like the one used here, audition dominates vision, especially if one has been presented with a presentation of the auditory component of the stimulus beforehand (i.e., a priming effect). This would explain a particular carry-over effect from the audio-only context, as the tempo anchoring effect is still evident in Block 2. There were no significant differences between the ratings for the audio versus the A+V stimuli at the 105 and 115 BPM levels (see Figure 3).

In Block 3 there was not a fixed relationship between the audio and video stimuli, and here a stronger influence of visual cues on tempo judgment is apparent. In those contexts, participants were readily able to distinguish between relaxed versus vigorous dance interpretations of the same song, even though both interpretations conveyed the same musical tactus in both audio and video stimuli. The three core BPM levels were also distinguished, though as in Block 1, the difference in tempo ratings between the 105 and 115 BPM levels is very small. Chiefly, in comparison with Block 1, in Block 3 vigorous dancing increased tempo ratings at all BPM levels, and both interpretations (relaxed vs. vigorous) modulated participants' responses at the 130 BPM level. At the 105 and 115 BPM levels the relaxed dance interpretations did not affect the tempo ratings in comparison with Block 1. Note again that in both Blocks 2 and 3 the movements of the dancers in all videos were phase locked to the beat level in the audio; this was not a case where the video information is temporally incongruent with the audio. Though not temporally incongruent, in some cases the videos could be described as gesturally incongruent.

As a reminder, the task involved a plausible real-world scenario: hearing and seeing a figure dancing to music and reporting a sense of the music's speed. In this context, we readily presume that the dancers motions are driven by the music, mirroring many aspects of the music's kinetic and expressive properties. That mirroring includes tempo, which one may take as a kind of kinematic default: in most instances, one moves at the same speed as the music. Of course, it is possible (and, we would conjecture, not unusual) to dance slowly to fast music, and our participants were presumably aware of this possibility. Thus in the context of the relaxed dance interpretations for music at a moderate tempos (105 and 115 BPM), they recognized that the dancers were gesturally incongruent--they were no longer mirroring the tempo of the music--and so participants adjusted their focus to the music, in accordance with the experimental prompt. By contrast, with the vigorous dance interpretations the video information appeared to be gesturally congruent. In those cases our participants once again relied on the normative association between movement and tempo, as well as a presumption that we do not normally move *faster* than the music. This is a reversed version of Schutz & Kubovy's "binding by causality" hypothesis: "sounds have a propensity to bind with the visible movements that could have caused them" (Schutz & Kubovy 2009, p. 1793). Here, visible movements are bound to the sounds that motivated them. Thus the vigorous interpretations, with their high kinetic energy and overall levels of movement, served to amplify the perceived tempo rate and systematically affected the musical tempo judgments.

4.3 Tempo judgments in the video-only context, and multisensory memory

Though it was not the primary objective of our study, we did want to give at least a preliminary examination of our participants' perception of tempo from the video stimuli alone. Here our results are mixed and our interpretation is also limited by our experimental design, as we did not counterbalance the order of the presentation blocks as would be required to fully determine the independent contributions of the video components to our participants' tempo judgments. Nonetheless, a number of significant results emerged from Blocks 4 and 5.

In Block 4, where there had previously been a unique pairing between the video and the auditory stimuli (Block 2), participant tempo ratings corresponded with ratings from the audio only (Block 1) and A+V (Block 2) presentations. The range for video-only ratings was somewhat compressed, but the same overall pattern of differentiation is evident, including lingering evidence of the song anchoring effect (see Figure 7). In Block 5, where contrasting dance interpretations had been previously paired with the same video, commensurate ratings with audio only (Block 1) and A+V (Block 3) ratings only occurred with the vigorous interpretation. There is no clear correspondence between Blocks 1 and 3 with the Relaxed videos in Block 5.

