Attention, Perception, & Psychophysics
https://doi.org/10.3758/s13414-024-02874-x

Audiovisual integration of rhythm in musicians and dancers

Tram Nguyen[1] · Rebekka Lagacé-Cusiac[1] · J. Celina Everling[1] · Molly J. Henry[2,3] · Jessica A. Grahn[1]

Accepted: 23 February 2024

Abstract
Music training is associated with better beat processing in the auditory modality. However, it is unknown how rhythmic training that emphasizes visual rhythms, such as dance training, might affect beat processing, nor whether training effects in general are modality specific. Here we examined how music and dance training interacted with modality during audiovisual integration and synchronization to auditory and visual isochronous sequences. In two experiments, musicians, dancers, and controls completed an audiovisual integration task and an audiovisual target-distractor synchronization task using dynamic visual stimuli (a bouncing figure). The groups performed similarly on the audiovisual integration tasks (Experiments 1 and 2). However, in the finger-tapping synchronization task (Experiment 1), musicians were more influenced by auditory distractors when synchronizing to visual sequences, while dancers were more influenced by visual distractors when synchronizing to auditory sequences. When participants synchronized with whole-body movements instead of finger-tapping (Experiment 2), all groups were more influenced by the visual distractor than the auditory distractor. Taken together, this study highlights how training is associated with audiovisual processing, and how different types of visual rhythmic stimuli and different movements alter beat perception and production outcome measures. Implications for the modality appropriateness hypothesis are discussed.

Keywords Audiovisual integration · Beat perception and production · Bimodal target-distractor synchronization task · Sensorimotor synchronization · Modality appropriateness hypothesis

Introduction

Listening to auditory rhythms often gives rise to the perception of a beat, which is a series of regularly recurring, salient psychological events (Cooper & Meyer, 1960; Large & Palmer, 2002; Parncutt, 1994). The perception of a beat, especially in auditory rhythms, may compel people to automatically synchronize through overt or covert movements such as head-bobbing, foot-tapping, or hand clapping (Burger et al., 2013; Merker, Madison, & Eckerdal, 2009; Su & Pöppel, 2012). However, visual rhythms rarely compel people to move in the same way. In fact, beat perception and beat synchronization are generally superior for audition than for vision (Grahn et al., 2011; Grahn, 2012; Lorås et al., 2012; Patel et al., 2005). Likewise, when participants tap in synchrony with isochronous auditory (tones) and visual (flashes) sequences, asynchrony is lower for auditory, meaning that taps are more aligned with the beat, compared to visual sequences (Chen et al., 2002; Jäncke et al., 2000; Kato & Konishi, 2006; Repp, 2003b; Repp & Penel, 2002, 2004).

According to the *modality appropriateness hypothesis*, perception is biased to the sensory modality best suited for the task at hand: audition for temporal processing and vision for spatial processing (Lederman & Klatzky, 2004; Welch & Warren, 1980). This bias is observed when processing bimodal stimuli such as in a target-distractor task.

Public significance statement Prior rhythm training in either the music or the dance domains biases attention to either auditory or visual information. During a perceptual task, auditory stimuli were easier to attend to (and harder to ignore) than visual stimuli regardless of music or dance expertise. However, during a synchronization task using finger-tapping, dance training biased movement toward the visual modality and music training biased movement toward the auditory modality.

✉ Jessica A. Grahn
jgrahn@uwo.ca

1 Brain and Mind Institute and Department of Psychology, University of Western Ontario, London, Ontario, Canada

2 Max Planck Institute for Empirical Aesthetics, Frankfurt, Germany

3 Department of Psychology, Toronto Metropolitan University, Toronto, Ontario, Canada

Published online: 01 April 2024

For example, when isochronous auditory and visual rhythms, such as sequences of tones and flashes, are presented simultaneously, and the participant is to synchronize to only the tones (target) and ignore the flashes (distractor) or vice versa, the distractor effect is greater for auditory distractors than for visual distractors (Repp, 2003a, 2004; Repp & Penel, 2002, 2004). Thus, there is considerable evidence that temporal processing is biased toward the auditory modality relative to the visual modality.

