



Metrical congruency and kinematic familiarity facilitate temporal binding between musical and dance rhythms

Yi-Huang Su¹

Published online: 15 May 2018
© Psychonomic Society, Inc. 2018

Abstract

Although music and dance are often experienced simultaneously, it is unclear what modulates their perceptual integration. This study investigated how two factors related to music–dance correspondences influenced audiovisual binding of their rhythms: the metrical match between the music and dance, and the kinematic familiarity of the dance movement. Participants watched a point-light figure dancing synchronously to a triple-meter rhythm that they heard in parallel, whereby the dance communicated a triple (congruent) or a duple (incongruent) visual meter. The movement was either the participant’s own or that of another participant. Participants attended to both streams while detecting a temporal perturbation in the auditory beat. The results showed lower sensitivity to the auditory deviant when the visual dance was metrically congruent to the auditory rhythm and when the movement was the participant’s own. This indicated stronger audiovisual binding and a more coherent bimodal rhythm in these conditions, thus making a slight auditory deviant less noticeable. Moreover, binding in the metrically incongruent condition involving self-generated visual stimuli was correlated with self-recognition of the movement, suggesting that action simulation mediates the perceived coherence between one’s own movement and a mismatching auditory rhythm. Overall, the mechanisms of rhythm perception and action simulation could inform the perceived compatibility between music and dance, thus modulating the temporal integration of these audiovisual stimuli.

Keyword Rhythm · Music · Dance · Audiovisual · Multisensory perception · Agency · Kinematics

Music and dance are inseparable with regard to their rhythm (Fitch, 2016). Humans often move spontaneously to music in ways that reflect the temporal structure of musical rhythm (Toiviainen, Luck, & Thompson, 2010)—for example, head nodding, foot tapping, or body bouncing periodically to a regular beat. In most dance repertoires, the choreographed movements also consist of patterns that temporally correspond to different metrical levels in music (Naveda & Leman, 2010). From a spectator’s perspective, we rarely observe dance without hearing the music to which the dancers move, and thus our experiences of music and dance are often coupled with each

other (Jola et al., 2013). Given these connections, surprisingly little is known about how the rhythms of music and dance interact in perception when experienced simultaneously. A few recent studies have suggested that, when presented concurrently in a synchronized manner, the visual rhythm of a street-dance-like bouncing movement can facilitate auditory beat perception (Su, 2014b) and that of a swing dance repertoire can impose metrical accents on auditory rhythms (Su, 2016b). Although these findings demonstrate visual modulation of auditory rhythm perception, it remains unclear what influences cross-modal binding between musical and dance rhythms, a question the present investigation was intended to address.

Our perceptual system binds co-occurring auditory and visual information on the basis of their temporal proximity (L. Chen & Vroomen, 2013) and contextual coherency (Nahorna, Berthommier, & Schwartz, 2012). In most scenarios, integration occurs when hearing the sounds produced by the observed action—for example, spoken words accompanied by lip movements (Nahorna et al., 2012) or instrumental sounds paired with musicians’ gestures (Chuen & Schutz, 2016). In

Electronic supplementary material The online version of this article (<https://doi.org/10.3758/s13423-018-1480-3>) contains supplementary material, which is available to authorized users.

✉ Yi-Huang Su
yihuang.su@tum.de

¹ Department of Movement Science, Faculty of Sport and Health Sciences, Technical University of Munich, Georg-Brauchle-Ring 62, 80992 Munich, Germany

some other situations, though, the brain binds concurrent audiovisual signals that are not causally linked, but rather are associated with each other by a perceived match in their content—referred to as *cross-modal correspondences* (Y.-C. Chen & Spence, 2017). The latter phenomenon involves cognitive mechanisms such as the belief that the two sources of information “belong together” (see also the ‘unity assumption’; Vatakis & Spence, 2008; Welch & Warren, 1980), and operates between multisensory inputs that tend to co-occur (Parise, Harrar, Ernst, & Spence, 2013). One example from speech research is that a speaker’s hand gestures (termed *beat gestures*) as visual signals can be perceptually integrated with the spoken sounds (Biau & Soto-Faraco, 2013). The (often rhythmic) hand movements mark points of semantic importance in the sentence, which modulates or even facilitates auditory speech processing. This investigation resembles the present study of music–dance integration, since both cases consider how a naturally occurring, largely rhythmic body movement may serve as visual cues to influence concurrent auditory perception.

