Aus dem Institut für Medizinische Psychologie der
Ludwig-Maximilians-Universität München
Vorstand: Prof. Martha Merrow, PhD

The influence of external and internal motor processes on human auditory rhythm perception

Dissertation
zum Erwerb des Doktorgrades der Humanbiologie
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München

vorgelegt von
Yi-Huang Su
aus Taipei, Taiwan
2012

Mit Genehmigung der Medizinischen Fakultät der Universität München

Berichterstatter: Prof. Dr. Ernst Pöppel

Mitberichterstatter: Priv. Doz. Dr. Maria Schuster
Prof. Dr. Adrian Danek
Prof. Dr. Markus Suckfüll

Mitbetreuung durch den promovierten Mitarbeiter:

Dekan: Prof. Dr. med. Dr. h.c. M. Reiser, FACR, FRCR

Tag der mündlichen Prüfung: 24. 05. 2012

Table of Content

1. ABSTRACT & DEUTSCHE ZUSAMMENFASSUNG

1.1 Abstract (English)

Musical rhythm is composed of organized temporal patterns, and the processes underlying rhythm perception are found to engage both auditory and motor systems. Despite behavioral and neuroscience evidence converging to this audio-motor interaction, relatively little is known about the effect of specific motor processes on auditory rhythm perception. This doctoral thesis was devoted to investigating the influence of both *external* and *internal* motor processes on the way we perceive an auditory rhythm. The first half of the thesis intended to establish whether overt body movement had a facilitatory effect on our ability to perceive the auditory rhythmic structure, and whether this effect was modulated by musical training. To this end, musicians and non-musicians performed a pulse-finding task either using natural body movement or through listening only, and produced their identified pulse by finger tapping. The results showed that overt movement benefited rhythm (pulse) perception especially for non-musicians, confirming the facilitatory role of external motor activities in hearing the rhythm, as well as its interaction with musical training. The second half of the thesis tested the idea that indirect, covert motor input, such as that transformed from the visual stimuli, could influence our perceived structure of an auditory rhythm. Three experiments examined the subjectively perceived tempo of an auditory sequence under different visual motion stimulations, while the auditory and visual streams were presented independently of each other. The results revealed that the perceived auditory tempo was accordingly influenced by the concurrent visual motion conditions, and the effect was related to the increment or decrement of visual motion speed. This supported the hypothesis that the internal motor information extracted from the visuomotor stimulation could be incorporated into the percept of an auditory rhythm. Taken together, the present thesis concludes that, rather than as a mere reaction to the given auditory input, our motor system plays an important role in contributing to the perceptual process of the auditory rhythm. This can occur via both external and internal motor activities, and may not only influence how we hear a rhythm but also under some circumstances improve our ability to hear the rhythm.

1.2 Deutsche Zusammenfassung

Musikalische Rhythmen bestehen aus zeitlich strukturierten Mustern akustischer Stimuli. Es konnte gezeigt werden, dass die Prozesse, welche der Rhythmuswahrnehmung zugrunde liegen, sowohl motorische als auch auditive Systeme nutzen. Obwohl sich für diese auditiv-motorischen Interaktionen sowohl in den Verhaltenswissenschaften als auch Neurowissenschaften übereinstimmende Belege finden, weiß man bislang relativ wenig über die Auswirkungen spezifischer motorischer Prozesse auf die auditive Rhythmuswahrnehmung. Diese Doktorarbeit untersucht den Einfluss externaler und internaler motorischer Prozesse auf die Art und Weise, wie auditive Rhythmen wahrgenommen werden. Der erste Teil der Arbeit diente dem Ziel herauszufinden, ob körperliche Bewegungen es dem Gehirn erleichtern können, die Struktur von auditiven Rhythmen zu erkennen, und, wenn ja, ob dieser Effekt durch ein musikalisches Training beeinflusst wird. Um dies herauszufinden wurde Musikern und Nichtmusikern die Aufgabe gegeben, innerhalb von präsentierten auditiven Stimuli den Puls zu finden, wobei ein Teil der Probanden währenddessen Körperbewegungen ausführen sollte und der andere Teil nur zuhören sollte. Anschließend sollten die Probanden den gefundenen Puls durch Finger-Tapping ausführen, wobei die Reizgaben sowie die Reaktionen mittels eines computerisierten Systems kontrolliert wurden. Die Ergebnisse zeigen, dass offen ausgeführte Bewegungen die Wahrnehmung des Pulses vor allem bei Nichtmusikern verbesserten. Diese Ergebnisse bestätigen, dass Bewegungen beim Hören von Rhythmen unterstützend wirken. Außerdem zeigte sich, dass hier eine Wechselwirkung mit dem musikalischen Training besteht. Der zweite Teil der Doktorarbeit überprüfte die Idee, dass indirekte, verdeckte Bewegungsinformationen, wie sie z.B. in visuellen Stimuli enthalten sind, die wahrgenommene Struktur von auditiven Rhythmen beeinflussen können. Drei Experimente untersuchten, inwiefern das subjektiv wahrgenommene Tempo einer akustischen Sequenz durch die Präsentation unterschiedlicher visueller Bewegungsreize beeinflusst wird, wobei die akustischen und optischen Stimuli unabhängig voneinander präsentiert wurden. Die Ergebnisse zeigten, dass das wahrgenommene auditive Tempo durch die visuellen Bewegungsinformationen beeinflusst wird, und dass der Effekt in Verbindung mit der Zunahme oder Abnahme der visuellen Geschwindigkeit steht. Dies unterstützt die Hypothese, dass internale Bewegungsinformationen, welche aus visuomotorischen Reizen extrahiert werden, in die Wahrnehmung eines auditiven Rhythmus integriert werden können. Zusammen genommen, zeigt die vorgestellte Arbeit, dass unser motorisches System eine wichtige Rolle im Wahrnehmungsprozess von auditiven Rhythmen spielt. Dies kann sowohl durch äußere als auch durch internale motorische Aktivitäten geschehen, und beeinflusst nicht nur die Art, wie wir Rhythmen hören, sondern verbessert unter bestimmten Bedingungen auch unsere Fähigkeit Rhythmen zu identifizieren.

2. GENERAL INTRODUCTION

2.1 Auditory rhythm perception

Human beings possess the ability to perceive auditory stimuli that are temporally complex, such as music and speech. One aspect of the temporal information we obtain from such input is its rhythm. The present work was derived mostly from the literature in the musical rhythm domain, and will focus on aspects of rhythm perception in the non-verbal and/or musical context. In the following subsections, theoretical definitions of (musical) rhythm and behavioral findings of rhythm perception will be presented.

2.1.1 What is "rhythm"?

Musical rhythm unfolds along time and comprises a hierarchy of organized temporal structures, of which three basic levels can be identified: pulse, meter, and rhythm (Cooper & Meyer, 1960). *Pulse* refers to the most basic hierarchy, consisting of a series of stable and undifferentiated *psychological* events arising endogenously in response to musical rhythms (Cooper & Meyer, 1960; Large, 2008). It is defined as a psychological percept of isochrony in the listener rather than a physical property of the stimuli. Individual pulses in the music are often regularly accentuated to give alternating strong and weak beats, and the measure of the regular beats gives rise to the percept of *meter,* such as the typical 4/4 (four beats per measure) march meter or 3/4 (three beats per measure) waltz meter in the western music (Cooper & Meyer, 1960; Lerdahl & Jackendoff, 1983; Large, 2008). The metrical grouping can be very diverse across different cultures, but most of them are based on a regular pulse (Jackendoff & Lerdahl, 2006; Arom, 1989; Humble, 2002). *Rhythm*, on the other hand, is the temporal pattern which emerges as a result of grouping local events against the metrical grid or the pulse, also defined as "the way in which one or more unaccented beats are grouped in relation to an accentuated one" (Cooper & Meyer, 1960). A rhythm can be either metrical or non-metrical, depending on the regularity of the accents it presents or induces in the listener (Povel & Essens, 1985): metrical rhythms are those rhythmic patterns where regularly alternating strong and weak beats can be perceived, while non-metrical rhythms do not possess this quality.

Although these elements in musical rhythm are theoretically defined, they are essentially perceptual phenomena in experience.

2.1.2 How do we perceive rhythm?

Much research on rhythm perception is devoted to understanding the behavioral mechanisms by which we as listeners process such nested temporal relations. The perception of a rhythm, be it a rhythmic pattern or a sequence, is typically assessed by perceptual tasks – such as the sensitivity to detect changes in a comparison pattern from a standard one (Monahan & Hirsch, 1990; Drake & Botte, 1993; McAuley & Jones, 2003; Miller & McAuley, 2005; Grahn & Brett, 2009), production tasks – such as producing or reproducing a given rhythmic pattern (Povel & Essens, 1985; Essens & Povel, 1985; Drake, 1993; Repp, Windsor, & Desain, 2002; Grahn & Brett, 2007), or sensorimotor synchronization tasks – such as tapping on the (perceived or given) beat along a rhythmic sequence (Parncutt, 1994; Repp, 2003; Patel et al., 2005; Keller & Repp, 2005; Repp & Doggett, 2006; Repp, Iversen, & Patel, 2008).

As the auditory rhythm is built upon concatenated temporal structures, one essential feature that makes it ‘perceivable’ is its embedded *periodicity*, a regularly occurring event or pattern to which the listener can follow as the rhythm unfolds. This periodicity can be perceived at different levels. It could be the regular pulse, for example, which constitutes the basic principle of communication by which human can synchronize to the auditory stimuli, often by means of motor behavior such as tapping the foot or clapping the hand along (Parncutt, 1994; Merker, Madison, & Eckerdal, 2009). It could also be the metrical or beat accents that are either physically present or induced by the interval relation in the temporal pattern: specifically, metrical as compared to non-metrical rhythms lead to better temporal representation in the listener (Grahn & Brett, 2009), are thus better reproduced (Povel & Essens, 1985; Essens & Povel, 1985; Grahn & Brett, 2007), and are easier to synchronize to by tapping their beat along (Keller & Repp, 2005; Patel et al., 2005; Repp, Iversen, & Patel, 2008).

Rhythm perception is essentially a process in the temporal domain. In time research, there has been a long tradition of explaining time perception by an *interval-based* timing mechanism, which works on a central pacemaker and pulse accumulator model (Buhusi & Meck, 2005). However, because of the perceivable periodicity embedded in a rhythmic pattern or sequence, the temporal representation of a rhythm may differ from that of a simple interval: rather than as a mere sum of several successive intervals, it can be represented as an organized temporal entity, in which each interval is represented as in relation to the perceived pulse or beat structure. This is proposed to engage an *entrainment,* or *beat-based* timing mechanism, as opposed to the *interval-based* mechanism (Jones & Boltz, 1989; Large & Jones, 1999; McAuley & Jones, 2003). The entrainment mechanism is based on a model of a self-sustained internal oscillator that can be tuned to the external stimulus frequency; in the presence of several concurrent frequencies such as the different temporal hierarchies in the rhythm, it may be tuned selectively to a favorable one (called the referent period), and may also switch between different levels (Drake, Jones, & Baruch, 2000; Large & Palmer, 2002; Large & Snyder, 2009). Supporting evidence from neurophysiological studies will be presented in a later section (2.2.2).

One aspect of rhythm perception that is particularly relevant to the present work is the interaction between the *sensory* (for now, auditory) and the *motor* processes. As mentioned earlier in this section, the sensorimotor coupling has often been probed by synchronization tasks where the motor output, typically finger tapping, is to be synchronized to the sensory input (for a review, see Repp, 2005). However, more central to the understanding of rhythm perception is the idea that the perceptual process of rhythm may necessarily entail the motor component, as rhythm and movement seem to be intrinsically linked (Bolton, 1894; Todd, 1985; Merker, Madison, & Eckerdal, 2009). For example, when we hear music that is richly rhythmic or has a strong emphasis on the beat (Madison et al., in press), the tendency to move along with it by foot tapping, head nodding, or even dancing along, feels intuitive to many of us. When we move naturally to the rhythm, we tend to move in a periodic manner which often (but perhaps not always) corresponds to what we feel as the tempo of the music (Parncutt, 1994; Drake, Jones, & Baruch, 2000; Moelant, 2002; McKinny & Moelant, 2006; see however London, 2011 and Repp, 2011). When the movement along the rhythm is instructed to be finger tapping, as in the case of most experimental paradigms, a range of preferred tapping rate has been reported at the inter-tap interval of 400 ms to 700 ms, with a peak around 600 – 700 ms (Parncutt, 1994; London, 2002) or around 500 ms (van Noorden & Moelants, 1999; Moelant, 2002). Interestingly, this movement frequency corresponds to the preferred frequency of human locomotion, around 2 Hz (Macdougall & Moore, 2005).

Todd and colleagues proposed a *sensorimotor theory* of temporal tracking (Todd, 1999; Todd, Lee, & O'Boyle, 2002), where the experience of rhythm emerges from both the sensory representation of the temporal information in the stimuli, as well as the motor representation of the musculoskeletal system in the form of a motor image of the body.

Quoted from the author (Todd, 1999, p.119), "*... if the spatiotemporal form of certain stimuli are matched to the dynamics of the motor system, then they may evoke a motion of an internal representation or motor image of the corresponding synergetic elements of the musculoskeletal system, even if the musculoskeletal system itself does not move.*" This theory appears to well capture the intrinsic – and likely obligatory – link between sensory and motor processes in the rhythm domain. However, as far as rhythm perception is concerned, the majority of studies demonstrate only how different parameters of sensory information can affect the motor output differently. Little is known about the reverse: whether and how the motor system could influence the sensory process in rhythm perception. Two studies came closest to answering these questions: Phillips-Silver & Trainor (2005; 2007) showed in both infants and adults that our metrical interpretation of an ambiguous rhythm could be shaped by the different ways we bounce to it. This evidence serves as an encouraging precursor especially for the first part of the present work (section 3.1: Su & Pöppel, 2012), and my doctoral thesis as a whole intended to pursue the central idea of motor influences on auditory perception. Before moving on to the present work, however, a survey of the neuroscience literature in this domain will be presented in the following section, supplementing converging arguments for the thesis.

2.2 Neural correlates of rhythm perception: audio-motor coupling

The link between sensory and motor components, as mentioned in 2.1.2, is not merely a behavioral observation. The relevant neurophysiological evidence will be covered in the following subsections.

2.2.1 Auditory timing in the brain

As rhythm perception entails timing of sensory input, a general understanding of the cortico-subcortical neural pathways underpinning time perception seems relevant. In terms of interval timing, two functionally distinct circuits have been proposed: The "automatic timing" system, which is sometimes called 'discrete event timing', is in charge of intervals in the sub-second range, and is subserved by the cerebellum and the primary and secondary motor cortices. The "cognitively-controlled timing" system, on the other hand, is responsible for longer, supra-second interval range and may require attentional modulation, implicating basal ganglia, parietal, and prefrontal areas (Lewis & Miall, 2003; Buhusi & Meck, 2005). Though supported at times, like most reported dissociations in functional neural mechanisms, the distinction between these two systems are not always clear (Diedrichsen, Ivry, & Pressing, 2003). It may be expected that both systems are involved most of the time, albeit to different extents.

Of more importance to the present work is the recent finding that interval-based timing and entrainment (or beat-based) timing mechanism appear to implicate separate neural circuitries (Teki et al., 2011; Grube et al., 2010). The former is deployed when no particular beat or pulse structure is detected in a sequence, and it activates the olivocerebellar network. The latter is concerned with timing sequential intervals where a regular beat can be perceived, relying on the striato-thalamo-cortical system of basal ganglia, thalamus, premotor cortex (PMC), supplementary motor area (SMA) and dorsal lateral prefrontal cortex (DLPFC). In whichever way the distinctions are drawn to differentiate between timing systems, it is clear that the subcortical and cortical motor systems are engaged in the process.

2.2.2 Rhythm / beat perception in the brain

The findings discussed in 2.1.2 are largely supported by the abundant neuroscience studies in recent years. First of all, our brain does respond to the *periodicity* of the auditory rhythm, evidenced by the cortical oscillatory activities. Neuronal oscillations are the high and low excitability states of a local neuronal ensemble alternating at different frequency ranges, e.g. alpha band: 8–12 Hz, beta band: 15–30 Hz, gamma band: > 30 Hz (Buzsáki & Draguhn, 2004). The different frequency ranges are hierarchically nested amongst each other, and the phases of excitability underpin the optimal phases for stimuli processing in the primary sensory cortices (Lakatos et al., 2005; Schroeder et al., 2008). Critically, the oscillations can be tuned (i.e. entrained) to the rhythmicity of the external stimuli as a result of attentional selection, thus optimizing the perceptual analysis of a given sensory input (Lakatos et al., 2008; Schroeder et al., 2008). As such, when the participants listened to a rhythmic tone sequence, their brainwaves in the gamma and sometimes beta bands are found to be phase-locked to the isochronous tone onset (Snyder & Large, 2005; Zanto et al., 2005; Zanto, Snyder, & Large, 2006; Will & Berg, 2007). The gamma band entrained to the isochronous tones would occur even at occasional tone omission, reflecting an internal temporal expectation as a result of rhythmic entrainment (Snyder & Large, 2005). The beta band is found to be modulated by metrical accents that are either physically present or mentally imposed, showing that the brain responds to the felt periodicity at a metrical level (Iversen, Repp, & Patel, 2009).

Single neuron recordings and imaging studies have also provided abundant information supporting the audio-motor interaction. The sensorimotor coupling in the brain had first been discovered in the visuomotor domain, using the classical grasping and reaching paradigm, where the premotor and supplementary motor cortices are activated by watching a learned action in the absence of action execution (Rizzolatti et al. 1996; Murata et al. 1997; Fadiga et al. 2000). This action representation in the brain elicited by sensory input was later observed in the audio-motor domain as well, where listening to the sound associated with a learned action could activate the premotor areas (Kohler et al., 2002; Lahav, Saltzman, & Schlaug, 2007). These findings support the idea that sensory and motor information – or, in another word: perception and action – may be commonly coded in the brain (Prinz, 1997).

Now, suppose it was the case that auditory rhythm, as reviewed in 2.1.2, should be intrinsically linked to movement and should elicit a motor representation in the listener, one may expect to find corresponding evidence of audio-motor coupling in the brain activities. This has indeed been found in many studies: when tapping to the beat of the rhythm, the dorsal premotor cortex (dPMC) was modulated by the metrical structure of the rhythm even when the motor output remained the same (Chen, Zatorre, & Penhune, 2006). When tapping a rhythmic pattern itself, the activities in the motor areas such as pre-SMA, SMA, dPMC, and cerebellum, as well as the connectivity within this network, also increased with increasing metrical complexity of the rhythm, again under the same motor output (Chen, Penhune, & Zatorre, 2008a). However, most interestingly, the motor activities have been observed in purely perceptual tasks. When listening to rhythmic patterns *without* any motor task, the commonly activated areas include pre-SMA, SMA, dPMC, basal ganglia (putamen), cerebellum, and superior temporal gyrus (STG) (Grahn & Brett, 2007; Chen, Penhune, & Zatorre, 2008b; Bengtsson et al., 2009). The activities in SMA and basal ganglia are especially modulated by the beat structure of the rhythm: a higher beat saliency or a stronger metrical accent leads to higher activations in these areas (Grahn & Brett, 2007), the putamen being crucial for beat perception (Grahn & Brett, 2009). In addition, the connectivity between putamen and the aforementioned auditory and motor areas is also modulated by the beat saliency (Grahn & Rowe, 2009). These findings seem consistent with the idea of an automatic audio-motor coupling necessary for perceiving auditory rhythm. Put in another word: the motor system seems obligatorily engaged in processing auditory rhythm. Based on these findings, as well as the anatomical connectivity between PMC and other auditory and motor areas, PMC has been proposed as one plausible candidate to mediate the transformation between auditory and motor information (Zatorre, Chen, & Penhune, 2007; Chen, Penhune, & Zatorre, 2009). The proposed role of PMC as a platform for sensorimotor integration has also been taken up by the present work, especially the part constituting the study of Su & Jonikaitis (2011) (see section 3.2).