However, recent research has suggested that visual beat perception and sensorimotor synchronization improves substantially if the stimuli are dynamic rather than static, such as a moving bar, a bouncing ball, or a bouncing point-light figure (Grahn, 2012; Hove et al., 2013; Hove & Keller, 2010; Hove et al., 2010; Su, 2014). For example, tap variability reduces when synchronizing with visual rhythms derived from apparent motion (i.e., a tapping finger) than static motion (i.e., a flashing light) (Hove & Keller, 2010). Moreover, tapping along to a bouncing ball (with a rectified sinusoidal velocity) reduces asynchrony compared to a flashing square (Iversen et al., 2015). In fact, synchronizing to a bouncing ball was equivalent to synchronizing to an auditory metronome, suggesting that dynamic stimuli enable better temporal prediction (Hove & Keller, 2010). Similarly, in bimodal distractor paradigms, dynamic visual stimuli, as opposed to static stimuli, begin to approach auditory stimuli in terms of distractor effects (Hove et al., 2013). These results somewhat contradict the modality appropriateness hypothesis.

Distractor effects can also be magnified by expertise or training that is concentrated in one modality. Musically trained individuals may be biased towards the auditory modality because music is defined by auditory rhythms (Repp & Penel, 2004). For example, Hove et al. (2013) found that, when a bouncing ball was pitted against an auditory metronome in a bimodal target-distractor synchronization task, the bouncing ball was more distracting than the metronome for visual experts (video gamers and ball players). For auditory experts (musicians), the metronome was more distracting than the bouncing ball. However, unlike musicians who have experience synchronizing to auditory rhythms, video gamers and ball players do not synchronize to visual rhythms. A more comparable visual expert population to musicians may be dancers, as both focus on synchronization with auditory and visual rhythms. While both musicians and dancers may attend to auditory and visual cues, each music and dance training differentially focus on fine-tuning motor actions in response to each modality: auditory for musicians and visual/proprioceptive for dancers (Ladda et al., 2020).

In addition, previous studies using the bimodal target-distractor synchronization paradigm have assessed sensorimotor synchronization with finger tapping, and often find a bias towards the auditory modality (Chen et al., 2002; Repp & Penel, 2002, 2004). Though sensorimotor synchronization is an important skill possessed by both musicians and dancers, the movements that musicians and dancers use to optimally synchronize with an external rhythm or beat may differ considerably. For example, musicians often rely on discrete effector-specific movements to produce music, whereas dancers often rely on gross whole-body movements to perform choreography (Karpati et al., 2016). Musicians show advantages in hand and finger movements compared to non-musicians (Fernandes & de Barros, 2012; Inui & Ichihara, 2001; Verheul & Geuze, 2004), whereas dancers show advantages in upper and lower limb movements compared to non-dancers (Buchanan et al., 2007; Sofianidis et al., 2012; Thullier & Moufti, 2004). Moreover, musicians tend to outperform dancers on tasks that involve effector-specific movements, but dancers tend to outperform or perform equally to musicians on tasks that involve whole-body movements (Karpati et al., 2016; Nguyen et al., 2022). Therefore, music and dance training may be comparable with regard to sensorimotor synchronization performance when dancers are tested with movements that are ecologically valid with respect to their training.

The current study

According to the modality appropriateness hypothesis, synchronization in a bimodal distractor task should be more biased by auditory distractors than by visual distractors. However, recent findings show this hypothesis may not hold in all situations. To our knowledge, no work has examined whether musicians and dancers show different modality biases when perceiving and synchronizing to competing auditory and visual rhythms using dynamic visual stimuli. Moreover, the bimodal target-distractor synchronization paradigm has not been used with whole-body sensorimotor synchronization. To address these gaps, we conducted two experiments to examine how music and dance training influence audiovisual integration and synchronization. In Experiment 1, audiovisual integration was measured using a variant of the "flash-beep" task (de Boer-Schellekens et al., 2013; Fiedler et al., 2011; Innes-Brown et al., 2011). Participants were presented with short audiovisual clips that paired a single auditory tone with the bouncing stick figure and judged whether the audio and video were synchronized. Audiovisual synchronization was measured using a variant of the bimodal target-distractor synchronization task (Chen et al., 2002; Repp & Penel, 2002, 2004). Participants tapped in synchrony with an isochronous auditory or visual (the bouncing stick figure) target sequence while a distractor sequence was presented in the other modality at one of nine temporal offsets. We tested how much the distractor sequences altered tapping synchrony to the target sequence.