There are a number of possible explanations for these participant ratings in Blocks 4 and 5. The simplest is that participants simply recovered tempo information from the videos alone, with no influence of the previous stimulus blocks. This would partially explain the differentiation and ranking of tempos in Block 4 and the Vigorous videos in Block 5; the failure of participants to make veridical rankings for the Relaxed videos may simply be because they lack enough salient tempo information to allow them to perform the task. Though parsimonious, this explanation seems implausible given the design of the experiment and the pattern of results in Block 4. The design primes participants for the dominance of audio in multimodal contexts by presenting them with audio-only stimuli first, and the instructions--repeated on every trial--also biased participants to the auditory modality. The audio stimuli are both highly memorable and were given repeated presentations in Block 1, albeit with time-stretching and/or video interpretation variants. At the same time the visual information that was paired with the audio in blocks 2 and 3 has a strong gestural congruence with the audio, as the music and its temporal structure is directly related to the motions of the dancer. This should engender a strong binding of auditory and visual features, whether or not a multisensory representation of the A+V stimuli is encoded in memory (Quak, London, & Talsma 2015). The Relaxed videos are the exception, as they lack the same gestural congruence between the BPM rate in the music and the level of motion in the video. If this is so, the results in Blocks 4 and 5 may be interpreted as follows: both due to the innate dominance of audition over vision for temporal information, and the block design of the experiment, participant tempo ratings were initially formed in Block 1. They were then reinforced in Blocks 2 and 3 in those contexts where there was a gestural congruence between the audio and video. This congruence enabled a strong binding of auditory and visual tempo features. Where there was a lack of congruence in the Relaxed videos in Block 3, that binding did not occur. As a result, similar tempo ratings persisted in Blocks 4 and 5 where strong binding had occurred in Blocks 2 and 3, as participants remembered their audio ratings even when the audio was muted. This would account for the recurrence of the tempo anchoring effect in Block 4. This interpretation, where a prior auditory stimulus affects subsequent multi-sensory experiences, complements the findings of Thelen, Talsma, & Murray (2015) who found prior multi-sensory experience could affect the discrimination of subsequently presented auditory objects. In both their and our research, this effect depends on congruence between the auditory and visual cues.

4.4 Conclusion

This study has investigated the interaction between auditory and visual cues for musical tempo, and has shown that our sense of tempo involves more than just beat rate (BPM). We have shown that even in a unimodal (auditory-only) context, tempo determination can involve both absolute and relative senses of the music's speed, the latter relying upon prior experience of the music. More broadly, we have explored a context where the integration of auditory and visual information takes place in a meaningful way. This is not a "semantic" coherence of information, but an ecological coherence of sights and sounds that normatively co-occur in many of our real-world experiences of music. As Spence (2015) has observed:

> When stimuli presented in different sensory modalities correspond, there may be perceptual interactions observed that are not present when the stimuli are incongruent. . . What is more, there is also a feeling of rightness that accompanies the pairing of stimuli that correspond cross-modally. . . Such correspondences need not be based on perceptual mapping, but they often are. What is more, they can often affect both *perceptual organization* and *awareness* [itals in original] (Spence 2015, in press).

The relative dominance of vision over audition has long been contested, but the growing consensus is that cross-modal interactions depend on the structure of the stimuli, one's perceptual goal, and the coherence of the auditory-visual object or dynamic array. In the particular case of musical tempo on the dance floor, the perceptual organization of the music is not affected by the dancers movements. Yet when experienced together, they give rise to a mutual awareness of musical and physical energy. Indeed, perhaps one reason that determinations of tempo are hard to pin down is that tempo itself may be a placeholder for a broader, energistic account of perceived musical motion. Our sense of tempo may be, at bottom, a sense of how much effort it takes for us to move with the music that we hear. When we see someone dancing to music, we get a clear indication of that effort, as we might imagine ourselves dancing with them, moving to the music together.

References

Alluri, V. & Toiviainen, P. (2010). Exploring perceptual and acoustic correlates of polyphonic timbre. *Music Perception*, 27(3): 223–241.

Arrighi, R., Alais, D., & Burr, D. (2006). Perceptual synchrony of audiovisual streams for natural and artificial motion sequences. *Journal of Vision 6*: 260–268.

Arrighi, R., Marini, F., & Burr, D. (2009). Meaningful auditory information enhances perception of visual biological motion. *Journal of Vision 9*: 1-7.

Boltz, M. G. (2011). Illusory tempo changes due to musical characteristics. *Music Perception 28*(4): 367-386.

Brittin, R. (1993). Discrimination of aural and visual tempo modulation. *Bulletin of the Council for Research in Music Education* 116: 23-32.