Lastly, a word should be mentioned concerning musical training and the audio-motor coupling. Playing a musical instrument entails quite sophisticated skills of sensorimotor transformation; it is thus reasonable to expect a greater degree of internal audio-motor binding in musicians compared to non-musicians, which has been supported by several studies (Bangert et al., 2006; Grahn & Brett, 2007; Grahn & Rowe, 2009; however, see also Chen, Penhune, & Zatorre, 2008a). This information is relevant to the first part of the present work (section 3.1: Su & Pöppel, 2012).

2.2.3 From motor to auditory processes?

Studies reviewed in 2.2.2 have augmented our understanding of rhythm perception and its neural mechanisms. Nonetheless, as is lacking in the behavioral findings (2.1.2), the causality and directionality of this audio-motor interaction at the cortical level is not yet clear. Again, most aforementioned studies were interested in how different temporal structures of the rhythm modulate the observed activities in the motor areas. However, as has been proposed in speech research (Hickok & Poeppel, 2004; 2007) and borrowed to parallel non-verbal rhythm perception (Zatorre, Chen, & Penhune, 2007; Chen, Penhune, & Zatorre, 2009), the neural projections connecting cortical auditory, motor, and sensorimotor integration areas appear to be bidirectional. There is evidence that the disruption of PMC can impair speech perception (Meister et al., 2007), suggesting a causal link from PMC to auditory perceptual analysis. The question thus remains: does the projection from the motor to the auditory area influence the perceptual process of non-verbal rhythm?

Several studies on cortical oscillations are supportive of this idea: as briefly reviewed in the beginning of 2.2.2, neuronal oscillations may underlie the mechanism of sensory perception especially in the case of rhythmic stimuli. The authors have argued that, for example, the oscillations in the primary auditory cortex can be reset by input from nonauditory areas so that the (rhythmic) auditory stimuli can be processed in the optimal phase (i.e. high excitability state) (Schroeder et al., 2008). They went further to propose that the flow of motor cortical oscillatory rhythms to the sensory cortices may subserve our perceptual selection, and that our perceptual experiences could be controlled by the motor sampling routines – a plausible mechanism of 'Active Sensing' in human (Schroeder et al., 2010). From this point of view, it seems reasonable to assume that cortical motor activities can entrain the auditory oscillations. However, as the authors pointed out, the behavioral evidence derived from speech research has not been consistent (Galantucci, Fowler, & Turvey, 2006; Meister et al., 2007). Further investigation of this hypothesis seems warranted, and, taken from here, the present work started out by testing a similar idea in the non-speech rhythm domain.

2.3 The present work

My doctoral research intended to answer a central question concerning the audio-motor interaction in rhythm perception: whether and how does the motor system influence the perception of an auditory rhythm? The thesis is divided into two parts, each addressing the effect of external (overt) and internal (covert) motor processes on rhythm perception, respectively.

2.3.1 Brief introduction of the thesis: Part I.

The first part, published as Su & Pöppel (2012), was designed to answer two questions: 1) Do *overt*, natural body movements – such as those we intuitively employ when listening to music: e.g. foot tapping, head nodding – have a positive effect on our ability to perceive the temporal structure of a rhythm? 2) If so, is the effect of body movement modulated by musical training? To this end, I developed a behavioral pulse-finding task where the participants were requested to listen to a sequence of tones occurring at pseudo-random intervals, and to search for their subjectively perceived pulse. The sequence was constructed with an underlying nominal pulse tempo that was not obvious on the rhythmic surface, and the pulse could thus be felt subjectively at different (sub)-harmonics of the nominal tempo. The perceived pulse was measured subsequently by finger tapping along the ongoing sequence. Twenty young, healthy, right-handed participants took part in this study (age range 20–35 years, mean age 24, SD 3.8). Ten were amateur musicians with a minimal training time of 8 years, and the other ten had never received any formal musical education.

The participants were divided into four groups based on musical training and the instructed task strategy: 1) musicians, movement, 2) non-musicians, movement, 3) musicians, no movement, and 4) non-musicians, no movement. In the movement groups, they were requested to search for the pulse by moving to the rhythm. In the no-movement groups, they were requested to sit still and searched for the pulse only through listening. I analyzed the data in terms of their response time, inter-tap intervals, produced pulse tempo, pulse stability, and synchronization stability. The results demonstrated an obvious contrast in the non-musicians between movement and no-movement groups, with movement yielding significantly better task performance. Musicians performed the task equally well with and without movement, though the produced pulse tempi differed between movement conditions. Overall this study argues for the facilitatory effect of overt motor component in auditory rhythm perception, and this effect is likely modulated by musical training.

Author contributions: Conceived and designed the experiment: Su & Pöppel. Implemented and performed the experiment: Su. Analyzed the data: Su. Result discussion: Su & Pöppel. Wrote the manuscript: Su.

2.3.2 Brief introduction of the thesis: Part II.

The second part of the thesis, published as Su & Jonikaitis (2011), is concerned with the idea of an indirect, *covert* motor influence on auditory rhythm perception. Specifically, I sought to find out whether the implicit motor information embedded in another independent sensory input – from the visual modality – could be integrated to the percept of an auditory tempo. The premise rested on the previous findings of sensorimotor representations in both visual and auditory domains (see section 2.2.2, paragraph 2). As such, the question in this study was: when presented concurrently, could the motor information transformed from the visual stimulation be incorporated to affect the way we perceive an auditory tempo? To this end, we designed an audiovisual paradigm where, in Experiment 1 and Experiment 2, the participants listened to a standard auditory sequence under different conditions of visual motion flow, and judged the comparison sequence tempo (presented without accompanying visual stimuli) relative to the standard. Critically, between the auditory and visual stimulations, there was no spatially or temporally corresponding event, thus ruling out any simple spatial or temporal capture of one stream by the other. In Experiment 1 (twelve participants, mean age 25, SD 3.1), the visual stimuli consisted of two conditions: 1) accelerating motion, implemented as a coherent flow of dots moving towards the observer in an accelerating manner, and 2) no motion, implemented as the same number of static dots, each changing its luminance independently of the others in an accelerating manner. The subjectively perceived auditory tempo was measured by the point of subjective equality (PSE) derived from the psychometric function under each visual motion condition. As expected, we found a faster perceived tempo when the sequence was accompanied by the accelerating motion. In Experiment 2 (twelve participants, mean age 25, SD 4.5), we employed the same auditory tempo judgement task under three visual motion conditions: 1) moving dots that became stepwise faster, 2) moving dots that became stepwise slower, and 3) moving dots without change of speed. A corresponding bias was found in the perceived auditory tempo: faster than actual, slower than actual, and no bias. In Experiment 3 (twelve participants, mean age 25, SD 3.5), we replicated the corresponding bias on auditory perception using a temporal reproduction task, validating the perceptual nature of this bias. Overall, the results support our hypothesis that the motor representation transformed from the visual motion can influence the perceptual process of an auditory rhythm – in this case, its tempo. Furthermore, this effect is related to the spatiotemporal properties of the visuomotor information. This study thus argues for the effect of covert motor component on the perception of an auditory rhythm.

Author contributions: Conceived and designed the experiment: Su & Jonikaitis. Implemented and performed the experiment: Su. Analyzed the data: Su. Wrote the manuscript: Su & Jonikaitis (10%).

In sum, the present thesis demonstrates that both external (overt) and internal (covert) motor activities can modulate the auditory perceptual process and influence the way we perceive an auditory rhythm. This furthers our understanding of the human sensorimotor interaction, which serves the basis not only for music cognition but also for other multimodal functions such as language acquisition. The information gained in this domain may also provide clinical implications for motor disorders such as Parkinson disease (e.g. Lim et al., 2005) or speech disorders such as developmental dyslexia (e.g. Thomson & Goswami, 2008; Corriveau & Goswami, 2009).

3. PUBLISHED SCIENTIFIC WORKS

3.1 Body movement enhances the extraction of temporal structures in auditory sequences.

Yi-Huang Su [a] & Ernst Pöppel [a,b]

Psychological Research, **2012, volume: 76, 373–382**

Accepted: 23 May, 2011

Published online: 22 Jun, 2011 (DOI: 10.1007/s00426-011-0346-3)

Published in print: 21 April, 2012

a. Institute of Medical Psychology, Ludwig-Maximilians-University Munich, Goethestr. 31, 80336 Munich, Germany

b. Human Science Center, Ludwig-Maximilians-University Munich

Psychological Research (2012) 76:373–382
DOI 10.1007/s00426-011-0346-3

ORIGINAL ARTICLE

Body movement enhances the extraction of temporal structures in auditory sequences

Yi-Huang Su · Ernst Pöppel

Received: 14 January 2011 / Accepted: 23 May 2011 / Published online: 22 June 2011

Abstract Auditory and motor systems interact in processing auditory rhythms. This study investigated the effect of intuitive body movement, such as head nodding or foot tapping, on listeners' ability to entrain to the pulse of an auditory sequence. A pulse-finding task was employed using an isochronous sequence of tones in which tones were omitted at pseudorandom positions. Musicians and non-musicians identified their subjectively fitting pulse either using periodic body movement or through listening only. The identified pulse was measured subsequently by finger tapping. Movement appeared to assist pulse extraction especially for non-musicians. The chosen pulse tempi tended to be faster with movement. Additionally, movement led to higher synchronization stabilities of the produced pulse along the sequence, regardless of musical training. These findings demonstrated the facilitatory role of body movement in entraining to auditory rhythms and its interaction with musical training.

Introduction

Imagine sitting in a live jazz concert: the listeners feel immersed in the music and start intuitively to move along with different parts of their body. The same is seen in the musicians: they automatically move their head or tap their feet in a periodic manner while playing some rhythmically engaging passages. Even in a classical concert where excessive gestures are not encouraged by convention, we often observe musicians rhythmically moving some parts of their body that are not engaged in playing the instruments, along with the music. Moving one's body periodically to the music, be it foot tapping or head nodding, is a frequent manifestation in listeners as well as in performing musicians. It presents a common example of audio–motor crosstalk in experiencing musical rhythms, and poses an interesting question about the nature of rhythm perception: do we move only because we react to the rhythm we hear, or does the movement itself contribute to the process of hearing the rhythm?

Perception and action are believed to share common representational mechanisms through which they interact (Prinz, 1997). Neurophysiological studies on rhythm perception have also concurred that processing auditory rhythms engages both auditory and motor areas of the brain (Chen, Penhune, & Zatorre, 2008b; Grahn & Brett, 2007; Bengtsson et al., 2009), and the motor system can be crucial in this act (Grahn & Brett, 2009). A direct behavioral link between 'hearing the rhythm' and 'moving the body' has been established in two studies where the interpretation of the same auditory rhythm was shaped by different patterns in which the listeners bounced their body (Phillips-Silver & Trainor, 2005, 2007). Given this audio–motor interplay, the present study further pursued the hypothesis that the use of body movement is not merely a reaction to

---

Electronic supplementary material The online version of this article (doi:10.1007/s00426-011-0346-3) contains supplementary material, which is available to authorized users.

Y.-H. Su (✉)
Institute of Medical Psychology,
Ludwig-Maximilians-University Munich,
Goethestr. 31, 80336 Munich, Germany
e-mail: jasmine.su@med.uni-muenchen.de

E. Pöppel
Human Science Center,
Ludwig-Maximilians-University Munich,
Munich, Germany

hearing rhythmic input, but could actively *assist* the processing of temporal structures in the auditory events.

The temporal structure of interest is the pulse.[1] In music, pulse is defined as a series of stable and undifferentiated *psychological events* arising endogenously in response to musical rhythms (Cooper & Meyer, 1960). Rather than a physical property of the stimuli, the pulse is a subjectively experienced isochrony. We chose to target this process because perceiving such isochrony is the basic principle of human entrainment to auditory stimuli (Merker, Madison, & Eckerdal, 2009). It serves as the subjective referent by which we experience complex temporal relations in musical rhythms (Large, 2008), and corresponds to the felt tempo. A relevant *sensorimotor theory* of temporal tracking has been proposed by Todd and colleagues (Todd, 1999; Todd, Lee, & O'Boyle, 2002). It incorporates importantly an internal motor representation of the body along with the sensory input and the motor output as coordinated mechanisms of tracking and synchronizing to an isochronous pulse. This theory emphasizes the relation between the embodied motor process and the percept of isochronous structure in the rhythm, which constitutes the central idea of the present study.

This idea is further supported by several complementary findings: human's ability to perceive a regular pulse seems innate (Winkler, Háden, Ladinig, Sziller, & Honing, 2009) and is proposed to arise from an endogenous neural oscillation entraining to rhythmic stimuli (Large & Snyder, 2009). Interestingly, such oscillations are also associated with motor tasks (Fujioka, Trainor, Large, & Ross, 2009; Salenius & Hari, 2003). Consistently, the premotor activation in the brain is enhanced by listening to rhythms at one's preferred tempo (Kornysheva, von Cramon, Jacobsen, & Schubotz, 2010), and human listeners' preferred tempo in music (Moelants, 2002) corresponds to the preferred frequency (~2 Hz) in locomotion (Macdougall & Moore, 2005). These findings seem to suggest that the percept of regular pulse, which also defines the tempo, could entail a motor component: forming a pulse by means of entrainment may require a motor process, at least internally (Grahn & Rowe, 2009).

In light of the audio–motor interaction, this study investigated whether an *external motor process* such as moving one's body to the rhythm—an intuitive behavior for many people—could actually assist the extraction of its pulse by means of facilitated entrainment. A pulse-finding task was employed where a tone sequence of no particular metrical or accent structure was presented continuously. The structure of the sequence could be seen as underlyingly isochronous (based on the nominal stimulus tempo) with tones omitted at pseudorandomly chosen positions, similar to one of the sequence types employed in the study of Patel, Iversen, Chen, and Repp (2005, sequence type 7: I-WM). The listeners first established their subjectively fitting pulse either using preferred periodic body movement or through listening only, and then produced their identified pulse by finger tapping. Critically, in establishing the pulse, the engagement of body movement was expected to initiate an overt motor activity *while* the listener searched for the regular pulse to which to entrain. Rather than a mere manifestation of the already established pulse, the movement should be adopted from the beginning of the entrainment process in order to assist finding a stable pulse. That is, the movement could be initially out of synchrony with any pulse period, but would gradually (or quickly, as we hypothesized the presence of movement to be facilitatory for the process) synchronize to the pulse of the sequence. Without movement, such entrainment process would have to be internally generated and might require more a cognitive strategy to analyze the temporal structure of the sequence.

As such, movement was expected to assist the listeners to more easily 'tune in' to the temporal information and to establish their pulse at one of several possible (sub)-harmonic frequencies (Large & Snyder, 2009). The stability of one's tapped pulse, preceded and accompanied by movement, was also expected to be higher as a result of enhanced sensorimotor integration (Chen, Penhune, & Zatorre, 2009). In addition, the movement effect was compared between musicians and non-musicians. Musicians were expected to be able to analyze the structure of the sequence (Chen, Penhune, & Zatorre, 2008a) and to generate the pulse internally (Grahn & Rowe, 2009) even in the absence of body movement, while non-musicians were expected to depend more on such overt motor activity to discover and entrain to the pulse.

Method

Participants

20 young, healthy, right-handed participants (range 20–35 years, mean age 24, SD 3.8) participated in the experiment via on-campus recruitment, and received payment in return. Ten were musically-trained (amateurs with at least 8 years training, 6 pianists and 4 violinists, 3 of whom were amateur orchestra members); the other ten had

---

[1] The terms pulse and beat are often used interchangeably in a musical context. However, beat implies a defined metrical organization based on the alternating strong and weak accentuation (Cooper & Meyer, 1960), which involves the perceptual grouping of pulse, e.g. groups of two or four as in a duple meter, or groups of three as in a waltz meter. Pulse itself, on the other hand, is not confined by metrical specifications; it exists as long as the isochrony is felt by the listener, and is generalizable in processing rhythms across different cultures and musical genres. Therefore, we prefer to use the term pulse here.

never received formal musical training. All reported music listening as leisure activities.

Stimuli and materials

Auditory stimuli were generated as wave files by the music software Logic 8 Express (Apple Inc. California) using a synthesized woodblock sound (as the instrument "clave"), with 42 ms tone duration. Each wave file was a 'building block' consisting of five isochronous time points; each point could be either occupied by a tone or not, resulting in 31 possible building blocks, excluding the block with no tone. The blocks were generated at six tempi: 60, 90, 120, 150, 180, and 210 beats per minute (BPM), corresponding to the shortest inter-onset interval (IOI) of 1000, 666.7, 500, 400, 333.3, and 285.7 ms. The experiment was carried out in Matlab® 2009a (Mathworks) using Psychophysics Toolbox extensions version 3 (Brainard, 1997), running on a MacBook Pro laptop computer. Participants sat in a comfortably-lit sound-proof room. The sound was delivered via headphones (Philips SBC HS900).

Design and procedure

The participants were divided into four groups based on the instructed task strategy and musical training: (1) movement, musicians, (2) movement, non-musicians, (3) no-movement, musicians; and (4) no-movement, non-musicians. There were five participants in each group.

Instruction

Pulse in the task was explained as (translated from German) 'the successive time points with equal intervals which are subjectively fitting tactus to the ongoing tone sequence. It should be as stable as possible throughout the trial and should not alternate between different levels.' Besides the verbal explanation, in order to ensure the same understanding of the task from both musicians and non-musicians, an instructional demonstration was carried out for each participant prior to the experiment. The demonstration differed between movement and no-movement groups as follows.

For the movement groups, the experimenter played an example tone sequence (as would be played in a real experimental trial) and demonstrated behaviorally with foot tapping where the pulse should be temporally. The experimenter demonstrated two different possibilities of the fitting pulse, one being the subharmonic of the other (i.e. twice as slow), which exemplified the notion of 'different pulse levels'.

For the no-movement groups, the experimenter played an example tone sequence which was accompanied by an additional sequence of low tones illustrating pulse. Two examples of such a combined sequence were played, the low tones in each example demonstrating a different (but fitting) pulse level. Crucially, there was no mention of the link between the present task and the everyday behavior such as 'tapping one's feet to the music'.

Experiment

In the beginning of each trial, 31 building blocks of one tempo were strung up in a randomized order, with the rule that all blocks were selected once, and the very first time point was occupied by a tone. The 31 concatenated blocks made up a long sequence that was looped within a trial. Each trial consisted of two consecutive phases: (a) pulse extraction, and (b) pulse production (Fig. 1).