Given different possible means of audio–visual mapping (Y.-C. Chen & Spence, 2017), what kinds of content compatibility could promote cross-modal binding between music and dance? The present research focused on two higher-level factors related to prominent features of the stimuli, one regarding their temporal structure and the other regarding the kinematics of the dance movement. For the former, because music and dance both have a defined rhythm, it seems reasonable that the match or mismatch in their rhythmic structures could affect whether they are perceived as a unity. A recent study showed that infants as young as 8–12 months could already discriminate whether the dance and the music matched in terms of synchrony (Hannon, Schachner, & Nave-Blodgett, 2017). Another study showed that, when adults observed a dance movement synchronized with the auditory rhythm, they could discriminate whether the *metrical accentuation* of the two matched each other (Su, 2017). Thus, congruency in rhythm—for example, synchronous music and dance communicating the same meter—may contribute to audiovisual binding. Next, nontemporal factors regarding the observed movement may also affect the perceived music–dance compatibility. For example, some movements appear more familiar than others because they contain kinematic features similar to those of the observer’s own motor style (Koul, Cavallo, Ansuini, & Becchio, 2016), and stimulus familiarity may increase the perceived audiovisual unity (Walker, Bruce, & O’Malley, 1995). Furthermore, we recently showed in a sensorimotor synchronization task that participants attempted more to temporally predict (or “simulate”) the rhythm of an observed dance movement if its kinematic style was more similar to their own (Su & Keller, 2018; see also the [supplementary material](#)). As such, it was hypothesized that dance movement containing more familiar kinematics—that is, one’s own

compared to another person’s movement—may be perceived as more compatible with the musical rhythm (due to facilitated simulation), which would in turn promote cross-modal binding.

The present study examined how these two factors influenced audiovisual temporal binding, using a paradigm developed by Su (2014a). In that study, participants observed a point-light figure (PLF; Johansson, 1973) bouncing periodically to an auditory rhythm of alternating downbeats and upbeats. The PLF bounced either downward (congruently) or upward (incongruently) with respect to the downbeat, and participants were less sensitive to a slightly perturbed auditory beat in the former condition. The result was interpreted as the congruent combination giving rise to stronger audiovisual binding (Nahorna et al., 2012), since sensitivity to cross-modal disparity is known to be inversely related to the strength of intersensory coupling (Parise et al., 2013). Besides, through audiovisual binding the visual stream may temporally “capture” the auditory deviant (i.e., shifting its perceived timing toward the visual stream, resembling a reversed “temporal ventriloquism”; Morein-Zamir, Soto-Faraco, & Kingstone, 2003), thus making a slight auditory temporal deviant less noticeable. In the present study, participants observed a PLF dancing synchronously to a triple-meter rhythm that they heard at the same time. The PLF danced in a triple (metrically congruent) or a duple (metrically incongruent) meter. On top of that, the observed dance was participants’ own movement (previously motion-captured) or the same movement performed by another participant. Participants attended to both the auditory and visual streams while detecting a temporal perturbation in one of the auditory beats. Following the logic of Su (2014a), if metrical match or kinematic familiarity promoted audiovisual binding, participants should be *less* sensitive to an auditory deviant in the metrically congruent condition or when the dance movement was their own.

Method

Participants

Nineteen young, healthy volunteers (14 females, five males; mean age = 25.9 years, $SD = 4.7$) participated. All were naïve to the study’s purpose, gave written informed consent prior to the experiment, and received an honorarium of €8 per hour. Twelve and 11 of the participants, respectively, had learned music (mean = 3.6 years, $SD = 3.9$) and dance (mean = 4.4 years, $SD = 5.3$; none in swing dance, however) as hobbies, amongst whom seven had trained in both. The study had been approved by the Ethics Commission of Technical University of Munich and was conducted in accordance with the ethical standards of the 1964 Declaration of Helsinki.

Stimuli and materials

Visual stimuli The visual stimuli consisted of a PLF performing the basic steps of the Charleston dance at a tempo of 150 beats per minute (interonset interval of 400 ms). As a departure from the authentic Charleston dance, the PLF performed these steps without moving its arms, with the hands placed upon the hips (see Su, 2016a). One cycle of the dance consisted of eight regular vertical bounces of the trunk (Pulses 1–8) and, in parallel, four lateral leg movements (Pulse 1, left leg backward; Pulse 3, left leg forward; Pulse 5, right leg forward; Pulse 7, right leg backward; see Fig. 1; note that the pulse number corresponded to the single frames illustrated in Fig. 1, whereas the beat number, denoted below the frames, referred to the visual “accents” communicated by the leg movement). The tempo of the legs was twice as slow as that of the bounce.

