Acta Psychologica 164 (2016) 70–80

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Speed on the dance floor: Auditory and visual cues for musical tempo☆

Justin London [a,*], Birgitta Burger [b], Marc Thompson [b], Petri Toiviainen [b]

[a] Carleton College, USA
[b] University of Jyväskylä, Finland

ARTICLE INFO

Article history:
Received 23 March 2015
Received in revised form 2 December 2015
Accepted 10 December 2015
Available online 30 December 2015

Keywords:
Music
Rhythm
Tempo
Audio-visual feature binding
Cross-modal perception

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 1.67–2 Hz. 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, 2011Drake, 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 & Kingstone, 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.

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 and 120 BPM (1.67–2 Hz). 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 600 ms, (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,

☆ 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.
* Corresponding author at: Department of Music, Carleton College, Northfield, MN 55057, USA.

http://dx.doi.org/10.1016/j.actpsy.2015.12.005
0001-6918/

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; Repp & Su, 2013; 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, 2012; Hove, Fairhurst, Kotz, & Keller, 2013). However, Hove, Iversen, Zhang, and Repp (2013) and Iversen, Patel, Nicodemus, and 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 and 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 and Wuerger (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 and 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 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.