Prior to the instruction, each participant in the *movement* groups had been asked to report their preferred means of body movement when they listened to music. In the extraction phase, they were requested to use their reported preferred movement (e.g. foot tapping, head nodding) from the start of the sequence to assist finding the pulse. As they started moving, their movement were usually not immediately in synchrony with any pulse of the sequence, but should be tuned to a subjectively fitting pulse level before they proceeded to the production phase. For the *no-movement* groups during the extraction phase, the participants were requested to try finding the pulse only by listening, strictly

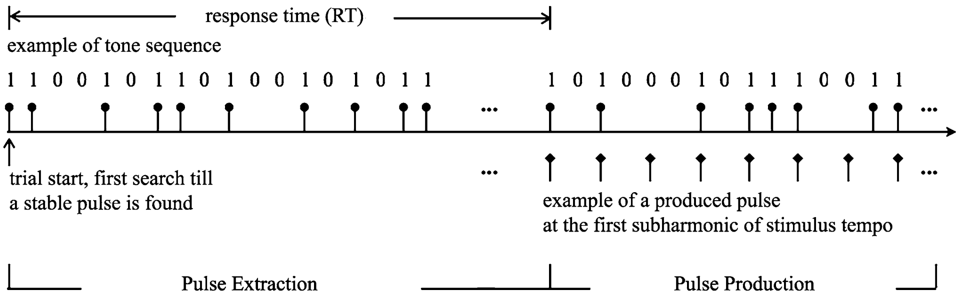

Fig. 1 Illustration of the trial procedure. The *upper panel* depicts an example of the stimulus sequence. The number 0s and 1s denote the theoretical positions of the isochronous pulse according to the nominal stimulus tempo, where (in the sequence) 1 is occupied by a tone, and 0 is not. The *lower panel* depicts an example of a pulse identified at the 1:2 subharmonic of the stimulus tempo

without any movement, until they felt a fitting stable pulse was found. The phase of pulse extraction was not speeded. When the participants felt sure of their identified pulse, they (in all groups) were requested to start the production phase by tapping their pulse on the computer key "B" along with the sequence in a synchronized manner. Participants in the no-movement groups were instructed to restrict the movement during the production phase to only the index finger, while those in the movement groups were not particularly requested to stop movement during tapping. This was meant to maximize the contrast between movement and no-movement groups throughout the task. 16 consecutive taps were recorded per trial (inter-tap intervals representing the identified inter-pulse intervals) before the next trial commenced. The time needed for pulse extraction (henceforth referred to as response time, RT)—the time between the start of stimulus presentation and the first pulse tap—was also recorded in each trial to index the subjective task difficulty.

The stimuli were presented in 6 tempi and 30 trials each, randomly assigned to 4 blocks. The whole experiment lasted 2.5–3 h depending on the individual speed, with breaks after each block. Before starting the experimental session, each participant underwent at least five practice trials and more if they did not show enough understanding of the task. One basic sign of the participant's understanding of the task was that, during practice, he or she did not produce taps that were simply time-locked (i.e. as a response) to the tones, but instead taps that exhibited certain degree of periodicity.

Data analyses and results

Percentages of stable and unstable pulse

For each trial, the mean inter-tap interval (ITI, in milliseconds) and the coefficient of variation (CV = within-trial standard deviation divided by mean ITI × 100%) were calculated, excluding always the first four taps. To index the task performance, each trial was first categorized as being *stable* or *unstable* by the following criteria:

Stable trials A criterion of CV $\leq$10% was first applied to identify trials with stable pulse series.[2] In order to reliably identify trials in which a pulse had really been found, as opposed to trials with stable taps around a mean ITI that was irrelevant to the correct pulse period, we applied an additional criterion on the mean ITI of every stable trial regarding its identified pulse period, within which the pulse was considered to be successfully found: $(N \times \mathrm{IOI}) \pm (N \times \mathrm{IOI}) \times 10\%$, $N = 0.25, 0.5, 1, 2, 3, 4$, etc. N represented the chosen pulse level in each trial (i.e. the mean ITI being around N times of the shortest stimulus IOI). This criterion filtered out the stable trials with a mean ITI that exceeded 10% deviation from the correct inter-pulse interval. As such, the stable trials were further divided into two sub-types: (1) stable pulse, and (2) stable, but not considered pulse.

Unstable trials Trials produced with CV >10% were labeled as unstable trials. Each unstable trial was further categorized as reflecting one of the three behaviors which most often cause a large within-trial CV[3]: (1) Type 1—constantly irregular and unstable ITIs, (2) Type 2—pulse switching between different (sub)-harmonic levels, and (3) Type 3—rarely occurring missing taps or a pause within an otherwise stable tap series.

The occurrence of unstable Type 2 was generally very low (average frequency <0.1%), so we excluded it from further analyses. Of the four analyzed pulse types—stable pulse, stable no pulse, unstable Type 1, and unstable Type 3—only the first one (stable pulse) represented the successful trials. The percentages of these four types were submitted to a mixed-model ANOVA with one within-subject factor: produced pulse type (4 levels), and two between-subject factors: movement (2 levels) and musical training (2 levels). It revealed a significant pulse type × movement × musical training interaction, $F(3, 48) = 4.85, p < 0.01, \eta_p^2 = 0.23$ (Fig. 2). Follow-up partial ANOVAs revealed that the three-way interaction resulted from a significant interaction between musical training and pulse types in the *no-movement* groups, $F(3, 24) = 6.73$, $p < 0.01$, $\eta_p^2 = 0.46$, but not in the movement groups, $F(3, 24) = 1.91$, $p > 0.15$. Post-hoc comparison (two-sample t test) revealed that for the *no-movement* groups, the percentage of stable pulse was different between musicians and non-musicians (78 vs. 29%), $p < 0.05$, $t(8) = 2.95$, and the percentage of Type 1 unstable pulse also differed between these two groups (11 vs. 60%), $p < 0.05$, $t(8) = 2.54$. For the *movement* groups, a main effect of pulse type was significant $F(2, 16) = 930$, $p < 0.001$, $\eta_p^2 = 0.99$, and the post-hoc comparisons (Tukey HSD) showed that the percentage of stable pulse was significantly higher than any of the three unsuccessful types (all ps < 0.001), while the percentages amongst these three types did not differ (all

---

[2] A criterion of ITI stability for constituting a 'pulse' has not been established in the literature, as it would depend on the task condition and the given stimuli. A study on the perceptual threshold of pulse attribution (Madison & Merker, 2002) found an average 8.6% deviation of the inter-tone intervals in the sequence, beyond which the participants were unable to identify the pulse. Considering the higher difficulty in the present task as the tones in a sequence did not appear regularly, and that the pulse was measured by production, a criterion of 10% was used. This, together with the criterion on inter-pulse interval, appeared to reflect the interaction between musical training and movement well (Fig. 2).

[3] For detailed classification criteria, see Figure S1 and the described procedure in the supplementary material.

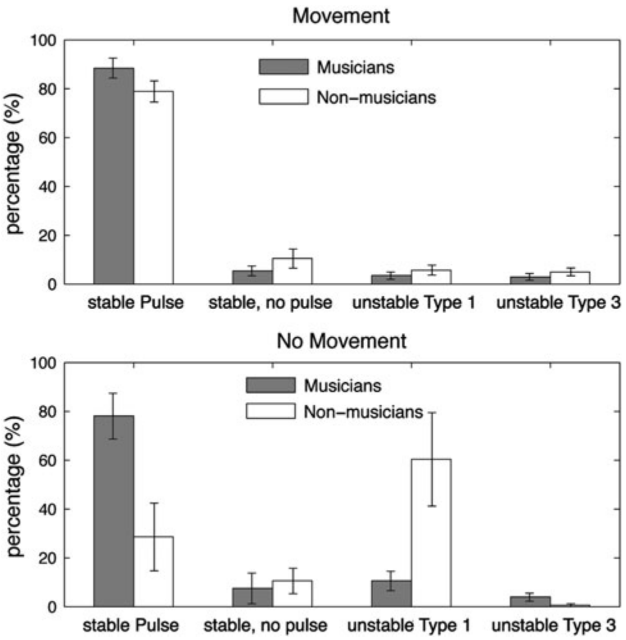

Fig. 2 Mean percentages of the four produced pulse types—stable pulse, stable but no pulse, unstable Type 1, and unstable Type3—from each of the four participant groups. *Error bars* represent standard errors of the mean

$ps > 0.5$). In short, non-musicians without movement produced a significantly higher percentage of unstable pulse than musicians without movement, and their unstable pulse mostly resulted from high variabilities of within-trial ITIs (Type 1). Between moving musicians and moving non-musicians, however, the distribution of produced pulse types did not differ, and they produced mostly stable pulse (88 and 79% in musicians and non-musicians.).

Similarly, the partial ANOVA between the two *musician* groups yielded no significant interaction between movement and pulse type, $F(3, 24) = 0.95$, $p > 0.4$, suggesting similar type distributions from musicians with and without movement. The partial ANOVA between the two *non-musician* groups, however, yielded a significant interaction between movement and pulse type, $F(3, 24) = 8.78$, $p < 0.001$, $\eta_p^2 = 0.52$. Post-hoc comparison (two-sample t test) revealed that non-musicians with movement produced a higher percentage of stable pulse than non-musicians without movement, $p < 0.01$, $t(8) = 3.44$ (79 vs. 29%, movement vs. no-movement).

Identified pulse tempo

The mean ITI from every stable pulse trial was transformed into the corresponding tempo (BPM) and then scaled as the *ratio* to the nominal stimulus tempo. Each resultant ratio was then logarithmically transformed before being plotted against the stimulus tempo. In this way the (sub)-harmonic relationship between the subjectively tuned-in pulse tempo (especially at slower subharmonics such as 1:2, 1:3, and 1:4) and the given stimulus tempo can be more clearly shown. Results from each participant group were plotted together, each cross representing a single trial (Fig. 3). It shows the tendency from each participant group to select certain pulse levels under each stimulus tempo. For the exact frequency of each cluster, see Figure S2 in the supporting information for detailed histogram distributions.

As seen in the scatterplot, in establishing pulse, the movement groups showed more focused tuning to the stimulus tempo and its 1:2 subharmonic (0.5 ratio). The no-movement groups tended more to scatter and shifted toward the 1:4 subharmonic (0.25 ratio) as the tempo increased. Musicians produced better-tuned tempi especially with movement. Non-musicians using movement could tune to similar pulse tempi as musicians with movement.

Time needed for pulse extraction

RTs were submitted to a mixed-model ANOVA with one within-subject factor, tempo (6 levels), and two between-subject factors, movement (2 levels) and musical training (2 levels). A main effect was found only for tempo, $F(5, 80) = 25.25$, $p < 0.001$, $\eta_p^2 = 0.61$, with longer RTs at slower tempi (Fig. 4a). Post-hoc comparisons (Tukey HSD) found significant differences between 60 BPM and all the other tempi (all $ps < 0.001$), and between 90 BPM and all the other tempi (all $ps < 0.05$). RTs appeared to decrease with increasing tempo until 120 BPM, above which they were not significantly differentiated by tempo. Interaction of movement × musical training was close to significant, $F(1, 16) = 3.55$, $p = 0.07$, $\eta_p^2 = 0.18$. As Fig. 4a shows, while both movement groups behaved similarly, non-musicians without movement seemed to need longer time than musicians without movement.

As an alternative, the RT data were also plotted not as the measured time but as the number of underlying pulse cycles (=RT/stimulus inter-pulse interval). The ANOVA naturally yielded the same between-group results as for RT, but the number of needed pulse cycles increased with the tempo (see Figure S3 in supplementary material).

Degree of synchronization

To measure the pulse stability by degrees of synchrony between the produced pulse and the sequence, the asynchrony was calculated between each tap and its theoretically correct position (based on the chosen pulse tempo). The variability was indexed as the within-trial standard deviation (SD) of the asynchronies—higher SD indicating lower stability—and submitted to a mixed-model ANOVA with one within-subject factor, tempo, and two between-subject factors, movement and musical training. Main

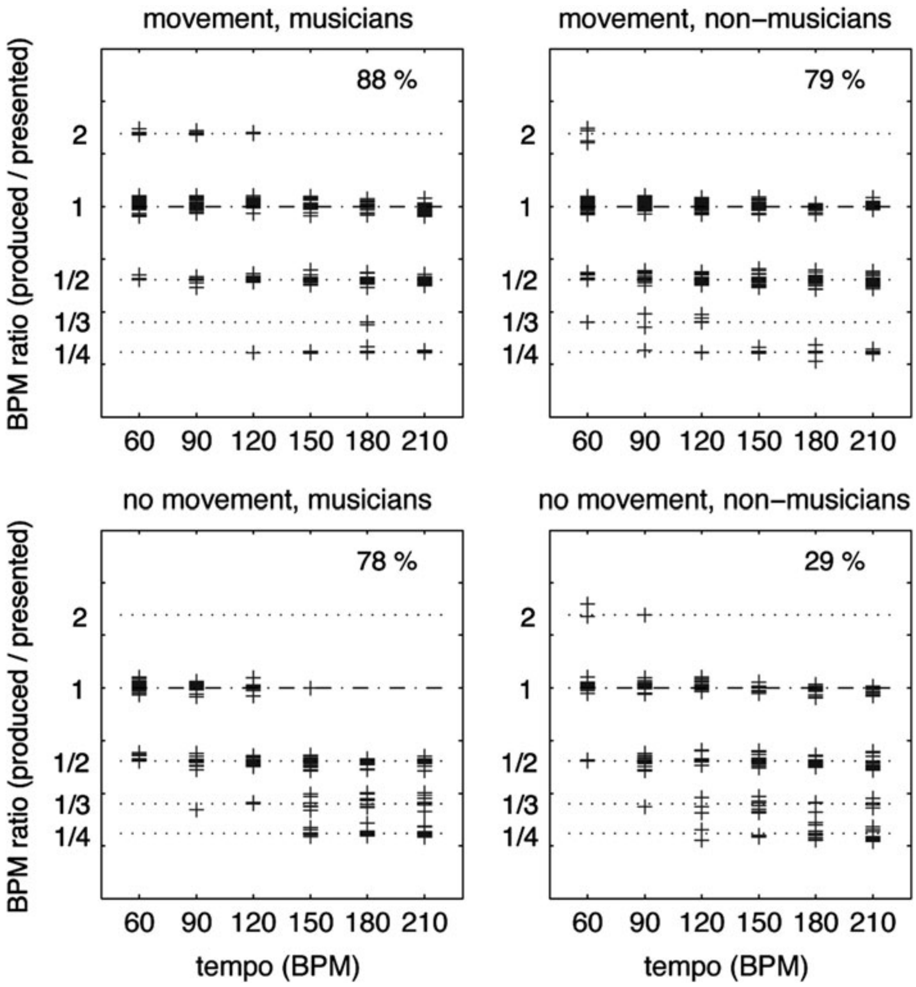

Fig. 3 Scatterplot of the produced pulse tempi as the ratio to the stimulus tempi, for each participant group separately. Each ratio was plotted as its logarithmic transformation for better viewing of pulse at slower subharmonics. *X* axis depicts each stimulus tempo condition. *Y* axis depicts the (sub)-harmonics of the stimulus tempo (1 = stimulus tempo, 2 = twice the stimulus tempo, 1/2, 1/3, and 1/4 = 0.5, 0.33, and 0.25 of the stimulus tempo). Only the tempi from *stable pulse* trials are plotted in this chart. Each *cross* represents a single trial. The *number in each chart* denotes the total percentage of *stable pulse* from this participant group

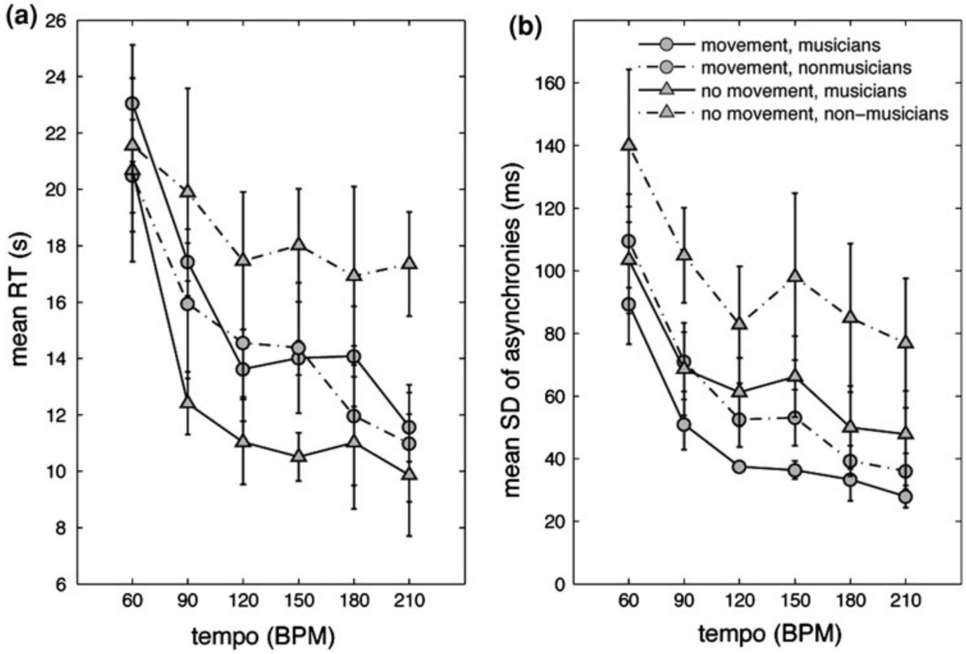

Fig. 4 **a** Mean RT as a function of the stimulus tempo, for each participant group. *Error bars* represent standard errors of the mean. **b** Mean standard deviation of asynchronies as a function of the stimulus tempo, for each participant group. *Error bars* represent standard errors of the mean

effects were found for movement $F(1, 16) = 10.11$, $p < 0.01$, $\eta_p^2 = 0.39$ (mean SD 53 vs. 97 ms, movement vs. no-movement), musical training $F(1, 16) = 7.51$, $p < 0.05$, $\eta_p^2 = 0.32$ (56 vs. 97 ms, musicians vs. non-musicians), and tempo $F(5, 80) = 44.42$, $p < 0.001$, $\eta_p^2 = 0.74$, without interactions (Fig. 4b). Post-hoc comparisons (Tukey HSD) found significant differences between the following tempi: 60 BPM versus all the other tempi; 90 versus 120, 180 and 210 BPM; 120 versus 210 BPM; 150 versus 180 and 210 BPM. The results showed that movement led to higher stability of synchronization in both musicians and non-musicians.

Discussion

Effect of movement on pulse extraction and entrainment

Our results highlight that moving one's body to an auditory sequence could indeed facilitate the extraction of the temporal structure such as the subjective pulse in a sequence.

The extent of this facilitation depended on musical training. Musicians are rhythmically trained and typically perform better in sensorimotor tasks (Chen et al., 2008a; Repp & Doggett, 2006; Franek, Mates, Radil, Beck, & Pöppel, 1994) and cross-modal timing tasks (Wöllner & Cañal-Bruland, 2010; Pecenka & Keller, 2009). It was not surprising that their training enabled them to analyze the temporal structure and establish stable pulse overall, even in the absence of movement. This proved to be much more challenging for non-musicians. With the assistance of body movement, however, non-musicians could find their pulse to a similar extent as the musicians.

What role does body movement play in this case, and what could account for its benefit? The use of body movement has been postulated as an intrinsic part of human entrainment to isochronous stimuli (Madison & Merker, 2002; Bolton 1894). Here, however, we tested the role of body movement in entraining to stimuli where the isochrony was implied but not explicitly or regularly given, and the pulse was thus more difficult to discover. Moreover, the movement we investigated was *not* a mere manifestation of the already extracted pulse such that the participants would first find the pulse internally and then start to move according to it. Instead, they started moving as soon as the sequence began, using the overt motor activities to facilitate the tuning to the pulse periodicity. In doing so, the exhibited movement for each sequence (as observed during the experimental session[4]) mostly did not start as being immediately in synchrony with the pulse, but rather went through a bit of adjustment before tuning to one of the fitting pulse levels. An interactive dynamic might be taking place during this process: the self-initiated movement frequency, which is not tuned-in at first, could be attracted to one of the underlying periodicities of the sequence (Repp, 2006), and in doing so leads the listener to start 'hearing' the pulse at that level, forming a positive audio–motor feedback loop. In the absence of overt movement, by contrast, this tuning process must then rely on the internal motor entrainment and/or the ability to analyze the sequence. Our results show that, unlike musicians, non-musician seemed to be lacking an effective internal motor simulation that entrained to the pulse when it was not regularly present at the rhythmic surface; nor did they possess additional musical knowledge as a compensatory strategy. They thus appeared to benefit much from the external motor process in order to entrain to the structure of the rhythm. This parallels the finding of Grahn and Rowe (2009) where, compared to non-musicians, musicians more often perceived the beat when it was less explicitly presented, and this was accompanied by higher connectivities between auditory and motor cortical areas, suggesting a higher level of internal audio–motor coupling.