Burger, B., Ahokas, R., Keipi, A., & Toiviainen, P. (2013, July). Relationships between spectral flux, perceived rhythmic strength, and the propensity to move. Proceedings of the *10th Sound and Music Computing Conference*. Stockholm, Sweden: 179–184.

Burger, B., London, J., Thompson, M., & Toiviainen, P. Moving to Motown: The effect of volitional and musical factors on movement characteristics and beat synchronization. MS in preparation.

Burger, B., Thompson, M. R., Saarikallio, S., Luck, G., & Toiviainen, P. (2014). Hunting for the beat in the body: on period and phase locking in music-induced movement. *Frontiers in Human Neuroscience 8*(903): 1–16.

Burger, B., & Toiviainen, P. (2013, August). MoCap toolbox – A Matlab toolbox for computational analysis of movement data. *10th Sound and Music Computing Conference,* Stockholm, Sweden: 172–178.

Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008). Listening to musical rhythms recruits motor regions of the brain. *Cerebral Cortex 18*(12): 2844-2854.

Clynes, M., & Walker, J. (1986). Music as time's measure. *Music Perception 4*(1): 85-120.

Drake, C., & Bertrand, D. (2003). The quest for universals in temporal processing in music. in *The Cognitive Neuroscience of Music,* I. Peretz & Ro. Zatorre, eds. New York, Oxford University Press: 21-31.

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

Drake, C., Penel, A., & Bigand, E. (2000). Why musicians tap slower than non-musicians. In *Rhythm Perception and Production,* P. Desain and W. L. Windsor, eds. Lisse, Swets & Zeitlinger: 245-248.

Elowsson, A., & Friberg, A. (2013). Modeling perception of speed in music audio. Proceedings of the Sound and Music Computing Conference 2013, Stockholm, Sweden: 735-741.

Fraisse, P. (1984). Perception and estimation of time. *Annual Review of Psychology 35:* 1-36.

Franěk, M., & Fabiánová, H. (2003). Short-term memory for tempo of metronomic sequences. In *Proceeings of the 5th Triennial ESCOM Conference*, R. Kopiez, A. C. Lehmann, I. Wolther, C. Wolf, Eds. http://www.epos.uos.de/music/templates/buch.php?id=49

Grahn, J. A., & Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. *Journal of Cognitive Neuroscience 19*(5): 893-906.

Grahn, J. A., Henry, M. J., & McAuley, J. D. (2012). FMRI investigation of cross-modal interactions in beat perception: Audition primes vision, but not vice versa. *NeuroImage 54*(2): 1231-1243.

Honing, H. (2006). Evidence for tempo-specific timing in music using a web-based experimental setup. *Journal of Experimental Psychology: Human Perception and Performance 32*(3): 780-786.

Honing, H. (2007). Is expressive timing relational invariant under tempo transformation?. *Psychology of Music 35*(2): 276-285.

Hove, M. J., Iversen, J. R., Zhang, A., & Repp, B. (2013). Synchronization with competing visual and auditory rhythms: Bouncing ball meets metronome. *Psychological Research 77*(4): 388-398.

Iversen, J. R., Patel, A. D., Nicodemus, B., & Emmorey, K. (2015). Synchronization to auditory and visual rhythms in hearing and deaf individuals. *Cognition* 134: 232-244.

Jamieson, S. (2004). Likert scales: How to (ab)use them. *Medical Education* 38: 1217–1218.

Janata, P., Tomic, S. T., & Haberman, J. M. (2012). Sensorimotor coupling in music and the psychology of the groove. *Journal of Experimental Psychology: General 141*(1): 54-75.

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review 96*(3): 459-491.

Levitin, D., & Cook, P. (1996). Memory for musical tempo: Additional evidence that auditory memory is absolute. *Perception & Psychophysics* 56: 414–423.

London, J. M. (2011). Tactus ≠ tempo: Some dissociations between attentional focus, motor behavior, and tempo judgment. *Empirical Musicology Review 6*(1): 43-55.

London, J. M., & Cogsdill, E. (2011, August). Movement Rate Affects Tempo Judgments for Some Listeners. Proceedings of biannual meeting of *The Society for Music Perception and Cognition,* Rochester, NY.