The PLF danced in two metrical variations (see the videos in the [supplementary material](#)). One was as described above, whose movement pattern fit a duple (4/4) musical meter. The other variation was modified such that one cycle of the dance corresponded to six bounces and three leg movements (i.e., without the last two bounces and the last leg movement of the former). This movement pattern fit a triple (3/4) musical meter (see Su, 2017).

For each participant, there were two “agent” conditions in the PLF stimuli: In one condition, the movement was the participant’s own, which had been recorded by a 3-D motion capture system (Qualisys Oqus, seven cameras at a sampling rate of 100 Hz, with 13 markers attached to the joints and the forehead) 4–5 months earlier, during which they had danced to a metronome. In the other, the movement was from another participant that was matched to each respective participant in terms of gender and kinematic parameters (i.e., peak velocity;

see the [supplementary material](#) for the procedure, and example videos can be found at <https://osf.io/t37pm/>). The sizes of the two PLF agents and the length of a single movement cycle from each agent were adjusted to be the same. Both the duple- and triple-meter movements had been performed and recorded across all participants.

The 3-D motion data were presented as point-light displays on a 2-D monitor. Every PLF was represented by 13 white discs against a black background, each of which subtended 0.4 deg of visual angle ($^{\circ}$). The whole PLF subtended approximately 5° (width) and 12° (height) when viewed at 80 cm. The PLF was displayed facing the observers, in a configuration as if the observers were watching from 20° left of its midline, which served to optimize depth perception of biological motion in a 2-D environment.

Auditory stimuli The auditory rhythm consisted of six isochronous discrete events with an interonset interval of 400 ms (i.e., the same tempo as the visual stimuli). The rhythm had a pattern of A–B–A–A–A–B, where A and B here were played by a snare drum and a high hat, respectively (similar to the stimuli in Phillips-Silver & Trainor, 2007). To avoid the rhythm being perceptually segmented in the same way across trials, two more variations were created by changing the starting phase of the rhythm: starting with the third event (A–A–A–B–A–B) or starting with the fifth event (A–B–A–B–A–A). The three rhythm patterns (not analyzed as a factor) occurred equally frequently across all the experimental conditions. All three variations were accentuated in a triple meter by a bass drum sound added on top of the rhythm. The accents always coincided with the first and fifth events (i.e., Beats 1 and 3 of the accented rhythm). The audio files can be found in the [supplementary material](#).

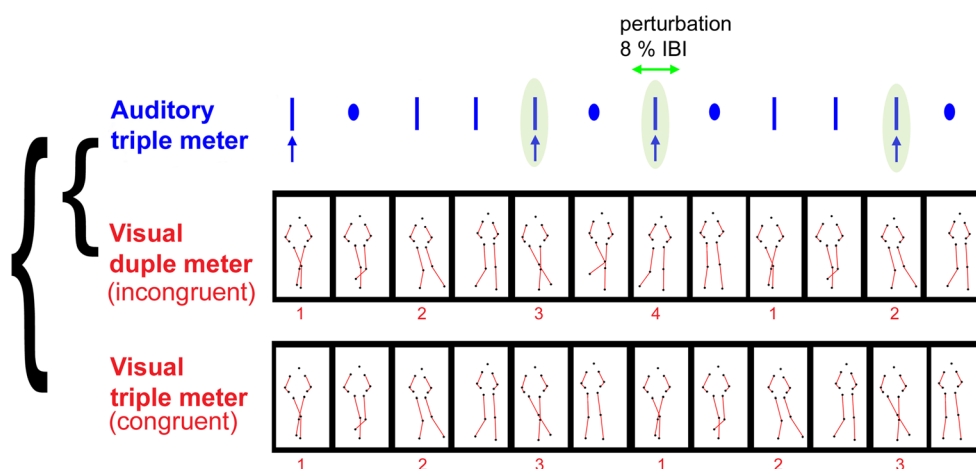


Fig. 1 Stimuli of the experiment. Two cycles of the auditory rhythm are shown, whereby the vertical bars and ovals at the top depict the snare drum and high-hat sounds, respectively. The arrows beneath the vertical bars depict the accents of an additional bass drum sound, occurring in the