Notably, our task required the search for a subjective temporal referent while no particular metrical accent was given, contrary to most people's experience of music listening. Meter has been defined as 'the measurement of the number of pulses between more or less regularly occurring accents' (Cooper & Meyer, 1960). While there are many cultural differences in meter, music from most cultures is pulse-based (Large, 2008; Arom, 1989; Humble, 2002). By not giving any metrical cues, we aimed to link body movement to a temporal process that was not strongly constrained by the previously-shaped listening experiences (e.g. Iversen, Patel, & Ohgushi, 2008). Namely, one did not necessarily need to recognize a particular meter (such as 2/4 or 3/4) before identifying the pulse. Although humans exhibit a preference for culturally familiar meters (Trehub & Hannon, 2009; Soley & Hannon, 2010) and might find it more difficult to follow some 'exotic' meters, our study demonstrated an approach that relied solely on the search for a pulse, regardless of metrical preferences. This search was found to be facilitated by the accompanying body movement—a potentially useful 'hearing by moving' strategy.

Effect of movement on pulse tempo

The presence and absence of movement as a pulse-search strategy seemed to lead to different preferred pulse levels. Movement was expected to predispose the chosen pulse tempi to a range of comfortable movement frequencies

---

[4] Every participant was observed for around 20–30 min during the first experimental block, and also for a shorter while in the beginning of each successive block.

(London, 2002; Macdougall & Moore, 2005), which appeared to be the case: with movement, the pulse was more often tuned to the nominal stimulus tempo, or to its 1:2 subharmonic when the stimulus tempo increased. The 1:4 subharmonic was rarely chosen, as it would have been too slow for continuous periodic movement. Without movement, they tended more to correspond to the slower subharmonics and especially more often to the 1:4 subharmonic as the stimulus tempo increased. We speculate that, in the absence of movement, the participants resorted more to a cognitive strategy to analyze the temporal structure of the sequence, especially in the case of musicians. Quite likely they would group the pulse automatically by imposing mental accents (Bolton, 1894; Repp, Iversen, & Patel, 2008), thus rendering the sequence to be heard as metrical in different ways. This would allow them to flexibly tune to different referent levels (Drake, Jones, & Baruch, 2000), though in the end they tended to opt for the slower subharmonics, as observed, because their internal pulse at a higher metrical level could be kept more stable against the irregular tones (Patel et al., 2005).

Therefore, complementary to the finding that different patterns of body bouncing can bias the metrical interpretation of a rhythm (Phillips-Silver & Trainor, 2005, 2007), the results here further demonstrate the differentiating role of the presence/absence of body movement in perceiving different pulse levels in an auditory sequence.

Time for pulse extraction

Though the time needed for pulse extraction—as measured in the experiment—may have been a function of both task difficulty and subjective readiness, it nevertheless revealed the between-group differences. That non-musicians without movement needed overall longer time than the other groups indicated the felt task difficulty, which paralleled the outcome of their pulse production. With movement, non-musicians needed similar amount of time as musicians with movement, arguing for the facilitatory effect of movement in the absence of compensating musical skills. In addition, RT decreased as the tempo increased up to about 120 BPM. The observation that RT did not decrease systematically above 120 BPM seems to reflect the relation between stimulus tempo and human's maximal pulse saliency around 80–100 BPM (London, 2002), outside of which it could be more difficult to feel the pulse.

Pulse synchronization

The degree of synchrony between the produced pulse and the sequence was also influenced by movement and musical training. Both musicians and non-musicians in the movement groups exhibited higher stability of synchronization than those in the no-movement groups. Since movement was present in the extraction phase and as observed also often in the production phase, it would be difficult to distinguish whether the pulse stability (as measured by the variability of asynchronies) benefited from the movement in either phase alone. Though the overall facilitatory effect on pulse entrainment should have derived from the movement prior to tapping, it seems reasonable to assume that concurrent body movement during tapping might play a positive role in tapping stability. It has been found that simultaneous bimanual tapping reduces the within-hand variabilities compared to tapping with only one hand, and this advantage is accounted for by the decreased variability in the central timing process (Wing & Kristofferson, 1973; Helmuth & Ivry, 1996) or the increased sensory reafference (Drewing & Aschersleben, 2003; Prinz, 1997). In this view, our result of reduced (single-handed) tapping variability could also be attributed to the concomitant larger-scale body movement. Further investigation is needed to elucidate whether different kinds of body movement leads to the same stabilization in finger tapping, and whether the improvement can be accommodated in the same theoretical framework. Our findings provide empirical support for the idea that musicians can indeed benefit from such natural body movement while playing the instrument, in keeping up a stable tempo.

Presence versus absence of movement

In interpreting our results as demonstrating the effect of body movement and musical training on pulse finding, two questions may arise: (1) Was the poor performance of non-musicians in the no-movement group attributable to the lack of understanding of the task, because no explanation linked to movement was given? (2) Did musicians in the no-movement condition perform better because of some micro-movements they used secretly though they were not supposed to?

The first question can be dismissed because of our instruction with the auditory demonstration, showing what the pulse was and where it should be temporally in relation to the tone sequence. The participants did not have to possess specific musical knowledge to understand the temporal nature of the task. They went through practice trials, and received feedback and explanations during practice until they showed sufficient understanding of the task. This ensured that the outcome of their performance was not due to lower understanding, but rather due to the task difficulty under the appointed experimental condition.

Regarding the second question, it is possible that musicians could potentially carry out some micro-movement, perhaps without being aware of it. If they had indeed moved secretly and constantly though they were not supposed to, the pattern of their results should have been very similar to that produced by musicians in the movement group. This was, however, not supported by our data: (1) without movement, musicians produced a rather different range of the chosen pulse tempi from musicians with movement (Fig. 3), which tended to be slower than would be naturally carried out by continuous movement; and (2) the stability of their produced pulse was also lower than that of the musicians with movement (Fig. 4b), signifying the absence of concurrent body movement to help stabilize the taps. Therefore, granted a higher tendency in musicians to carry out micro-movement in no-movement condition, our observed results suggested that this possibility either did not occur, or even if it had, its effect was both qualitatively and quantitatively different from that of the natural overt movement, and more similar to that of the no-movement condition. Two points may distinguish such potential micro-movement from the overt body movement in terms of its effect on pulse entrainment: it may have occurred not in a continuously periodic manner, or may have involved less motor activation in the brain, or both. If we were to explain the observed data from musicians without movement as a result of using secret micro-movement, it would still suggest that such micro-movement must function differently and less effectively as overt movement. As opposed to the external motor entrainment initiated by overt body movement, micro-movement might be a natural manifestation of the *internal* motor engagement. This explanation would not contradict our interpretation of facilitation by overt body movement and its interaction with musical training, but would rather point out the unique advantage of overt natural movements compared to less intuitive and much smaller-scale ones during the entrainment process.

What kind of movement?

The aim of our study was to investigate the effect of the presence and absence of movement on entraining to the pulse, and the movement of interest is the kind that a listener would naturally employ in a real-life scenario such as when listening to the music. The kinds of movement reported and performed subsequently by the participants in the task included most often head nodding (often involving the neck and the upper back) and foot tapping (often accompanied by slight head movement). In one case it was elected to be arm swiveling with foot tapping.[5] Indeed finger tapping was also a movement, though a smaller one and not so commonly observed as a natural listening habit. In order to register the pulse, it had to be performed by all participants including those in the no-movement groups. However, since none of the participants reported nor chose to use finger tapping itself during pulse extraction phase, the result was not confounded with a ‘practice effect’ of finger tapping. We can, therefore, attribute the observed effects to the opted larger body movements during the pulse discovery/entrainment process, which possibly also stabilized pulse tapping. Future studies might attempt to reveal whether different scales of movement, e.g., larger body movement versus smaller one such as finger tapping, would lead to different effects of motor simulation for entraining to the rhythm.

Overall our study demonstrated that overt body movement assisted the extraction of the underlying pulse in a non-isochronous sequence. It also led to better tuning to the sequence tempo and better synchronization to the sequence. The results provide empirical evidence of body movement as being a useful strategy especially for untrained listeners to approach auditory rhythms, and when musicians intuitively move their head or tap their feet while playing an instrument, it could help them keep up a stable tempo.

Acknowledgments This work was supported by the doctoral scholarship to the first author from the Bayerische Forschungsstiftung, and the experimental expenses were supported by the Andrea von Braun Foundation. The authors thank Marc Wittmann, Dragan Rangelov, and Björn Merker for earlier discussions of the work, as well as Bruno Repp and an anonymous reviewer for very useful inputs on the manuscript.

Conflict of interest The authors declare no conflict of interest.

References

Arom, S. (1989). Time structure in the music of central Africa: periodicity, meter, rhythm and polyrhythmics. *Leonardo, 22*, 91–99.

Bengtsson, S. L., Ullen, F., Ehrsson, H. H., Hashimoto, T., Kito, T., Naito, E., et al. (2009). Listening to rhythms activates motor and premotor cortices. *Cortex, 45*, 62–71.

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

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436.

Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008a). Moving on time: brain network for auditory–motor synchronization is modulated by rhythm complexity and musical training. *Journal of Cognitive Neuroscience, 20*, 226–239.

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

---

[5] Often two kinds of movement were adopted together by the same participant. The frequency of each employed movement was as follows: head nodding (6), foot tapping (9), and arm swiveling (1).

Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2009). The role of auditory and premotor cortex in sensorimotor transformations. *Annals of the New York Academy of Sciences, 1169*, 15–34.

Cooper, G., & Meyer, A. B. (1960). *The rhythmic structure of music*. London: The University of Chicago Press.

Drake, C., Jones, M. R., & Baruch, C. (2000). The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. *Cognition, 77*, 251–288.

Drewing, K., & Aschersleben, G. (2003). Reduced timing variability during bimanual coupling: a role for sensory information. *The Quarterly Journal of Experimental Psychology Section A, 56*, 329–350.

Franek, M., Mates, J., Radil, T., Beck, K., & Pöppel, E. (1994). Sensorimotor synchronization: motor responses to pseudoregular auditory patterns. *Perception & Psychophysics, 55*, 204–217.

Fujioka, T., Trainor, L. J., Large, E. W., & Ross, B. (2009). Beta and gamma rhythms in human auditory cortex during musical beat processing. *Annals of New York Academy of Science, 1169*, 89–92.

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

Grahn, J. A., & Brett, M. (2009). Impairment of beat-based rhythm discrimination in Parkinson's disease. *Cortex, 45*, 54–61.

Grahn, J. A., & Rowe, J. B. (2009). Feeling the beat: premotor and striatal interactions in musicians and nonmusicians during beat perception. *Journal of Neuroscience, 29*, 7540–7548.

Helmuth, L. L., & Ivry, R. B. (1996). When two hands are better than one: reduced timing variability during bimanual movements. *Journal of Experimental Psychology Human Perception and Performance, 22*, 278–293.

Humble, M. (2002). The development of rhythmic organization in Indian classical music. Unpublished dissertation, University of London.

Iversen, J. R., Patel, A. D., & Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. *Journal of Acoustic Society America, 124*(4), 2263–2271.

Kornysheva, K., von Cramon, D. Y., Jacobsen, T., & Schubotz, R. I. (2010). Tuning-into the beat: aesthetic appreciation of musical rhythms correlates with a premotor activity boost. *Human Brain Mapping, 31*, 48–64.

Large, E. W. (2008). Resonating to musical rhythm: theory and experiment. In S. Grondin (Ed.), *The psychology of time* (pp. 189–231). United Kingdom: Emerald.

Large, E. W., & Snyder, J. S. (2009). Pulse and meter as neural resonance. *Annals of New York Academy of Science, 1169*, 46–57.

London, J. (2002). Cognitive constraints on metric systems: some observations and hypotheses. *Music Perception, 19*, 529–550.

Macdougall, H. G., & Moore, S. T. (2005). Marching to the beat of the same drummer: the spontaneous tempo of human locomotion. *Journal of Applied Physiology, 99*, 1164–1173.

Madison, G., & Merker, B. (2002). On the limits of anisochrony in pulse attribution. *Psychological Research, 66*, 201–207.

Merker, B., Madison, G., & Eckerdal, P. (2009). On the role and origin of isochrony in human rhythmic entrainment. *Cortex, 45*, 4–17.

Moelants, D. (2002). Preferred tempo reconsidered. In C. Stevens, D. Burnham, G. McPherson, E. Schubert, & J. Renwick (Ed.), Proceedings of the 7th International Conference on Music Perception and Cognition (pp. 580–583). Sydney.

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

Pecenka, N., & Keller, P. (2009). Auditory pitch imagery and its relationship to musical synchronization. *Annals of New York Academy of Science, 1169*, 282–286.

Phillips-Silver, J., & Trainor, L. J. (2005). Feeling the beat: movement influences infant rhythm perception. *Science, 308*, 1430.

Phillips-Silver, J., & Trainor, L. J. (2007). Hearing what the body feels: auditory encoding of rhythmic movement. *Cognition, 105*, 533–546.

Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology, 9*, 129–154.

Repp, B. (2006). Does an auditory distractor sequence affect self-paced tapping. *Acta Psychologica*, *121*(1), 81–107. doi: 10.1016/j.actpsy.2005.06.006.

Repp, B. H., & Doggett, R. (2006). Tapping to a very slow beat: a comparison of musicians and nonmusicians. *Music Perception, 24*, 367–376.

Repp, B. H., Iversen, J. R., & Patel, A. D. (2008). Tracking an imposed beat within a metrical grid. *Music Perception, 26*, 1–18.

Salenius, S., & Hari, R. (2003). Synchronous cortical oscillatory activity during motor action. *Current Opinion in Neurobiology, 13*, 678–684.

Soley, G., & Hannon, E. E. (2010). Infants prefer the musical meter of their own culture: a cross-cultural comparison. *Developmental Psychology, 46*, 286–292.

Todd, N. P. M. (1999). Motion in music: a neurobiological perspective. *Music Perception, 17*, 115–126.

Todd, N. P. M., Lee, C. S., & O'Boyle, D. J. (2002). A sensorimotor theory of temporal tracking and beat induction. *Psychological Research, 66*, 26–39.

Trehub, S. E., & Hannon, E. E. (2009). Conventional rhythms enhance infants' and adults' perception of musical patterns. *Cortex, 45*, 110–118.

Wing, A. M., & Kristofferson, A. B. (1973). Response delays and the timing of discrete motor responses. *Perception & Psychophysics, 14*, 5–12.

Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., & Honing, H. (2009). Newborn infants detect the beat in music. *Proceedings of the National Academy of Science, 106*, 2468–2471.

Wöllner, C., & Cañal-Bruland, R. (2010). Keeping an eye on the violinist: motor experts show superior timing consistency in a visual perception task. *Psychological Research, 74*, 579–585.

3.2 Hearing the speed: visual motion biases the perception of auditory tempo.

Yi-Huang Su [a] & Donatas Jonikaitis [b]

***Experimental Brain Research*, 2011, volume: 214, 357–371**

Accepted: 3 August 2011

Published in print: 25 September, 2011

a. Institute of Medical Psychology, Ludwig-Maximilians-University Munich, Goethestr. 31, 80336 Munich, Germany

b. Allgemeine und Experimentelle Psychologie, Ludwig-Maximilians-University, Munich

Exp Brain Res (2011) 214:357–371
DOI 10.1007/s00221-011-2835-4

RESEARCH ARTICLE

Hearing the speed: visual motion biases the perception of auditory tempo

Yi-Huang Su · Donatas Jonikaitis

Received: 12 April 2011 / Accepted: 3 August 2011 / Published online: 13 August 2011

Abstract The coupling between sensory and motor processes has been established in various scenarios: for example, the perception of auditory rhythm entails an audiomotor representation of the sounds. Similarly, visual action patterns can also be represented via a visuomotor transformation. In this study, we tested the hypothesis that the visual motor information, such as embedded in a coherent motion flow, can interact with the perception of a motor-related aspect in auditory rhythm: the tempo. In the first two experiments, we employed an auditory tempo judgment task where participants listened to a standard auditory sequence while concurrently watching visual stimuli of different motion information, after which they judged the tempo of a comparison sequence related to the standard. In Experiment 1, we found that the same auditory tempo was perceived as faster when it was accompanied by accelerating visual motion than by non-motion luminance change. In Experiment 2, we compared the perceived auditory tempo among three visual motion conditions, increase in speed, decrease in speed, and no speed change, and found the corresponding bias in judgment of auditory tempo: faster than it was, slower than it was, and no bias. In Experiment 3, the perceptual bias induced by the change in motion speed was consistently reflected in the tempo reproduction task. Taken together, these results indicate that between a visual spatiotemporal and an auditory temporal stimulation, the embedded motor representations from each can interact across modalities, leading to a spatial-to-temporal bias. This suggests that the perceptual process in one modality can incorporate concurrent motor information from cross-modal sensory inputs to form a coherent experience.

Keywords Sensory-motor coupling · Audiovisual interaction · Visual motion · Auditory rhythm perception · Cross-modal interaction

Introduction

Integrating concurrent information across sensory modalities is important in how we experience our environment as a continuous and coherent entity. In many cases, information reaching different modalities comes from the same source: moving sounds of the forthcoming car or words spoken by the person we see. Sometimes, however, we receive inputs in different modalities that do not originate from the same source of action, but we habitually incorporate them together as one perceptual experience: for example, we listen to the music accompanying a scene in the movie, or we watch the video accompanying a song. It seems in this case that two streams of information—coming from different sources and processed by two sensory modalities—might interact. This study was concerned with the latter scenario, and we investigated which particular aspects of visual and auditory information lead to the interaction between the two modalities.

An extensive body of literature has investigated how our perceptual system combines or reconciles inputs from different modalities. When visual and auditory stimuli are presented in temporal or spatial proximity, the *modality-appropriateness hypothesis* (Welch and Warren 1980)

Y.-H. Su (✉)
Institute of Medical Psychology,
Ludwig-Maximilians-University, Goethestr. 31,
80336 Munich, Germany
e-mail: jasmine.su@med.uni-muenchen.de

D. Jonikaitis
Allgemeine und Experimentelle Psychologie,
Ludwig-Maximilians-University, Munich, Germany

predicts that the more competent modality for the task-relevant property wins over the other. This typically means when the task entails judgments of a spatial nature, such as localizing concurrent visual items and sounds, vision tends to bias audition as in the common example of ventriloquism (Warren et al. 1981; Kitagawa and Ichihara 2002; Soto-Faraco et al. 2002; Alais and Burr 2004; Alink et al. 2008). On the other hand, when the task requires temporal judgment of the rate of the events or the time of the event occurrence, audition appears to be more sensitive and dominant, and it often biases the judgment of the concurrent visual items (Shipley 1964; Shams et al. 2000; Fendrich and Corballis 2001; Recanzone 2003; Morein-Zamir et al. 2003; Guttman et al. 2005; Burr et al. 2009).