Luck, G., & Sloboda, J. A. (2009). Spatio-temporal cues for visually mediated synchronization. *Music Perception 26*(5): 465-473.

Manning, F., & Schutz, M. (2013). Moving to the beat improves timing perception. *Psychonomic Bulletin & Review 20*(6): 1133-1139.

Martens, P. A. (2005). Beat-finding, listener strategies, and musical meter. Ph.D. Thesis, University of Chicago.

Merchant, H., & Honing, H. (2014). Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. *Frontiers in Neuroscience 7*(274). doi: 10.3389/fnins.2013.00274

Meyer, G. F. & Wuerger, S. M. (2001). Cross-modal integration of auditory and visual motion signals. *NeuroReport 12*(11): 2557-2560.

Miller, N. S., & McAuley, J. D. (2005). Tempo sensitivity in isochronous tone sequences: The multiple-look model revisited. *Perception & Psychophysics 67*(7): 1150-1160.

Molinari, M., Leggio, M. G., De Martin, M., Cerasa, A., & Thaut, M. (2003). Neurobiology of rhythmic motor entrainment. In *The Neurosciences and Music*, G. Avanzini, C. Faienza, L. Lopez, M. Majno, & D. Minciacchi, eds. *Annals of the New York Academy of Sciences 999*: 313-321.

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Science Education 15*: 625–632.

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

Patel, A. D., Iversen, J. R., Chen, Y., & Repp, B. H. (2005). The influence of metricality and modality on synchronization with a beat. *Experimental Brain Research* 163: 226-238.

Penel, A., Rivenez, M., and Drake, C. (2001). Estimates of sequence acceleration and deceleration support the synchronization of internal rhythms. *Annals of the NY Academy of Sciences: The Biological Foundations of Music 930:* 412-413.

Phillips-Silver, J., & Keller, P. E. (2012). Searching for roots of entrainment and joint action in early musical interactions. *Frontiers in Human Neuroscience 6*(26). doi: 10.3389/fnhum.2012.00026

Phillips-Silver, J., Aktipis, C. A., & Bryant, G. A. (2010). The ecology of entrainment: Foundations of coordinated rhythmic movement. *Music Perception 28*(1): 3-14.

Quak, M., London, R.E., & Talsma, D. (2015). A multisensory perspective of working memory. *Frontiers in Human Neuroscience* 9(197), doi: 10.3389/fnhum.2015.00197

Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review 12*(6): 969-992.

Repp, B, H., & Su, Y. (2013). Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonomic Bulletin and Review 20*(3): 403–52.

Schutz, M., & Kubovy, M. (2009). Causality and cross-modal integration. *Journal of Experimental Psychology: Human Perception and Performance 35*(6): 1791-1810.

Schutz, M., & Lipscomb, S. (2007). Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception 36*(6): 888-897.

Schwartze, M., Keller, P. E., Patel, A. D., & Kotz, S. A. (2011). The impact of basal ganglia lesions on sensorimotor synchronization, spontaneous motor tempo, and the detection of tempo changes. *Behavioural Brain Research 216*(2): 685-691.

Shams, L., Kamitani, Y., & Shimojo, S. (2004). Modulations of visual perception by sound. In *The Handbook of Multisensory Processes,* G. A. Calver, C. Spence, & B. E. Stein, eds. MIT Press: 27-33.

Snyder, J., & Krumhansl, C. L. (2001). Tapping to ragtime: Cues to pulse finding. *Music Perception 18*(4): 455-489.

Soto-Faraco, S., & Kingstone, A. Multisensory integration of dynamic information. In *The Handbook of Multisensory Processes,* G. A. Calver, C. Spence, & B. E. Stein, eds. MIT Press: 49-67.

Spence, C. (2015). Cross-modal perceptual organization. In the *Oxford Handbook of Perceptual Organization,* J. Wagemans, ed. Oxford University Press, in press.

Thelen, A., Talsma, D., & Murray, M. M. (2015). Single-trial multisensory memories affect later auditory and visual object discrimination. *Cognition 138*: 148-160.

Toiviainen, P., Luck, G., & Thompson, M. (2010). Embodied meter: Hierarchical eigenmodes in music-induced movement. *Music Perception 28*(1): 59-70.

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

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