Beat 1 and Beat 3 positions of the triple-meter rhythm. The numbers under the visual stimuli represent the beat count, derived from the leg movement for each metrical variation of the dance. Possible positions for an auditory perturbation are marked in the surrounding ovals at top

Note that, when the visual dance movement (both metrical versions) and the auditory rhythm were presented concurrently, they were synchronized at the pulse (bounce) level—that is, every bounce coincided with an auditory event. The leg movements communicated accents that could be grouped by threes (triple-meter version) or fours (duple-meter version), the former of which matched the grouping pattern of the auditory accents.

Procedure

The stimuli and experimental program were controlled by a customized Matlab (version R2012b; The MathWorks) script and Psychophysics Toolbox, version 3, routines (Brainard, 1997) running on a Mac OSX environment. The visual stimuli were displayed on a 17-in. CRT monitor (Fujitsu X178 P117A) with a frame frequency of 100 Hz at a spatial resolution of $1,024 \times 768$ pixels. Participants sat with a viewing distance of 80 cm, and sounds were delivered by closed studio headphones (AKG K271 MKII).

On each trial, participants listened to a rhythm while watching a PLF dance the Charleston basic steps simultaneously. The auditory rhythm was always accentuated in a triple meter. The PLF danced in a triple or duple meter that, when combined with the auditory rhythm, yielded metrically congruent or incongruent audiovisual rhythms, respectively. For each participant, the PLF movement was either their own or that of another participant. Each presentation contained two cycles of the auditory rhythm and the same length of visual dance synchronized with it. In half of the trials there was a perturbation in the auditory rhythm, such that one of the auditory events was delayed or advanced (equally often) by 8% of the interonset interval (32 ms). Within the 12 auditory events, the perturbation could occur in the fifth, seventh, or 11th event (equally often). Participants were instructed to listen to the rhythm while observing the dance and to detect whether one of the auditory events was temporally perturbed or not in a two-alternative forced choice (2AFC) task (Fig. 1). They made their response at the end of each trial.

To ensure participants' visual attention, a secondary task was imposed that required them to judge—after each response of perturbation detection—whether or not the auditory rhythm and the visual dance were synchronous (2AFC). Participants were explicitly instructed that perturbation detection was the more important task, and that the synchrony judgment should only be based on their subjective impression (as in Su, 2014a). Note that, although audiovisual synchrony was not manipulated in the stimuli, the synchrony response could additionally indicate whether the effects of interest were related to differences in the perceived synchrony (e.g., due to different kinematics between the agents).

The 2 (audiovisual metrical congruency¹) $\times 2$ (movement agent) $\times 2$ (auditory perturbation) conditions, each with 18 repetitions, were presented in six blocks of 24 trials each. All the conditions were balanced across blocks, and the order of conditions randomized within a block. Participants practiced eight trials before starting the experiment, which was completed in around 1 h.

Additional self-recognition task

To address whether task performance was related to each participant's ability to recognize their own visual stimulus (Keller, Knoblich, & Repp, 2007), an additional self-recognition task was conducted 1–2 months following the main experiment. A total of 16 participants returned to complete this task, in which they observed (without auditory stimuli) on each trial one cycle of their own movement or the movement of their matched "other," for each metrical variation separately. Participants answered whether or not the movement was their own (2AFC). Ten repetitions were included for each of the 2 (agent) $\times 2$ (metrical variation) conditions.

Data analysis

For the main experiment, sensitivity to the deviant (d') and the response criterion (C) were computed individually following signal detection theory (Stanislaw & Todorov, 1999). d' was calculated as the z -transformed hit rate minus the z -transformed false alarm rate, and a greater d' value indicated greater sensitivity to the target. C was the average of the z -transformed hit rate and the z -transformed false alarm rate, multiplied by -1 . A more positive C value indicated more conservative responses. For the self-recognition task, d' was also computed individually for each metrical variation in order to indicate sensitivity to one's own movement. Here, a hit was a correct answer of "Self" when seeing one's own movement, whereas a false alarm was an incorrect answer of "Self" when seeing the movement of the matched "other".