The temporal dominance of audition over vision is further documented in another scenario: rhythm perception. Rhythm commonly refers to the 'organization of events in time' (Large 2008) or how the discrete events are grouped by the stimulus structure or by the listener's percept. The rhythm tasks are essentially of a temporal nature; they require not merely simple interval timing but rather forming an organized representation of successive intervals (Guttman et al. 2005; Grondin and McAuley 2009), perceiving the regular beat based on the given or implied meter (Patel et al. 2005; McAuley and Henry 2010) or synchronizing to a metronomic sequence (Repp and Penel 2002; Repp 2003). Findings from these studies generally point to the relative ease of rhythm perception in audition compared to in vision, yielding further evidence of the auditory advantage in processing complex temporal relations.

The auditory dominance over vision in the rhythm domain could be explained by the mechanism that rhythm perception entails: a sensory-motor coupling. It has been demonstrated that hearing sounds associated with an action activates the motor network of the frontoparietal and premotor cortex even in the absence of action execution (Kohler et al. 2002; Lahav et al. 2007). This audiomotor representation of sounds seems important in processing auditory rhythms, as the relevant premotor and supplementary motor areas are activated by simply listening to rhythmic stimuli or performing auditory perceptual tasks (Chen et al. 2008, 2009; Bengtsson et al. 2009). Other motor areas such as basal ganglia and cerebellum are also recruited when the temporal structure of the auditory rhythm is processed (Grahn and Brett 2007; Grahn and Rowe 2009), further confirming the audiomotor coupling in rhythm-related processes. On the other hand, these neural correlates are not exclusively found in auditory tasks: *both* auditory and visual rhythms—presented as successive discrete events in time—seem to engage the same network encompassing posterior superior temporal gyri, premotor cortex, and supplementary motor area (Karabanov et al. 2009). Despite the shared neural basis for processing auditory and visual rhythms, it has been proposed that the visual events should be first *recoded* into an audiomotor representation in order to be processed as 'rhythm' in the relevant neural network (Guttman et al. 2005; Karabanov et al. 2009). Following this argument, the consistently observed lower performance in visual rhythm tasks seems to suggest that the sensory-motor recoding of discrete visual events is not as effective as the direct sensory-motor coding in audition. This is further supported by the finding that the same rhythm task leads to higher activities in relevant brain areas, such as superior temporal gyri and basal ganglia, when it is presented in the auditory than in the visual modality (Grahn et al. 2010). In addition, the same study demonstrates that a visual rhythm following an auditory one elicits higher activity in putamen (a structure associated with beat perception) than when it is not preceded by an auditory rhythm; on the contrary, activity elicited by an auditory rhythm is unaffected by the preceding visual rhythm. This is taken to indicate that the internal rhythm representation is primed by auditory rhythmic events, but not by visual ones. To sum up, the visual analog of auditory rhythm appears to be inefficient in terms of sensory-motor coupling.

However, people become more sensitive to the rhythmic properties when the visual rhythms are presented with visual motion (Hove et al. 2010; Brandon and Saffran 2011): motion information seems to enhance the visual 'rhythm salience'. Besides, similar to the audiomotor association of sounds, visuomotor information such as the association between the observed movement and its action is also represented in the premotor area (Rizzolatti et al. 1996; Murata et al. 1997; Fadiga et al. 2000). It leads to the idea that, despite the typical auditory dominance in the temporal domain, certain aspects of visual stimulation related to its movement information might interact with auditory rhythms at the level of sensory-motor representation. This idea is further encouraged by the finding that an optic flow moving in space with coherent and perceivable direction activates dorsal premotor cortex (van der Hoorn et al. 2010), an area that is tightly associated with rhythm perception. It seems such visual motion can be readily transformed into action patterns which may suitably interact with the motor aspect of auditory rhythm. In addition, it has been observed that a single auditory interval is perceived to be longer when it is accompanied by an expanding disk that appears as moving toward the observer, compared to that accompanied by a static disk (van Wassenhove et al. 2008). This finding further supports that a vision-to-audition bias in the temporal domain can be induced by visual stimuli containing motion.

Taken from here, the present study investigated whether visual motor information embedded in a coherent motion flow could influence concurrent auditory rhythm perception, specifically the tempo. Auditory tempo, as part of the rhythmic percept, also entails premotor cortical activities (Kornysheva et al. 2010), and we expected it to be susceptible to the influence of visual motion information. To this end, we employed first an auditory tempo judgment task (Drake and Botte 1993; Miller and McAuley 2005) in two experiments and measured the subjectively perceived auditory tempo under different visual motion conditions. In Experiment 1, we investigated whether a visual array containing motion influenced auditory tempo perception, compared with an array containing no visual motion. We hypothesized that the same auditory tempo accompanied by accelerating visual motion would be perceived as faster than that accompanied by temporally changing stimuli which contained no motion information. In Experiment 2, we further tested whether the influence of visual motion on auditory tempo was directly related to the change of motion speed. Here, the visual motion started at the same baseline speed and then increased, decreased, or did not change, and we predicted a corresponding bias in auditory tempo: faster, slower, and no bias. In Experiment 3, we further validated the perceptual bias using a reproduction task. Here, in two-thirds of the trials, the speed started as faster or slower and changed to the same baseline speed, and in one-third of the trials, the speed did not change. We predicted that the tempo reproduction would again reflect the perceptual bias induced by the speed change.

Experiment 1: Visual static versus visual motion

In this experiment, we investigated whether the presence of visual motion information, such as dots moving coherently with accelerating speed, would influence the concurrent auditory tempo perception. This condition was compared to that in which temporal information without concurrent motion was presented, implemented by static dots with incoherently accelerating changes of luminance. We predicted a bias on auditory tempo perception toward the faster side when it was accompanied by accelerating moving dots.

Method

Participants

Eleven young, healthy participants and the first author participated in this experiment (mean age 25 years, SD 3.1). The participants registered via on-campus recruitment and received payment in return for their participation. Musical training ranged from none to 8 years (amateur). Informed consent was obtained from all participants prior to the experiment.

Stimuli and materials

Auditory stimuli The auditory stimuli consisted of five tones in an isochronous sequence. The tones were generated by the music software Logic 8 Express (Apple Inc. California) using a synthesized woodblock sound (the instrument “clave”), with 42-ms tone duration. Two kinds of sequences were presented in the experiment: a standard sequence and a comparison sequence. The standard sequence was generated in two different tempi: 100 beats per minute (BPM) and 150 BPM, corresponding to the inter-onset interval (IOI) of 600 and 400 ms, respectively. The comparison sequence consisted also of five isochronous tones, and their IOIs were ±4, ±8, or ±12% of the respective standard IOI. The sound was delivered via headphones (Philips SBC HS900).

Visual stimuli The visual stimuli were generated by Matlab® 2009a (Mathworks) using Psychophysics Toolbox extensions version 3 (Brainard 1997) and presented on a MacBook Pro laptop computer, with a display resolution of 1,280 × 800 pixels. The display carried a frame frequency of 61 Hz. Two conditions were presented: a motion condition and a static condition.

Motion condition. Fifty white dots (size 0.15° of visual angle) moved against a black background in an area defined by an invisible 15° radius circle. Dots moved from the fixation point in the center of the screen along the radius of the circle, and all the dots on the screen moved with the same speed at any given time, forming a percept of coherent optic flow in the direction of the observer. A dot disappeared when it reached the edge of the circle, and a new dot started moving from the center. The dots were moving in a constantly accelerated manner, and the final speed was always twice the starting speed. When presented concurrently with the auditory sequence at 150 and 100 BPM, respectively, the starting speed was 35 and 23.33 deg/s, at a constant acceleration of 0.23 and 0.2 deg/frame. The ratio between the slower and the faster conditions was thus the same for the auditory tempo and for the dot speed.

Static condition. Fifty dots (size 0.15°) were presented against a black background in random locations within 15° of the radius from the screen center. Every dot changed its luminance continuously throughout a trial presentation. The change in luminance was programmed along a gradual scale between white (255 255 255 in RGB color) and black (0 0 0 in RGB color), in 8 steps. For each single dot, the starting luminance was randomly chosen from one of the 8 steps along the scale, and the luminance changed systematically toward black and then backwards to white and again toward black and so on. The speed of the luminance change was controlled by the time interval between two luminance steps, Δt_{lum} (i.e., the duration for which each dot luminance stayed on the screen before it changed to the next). To be comparable to the acceleration in the motion condition, the luminance also changed in an accelerating manner. The acceleration was implemented by shortening each successive Δt_{lum} by the same amount, till the last Δt_{lum} was half of the initial Δt_{lum}. The initial Δt_{lum} was 200 and 133 ms when presented concurrently with the auditory sequence at 100 and 150 BPM, respectively. As each dot had a randomly chosen luminance in the beginning and was changing luminance independently of the others, there was no particular perceptual coherency from the visual stimulation as a whole.

Design and procedure

Participants sat in a dimly lit room with 60 cm viewing distance from the screen while listening to sounds delivered via headphones. Figure 1 depicts the trial procedure. In each trial, participants were instructed to listen to the auditory standard sequence and to pay attention to its tempo while watching concurrent visual stimuli on the screen.[1] They were requested to fixate at the center of the screen. The visual stimuli stopped one IOI later than the onset of the last tone, so that the cessation of visual stimulation was felt less abrupt in relation to the auditory sequence. Following that, the auditory comparison sequence was presented without accompanying visual stimuli. After the end of the comparison sequence, the participants were requested to make a two-alternative forced-choice (TAFC) judgment whether the tempo of the comparison sequence was ‘faster’ or ‘slower’ than the tempo of the standard sequence. The response was made by pressing one of the two designated keys. A short interval was inserted between the end of the standard and the beginning of the comparison sequence, which varied randomly from trial to trial between 1,000 and 1,500 ms in steps of 10 ms. This was meant to prevent the participants from hearing the two sequences as a whole and making their judgment based on the within-sequence tempo change. The next trial commenced 2.5 s after the response was collected.

The experiment followed a 2 (auditory tempo in standard sequence) × 2 (motion vs. static visual stimuli) design. Within each condition, there were six variations of the comparison sequence (see “Stimuli and Material”), each with 20 repetitions. The resultant 480 trials were randomly assigned to 5 blocks, and all the conditions balanced across the blocks. The participants underwent a practice block of at least 8 trials prior to the experiment and received a break after each block. The whole experiment lasted between 1.5 and 2 h.

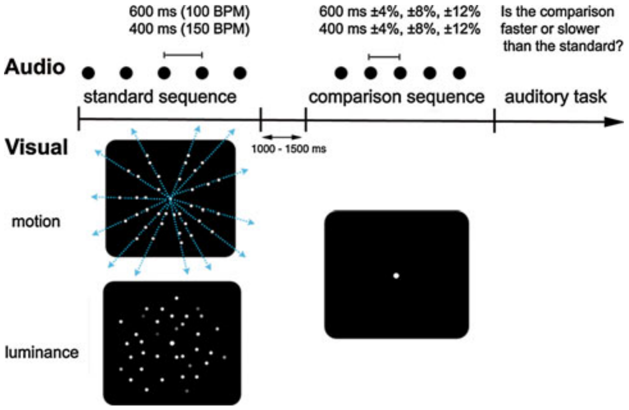

Fig. 1 Illustration of the trial procedure in Experiment 1. The auditory standard sequence (at 100 or 150 BPM) and the visual stimuli (acceleratingly moving *dots* or *dots* changing luminance) started at the same time, after which the visual stimuli stopped and the auditory comparison sequence was presented. A two-alternative forced-choice (TAFC) response was requested in the end to judge whether the comparison sequence was faster or slower than the standard sequence. *Arrows* on the visual dot display represent the invisible radii along which the dots moved (the invisible *circle* itself is not drawn here)

Data analysis and results

Perceived tempo measured by point of subjective equality

Data from one participant were excluded from the analysis due to low auditory discrimination. For the other 11 participants, the data were first computed individually: under each of the four (2 auditory × 2 visual) conditions, the percentages of responding ‘faster’ were plotted on the Y-axis against the six measurement points, ±4, ±8, and ±12%, (i.e., the comparison conditions) on the X-axis. The data were then fitted to a Weibull distribution, yielding four psychometric curves for each individual participant. The goodness of the fit was indexed by R^2, and we used the criterion of $R^2 > 0.8$ as determining that the fit described the data well enough. All the R^2 values ranged between 0.8 and 0.99. Figure 2 shows the fitted curves of the *mean* data points across all participants.

The point of subjective equality (PSE) was calculated, for each participant in each experimental condition, as the point on the X-axis corresponding to 50% probability on the Y-axis of the individually fitted curve. The PSE here was a relative PSE, representing the *percent of deviation* of the perceived IOI from the standard IOI, under the

---

[1] In order to facilitate the percept of forward motion, during instruction the experimenter mentioned that the appearance of motion condition might resemble that of the situation while they were sitting in a forward-moving car, with the scenes outside moving towards them. All participants were able to relate to this sensation while watching visual motion stimuli.

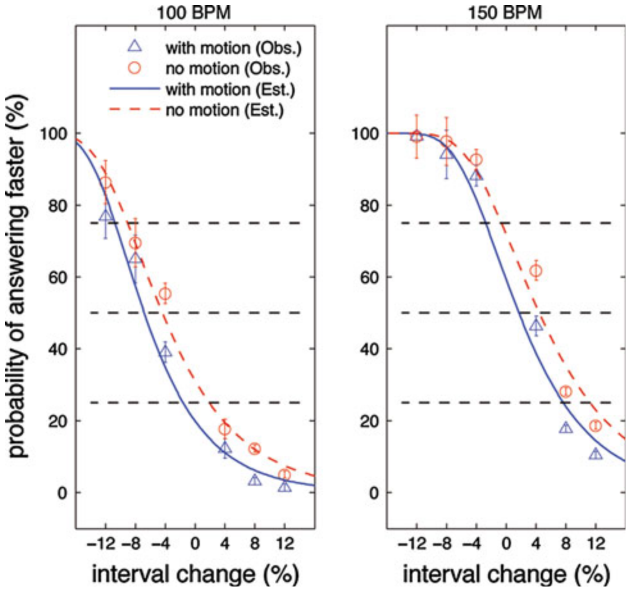

Fig. 2 Data plotted across all participants in Experiment 1, showing the mean observed response probabilities as the scattered triangles and circles, and the fitted psychometric curve under each visual and auditory tempo condition. *Error bars* represent standard error of the means. Note that while the fitted curves of mean data points are shown here, the PSEs were calculated for every participant separately by individually fitted curves, PSE determined by the point on the *X*-axis corresponding to 50% on the *Y*-axis. *Obs* observed; *Est* estimated

accompanying visual condition.[2] A positive PSE shift suggested the perceived tempo of the standard sequence to be slower than it actually was, while a negative shift suggested the perceived tempo to be faster than actual.

The resultant PSEs were submitted to a 2 (auditory tempo) × 2 (visual condition) within-subject ANOVA. A main effect was found for tempo, $F(1,10) = 58.51$, $P < 0.001$, $\eta_p^2 = 0.85$ (100 vs. 150 BPM: −6.35% vs. 3.38%), and for visual condition, $F(1,10) = 6.38$, $P < 0.05$, $\eta_p^2 = 0.40$ (motion vs. static: −2.82% vs. −0.15%). The interaction between auditory tempo and visual condition was not significant, $P > 0.6$ (Fig. 3). In summary, two findings were shown here: first, the standard tempo at 100 BPM was generally perceived as faster than it was, while the standard tempo at 150 BPM was perceived as slower than it was. This could be explained by Vierordt's law (1868), which predicts that, when presented in the same experiment, shorter intervals tend to be overestimated and longer intervals underestimated. This is also in accord with the effect of global temporal context on tempo judgment (Jones and McAuley 2005; McAuley and Miller 2007). Secondly, and more importantly, even under the influence of a global pace, the subjective tempo was still consistently biased by the visual condition in our predicted fashion: the same auditory tempo accompanied by acceleratingly moving dots was perceived as faster than that accompanied by static dots with changing luminance. Results showed that the visual motion information influenced the auditory tempo perception, and the direction of auditory bias was in accord with the increase in visual spatial frequencies.

Tempo sensitivity measured by just noticeable difference

To assess the sensitivity to the auditory tempo under our experimental conditions, the just noticeable difference (JND) in each condition for each participant was calculated as half of the distance between the points on *X*-axis corresponding to the 25th and 75th percentiles on the *Y*-axis. The yielded JNDs were submitted to a 2 (auditory tempo) × 2 (visual condition) within-subject ANOVA. No significant effect was found, suggesting that the sensitivity to the auditory tempo did not differ among conditions. The overall mean JND across conditions was 5.2%. The result showed that the two kinds of visual stimuli did not impose different levels of sensory interference on the sensitivity to the auditory tempo.

Discussion

Experiment 1 attempted to verify whether concurrent visual motion information could interact with auditory tempo perception. The results confirmed our prediction that, compared to visual stimulation with temporally changing but non-motion information, accelerating visual motion led to a faster subjective auditory tempo. This serves as a first evidence that the accompanying visual motion can influence the judgment of auditory tempo, and

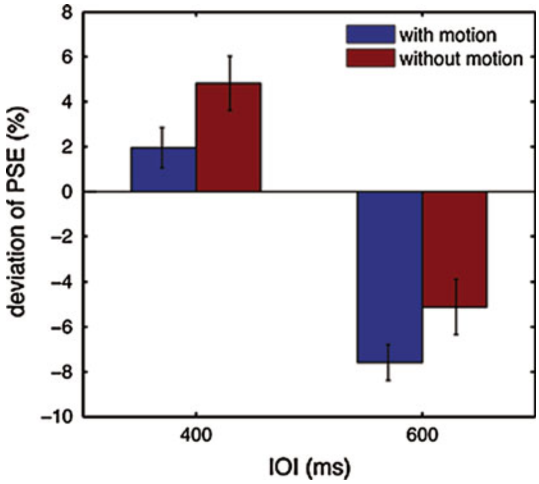

Fig. 3 The mean percentage of deviation in the point of subjective equality (PSE) under each visual and tempo condition in Experiment 1. *Error bars* represent standard error of the means

---

[2] Our PSE as percent of deviation from standard IOI is the same index as the relative constant error (CE) in the studies of Miller and McAuley (2005) and McAuley and Miller (2007).

the bias seems in accord with the increasing motion speed. Since the visual information was *not* presented as discrete rhythmic events, the observed visual influence on auditory rhythm could not have resulted from a recoding of visual rhythm into an audiomotor pattern (Guttman et al. 2005; Karabanov et al. 2009). Furthermore, there was no direct parameter-specific relation between the stimulation in two modalities—it was not both temporal or both spatial; the result was therefore not easily explainable by a 'visual driving' in spatial or motion processes (e.g., Kitagawa and Ichihara 2002; Alais and Burr 2004; Alink et al. 2008), nor by a visual temporal capture of audition (Fendrich and Corballis 2001, Experiment 2). It is more likely, as in our interpretation, that another aspect of the stimulus property was shared between the two modalities: the motor representation derived from the sensory input. Findings on the neural correlates of the audiomotor (Lahav et al. 2007; Kohler et al. 2002; Chen et al. 2009) and visuomotor representation (Murata et al. 1997; Fadiga et al. 2000; Lee and van Donkelaar 2006; van der Hoorn et al. 2010) support the idea that both the auditory and visual transformations into action patterns are coded in the premotor cortical area, making it a plausible platform for our investigated interaction.

