Cognition 131 (2014) 330–344

Contents lists available at ScienceDirect

Cognition

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

Peak velocity as a cue in audiovisual synchrony perception of rhythmic stimuli

Yi-Huang Su *

Technical University of Munich, Germany

ARTICLE INFO

Article history:
Received 14 June 2013
Revised 5 February 2014
Accepted 14 February 2014
Available online 13 March 2014

Keywords:
Audiovisual synchrony perception
Rhythm perception
Biological motion
Point-light figure
Visual beat

ABSTRACT

This study investigated audiovisual synchrony perception in a rhythmic context, where the sound was not consequent upon the observed movement. Participants judged synchrony between a bouncing point-light figure and an auditory rhythm in two experiments. Two questions were of interest: (1) whether the reference in the visual movement, with which the auditory beat should coincide, relies on a position or a velocity cue; (2) whether the figure form and motion profile affect synchrony perception. Experiment 1 required synchrony judgment with regard to the same (lowest) position of the movement in four visual conditions: two figure forms (human or non-human) combined with two motion profiles (human or ball trajectory). Whereas figure form did not affect synchrony perception, the point of subjective simultaneity differed between the two motions, suggesting that participants adopted the peak velocity in each downward trajectory as their visual reference. Experiment 2 further demonstrated that, when judgment was required with regard to the highest position, the maximal synchrony response was considerably low for ball motion, which lacked a peak velocity in the upward trajectory. The finding of peak velocity as a cue parallels results of visuomotor synchronization tasks employing biological stimuli, suggesting that synchrony judgment with rhythmic motions relies on the perceived visual beat.

1. Introduction

In the course of multisensory perception, we often observe an action and hear the sounds produced by it (e.g., the footsteps of someone we see walking, or the music played by the musician we are watching). In some situations, however, we observe actions that are coordinated with extraneous sounds, as in dancing. Whereas multisensory perception in the former scenario has been extensively examined (Arrighi, Alais, & Burr, 2006; Petrini, Dahl, et al., 2009; Petrini, Russell, & Pollick, 2009; Saygin, Driver, & de Sa, 2008; van Wassenhove, Grant, & Poeppel, 2007), the perceptual mechanisms engaged by the latter scenario are less well understood. For example, how do we perceive whether an observed dancer is moving in synchrony with the music? Focusing on the visual part of the question, the present study investigated synchrony perception between rhythmic auditory and visual streams, and the effects of different visual parameters on this perception.

Judgment of audiovisual synchrony with continuous stimuli, such as a lip movement paired with a speech sound (van Wassenhove et al., 2007; Vatakis & Spence, 2006) or a drumming movement paired with an impact sound (Arrighi et al., 2006; Love, Petrini, Cheng, & Pollick, 2013; Petrini, Russell, et al., 2009), is found to be more difficult than for simple discrete stimuli, such as a flash paired

* Address: Department of Movement Science, Faculty of Sport and Health Science, Technical University of Munich, Georg-Brauchle-Ring 62, 80992 Munich, Germany.
E-mail address: yihuang.su@tum.de

http://dx.doi.org/10.1016/j.cognition.2014.02.004
0010-0277/

with a beep (Hirsh & Sherrick, 1961). This may be because humans are relatively insensitive to the physical asynchrony between two sensory streams when one or both consist of temporally and spatially varying features (Van Der Burg, Cass, Olivers, Theeuwes, & Alais, 2010; see Vroomen & Keetels, 2010, for a review). In a continuous movement such as drumming, the visual stream is more complex as it consists of spatial and temporal information, whereas the auditory stream contains mainly temporal events. Despite the visual stimulus complexity, synchrony judgment (SJ) in these cases is aided by cross-modal predictability: The trajectory of the movement predicts the temporal occurrence of the sound, which can be compared with its actual timing.

The advantage afforded by stimulus predictability is not restricted to audiovisual information that is causally linked. SJ with artificial bimodal stimuli (e.g., a sequence of tones and flashes, or a moving disc and a sound) can also be aided by the intra-modal predictability, in which either or both streams follow a regular temporal or spatiotemporal pattern (L. A. Cook, Van Valkenburg, & Badcock, 2011; Vroomen & Stekelenburg, 2010). Likewise, in a rhythmic human movement that accompanies rhythmic sounds, such as in dancing to the music, the periodicity of the movement corresponds to the periodicity of the sounds (Leman & Naveda, 2010). As such, both the temporal structure of the auditory events and the spatiotemporal structure of the observed movement are predictive of subsequent sounds and movements, due to their shared periodicity. For an observer/listener, the shared periodicity combined with synchrony gives rise to the percept of a coherent audiovisual rhythm (Nozaradan, Peretz, & Mouraux, 2011). The present study is concerned with audiovisual synchrony perception in such a rhythmic context, where SJ between the two streams must be based on the coincidence of the perceived 'reference points' – a recurrent event or feature that defines its periodicity – in each sensory rhythm.