In both the audiovisual integration and the synchronization tasks, a bouncing stick figure was used as a dynamic visual rhythm. Like the bouncing ball, the bouncing stick figure has a continuous motion, which consisted of a repetitive knee-bending motion generated from a dancer's bouncing trajectory. Thus, the movement was both continuous and biologically valid. Experiment 2 replicated Experiment 1, except that in the bimodal target-distractor synchronization task, participants bounced (instead of tapping their finger) in synchrony with the target sequence while regarding the other modality as a distractor. This whole-body movement, compared to tapping, may be more similar to the movements that dancers train on, thus music and dance expertise may influence task performance differently for bouncing than tapping.

Table 1 Participant characteristics for Experiment 1

	Musicians	Dancers	Controls
Sex	14 females 6 males	19 females 1 male	14 females 6 males
Age range, y	19 to 26	19 to 30	18 to 29
Age, y (M ± SD)	21.00 ± 1.97	22.25 ± 3.19	22.80 ± 3.76
Music training, y (M ± SD)	13.05 ± 3.47	0.45 ± 1.15	0.42 ± 1.18
Dance training, y (M ± SD)	0.20 ± 0.70	12.95 ± 4.26	0.30 ± 0.80
Starting age, y (M ± SD)	7.95 ± 2.98	8.35 ± 3.84	NA
Weekly practice, h (M ± SD)	2.60 ± 1.93	4.15 ± 3.72	NA

Experiment 1

The objective of Experiment 1 was to examine whether musicians and dancers show differential biases to the auditory and visual modalities, and how that affects audiovisual integration (perception) and synchronization (production). To understand the interaction between expertise and modality on audiovisual integration and synchronization, we tested musicians, dancers, and non-musician/non-dancer controls. For the audiovisual integration task, we expected musicians and dancers to outperform the controls, as evidenced by smaller windows of perceived simultaneity. Specifically, musicians and dancers were expected to detect asynchrony for smaller asynchronies than controls. For the audiovisual synchronization task, we predicted that musicians would be more influenced by the auditory modality because of their experience with auditory rhythms, whereas dancers would be more influenced by the visual modality, or at the least have a smaller auditory effect, because of their experience with visual rhythms. We predicted that controls would be more influenced by the auditory modality given the general bias towards the auditory modality on temporal tasks, but to a lesser degree than musicians (Welch & Warren, 1980).

Methods

Participants

Three groups of participants were tested: musicians, dancers, and non-musician/non-dancer controls. There were 20 participants in each group, for a total of 60. Age ranged between 18 and 30 years (M = 22.02 years, SD = 3.11 years). Table 1 summarizes the demographic characteristics of the sample. To be classified as a musician or a dancer, an individual needed at least 5 cumulative years of formal training in either music or dance, and to be currently playing or dancing. Individuals with both music and dance training that exceeded 5 years were excluded. Controls required less than 5 years of formal training in music and dance. All participants reported normal hearing and normal or corrected-to-normal vision. Participants received either two research credits or $20.00 (CAD) for their participation. All participants provided informed consent in accordance with the guidelines approved by the University of Western Ontario Non-Medical Research Ethics Board.

Procedure

Audiovisual integration was tested using a task analogous to the "flash-beep" paradigm (de Boer-Schellekens et al., 2013; Fiedler et al., 2011; Innes-Brown et al., 2011), while audiovisual synchronization was tested using a variant of the bimodal target-distractor synchronization task (Chen et al., 2002; Repp & Penel, 2002, 2004). The tasks were administered on a MacBook Pro using Psychtoolbox (Kleiner et al., 2007) in MATLAB R2016a (The MathWorks Inc., Natick, MA, USA). All auditory stimuli for the tasks were delivered through Sennheiser HD 280 headphones at a comfortable volume. All participants completed the audiovisual integration task followed by the audiovisual synchronization task in one session. The entire session took approximately 2 h. All participants were fully debriefed after the study.