Alternatively, since tempo discrimination may entail a similar mechanism to basic interval timing (Fraisse 1963), the observed cross-modal interaction could be explained in terms of the timing mechanism related to a common internal clock (e.g., Buhusi and Meck 2005). Within the visual modality, it was found that motion stimuli lead to a dilated time estimation, and the influence was best described by the temporal frequency instead of speed, spatial frequency, or motion coherency (Kanai et al. 2006). This would have led to the opposite result between our conditions: the moving and accelerating dots should have led to greater time dilation. This would in turn have resulted in a slower perceived tempo in the motion than in the no-motion condition, which was not the case. Therefore, a cross-modal time dilation by dynamic visual stimuli cannot account for the inter-sensory interaction in our scenario. We postulate that the bias should be attributed to an integrated representation of the comparable information from both modalities: motion in visual movement and motion in auditory tempo.

The results here were, however, affected by the global temporal context, which drew the general perception of two standard tempi toward their arithmetic mean (Jones and McAuley 2005). In addition, compared to the static condition with temporal information, the motion condition contained not only additional spatial information but also a constant increase in the spatial frequency (i.e., the acceleration of speed). This was meant to maximize the difference between the two conditions but might make it somewhat difficult to interpret whether the bias in the motion condition resulted from the mere presence of motion or from the increasing speed. Furthermore, if an increase in speed was needed to induce the observed bias, did it necessarily require a constant acceleration? In the following experiment, we addressed these questions by manipulating the speed of the motion and testing the effect of different motion speed changes on the perception of one auditory tempo.

Experiment 2: Visual motion: changing speed versus constant speed

In Experiment 2, we specifically tested the effect of speed properties of visual motion on auditory tempo perception; meanwhile, we intended to suppress the global context effect by presenting mainly one auditory tempo. Three visual motion conditions were employed: moving dots that increased, decreased, or did not change their speed. We assumed that the bias in Experiment 1 was caused by the speed increase instead of the mere presence of motion, which we intended to validate here. As such, we predicted that the perceived auditory tempo in Experiment 2 would be accordingly biased by the change in the visual motion speed (to be faster or slower than actual) in the conditions where the speed changed, but not when the speed did not change. In this way, we could reveal whether the mere presence of visual motion information would suffice or whether a change of speed was necessary to introduce the auditory bias. In addition, the increase or decrease in speed was implemented as a step-up change instead of constant acceleration or deceleration. This would tell us whether other forms of speed change could also impose the bias.

Method

Participants

Eleven young, healthy participants and the first author participated in this experiment (mean age 25 years, SD 4.5). The participants registered via on-campus recruitment and received payment in return for their participation. Except for the first author, none of them had participated in the first experiment, and the majority (9 out of 11) had no formal musical training. Informed consent was obtained from all participants prior to the experiment.

Stimuli and materials

Auditory stimuli The same auditory sequences as used in Experiment 1 were employed here. The only difference was that both the standard and the comparison sequences consisted now of four tones instead of five. This was done to reduce the duration of the experiment, while reducing the number of IOIs in both standard and comparison sequences from four to three should not alter the tempo sensitivity significantly (Drake and Botte 1993; Miller and McAuley 2005). The sequences at 150 BPM constituted the main auditory condition in this experiment; the sequences at 100 BPM occurred only in 20% of the trials, serving as sparse distractors to prevent the participants from memorizing the standard tempo during the session.

Visual stimuli The visual stimuli were generated by Matlab® 2009a (Mathworks) using Psychophysics Toolbox extensions version 3 (Brainard 1997) and presented on a 21-inch CRT monitor (Sony GDM-F500R) with a frame frequency of 85 Hz, at a spatial resolution of 1,024 × 768 pixels.

Three visual motion conditions were presented in this experiment: (1) increase in speed, (2) decrease in speed, and (3) no speed change. The stimuli in all three conditions resembled generally the moving dots presented in the 'motion condition' of Experiment 1: 50 dots of the same size moved radially from the screen center toward the observer. As in Experiment 1, two auditory tempi—100 and 150 BPM—were employed (though the latter was presented much more frequently), and we manipulated the accompanying visual motion speed for each auditory tempo separately. In the *increasing-speed* condition, dots accompanying 150 BMP auditory tempo moved at 35 deg/s for 1,600 ms (corresponding to 4 × IOI) and then switched to a speed which was 40% faster and moved for another 2,400 ms (6 × IOI). Dots accompanying 100 BPM auditory tempo moved at a constant speed of 23.33 deg/s for 2,400 ms and then switched to a speed which was 40% faster and lasted for 3,600 ms. In the *decreasing-speed* condition, the initial speed and all the corresponding durations were the same, except that after the switch, the speed changed to 40% slower than the initial. In the *no speed change* condition, the parameters were the same as above, except that the speed remained constant (as the initial speed) throughout the presentation. In all conditions, the visual stimulation initially preceded the auditory sequence (for details, see Procedure below and Fig. 4).

The speed change in all cases was introduced as an easily perceivable step-up change, not a gradually accelerating process as in Experiment 1. This was meant to present the observer with an obvious percept of the motion speed change from a 'baseline' speed, and in doing so, we would be able to attribute the differential bias on audition, if any, to the difference between the two visual speeds. No participant reported difficulty in perceiving the speed change during practice.

Design and procedure

Participants sat in a dimly lit room with 80-cm viewing distance from the screen while listening to sounds delivered via headphones. The head position was fixed by a chin rest. Besides the auditory tempo judgment task as in Experiment 1, we introduced here a simple secondary task of detecting visual motion change, with the purpose of keeping the participants attending to both visual and

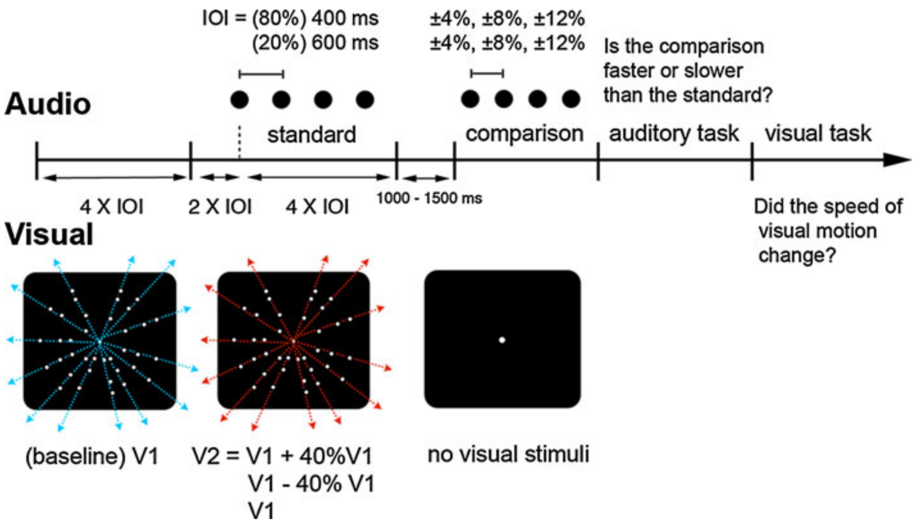

Fig. 4 Trial procedure in Experiment 2. The visual motion started first at a baseline speed (V1) and changed to speed V2, shortly after which the standard tone sequence was presented (at 150 BPM in 80% of all the trials and at 100 BPM in the rest), accompanied by visual motion at speed V2. The comparison sequence followed without accompanying motion, after which the auditory and visual task responses were collected in succession. *Arrows* on the visual *dot* display represent the invisible radius along which the *dots* moved (the invisible *circle* itself is not drawn here)

auditory stimulations throughout the presentation. Figure 4 depicts the trial procedure. In each trial, the dots first appeared on the screen and started moving at a constant speed V1 for 4 × IOI amount of time, after which they immediately switched to moving at a constant speed V2 for 6 × IOI amount of time. After a lag of 2 × IOI from the onset of V2, the standard auditory sequence started playing, accompanied by the dots moving at the speed V2. This lag was introduced to provide enough time for participants to have been aware of the visual speed change before concentrating on the auditory tempo. As in Experiment 1, the visual stimuli stopped one IOI later than the onset of the last tone. After an interval varying (in steps of 10 ms) between 1,000 and 1,500 ms, the comparison sequence was presented without concurrent visual stimuli. The participants had to give two responses in succession in the end of a trial: (1) first, whether the second auditory tempo was slower or faster than the first one (by pressing one of the two keys corresponding to 'faster' or 'slower') and then, (2) whether the visual motion had changed during the presentation or not (by pressing one of the two keys corresponding to 'yes' or 'no'). An error feedback was given for an incorrect response in the visual task in order to facilitate learning and to maintain concentration on the visual modality. The next trial commenced 2.5 s after the last response or the error feedback (when present).

For 150 BPM, 20 repetitions were included for each comparison sequence and each visual condition, yielding 360 trials. For 100 BPM, there were 5 repetitions each, so they occurred in only 20% of all the trials. The resultant 450 trials were randomly assigned to 5 blocks, and all the conditions balanced across the blocks. The participants underwent a practice block of at least 8 trials prior to the experiment and received a break after each block. The whole experiment lasted about 2–2.5 h.

Data analysis and results

Perceived tempo measured by PSE

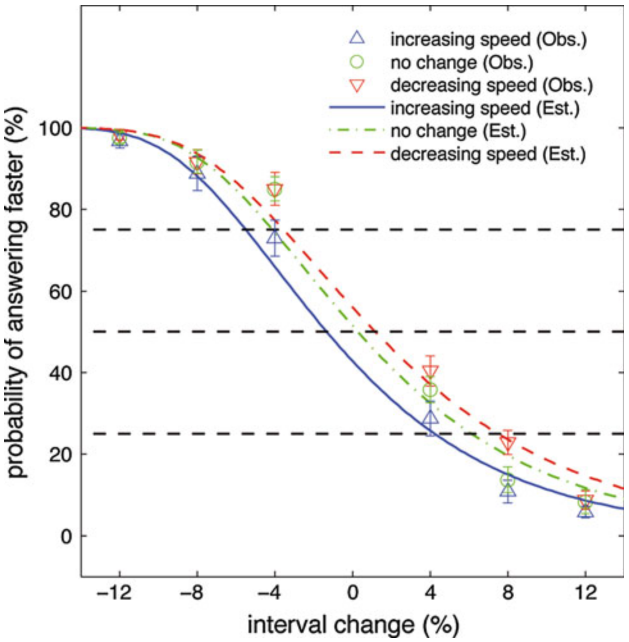

Fig. 5 Data plotted across all participants in Experiment 2, showing the scattered triangles as the mean observed response probabilities, and the fitted psychometric curve for each motion condition at tempo 150 BPM. *Error bars* represent standard error of the means. *Obs* observed; *Est* estimated

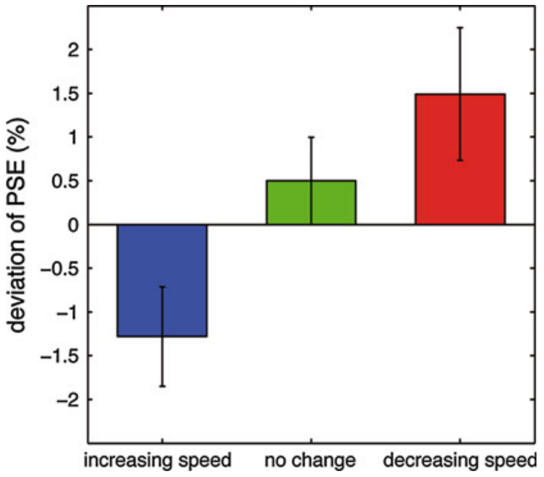

Fig. 6 The mean PSE (as percent of deviation from the standard IOI) under each visual motion condition in Experiment 2. *Error bars* represent standard error of the means

We analyzed data from 150 BPM under three visual motion conditions. The same procedure of psychometric curve fitting as in Experiment 1 was applied to individual data, yielding three (relative) PSEs for each participant. The R^2 values ranged between 0.9 and 1.0. Figure 5 shows the fitted curves of the mean data points across all participants. The individual PSEs (also as percents of deviation from the standard IOI) were submitted to a one-way within-subject ANOVA with the variable of visual condition (3 levels: increasing speed, decreasing speed, and no speed change). A significant effect was found, $F(2, 22) = 17.5$, $P < 0.001$, $\eta_p^2 = 0.78$. Post hoc comparisons (Tukey HSD) revealed the PSEs between 'increasing speed' and 'decreasing speed' to be significantly different, $P < 0.01$ (increasing speed vs. decreasing speed: −1.3% vs. 1.5%). As shown in Fig. 6, a negative shift of PSE was observed when the visual motion increased in speed, suggesting the concurrent tempo was perceived as faster than it was. The opposite was observed in the 'decreasing-speed' condition, with a positive shift of PSE, suggesting the perceived tempo to be slower than it was. One-sample *t* test confirmed that the PSE shift in the 'no speed change' condition was not significantly different from 0, $t(11) = 1.0$, $P = 0.34$. Results here demonstrated the differential influence of visual speed change on auditory tempo perception: when the motion was perceived to *have become* faster or slower, the tempo of the concurrent auditory sequence was perceived as faster or slower than it actually was. When the motion speed remained constant, it did not impose any particular bias on the auditory tempo.

Tempo sensitivity measured by JND

The tempo sensitivity in each condition for each participant was also computed as in Experiment 1, indexed by the JND. The one-way ANOVA of JNDs yielded no significant effect of visual condition, $P > 0.6$, suggesting again the auditory tempo sensitivity was comparable among three visual conditions. The overall mean JND across conditions was 4.6%.

Auditory and visual task accuracy

To assess the potential trade-off between the auditory and visual task performance, the correlation between the overall accuracy of the two tasks was computed across participants, which was not significant, $r = 0.51$, $P = 0.09$. A multivariate analysis of variance (MANOVA), using Wilk's λ, with one independent variable—visual condition—and two dependent variables—accuracy of auditory task and accuracy of visual task—yielded no significant effect of visual conditions on the two task accuracies, $P = 0.19$. The mean accuracy was 83% for the auditory task and 86% for the visual task. We concluded that (1) the auditory and visual response accuracy was equal across three motion conditions and (2) the response accuracies in auditory and in visual task were rather independent of each other, as no interaction of the dependent variables was found. On a side note, the accuracy in the visual task was not close to 100%, though the change in speed was made rather obvious. This may have been due to the fact that participants sometimes forgot the answer to the visual task after they responded to the auditory task first.

Discussion

Results of Experiment 2 further demonstrated the influence of visual motion information on the perception of auditory tempo, which depended on the direction of the speed change in the former. Critically, it showed that visual motion was necessary but not sufficient: a change in the spatial frequency, the speed, was a prerequisite for eliciting a bias in the auditory tempo, as no bias was observed in the constant speed condition. However, two questions may arise from here: (1) Did the bias indeed result from the *change* in visual speed and not from the different visual speeds accompanying the standard sound sequence? (2) Might the observed bias have occurred at a post-perceptual, semantic level? That is, when an auditory sequence was accompanied by faster/slower visual speed, did the participants *hear* it as faster/slower, or did they simply label it as 'faster/slower', thus rendering the decision at the response stage to be semantically biased? First and notably, regarding response bias, it was *not* the case that participants simply responded 'faster' (or 'slower') for the auditory task when the visual stimuli moved faster (or slower). Rather, they tended to answer 'slower' for the auditory task under faster visual motion condition, and vice versa, as the task was to judge the comparison sequence relative to the standard. Nevertheless, the question of semantic bias involving verbal judgment remains when making such a comparison. Before accepting and discussing the observed effect, we designed a third experiment to answer these questions.

Experiment 3: Auditory tempo reproduction in three visual motion conditions

In this experiment, we aimed to further validate two specifics of the observed motion-to-tempo bias in Experiment 2: (1) The bias on auditory tempo should result from the 'change' in the visual speed, not from the speed itself that accompanies the auditory sequence. (2) The bias should occur at a perceptual level, not as a result of semantic labeling being transferred from the visual stimulation. To this end, we modified the experimental design of Experiment 2 as follows: (1) The speed change in each motion condition was reversed: it could increase or decrease to *the same speed*, or the speed did not change. In this way, the visual stimuli moved at identical speed during the single auditory sequence in all conditions. (2) The subjectively perceived auditory tempo was measured by a tapping reproduction task instead of a perceptual judgment task. This way, we could rule out the potential semantic bias at the stage of response selection. Besides, tapping reproduction of a tone sequence has been found to reflect the subjectively perceived tempo (Repp 2008; Repp and Bruttomesso 2009). We predicted that the reproduced tempo—reflecting the perceived tempo—would again be accordingly biased by the *change* of visual speed, despite the same speed that always accompanied the auditory sequence.

Method

Participants

Twelve young, healthy participants took part in this experiment (ten volunteers and the two authors, mean age 25 years, SD 3.5). The volunteers registered via on-campus recruitment and received payment in return for their participation. Musical training varied widely, ranging from none to 10 years (amateur). Except for the first author, none of them had participated in the previous experiments. Informed consent was obtained from all participants prior to the experiment.

Stimuli and materials

Auditory stimuli The auditory stimuli were generated by the same programs as previously, and each stimulus consisted of a four-tone sequence at one of seven possible IOIs corresponding to ±4, ±8, ±12, and 0% deviation from 400 ms (i.e., 352, 368, 384, 400, 416, 432, and 448 ms). The range of sequence tempo here was thus comparable to the standard (at 150 BPM) and comparison tempi in Experiment 2.

Visual stimuli The visual stimuli were generated by the same program as in Experiment 2. Similarly, three motion conditions were included: (1) increase in speed, (2) decrease in speed, and (3) no speed change. As in Experiment 2, the baseline speed was 35 deg/s. A major difference was introduced here: in the *increasing-speed* condition, dots started moving at a speed which was 40% *slower* than the baseline speed for an amount of time (corresponding to 4 × IOI of the respective auditory sequence in each trial) and then switched to the baseline speed and moved for 6 × IOI amount of time. This was in contrast to Experiment 2 where baseline speed was presented first and then the speed increased. In the *decreasing-speed* condition, dots started moving at a speed that was 40% *faster* than the baseline speed and then switched to the baseline speed. In the no speed change condition, dots moved constantly at the baseline speed throughout the presentation. In all conditions, the visual stimulation initially preceded the auditory sequence.

Design and procedure

The same setup as in Experiment 2 was employed, and there were two tasks within each trial: first, the reproduction of the auditory tempo by finger tapping and then as a secondary task, the detection of motion speed change. The participants started every trial by pressing the space key whenever they were ready. The time course of visual and auditory stimuli was similar to that in Experiment 2, except that now there was only one auditory sequence (which appeared at the same position as the standard sequence in Experiment 2). In each trial, the dots first appeared on the screen and started moving at a constant speed V1 for 4 × IOI amount of time, after which they immediately switched to moving at a constant speed V2 for 6 × IOI amount of time. V1 could be either 40% faster or 40% slower or the same as V2. After a lag of 2 × IOI from the onset of V2, the auditory sequence started playing, accompanied by the dots moving at the speed V2. The visual stimuli stopped one IOI later than the onset of the last tone. The participants were first requested to reproduce the exact auditory sequence tempo they had just heard by tapping on the key ‘B’ (four taps per trial). If the coefficient of variation (CV) of a tap series exceeded 15% or the mean inter-tap interval (ITI) deviated more than 20% from the IOI, the trial was repeated. Following the taps, the participants were requested to answer whether the visual motion had changed or not during the presentation by pressing one of the two keys. An error feedback was given for an incorrect response in the visual task. During the instruction, the participants were especially reminded that the two tasks were independent of each other and that they should perform each as accurately as possible.

The experiment followed a 7 (auditory tempo) × 3 (visual motion) design, each with 10 repetitions. The resultant 210 trials were randomly assigned to 2 blocks, with all conditions balanced. The whole session lasted around 1 h.