Results

Auditory perturbation detection

The 2 (audiovisual metrical congruency) $\times 2$ (movement agent) repeated measures analysis of variance (ANOVA) on d' revealed main effects of audiovisual metrical congruency, $F(1, 18) = 7.30, p = .015, \eta_p^2 = .29$, and of movement agent,

¹ Note that the same triple-meter auditory rhythms were paired with triple-meter (congruent) or duple-meter (incongruent) visual dance stimuli, and thus that audiovisual metrical congruency arose from the visual manipulation.

$F(1, 18) = 7.48, p = .014, \eta_p^2 = .29$, with no interaction between the two, $F(1, 18) = 0.27, p = .61, \eta_p^2 = .01$ (Fig. 2a). The results showed lower d' for the audiovisual congruent condition ($M = 1.00$ vs. 1.16), as well as for the condition of one's own movement ($M = 0.98$ vs. 1.19). The 2×2 repeated measures ANOVA on C did not identify any significant main effect or interaction, all p values $> .4$. Thus, the response criterion remained comparable across all the conditions (Fig. 2b).

Audiovisual synchrony response

The 2×2 repeated measures ANOVA on the individual percentages of synchrony response revealed only a main effect of audiovisual metrical congruency, $F(1, 18) = 7.48, p = .014, \eta_p^2 = .29$. It appeared that metrically congruent audiovisual rhythms were judged to be more synchronous than metrically incongruent ones. The movements of different agents did not affect the perceived synchrony, $F(1, 18) = 0.66, p = .43, \eta_p^2 = .04$ (Fig. 2c).

Correlation with self-recognition

Individual d' scores for the “self” condition of the perturbation detection task and those for the self-recognition task (both z -transformed) were entered in a Spearman's correlational analysis ($N = 16$), for each movement metrical variation separately. A significant negative correlation between the two measures was found for the duple-meter movement (corresponding to the audiovisual incongruent condition), $r_s = -.64, p = .008$, but not for the triple-meter movement (corresponding to the audiovisual congruent condition), $r_s = .15, p = .58$ (Fig. 3). Thus, for the metrically *incongruent* audiovisual rhythms involving one's own movement, lower sensitivity to the auditory perturbation was correlated with better self-recognition of the movement stimuli. Within the self-recognition task, a paired t test showed that the d' scores were comparable for both movement variations, $t(15) = 0.88, p = .39$.

Discussion

This study showed that participants were less sensitive to a slightly perturbed auditory beat when the observed dance was metrically congruent to the auditory rhythm, as well as when the movement was the participant's own, suggesting that in these conditions the auditory rhythm was more temporally bound to the visual one. Moreover, lower sensitivity to the auditory deviant in the metrically incongruent condition involving one's own movement—indicating greater audiovisual binding—was correlated with more successful self-recognition of the movement stimuli.

The effect of audiovisual metrical congruency on temporal binding was consistent with that obtained in Su (2014a) with a simpler metrical structure—a downbeat–upbeat paired with a down–up movement. The present results replicated and extended the finding to an ecological scenario encompassing more complex rhythms. Because the visual stimuli were essentially irrelevant to the auditory detection task, the effect of the visual meter argued for possibly automatic audiovisual binding on the basis of the perceived content match. Namely, music and dance stimuli that appear more rhythmically compatible (here, in terms of metrical accentuation) are more likely to be temporally integrated in perception. The binding mechanism may have reduced the saliency of a slight irregularity in one stream, due to a stable and coherent bimodal percept (Parise et al., 2013). In addition, the binding of audiovisual rhythms may enable the visual stream to temporally “capture” the auditory deviant, making it less distinct in perception. Similar findings have been reported in the speech domain, where semantically congruent auditory and visual syllables are more likely to be integrated at a greater temporal disparity (ten Oever, Sack, Wheat, Bien, & van Atteveldt, 2013). Conversely, a preceding incongruent audiovisual context (e.g., auditory syllables paired with a video of irrelevant spoken sentence) has been shown to reduce subsequent audiovisual binding (Nahoma et al., 2012). The present result seems consistent with such context effects in speech, thus