"Bounce-beep" integration task

Participants watched a bouncing stick figure video (Fig. 1), during which a single auditory tone was presented (500 Hz, 10 ms long, onset and offset ramps of 5 ms), to determine the threshold of their audiovisual integration. Participants judged whether the auditory tone and the bounce of the stick figure occurred simultaneously, that is, whether the beep occurred when the stick figure was at the bottom of the bounce (lowest point, or the deepest knee-bend). The

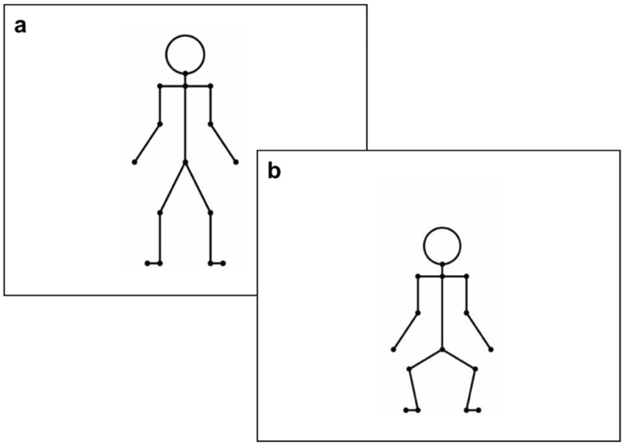

Fig. 1 Visual representation of the bouncing stick figure. (**a**) The stick figure in the most upright position. (**b**) The stick figure in the lowest, most bent-knee position

bouncing stick figure was programmed in MATLAB R2014a (The MathWorks Inc., Natick, MA, USA). The degree of asynchrony between timing of the bottom of the bounce and the onset of beep was altered from trial to trial using an adaptive tracking procedure with four separate tracks. Trials for each track were randomly interleaved. Two tracks were audio-leading: the first trial started with the audio in synchrony with the video (beeps aligned with bottom of bounces), with asynchrony increased by advancing the audio 5% per trial until participants responded that the beeps and bounces were no longer in synchrony [A0], then decreasing the audio advance by 5% until participants responded that beeps and bounces were in synchrony, and so on, until 12 reversals were made. The second track started with the audio advanced by 50%, relative to the video [A50], with asynchrony decreasing in 5% steps on each trial until participants indicated synchrony, then increasing, again until 12 reversals were made. The other two tracks were comparable, but with the video leading, with the first track starting in perfect synchrony [V0], and the second track starting with the video advanced by 50% [V50]. All responses were made on the laptop keyboard.

Bimodal target-distractor synchronization task

Participants tapped in synchrony with a target isochronous auditory sequence (500 Hz, 10-ms tones, with onset and offset ramps of 5 ms) or a visual sequence (bouncing stick figure) while a distractor sequence was presented in the other modality at one of nine temporal offsets. Both target and distractor sequences consisted of 32 events with an inter-onset interval (IOI) of 625 ms. The nine temporal phase displacements between the target and distractor sequences ranged from –50% to +50% of the IOI: 0%, ± 12.5%, ± 25%, ± 37.5%, ± 50%. The task consisted of 180 trials; each of the nine temporal offsets occurred with the auditory sequence as the distractor and with the visual sequence as the distractor, and repeated ten times each. The order of the 180 trials was randomized for each participant and broken into five blocks of 36 trials to prevent fatigue. Participants were instructed to tap in synchrony with the target sequence on the spacebar of a laptop keyboard, starting at the third event of each target sequence, and to continue until the end of the trial, while ignoring the distractor sequence. Although greater temporal precision may have been possible with a response box or MIDI drum pad, the additional variability of the keyboard introduces no bias, thus does not affect the primary outcome of interest (i.e., distractor effects explained below).