Data analysis and results

Reproduced auditory tempi

For each participant, the mean was calculated for all the within-trial mean ITIs[3] under each auditory tempo and visual motion condition. The resultant mean ITIs were transformed into percentages of deviation from the respective IOI and were submitted to a 7 (auditory tempo) × 3 (visual condition) within-subject ANOVA. A significant main effect was found only for visual condition, $F(2, 22) = 5.64$, $P = 0.01$, $\eta_p^2 = 0.34$. Post hoc comparisons (Tukey HSD) revealed that the mean ITI (as deviation from IOI) was different between the ‘increasing-speed’ and the ‘decreasing-speed’ conditions (−3.7% vs. −2.0%), $P = 0.03$, while the mean reproduction in the ‘no-change’ condition (−2.5%) did not differ significantly from that in the ‘increasing-speed’ ($P = 0.2$) or in the ‘decreasing-speed’ condition ($P = 0.8$). The interaction between auditory tempo and visual condition was not significant, $P > 0.1$. Figure 7 shows the reproduction result under each condition. Notably, all the auditory sequences were reproduced somewhat faster than their actual tempo, a phenomenon that has been observed in other studies with

---

[3] When a mean ITI exceeded three standard deviations from the average of mean ITIs under the same condition, the trial was excluded from the analysis. The occurrence of exclusion was generally very low, never exceeding 2% in any participant.

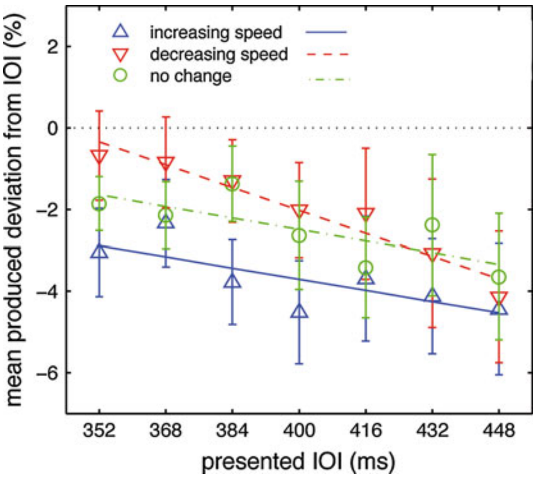

Fig. 7 Mean reproduced ITI (as percent of deviation from IOI) as a function of presented auditory IOI in each visual condition of Experiment 3. *Error bars* represent standard error of the means. Each *line* represents the linear regression of the data points under the corresponding visual condition. The *dotted horizontal line* represents zero deviation from presented IOI

reproduction of empty-interval sequences especially in the case of non-musicians (Repp 2008; Repp and Bruttomesso 2009). Despite this general tendency, the bias in the reproduced tempi remained between the two visual conditions with speed change. More importantly, and consistent with Experiment 2, the bias was in accord with the direction of the speed change, while the accompanying speed itself was the same in all conditions.

Visual task accuracy

The visual task accuracy was submitted to a one-way within-subject ANOVA with the variable of visual condition (3 levels: increasing speed, decreasing speed, and no speed change). A main effect was found of visual condition, $F(2, 22) = 14.44$, $P < 0.001$, $\eta_p^2 = 0.57$, and post hoc comparisons (Tukey HSD) identified only that the mean accuracy was significantly higher in the increasing-speed (97%) than in the decreasing-speed (82%) and the constant-speed conditions (83%), both $P\text{s} < 0.001$. Indeed, after the experiment, several participants commented on the relative ease to notice the speed change when it sped up than when it slowed down (although the magnitude of speed change was the same from the stimulus parameters). To reveal whether the effect on the reproduced tempo paralleled the subjective saliency of visual speed change, we computed the correlation between the following two measurements at the individual level: (1) the difference in the deviation of the reproduced tempo between increasing- and decreasing-speed conditions and (2) the difference in visual response accuracy between increasing- and decreasing-speed conditions. No significant correlation was found between these two measurements of difference, $r = 0.02$, $P = 0.9$. It seems the perceived saliency of speed change, as reflected in the response accuracy, did not have any direct influence on the reproduced auditory tempi.

Discussion

Experiments 2 and 3 further confirmed the bias in auditory tempo perception according to the direction of speed change in visual motion: speed that changed to faster led the auditory tempo to be perceived as faster than it actually was, and speed that changed to slower led to a slower perceived tempo than it was. This argues for a vision-to-audition bias that is qualitatively related to the motion information in the visual display.

While in Experiment 2, the observed bias could have been interpreted as resulting from the accompanying visual speed itself that differed across conditions, in Experiment 3, we provided data that argued against this interpretation, showing a bias consistent with the speed change while the auditory sequence was always accompanied by the same visual speed. Besides, the perceived tempo was measured by reproduction, thus ruling out the possibility of a semantic bias. Due to the overall underestimation in reproducing empty-interval sequences (Repp 2008; Repp and Bruttomesso 2009), it is difficult to prove in Experiment 3 whether there was indeed no bias in the constant speed condition. Nevertheless, in comparison to this condition, we did find again a bias differentiated by the preceding speed.

Overall, the results reveal several specifics of this visual-spatiotemporal to auditory temporal influence. First, we discovered that for visual motion to bias auditory tempo, a constant acceleration is not a prerequisite. The bias can occur also when the auditory sequence itself is accompanied by visual motion at a constant speed, provided that a slower or faster motion immediately precedes the auditory visual pair. Secondly, a constant speed without change would not suffice either, as the no-change condition did not seem to lead to such bias. That is, the observed cross-modal effect of slower or faster subjective auditory tempo seems to come from a perceived visual speed *in relation to* its preceding speed. Among several possible mechanisms of intersensory integration, the temporal proximity within which information from various modalities is received should play a determining role in the likelihood of the integration (Lewald and Guski 2003). Since the preceding speed in our case was rather temporally close to the paired audiovisual stimulation, the visual motion seemed to be integrated with auditory tempo as result of its *relative* rather than absolute speed. An interesting aspect for further research would be to elucidate the effective temporal distance between the speed change and the onset of the auditory sequence in order to induce this bias.

Regarding the possibility of time dilation in a dynamic visual display involving motion and speed, Kaneko and Murakami (2009) found that the time interval dilation increases with the speed of visual motion stimuli. Similarly, time dilation has been visually induced by the object appearing as moving toward the observer (Wittmann et al. 2010). In addition, the speedup of the 'internal clock' under higher arousal state has been documented, which leads to overestimation of time intervals (Treisman et al. 1990; Droit-Volet and Wearden 2002). Arguably, viewing increasing motion speed would potentially elicit higher arousal state in the observer. If any of these explanations applied to our study, then one would expect the same auditory IOI to be judged as longer when the visual speed is higher—meaning a slower perceived tempo—and as shorter when visual speed is lower—meaning a faster perceived tempo. However, we have observed the opposite in the results, suggesting that the inter-sensory interaction here does not originate from a general clock quickening. Even if the visual time dilation had indeed occurred and had been systematically transferred to audition (van Wassenhove et al. 2008), the resultant bias in the perceived auditory tempo would have to be attributed to a different cross-modal influence, which we propose to be based on the integrated motor information.

General discussion

From visual motion to auditory tempo

Our findings present a somewhat complicated inter-sensory interaction across modalities and parameters. Indeed, one could intuitively associate the idea of 'speed', as the speed of visually moving objects, with the speed of a passage of sounds: tempo in the context of music is often referred to as 'fast' or 'slow'. Seen in this light, the inputs from two modalities can possibly be integrated in terms of their 'speed'. However, it is not clear how the brain merges the percept of visual spatiotemporal speed and auditory temporal speed, especially in the absence of corresponding events from each modality. The human perceptual system can combine signals across modalities in an optimal fashion to form a fused percept (Alais and Burr 2004; Shams et al. 2005). In our case, though, the observed bias indicates that relevant and comparable signals have to be extracted from the spatiotemporal and temporal properties of the visual and auditory stimulation, respectively. This suggests a cross-modal spatiotemporal mapping, which we assume to occur not at the lower level of spatial or temporal properties per se, but more likely via the action pattern embedded in the visual motion and in the auditory tempo. A common platform for sensory-motor representation may exist (Prinz 1997) that is not necessarily sensory-modality specific, where spatial and temporal dynamics of all present information could interact. As mentioned before, both audiomotor and visuomotor transformations are found to implicate the premotor areas (Kohler et al. 2002; Lahav et al. 2007; Chen et al. 2009; Rizzolatti et al. 1996; Murata et al. 1997; Fadiga et al. 2000). Specifically, rather than direct mapping of sensory input onto executed action, the dorsal premotor area codes abstract, higher-order information as sensory *cues* for potential action planning without immediate indication of action (Chen et al. 2009): for example, it reacts to the metrical organization of an auditory sequence during passive listening (the premotor activities reflecting the extent of auditory information that could be *potentially* used to facilitate motor synchronization), or it conditionally associates certain visual cues with motor planning (Hoshi and Tanji 2006). Along this line, dorsal premotor area is also involved in forward motion perception, a manifest of visuomotor transformation in human locomotion (van der Hoorn et al. 2010) which was likely induced by our visual motion stimuli as well. We thus find it a plausible neural correlate here for integrating concurrent (albeit possibly abstract) motor information that can be extracted from different sensory modalities.

In addition, our interpretation of cross-modal interaction based on motor representations, at least in the rhythm domain, is consistent with the *audiovisuomotor mechanism* in rhythm perception proposed by Todd and colleagues (Todd et al. 2002). Quoted from Todd (1999, pp. 119), "… if the spatiotemporal forms of certain stimuli are matched to the dynamics of the motor system, then they may evoke a motion of an internal representation or motor image of the corresponding synergetic elements of the musculoskeletal system, even if the musculoskeletal system itself does not move". Our findings provide empirical evidence that visual input containing motion information can indeed elicit effective (internal) motor patterns that may be shared with the auditory rhythm, and it proved to affect the judgment of auditory tempo even in a simple isochronous sequence.

Concurrent entrainment to visual motion and auditory tempo

Although several studies have investigated the subjective time distortion in different visual motion conditions (Kanai et al. 2006; Kaneko and Murakami 2009; van Wassenhove et al. 2008; Wittmann et al. 2010) and reported consistently overestimation of time intervals presented by moving visual stimuli, we observed the opposite influence of visual motion speed on subjective auditory tempo. Another study on duration perception of different visual speeds (Matthews 2011) did find that the same interval occupied by accelerating visual objects was judged to last shorter than by decelerating ones, which seems to agree with our results. However, both accelerating and decelerating objects there were judged to last shorter than constant-speed ones, which was not observed here. Furthermore, the perceptual difference between acceleration and deceleration was not observed in subsequent reproduction tasks in the same study, while we found consistent bias across perceptual and reproduction tasks.

A critical difference exists between the present study and the aforementioned ones: while they targeted how different visual information affected the perception of a single duration, we were interested in the motion influence on the perceived tempo of an auditory sequence. Temporal perception in a rhythmic context may be explained not only by a simple interval timing mechanism but also by entrainment processes (McAuley and Jones 2003). Hearing the tempo of a simple isochronous sequence, as in our task, is an example of entraining to a given auditory pulse frequency (Merker et al. 2009). The neural oscillations underlying auditory entrainment, such as gamma and beta bands, have been indicated to link auditory and motor cortical activities in response to rhythmic stimuli (Large and Snyder 2009; Iversen et al. 2009). Similar oscillations have also been found as an attentional mechanism in visual perception (Fries et al. 2001; Lakatos et al. 2008). While auditory entrainment entails temporal periodicities in the stimuli, visual entrainment might be based on spatially defined periodicities: for example, a visual stimulus moving in an oscillatory manner (sinusoidally moving leftwards and rightwards) is found to entrain concurrent limb movement (Schmidt et al. 2007), providing evidence of sensory-motor coupling through visual entrainment to moving objects. Our visual motion did not possess oscillatory movement but nevertheless carried an underlying spatiotemporal regularity (the speed). We speculate that in our case, the entrainment to the auditory rhythm and the entrainment to the visual motion could be both linked to the neural resonance in the internal motor system, and the visually induced motor information—as relatively faster or slower speed—might reciprocally modulate the auditory temporal process (Iversen et al. 2009, p. 70), leading to the biased tempo perception.

Motion properties in visual stimuli

We presented stimuli of coherent global motion, which should engage a later stage (MT/V5) of visual motion processing (e.g., Kanai et al. 2004), where audiovisual motion integration is also found to take place (Alink et al. 2008; Scheef et al. 2009). However, the auditory motion in those studies consisted of sounds moving in space, and the (in)congruency with visual motion was mostly spatially based. As our auditory stimuli contained temporal instead of spatial properties, we find it unlikely (though not impossible) that the interaction should occur within the multimodal motion-specific area. This would still have to be verified with relevant imaging techniques.

Our motion stimuli did not resemble strictly those in the studies in which observers' forward motion and distinct percept of depth was induced by the enlarging objects moving along the radius (Kovács et al. 2008; van der Hoorn et al. 2010), which critically led to cortical activation in the dorsal premotor area. We did, however, intentionally 'prime' the participants during instruction that the motion stimuli they were going to receive might bring out a similar sensation as what they experience when sitting in a forward-moving car. All the participants reported immediately they could relate to this feeling while watching the stimuli. An internal viewpoint in the observer can enhance the dorsal premotor activity during action observation (Pilgramm et al. 2010). We believe our visual stimulation, by means of its motion properties and the observer's intention, has the potential to go through the premotor network of interest. Our result of bias in auditory tempo further supports this premise.

Attention allocation to auditory and visual stimulation

The temporal judgment of visual and auditory stimuli can be influenced by the differently weighted attention orientating to the presented events (Tse et al. 2004), which is also found in the context of rhythmic attending to the tempo (*Dynamic Attending Theory:* Jones and Boltz 1989; Drake et al. 2000). A recent study (Chapin et al. 2010) shows that, when presented with concurrent auditory and visual stimulation, the activities in motor areas of the brain that respond to the auditory rhythmic structure are enhanced by selective attention to the auditory stimuli. Possibly, in our case, the attention to the tones was to some extent diverted to the concurrent visual stimuli, leading to a lesser degree of motor activation in response to the auditory tempo and thus a 'capture' of auditory tempo by visual motion. However, it does not appear to accommodate the whole picture: in Experiment 1, although we asked the participants to attend to both visual and auditory stimuli, the task was only auditory. The listeners could have well chosen to ignore the task-irrelevant visual stimuli; yet there was an effect. Furthermore, the effect was specific to motion stimuli, and not to stimuli changing luminance. The visual influence here seems rather obligatory, and the resultant bias seems more attributable to a forced fusion of inter-sensory signals (van Wassenhove et al. 2008) rather than to an attentional strategy. In Experiment 2, though we did include a secondary task of visual judgment in addition to the auditory task, we did not find a trade-off between auditory and visual task performances; both tasks seemed equally attended to. Furthermore, the measured JND (4.6%) was even slightly lower than that found in Experiment 1 (5.2%), arguing against the possibility of interference (or lower attentional sources) due to the secondary task. Future studies differentially targeting the attentional source and the *direction* of the visual influence (faster or slower) will be needed to elucidate the role of attentional modulation in our investigated scenario.

In summary, the present study proved that dynamic visual-spatiotemporal information, such as in the speed of the motion flow, can influence the perception of concurrent non-spatial auditory tempo. The observed effect of 'visual spatial speed' on 'auditory temporal speed' reflects a cross-modal spatial–temporal integration, which likely occurs through the motor representation embedded in each stimulation. The findings suggest that the perceptual process in at least the auditory modality can incorporate relevant motor information from various sensory inputs to form a coherent experience.

Acknowledgments This work was supported by a doctoral scholarship to the first author from the Bayerische Forschungsstiftung, and the experimental expenses were additionally supported by the Andrea von Braun Stiftung. The authors thank Bruno Repp for very helpful comments on earlier versions of the manuscript.

References

Alais D, Burr D (2004) The ventriloquist effect results from near-optimal bimodal integration. Curr Biol 14:257–262

Alink A, Singer W, Muckli L (2008) Capture of auditory motion by vision is represented by an activation shift from auditory to visual motion cortex. J Neurosci 28:2690–2697

Bengtsson SL, Ullen F, Ehrsson HH, Hashimoto T, Kito T, Naito E, Forssberg H, Sadato N (2009) Listening to rhythms activates motor and premotor cortices. Cortex 45:62–71

Brainard DH (1997) The psychophysics toolbox. Spat Vis 10:433–436

Brandon M, Saffran JR (2011) Apparent motion enhances visual rhythm discrimination in infancy. Atten Percept Psychophys. doi:10.3758/s13414-011-0106-x

Buhusi C, Meck WH (2005) What makes us tick? Functional and neural mechanisms of interval timing. Nat Rev Neurosci 6:755–765

Burr D, Banks M, Morrone M (2009) Auditory dominance over vision in the perception of interval duration. Exp Brain Res 198:49–57

Chapin HL, Zanto T, Jantzen KJ, Kelso SJA, Steinberg F, Large EW (2010) Neural response to complex auditory rhythms: the role of attending. Front Psychol 1:1–18

Chen JL, Penhune VB, Zatorre RJ (2008) Listening to musical rhythms recruits motor regions of the brain. Cereb Cortex 18:2844–2854

Chen JL, Penhune VB, Zatorre RJ (2009) The role of auditory and premotor cortex in sensorimotor transformations. Ann NY Acad Sci 1169:15–34

Drake C, Botte MC (1993) Tempo sensitivity in auditory sequences: evidence for a multi-look model. Percept Psychophys 54: 277–286

Drake C, Jones MR, Baruch C (2000) The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. Cognition 77:251–288

Droit-Volet S, Wearden JH (2002) Speeding up an internal clock in children? Effects of visual flicker on subjective duration. Q J Exp Psychol B 55:193–211

Fadiga L, Fogassi L, Gallese V, Rizzolatti G (2000) Visuomotor neurons: ambiguity of the discharge or 'motor' perception? Int J Psychophysiol 35:165–177

Fendrich R, Corballis PM (2001) The temporal cross-capture of audition and vision. Percept Psychophys 63:719–725

Fraisse P (1963) Psychology of time. Harper & Row, New York

Fries P, Reynolds JH, Rorie AE, Desimone R (2001) Modulation of oscillatory neuronal synchronization by selective visual attention. Science 291:1560–1563

Grahn JA, Brett M (2007) Rhythm and beat perception in motor areas of the brain. J Cogn Neurosci 19:893–906

Grahn JA, Rowe JB (2009) Feeling the beat: premotor and striatal interactions in musicians and nonmusicians during beat perception. J Neurosci 29:7540–7548

Grahn JA, Henry MJ, McAuley JD (2010) FMRI investigation of cross-modal interactions in beat perception: audition primes vision, but not vice versa. Neuroimage. doi:10.1016/j.neuroimage.2010.09.033

Grondin S, McAuley D (2009) Duration discrimination in crossmodal sequences. Perception 38:1542–1559

Guttman SE, Gilroy LA, Blake R (2005) Hearing what the eyes see: auditory encoding of visual temporal sequences. Psychol Sci 16:228–265

Hoshi E, Tanji J (2006) Differential involvement of neurons in the dorsal and ventral premotor cortex during processing of visual signals for action planning. J Neurophysiol 95:3596–3616

Hove MJ, Spivey MJ, Krumhansl CL (2010) Compatibility of motion facilitates visuomotor synchronization. J Exp Psychol Hum Percept Perform 36:1525–1534

Iversen JR, Repp BH, Patel AD (2009) Top-down control of rhythm perception modulates early auditory responses. Ann NY Acad Sci 1169:58–73

Jones MR, Boltz M (1989) Dynamic attending and responses to time. Psychol Rev 96:459–491