Two main questions were of interest. First, what are the reference points in a rhythmic movement, with which the auditory reference points should coincide? The reference in an auditory rhythm is of a temporal nature, typically the perceived regular 'beat' that marks the most salient level of periodicity in the rhythm (Large, 2008). A continuous visual movement, however, contains varying temporal and spatial information, and the reference may be defined by two parameters in the trajectory. One is a distinct spatial position, such as the path reversal, that can be used to perceptually segment the movement (e.g., the lowest or highest position in a vertical movement). Alternatively, a salient spatiotemporal feature in the trajectory, such as the peak velocity or acceleration, might serve as the reference. This possibility has received support from visuomotor synchronization studies (Luck & Sloboda, 2008, 2009; Wöllner, Deconinck, Parkinson, Hove, & Keller, 2012). In these studies, when participants were instructed to synchronize finger taps to a point-light biological motion (a conductor's beating gesture), their taps tended to synchronize to the point of maximal vertical acceleration and, when the movement curvature was small, to the maximal instantaneous velocity in the trajectory. Luck & Sloboda proposed that points of peak velocity or acceleration are perceived as the visual 'beat' in a rhythmic movement. These considerations suggest two contrasting hypotheses: An auditory rhythm and an observed movement might be perceived as synchronous if the auditory beat is perceived to coincide with either (1) a defined position, or (2) a salient velocity feature, in the trajectory. The first aim of the present study was to identify which cue is taken as the visual reference in an audiovisual SJ task.

A second aim was to investigate whether human observers are more inclined to perceive synchrony when the moving figure resembles a human, andor when the movement follows human kinematics, than when either or both are not humanlike. This question was inspired by findings showing an intrinsic link between auditory rhythm perception and human movement, often manifested as people moving spontaneously to the beat of the rhythm (Phillips-Silver & Trainor, 2007; Su & Pöppel, 2012; Todd, 1999). Moreover, auditory rhythms yielding a clear sense of beat elicit a strong action representation in our motor system (for a review, see Section 4.1.2 in Repp & Su, 2013), of which the commonly activated area, premotor cortex (PMC), is also engaged in processing biological motion performed by a human figure (Saygin, 2007; Saygin, Wilson, Hagler, Bates, & Sereno, 2004; Stadler, Ott et al., 2012; van Kemenade, Muggleton, Walsh, & Saygin, 2012) and in integrating audiovisual information in a human motion (Wuerger et al., 2012). Compared to otherwise similar non-human stimuli, the appearance of a human agent (Saygin & Stadler, 2012), as well as human kinematics (Casile et al., 2010; Di Dio et al., 2013; Press, Cook, Blakemore, & Kilner, 2011; Stadler, Springer, Parkinson, & Prinz, 2012), is found to elicit a stronger action representation in the observer. As such, visual - stimuli suggestive of human appearance and kinematics may be perceived as more qualitatively congruent with an auditory rhythm, due to their common link to human actions. As content-related congruency between auditory and visual streams is associated with a lower sensitivity to their asynchronies (Petrini, Dahl, et al., 2009; van Wassenhove et al., 2007), sensitivity to audiovisual asynchronies was hypothesized to be lower for movements carried out by a human figure or following human kinematics, compared to non-human variants.

Two experiments were conducted in which participants judged synchrony between a vertically bouncing point-light figure (PLF) (Johansson, 1973) and an auditory rhythm. Experiment 1 intended to answer the two posed questions by manipulating the figure form (human or non-human) and the motion profile (human or non-human trajectory) of the visual stimulus in the SJ task. Experiment 2 further explored the effect of visual velocity and position cues on SJ by varying the position of the movement, for which SJ was required, as well as the motion profile (as in Experiment 1). In both experiments, the non-human trajectory was implemented as the figure moving along the same path of human trajectory but with a bouncing-ball velocity profile. This manipulation was chosen for two reasons: First, while the path and the periodicity were identical between the two motions, the ball trajectory had a different location of a potential velocity cue (i.e.,