To ensure that participants did not deliberately close their eyes or look away from the video in the auditory target-visual distractor (A-V) conditions, they were required to report at the end of each trial whether the joints of the bouncing stick figure had briefly changed colour (from black to red). This attention check occurred on half of the A-V trials and was randomized to occur between events 13 and 21. Participants also performed a similar attention check for the visual target-auditory distractor (V-A) conditions, reporting whether an auditory tone changed in pitch (from 500 Hz to 700 Hz), which was randomized to occur between event 13 and 21. Half of the V-A conditions contained a pitch change.

Statistical analyses

For all analyses, follow-up pairwise t-tests were conducted where appropriate to determine the nature of any interactions. All hypothesis tests used $\alpha = .05$ for significance. The main analyses pertaining to the experiment were supplemented with Bayesian analyses (i.e., repeated-measures Bayesian ANOVA and Bayesian t-test) to quantify the amount of evidence for each effect. Bayes factors for the inclusion of a given effect (BF_{inclu}) across matched models are reported, rather than the BF for a given model. In this context, BFs equal to 3, 10, 30, and 100 are associated with moderate, strong, very strong, and extreme evidence for the effect of a predictor (Faulkenberry et al., 2020). BFs below 1 are interpreted as evidence against the effect of a predictor. Demographic information was analyzed with SPSS (23.0) software, frequentist analyses were conducted in R (version 4.2.2), and Bayesian analyses were conducted in JASP (version 0.17.3).

Group demographics

First, a one-way ANOVA (*group*: musicians, dancers, and non-musician/non-dancer controls) was conducted on years of music and dance training. Separate t-tests comparing musicians and dancers were also conducted on years of training, starting age of training, and hours of weekly training to ensure that the three groups only differed on expertise type (music vs. dance) and not quantity of training.

"Bounce-beep" integration accuracy

Performance on the "bounce-beep" integration task was analyzed by averaging the percentage of asynchrony at the last four reversals for all four tracks to obtain the audiovisual integration threshold. The audiovisual integration thresholds for the A0 and A50 tracks were then averaged to get the audiovisual integration threshold value for the audio-leading tracks. Similarly, the audiovisual integration thresholds for the V0 and V50 tracks were averaged to get the audiovisual integration threshold value for the video-leading tracks. A 3 (*group*: musicians, dancers, and controls) × 2 (*track*: audio-leading and video-leading) mixed ANOVA with group as the between-subjects variable and track as the within-subject variable was conducted to assess the interaction between group and track on audiovisual integration thresholds.

Bimodal target-distractor synchronization task

Attention check accuracy For the target-distractor synchronization task, the percentage of correct responses was calculated for detection of colour and pitch changes in the distractor sequences. Participants scoring below 85% accuracy were excluded to ensure that participants included in the final analyses were attending to the distractor sequences while synchronizing to the target sequences. A one-way ANOVA (*group*: musicians, dancers, and controls) was conducted to assess differences in attention check accuracy for colour and pitch changes in the distractor sequences.

Relative asynchrony Relative asynchrony measures the participant's accuracy in synchronizing with the target sequence while ignoring the distractor sequence. To measure relative asynchrony, the mean difference between each tap time and the nearest time in the target sequence was calculated and divided by the mean IOI (625 ms; see Eq. 1). The relative asynchrony values were then averaged across ten trials to obtain a single relative asynchrony score for each participant at each of the nine temporal offsets.

$$\text{RELATIVE ASYNCHRONY} = \frac{\text{MEAN}_{(\text{TAP-TARGET})}}{\text{MEAN}_{\text{IOI}}} \quad (1)$$

A distractor effect is measured by calculating the change in relative asynchrony between responses made at different offsets of the distractor sequence. If a distractor effect was present, responses should occur earlier than the target when the distractor preceded the target, resulting in greater negative asynchrony scores relative to trials in which the target and distractor were in synchrony. When the distractor followed the target, responses should occur later than the target, resulting in greater positive asynchrony scores relative to trials in which the target and distractor were in synchrony.

For each participant, the magnitude of the distractor effect was calculated as the difference between the maximum and minimum mean relative asynchrony scores obtained from the nine mean relative asynchrony scores (one score for each offset). The magnitude of the distractor effect for each group was analyzed with a 3 (*group*: musicians, dancers, and controls) × 2 (*distractor modality*: audition and vision) mixed ANOVA with group as the between-subjects variable and distractor modality as the within-subject variable to assess whether the distractor effect differed between groups and modalities.