Jones MR, McAuley JD (2005) Time judgments in global temporal contexts. Percept Psychophys 67:398–417

Kanai R, Paffen CL, Gerbino W, Verstraten FA (2004) Blindness to inconsistent local signals in motion transparency from oscillating dots. Vision Res 44:2207–2212

Kanai R, Paffen CLE, Hogendoorn H, Verstraten FAJ (2006) Time dilation in dynamic visual display. J Vis 9:1421–1430. doi: 10.1167/6.12.8

Kaneko S, Murakami I (2009) Perceived duration of visual motion increases with speed. J Vis 9:1–12. doi:10.1167/9.7.14

Karabanov A, Blom Ö, Forsman L, Ullén F (2009) The dorsal auditory pathway is involved in performance of both visual and auditory rhythms. Neuroimage 44:480–488

Kitagawa N, Ichihara S (2002) Hearing visual motion in depth. Nature 416:172–174

Kohler E, Keysers C, Umiltà MA, Fogassi L, Gallese V, Rizzolatti G (2002) Hearing sounds, understanding actions: action representation in mirror neurons. Science 297:846–848

Kornysheva K, von Cramon DY, Jacobsen T, Schubotz RI (2010) Tuning-into the beat: aesthetic appreciation of musical rhythms correlates with a premotor activity boost. Hum Brain Mapp 31:48–64

Kovács G, Raabe M, Greenlee MW (2008) Neural correlates of visually induced self-motion illusion in depth. Cereb Cortex 18:1779–1787

Lahav A, Saltzman E, Schlaug G (2007) Action representation of sound: audiomotor recognition network while listening to newly acquired actions. J Neurosci 27:308–314

Lakatos P, Karmos G, Mehta AD, Ulbert I, Schroeder CE (2008) Entrainment of neuronal oscillations as a mechanism of attentional selection. Science 320:110–113

Large EW (2008) Resonating to musical rhythm: theory and experiment. In: Grondin S (ed) The psychology of time. Emerald, UK, pp 189–231

Large EW, Snyder JS (2009) Pulse and meter as neural resonance. Ann NY Acad Sci 1169:46–57

Lee J, van Donkelaar P (2006) The human dorsal premotor cortex generates on-line error corrections during sensorimotor adaptation. J Neurosci 26:3330–3334

Lewald J, Guski R (2003) Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. Cogn Brain Res 16:468–478

Matthews WJ (2011) How do changes in speed affect the perception of duration? J Exp Psychol Human. doi:10.1037/a0022193

McAuley JD, Henry MJ (2010) Modality effects in rhythm processing: auditory encoding of visual rhythms is neither obligatory nor automatic. Atten Percept Psychophys 72:1377–1389

McAuley JD, Jones MR (2003) Modeling effects of rhythmic context on perceived duration: comparison of interval and entrainment approaches to short-interval timing. J Exp Psychol Hum 29:1102–1125

McAuley JD, Miller NS (2007) Effects of global temporal context on sensitivity to the tempo of auditory sequences. Percept Psychophys 69:709–718

Merker BH, Madison GS, Eckerdal P (2009) On the role and origin of isochrony in human rhythmic entrainment. Cortex 45:4–17

Miller NS, McAuley JD (2005) Tempo sensitivity in isochronous tone sequences: the multiple-look model revisited. Percept Psychophys 67:1150–1160

Morein-Zamir S, Soto-Faraco S, Kingston A (2003) Auditory capture of vision: examining temporal ventriloquism. Cogn Brain Res 17:154–163

Murata A, Fadiga L, Fogassi L, Gallese V, Raos V, Rizzolatti G (1997) Object representation in the ventral premotor cortex (area F5) of the monkey. J Neurophysiol 78:2226–2230

Patel AD, Iversen JR, Chen Y, Repp BH (2005) The influence of metricality and modality on synchronization with a beat. Exp Brain Res 163:226–238

Pilgramm S, Lorey B, Stark R, Munzert J, Vaitl D, Zentgraf K (2010) Differential activation of the lateral premotor cortex during action observation. BMC Neurosci 11:89. doi:10.1186/1471-2202-11-89

Prinz W (1997) Perception and action planning. Eur J Cogn Psychol 9:129–154

Recanzone GH (2003) Auditory influences on visual temporal rate perception. J Neurophysiol 89:1078–1093

Repp BH (2003) Rate limits in sensorimotor synchronization with auditory and visual sequences: the synchronization threshold and the benefits and costs of interval subdivision. J Mot Behav 35:355–370

Repp BH (2008) Metrical subdivision results in subjective slowing of the beat. Music Percept 26:19–39

Repp BH, Bruttomesso M (2009) A filled duration illusion in music: effects of metrical subdivision on the perception and production of beat tempo. Adv Cogn Psychol 5:114–134

Repp BH, Penel A (2002) Auditory dominance in temporal processing: new evidence from synchronization with simultaneous visual and auditory sequences. J Exp Psychol Hum Percept Perform 28:1085–1099

Rizzolatti G, Fadiga L, Gallesi V, Fogassi L (1996) Premotor cortex and the recognition of motor actions. Brain Res Cogn Brain Res 3:131–141

Scheef L, Boecker H, Daamen M, Fehse U, Landsberg MW, Granath DO, Mechling H, Effenberg AO (2009) Multimodal motion processing in area V5/MT: evidence from an artificial class of audio-visual events. Brain Res 1252:94–104

Schmidt RC, Richardson MJ, Arsenault C, Galantucci B (2007) Visual tracking and entrainment to an environmental rhythm. J Exp Psychol Human 33:860–870

Shams L, Kamitani Y, Shimojo S (2000) Illusions. What you see is what you hear. Nature 408:788

Shams L, Ma WJ, Beierholm U (2005) Sound-induced flash illusion as an optimal percept. Neuroreport 16:1923–1927

Shipley T (1964) Auditory flutter-driving of visual flicker. Science 145:1328–1330

Soto-Faraco S, Lyons J, Gazzaniga M, Spence C, Kingston A (2002) The ventriloquist in motion: illusory capture of dynamic information across sensory modalities. Cogn Brain Res 14: 139–146

Todd NPM (1999) Motion in music: a neurobiological perspective. Music Percept 17:115–126

Todd NPM, Lee CS, O'Boyle DJ (2002) A sensorimotor theory of temporal tracking and beat induction. Psychol Res 66:26–39

Treisman M, Faulkner A, Naish PLN, Brogan D (1990) The internal clock: evidence for a temporal oscillator underlying time perception with some estimates of its characteristic frequency. Perception 19:705–743

Tse PU, Intriligator J, Rivest J, Cavanagh P (2004) Attention and the subjective expansion of time. Percept Psychophys 66:1171–1189

van der Hoorn A, Beudel M, de Jong BM (2010) Interruption of visually perceived forward motion in depth evokes a cortical activation shift from spatial to intentional motor regions. Brain Res 1358:160–171

van Wassenhove V, Buonomano D, Shimojo S, Shams L (2008) Distortions of subjective time perception within and across senses. Plos One 3:e1437. doi:10.1371/journal.pone.0001437

Vierordt K (1868) Der Zeitsinn nach Versuchen. Tübingen, Laupp

Warren DH, Welch RB, McCarthy TJ (1981) The role of visual-auditory "compellingness" in the ventriloquism effect: implications for transitivity among the spatial senses. Percept Psychophys 30:557–564

Welch RB, Warren DH (1980) Immediate perceptual response to intersensory discrepancy. Psychol Bull 88:638–667

Wittmann M, van Wassenhove V, Craig AD, Paulus MP (2010) The neural substrates of subjective time dilation. Front Hum Neurosci 4. doi:10.3389/neuro.09.002.2010

4. REFERENCES

Arom, S. (1989). Time structure in the music of central Africa: periodicity, meter, rhythm and polyrhythmics. *Leonardo*, 22, 91–99.

Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H., Heinze, H. J., & Altenmüller, E. (2006). Shared networks for auditory and motor processing in professional pianists: evidence from fMRI conjunction. *Neuroimage*, 30, 917–926.

Bengtsson, S. L., Ullen, F., Ehrsson, H. H., Hashimoto, T., Kito, T., Naito, E., et al. (2009). Listening to rhythms activates motor and premotor cortices. *Cortex*, 45, 62–71.

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

Buhusi, C., & Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nature Reviews Neuroscience,* 6, 755–765.

Buzsaki, G. & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, 304, 1926–1929.

Chen, J. L., Zatorre, R. J., & Penhune, V. B (2006). Interactions between auditory and dorsal premotor cortex during synchronization to musical rhythms. *NeuroImage*, 32, 1771–1781.

Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008a). Moving on time: brain network for auditory–motor synchronization is modulated by rhythm complexity and musical training. *Journal of Cognitive Neuroscience*, 20, 226–239.

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

Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2009). The role of auditory and premotor cortex in sensorimotor transformations. *Annals of the New York Academy of Sciences*, 1169, 15–34.

Cooper, G., & Meyer, A. B. (1960). *The rhythmic structure of music*. London: The University of Chicago Press.

Corriveau, K. H., & Goswami, U. (2009). Rhythmic motor entrainment in children with speech and language impairments: Tapping to the beat. *Cortex*, 45, 119–130.

Diedrichsen, J., Ivry, R. B., & Pressing, J. (2003). Cerebellar and basal ganglia contributions to interval timing. In W. Meck (Ed.), *Functional and neural mechanisms of interval timing*. CRC Press LLC.

Drake, C. (1993). Reproduction of musical rhythms by children, adult musicians, and adult nonmusicians. *Perception & Psychophysics,* 53, 25–33.

Drake, C., & Botte, M. C. (1993). Tempo sensitivity in auditory sequences: evidence for a multi-look model. *Perception & Psychophysics,* 54, 277–286.

Drake, C., Jones, M. R., & Baruch, C. (2000). The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. *Cognition*, 77, 251–288.

Essens, P. J., & Povel, D. J. (1985). Metrical and nonmetrical representations of temporal patterns. *Perception & Psychophysics,* 37, 1–7.

Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (2000). Visuomotor neurons: ambiguity of the discharge or 'motor' perception? *International Journal of Psychophysiology*, 35, 165–177.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin Review*, 2006, 13, 361–377.

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

Grahn, J. A., & Brett, M. (2009). Impairment of beat-based rhythm discrimination in Parkinson's disease. *Cortex*, 45, 54–61.

Grahn, J. A., & Rowe, J. B. (2009). Feeling the beat: premotor and striatal interactions in musicians and nonmusicians during beat perception. *Journal of Neuroscience*, 29, 7540–7548.

Grube, M, Cooper, F. E., Chinnery, P. F., Griffiths, T. D. (2010). Dissociation of duration-based and beat-based auditory timing in cerebellar degeneration. *Proceedings of National Academy of Science USA*, 107, 11597–11601.

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, 92, 67–99.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8, 393–402.

Humble, M. (2002). The development of rhythmic organization in Indian classical music. Unpublished dissertation, University of London.

Iversen, J. R., Repp, B. H., & Patel, A. D. (2009). Top-down control of rhythm perception modulates early auditory responses. *Annals of New York Academy of Science,* 1169, 58–73.

Jackendoff, R., & Lerdahl, F. (2006). The capacity for music: What is it, and what's special about it? *Cognition*, 100, 33–72.

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

Keller, P. E., & Repp, B. H. (2005). Staying offbeat: Sensorimotor syncopation with structured and unstructured auditory sequences. *Psychological Research*, 69, 292–309.

Kohler, E., Keysers, C., Umilta, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science*, 297, 846–848.

Lahav, A., Saltzman, E., & Schlaug, G. (2007). Action representation of sound: audiomotor recognition network while listening to newly acquired actions. *Journal of Neuroscience*, 27, 308–314.

Lakatos, P., Shah, A.S., Knuth, K.H., Ulbert, I., Karmos, G., & Schroeder, C.E. (2005). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *Journal of Neurophysiology*, 94, 1904–1911.

Lakatos, P., Karmos, G., Mehta, A.D., Ulbert, I., & Schroeder, C.E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science*, 320, 110–113.

Large, E. W. (2008). Resonating to musical rhythm: theory and experiment. In S. Grondin (Ed.), *The psychology of time* (pp. 189–231). United Kingdom: Emerald.

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

Large, E. W., & Palmer, C. (2002). Perceiving temporal regularity in music. *Cognitive Science*, 26, 1–37.

Large, E. W., & Snyder, J. S. (2009). Pulse and meter as neural resonance. *Annals of New York Academy of Science*, 1169, 46–57.

Lerdahl, F., & Jackendoff, R. A. (1983). *Generative Theory of Tonal Music*. Cambridge: MIT Press.

Lewis, P. A., & Miall, R. C. (2003). Distinct systems for automatic and cognitively controlled time measurement: Evidence from neuroimaging. *Current Opinion in Neurobiology*, 13, 250–255.

Lim, I., van Wegen, E., de Goede, C., Deutekom, M., Nieuwboer, A., Willems, A., Jones, D., Rochester, L., & Kwakkel, G. (2005). Effects of external rhythmical cueing on gait in patients with Parkinson's disease: a systematic review. *Clinical Rehabilitation*, 19, 695–713.

London, J. (2002). Cognitive constraints on metric systems: some observations and hypotheses. *Music Perception*, 19, 529–550.

London, J. (2011). Tactus ≠ Tempo: Some Dissociations Between Attentional Focus, Motor Behavior, and Tempo Judgment. *Empirical Musicology Review*, 6, 43–55.

Macdougall, H. G., & Moore, S. T. (2005). Marching to the beat of the same drummer: the spontaneous tempo of human locomotion. *Journal of Applied Physiology*, 99, 1164–1173.

Madison, G., Gouyon, F., Ullén, F., & Hörnström, K. (in press). Modeling the tendency for music to induce movement in humans: First correlations with low-level audio descriptors across music genres. *Journal of Experimental Psychology: Human Perception and Performance*. doi: 10.1037/a0024323.

McAuley, J. D., & Jones, M. R. (2003). Modeling effects of rhythmic context on perceived duration: A comparison of interval and entrainment approaches to short-interval timing. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 1102–1125.

McKinney, M. F., & Moelants, D. (2006). Ambiguity in tempo perception: What draws listeners to different metrical levels? *Music Perception*, 24, 155–166.

Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology*, 17, 1692–1696.

Merker, B., Madison, G., & Eckerdal, P. (2009). On the role and origin of isochrony in human rhythmic entrainment. *Cortex*, 45, 4–17.

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

Moelants, D. (2002). Preferred tempo reconsidered. In C. Stevens, D. Burnham, G. McPherson, E. Schubert, & J. Renwick (Ed.), Proceedings of the 7th International Conference on Music Perception and Cognition (pp. 580–583). Sydney.

Monahan, C.B., & Hirsh, I. J. (1990). Studies in auditory timing: 2. Rhythm patterns. *Perception & Psychophysics*, 47, 227–242.

Murata, A., Fadiga, L., Fogassi, L., Gallese, V., Raos, V., & Rizzolatti, G. (1997). Object representation in the ventral premotor cortex (area F5) of the monkey. *Journal of Neurophysiology*, 78, 2226–2230.

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

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

Phillips-Silver, J., & Trainor, L. J. (2005). Feeling the beat: movement influences infant rhythm perception. *Science*, 308, 1430.

Phillips-Silver, J., & Trainor, L. J. (2007). Hearing what the body feels: auditory encoding of rhythmic movement. *Cognition*, 105, 533–546.

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

Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology*, 9, 129–154.

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

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

Repp, B. H. (2011). Comments on “Tactus ≠ Tempo: Some Dissociations Between Attentional Focus, Motor Behavior, and Tempo Judgment” by Justin London. *Empirical Musicology Review*, 6, 56–61.

Repp, B. H., & Doggett, R. (2006). Tapping to a very slow beat: a comparison of musicians and nonmusicians. *Music Perception*, 24, 367–376.

Repp, B. H., Iversen, J. R., & Patel, A. D. (2008). Tracking an imposed beat within a metrical grid. *Music Perception*, 26, 1–18.

Repp, B. H., Windsor, W. L., & Desain, P. (2002). Effects of tempo on the timing of simple musical rhythms. *Music Perception*, 19, 565–593.

Rizzolatti, G., Fadiga, L., Gallesi, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3, 131–141.

Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends in Cognitive Sciences*, 12, 106–113.

Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H., & Lakatos, P. (2010). Dynamics of Active Sensing and perceptual selection. *Current Opinion in Neurobiology*, 20, 172–176.

Snyder, J. S., & Large, E. W. (2005). Gamma-band activity reflects the metric structure of rhythmic tone sequences. *Cognitive Brain Research*, 24, 117–126.

Su, Y.-H., & Jonikaitis, D. (2011). Hearing the speed: visual motion biases the perception of auditory tempo. *Experimental Brain Research*, 214, 357–371.

Su, Y.-H., & Pöppel, E. (2012). Body movement enhances the extraction of temporal structures in auditory sequences. *Psychological Research*, 76, 373–382.

Teki, S., Grube, M., Kumar, S., & Griffiths, T. D. (2011). Distinct neural substrates of duration-based and beat-based auditory timing. *The Journal of Neuroscience*, 31, 3805–3812.

Thomson, J. M., & Goswami, U. (2008). Rhythmic processing in children with developmental dyslexia: auditory and motor rhythms link to reading and spelling. *Journal of Physiology – Paris*, 102, 120–129.

Todd, N. P. M. (1985). A model of expressive timing in tonal music. *Music Perception*, 3, 33–57.

Todd, N. P. M. (1999). Motion in music: a neurobiological perspective. *Music Perception*, 17, 115–126.

Todd, N. P. M., Lee, C. S., & O'Boyle, D. J. (2002). A sensorimotor theory of temporal tracking and beat induction. *Psychological Research*, 66, 26–39.

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

Will, U., & Berg, E. (2007). Brain wave synchronization and entrainment to periodic acoustic stimuli. *Neuroscience Letters*, 424, 55–60.

Zanto, P., Large, E. W., Fuchs, A., & Kelso, J. A. S. (2005). Gamma-band responses to perturbed auditory sequences: evidence for synchronization of perceptual processes. *Music Perception*, 22, 531–547.

Zanto, P., Snyder, J. S., & Large, E. W. (2006). Neural correlates of rhythmic expectancy. *Advances in Cognitive Psychology*, 2, 221–231.

Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: auditory–motor interactions in music perception and production. *Nature Reviews Neuroscience*, 8, 547-559.

5. ACKNOWLEDGEMENTS

My doctoral research has been supported by the Bayerische Forschungsstiftung (DOK 109-08) and the Andrea von Braun Stiftung, to both of whom I owe my utmost gratitude for making my scientific pursuit possible.

Many thanks, more than word can express, go to the following people:

my advisor, Prof. Dr. Ernst Pöppel, who gave me the opportunity to accomplish a doctoral project, trusted me to conduct independent research on a topic I am truly interested in, and encouraged me to achieve scientific publications;

a great mentor and friend Bruno H. Repp, who, besides giving me the most helpful input on my articles, has patiently answered each of my questions, from whom I have learned so much and with whom I have always exchanged wonderful discussions on our works;

my husband Daniel Su-Schroll, who has brought the whole world of fascinating rhythms into my life, and whose unwavering love has been of great support for me to continue in times of difficulty;

my parents and sister who, though far away during this time, have loved and supported me all the same, believed in my potentials in the chosen career, and whom I could never thank enough;

several wonderful peers who I cherish as friends and respect as scientists, who gave me great encouragement in the scientific pursuit, and from whom I have learned a lot along the way: Donatas Jonikaitis, Dragan Rangelov, Lana Kambeitz-Ilankovic, Tanja Radic, Marc Wittmann, and Hui Li.