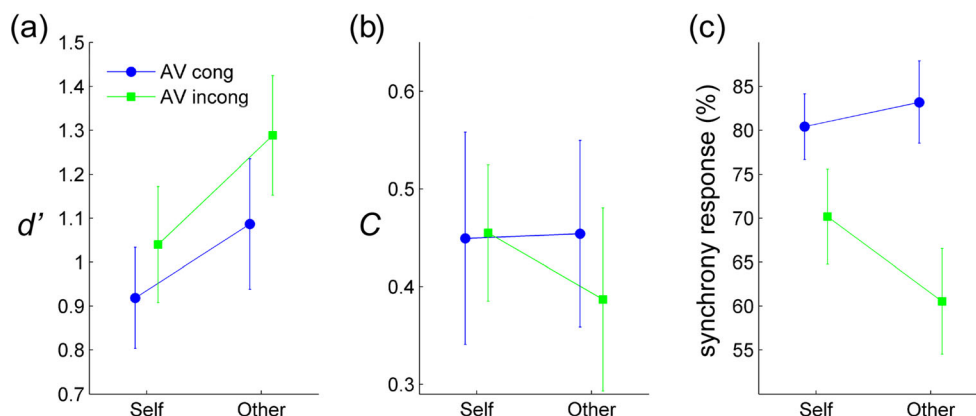


Fig. 2 Results of the main experiment, showing the mean data for (a) sensitivity to the deviant, d' ; (b) response criterion, C ; and (c) percentage of synchrony response. Error bars represent standard errors of the means

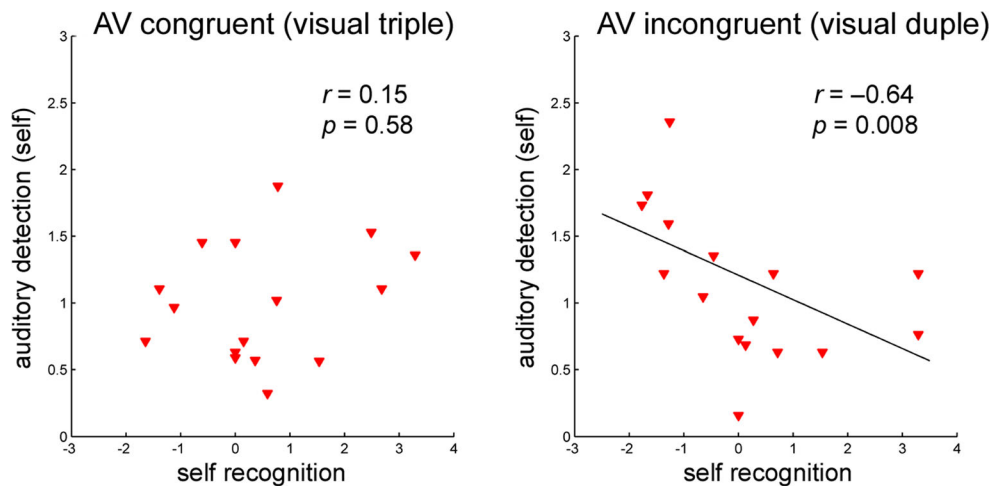


Fig. 3 Correlations between auditory perturbation detection for the self condition and self-recognition (both d' scores), for each movement metrical variation separately. Data points are z -scored. The black line represents the linear regression across data points whose correlation is significant

demonstrating similar mechanisms underlying the multisensory processing of speech and musical stimuli (Su, 2014b).

Regarding the effect of agency, observing one's own dance movements, which contained more familiar kinematic cues than those of others (Sevdalis & Keller, 2010), also promoted temporal binding between musical and dance rhythms. This result can be accommodated by the unity assumption (Y.-C. Chen & Spence, 2017)—namely, visual familiarity modulates the perceived fit between the auditory and visual streams (Walker et al., 1995; see also Hein et al., 2007, for findings related to object familiarity). Although different classes of stimuli—for example, speech as being more familiar than musical stimuli—had been previously suggested as a possible higher-level attribute that influences unity perception (Vatakis & Spence, 2008), the effect of stimulus familiarity within a specific class of musical stimuli—due to expertise—may be more pronounced when the differences between the matching and mismatching conditions are smaller (Chuen & Schutz, 2016). This seems consistent with the present effect of kinematic familiarity in the movement stimuli, which did not pertain to different classes of movement, but rather arose from each participant's long-term exposure to their own kinematic parameters. As such, the differences between familiar (self) and unfamiliar (other) stimuli arguably resided in lower-level characteristics than the differences between two categorically different movements. Moreover, dance observation likely engages covert motor simulation, especially as participants watched movements they had previously learned (Cross, Hamilton, & Grafton, 2006). Watching one's own dance movement (as compared to those of others) along with a musical rhythm should evoke greater motor resonance for simulation (Decety & Chaminade, 2003), which may in turn lead to greater perceived match in bimodal stimuli (i.e., “this fits how I would move to the rhythm” vs. “this is not so much how I would move to the rhythm”). As such, the perceived

compatibility in an audiovisual action can be based on the visual stimulus's similarity to each individual's motor style (Koul et al., 2016), which is a new finding that incorporates mechanisms of action–perception coupling into the familiarity aspect of multisensory integration. Future investigations could examine whether comparable results could also be achieved by manipulating the familiarity of auditory action—for example, when hearing self- versus other-generated rhythms (Keller et al., 2007).

At a group level, metrical congruency and movement agency did not interact with each other, suggesting that these two factors operated independently to modulate audiovisual binding. On an individual level, though, a correlation emerged between the degree of self-recognition and that of integration involving self-generated visual stimuli—notably, only for the metrically incongruent condition. This pattern was not related to different degrees of self-recognition between the two visual metrical variations, which were found to be comparable. The correlation showed that individuals who were better at recognizing their own movements also tended more to bind their own visual movement rhythm with an incongruent auditory rhythm. This result indicated that action simulation may underlie both the recognition of self-generated dance movement and the tendency to perceive as coherent a self-generated movement rhythm and a metrically mismatched auditory counterpart. One possible explanation is that there might be a general tendency (or preference) to perceive one's own movement as compatible with the auditory rhythm. Given metrically incompatible audiovisual combinations, which conflicted with this preference, the simulation mechanism ensured that those who were more sensitive to their own motion profiles were also more inclined to bind their own movement stimuli with the mismatching auditory rhythm.

In conclusion, two factors pertaining to cross-modal correspondences between music and dance were found to influence

temporal binding between music and dance across modalities: the match in their rhythms, and the kinematic familiarity of the movement. When the rhythms of music and dance are bound to form a coherent bimodal percept, a slight irregularity in one stream may become less noticeable. The present findings link accounts of rhythm perception and motor simulation with multisensory temporal integration, with implications for designing individually fitted multisensory stimulations to shape sensorimotor behaviors.

Author note This work and the author were supported by a grant from the German Research Foundation (DFG), SU782/1-2. The author thanks Anika Berg for data collection.

References

- Biau, E., & Soto-Faraco, S. (2013). Beat gestures modulate auditory integration in speech perception. *Brain and Language*, *124*, 143–152.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. <https://doi.org/10.1163/156856897X00357>
- Chen, L., & Vroomen, J. (2013). Intersensory binding across space and time: A tutorial review. *Attention, Perception, & Psychophysics*, *75*, 790–811. <https://doi.org/10.3758/s13414-013-0475-4>
- Chen, Y.-C., & Spence, C. (2017). Assessing the role of the “unity assumption” on multisensory integration: A review. *Frontiers in Psychology*, *8*, 445. <https://doi.org/10.3389/fpsyg.2017.00445>
- Chuen, L., & Schutz, M. (2016). The unity assumption facilitates cross-modal binding of musical, non-speech stimuli: The role of spectral and amplitude envelope cues. *Attention, Perception, & Psychophysics*, *78*, 1512–1528. <https://doi.org/10.3758/s13414-016-1088-5>
- Cross, E. S., de Hamilton, A. F. C., & Grafton, S. T. (2006). Building a motor simulation de novo: Observation of dance by dancers. *NeuroImage*, *31*, 1257–1267.
- Decety, J., & Chaminade, T. (2003). When the self represents the other: A new cognitive neuroscience view on psychological identification. *Consciousness and Cognition*, *12*, 577–596.
- Fitch, W. T. (2016). Dance, music, meter and groove: A forgotten partnership. *Frontiers in Human Neuroscience*, *10*, 64. <https://doi.org/10.3389/fnhum.2016.00064>
- Hannon, E. E., Schachner, A., & Nave-Blodgett, J. E. (2017). Babies know bad dancing when they see it: Older but not younger infants discriminate between synchronous and asynchronous audiovisual musical displays. *Journal of Experimental Child Psychology*, *159*, 1–16.
- Hein, G., Doehrmann, O., Müller, N. G., Kaiser, J., Muckli, L., & Naumer, M. J. (2007). Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *Journal of Neuroscience*, *27*, 7881–7887.
- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, *14*, 201–211. <https://doi.org/10.3758/BF03212378>
- Jola, C., McAleer, P., Grosbras, M.-H. L. N., Love, S. A., Morison, G., & Pollick, F. E. (2013). Uni- and multisensory brain areas are synchronised across spectators when watching unedited dance recordings. *i-Perception*, *4*, 265–284.
- Keller, P. E., Knoblich, G., & Repp, B. H. (2007). Pianists duet better when they play with themselves: On the possible role of action simulation in synchronization. *Consciousness and Cognition*, *16*, 102–111.
- Koul, A., Cavallo, A., Ansuini, C., & Becchio, C. (2016). Doing it your way: How individual movement styles affect action prediction. *PLoS ONE*, *11*, 165297:1–14. <https://doi.org/10.1371/journal.pone.0165297>
- Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research*, *17*, 154–163.
- Nahoma, O., Berthommier, F., & Schwartz, J.-L. (2012). Binding and unbinding the auditory and visual streams in the McGurk effect. *Journal of the Acoustical Society of America*, *132*, 1061–1077.
- Naveda, L., & Leman, M. (2010). The spatiotemporal representation of dance and music gestures using Topological Gesture Analysis (TGA). *Music Perception*, *28*, 93–111.
- Parise, C. V., Harrar, V., Ernst, M. O., & Spence, C. (2013). Cross-correlation between auditory and visual signals promotes multisensory integration. *Multisensory Research*, *26*, 307–316.
- Phillips-Silver, J., & Trainor, L. J. (2007). Hearing what the body feels: Auditory encoding of rhythmic movement. *Cognition*, *105*, 533–546.
- Sevdalis, V., & Keller, P. E. (2010). Cues for self-recognition in point-light displays of actions performed in synchrony with music. *Consciousness and Cognition*, *19*, 617–626.
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*, 137–149. <https://doi.org/10.3758/BF03207704>
- Su, Y.-H. (2014a). Content congruency and its interplay with temporal synchrony modulate integration between rhythmic audiovisual streams. *Frontiers in Integrative Neuroscience*, *8*, 92:1–13. <https://doi.org/10.3389/fmint.2014.00092>
- Su, Y.-H. (2014b). Visual enhancement of auditory beat perception across auditory interference levels. *Brain and Cognition*, *90*, 19–31. <https://doi.org/10.1016/j.bandc.2014.05.003>
- Su, Y.-H. (2016a). Sensorimotor synchronization with different metrical levels of point-light dance movements. *Frontiers in Human Neuroscience*, *10*, 186. <https://doi.org/10.3389/fnhum.2016.00186>
- Su, Y.-H. (2016b). Visual enhancement of illusory phenomenal accents in non-isochronous auditory rhythms. *PLoS ONE*, *11*, e0166880. <https://doi.org/10.1371/journal.pone.0166880>
- Su, Y.-H. (2017). *Rhythm of music seen through dance: Probing music–dance coupling by audiovisual meter perception*. Retrieved from [psyarxiv.com/ujkq9](https://arxiv.org/abs/1708.09999)
- Su, Y.-H., & Keller, P. E. (2018). *Your move or mine? Music training and kinematic compatibility modulate synchronization with self- versus other-generated dance movement*. *Psychological Research*. <https://doi.org/10.1007/s00426-018-0987-6>.
- ten Oever, S., Sack, A. T., Wheat, K. L., Bien, N., & van Atteveldt, N. (2013). Audio–visual onset differences are used to determine syllable identity for ambiguous audio–visual stimulus pairs. *Frontiers in Psychology*, *4*, 331:1–13. <https://doi.org/10.3389/fpsyg.2013.00331>
- Toiviainen, P., Luck, G., & Thompson, M. R. (2010). Embodied meter: Hierarchical eigenmodes in music-induced movement. *Music Perception*, *28*, 59–70.
- Vatakis, A., & Spence, C. (2008). Evaluating the influence of the “unity assumption” on the temporal perception of realistic audiovisual stimuli. *Acta Psychologica*, *127*, 12–23.
- Walker, S., Bruce, V., & O’Malley, C. (1995). Facial identity and facial speech processing: Familiar faces and voices in the McGurk effect. *Perception & Psychophysics*, *57*, 1124–1133.
- Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, *88*, 638–667. <https://doi.org/10.1037/0033-2909.88.3.638>