Power analysis

Results from a power analysis indicated that a sample size of 60 participants was sufficient to detect with a power of .8 a within-between subject interaction with a large effect size ($\eta_p^2 = 0.15$). The power analysis was conducted using G*Power (version 3.1.9.4; Faul et al., 2007, 2009). In addition, a simulation-based power analysis was conducted in R with the hypothetical values reflecting expected effects, and showed that a within-between subject interaction could be detected with a power of .8 with 60 participants (Lakens, 2022).

Results

Group demographics

One-way ANOVAs conducted on years of music and dance training revealed expected significant differences for both music, $F(2, 57) = 215.48$, $p < .001$, $\eta_p^2 = .88$, and dance training, $F(2, 57) = 167.31$, $p < .001$, $\eta_p^2 = .85$, between the groups. Post hoc comparisons confirmed that musicians and controls significantly differed in years of music training, $t(38) = 15.39$, $p < .001$. Musicians and dancers also significantly differed, $t(38) = 15.42$, $p < .001$. However, there were no significant differences in years of music training between controls and dancers, $t(38) = .07$, $p = .95$. Similarly, post hoc comparisons confirmed that dancers and controls significantly differed in years of dance training, $t(38) = 13.05$, $p < .001$. Dancers and musicians also significantly differed, $t(38) = 13.21$, $p < .001$, but there were no significant differences in years of dance training between controls and musicians, $t(38) = .42$, $p = .68$.

Independent-samples t-tests between musicians and dancers were conducted on years of training, starting age of training, and hours of practice per week in their respective expertise. Musicians and dancers did not significantly differ in the years of training in their respective expertise, $t(38) = .1$, $p = .94$, $\eta_p^2 = .0002$, or starting age, $t(38) = .37$, $p = .72$, $\eta_p^2 = .004$. Finally, musicians and dancers did not significantly differ on the number of hours they practiced per week, $t(38) = 1.66$, $p = .11$, $\eta_p^2 = .07$.

"Bounce-beep" integration task

Performance on the "bounce-beep" integration task did not significantly differ between groups, $F(2, 57) = .81$, $p = .45$, $\eta_p^2 = .03$, $BF_{inclu} = 0.17$. Musicians ($M = 24.59\%$, $SE = 2.05\%$), dancers ($M = 26.24\%$, $SE = 1.55\%$), and controls ($M = 27.68\%$, $SE = 1.85\%$) all performed similarly on the task. However, performance did significantly differ for track type, $F(1, 57) = 8.34$, $p = .005$, $\eta_p^2 = .13$, $BF_{inclu} = 19.32$. For all groups, the window of perceived simultaneity was smaller when the audio led the video ($M = 23.11\%$, $SE = 1.55\%$) compared to when the video led the audio ($M = 29.23\%$, $SE = 1.32\%$), suggesting that there is a larger window of audio-visual integration when vision leads audition. The interaction between group and track type was not significant, $F(2, 57) = .33$, $p = .72$, $\eta_p^2 = .01$, $BF_{inclu} = 0.18$ (Fig. 2).

Bimodal target-distractor synchronization task

Attention check accuracy No participants were excluded as all scored at least 85%. Average response accuracy was 94.9%. Accuracy did not significantly differ between groups, $F(2, 57) = 1.05$, $p = .36$, $\eta_p^2 = .04$ (musicians: $M = 94\%$, $SE = 9.72\%$; dancers: $M = 96\%$, $SE = 9.43\%$; controls: $M = 95\%$, $SE = 8.00\%$). Thus, participants were still attending to the distractor sequences while synchronizing to the target sequences.

Relative asynchronies

Comparison between Groups The 3 × 2 mixed ANOVA with group as the between-subjects variable and distractor modality as the within-subject variable on the magnitude of the distractor effect produced a significant main effect of group, $F(2, 57) = 4.38$, $p = .02$, $\eta_p^2 = .13$, $BF_{inclu} = 1.4$. Interpretation of this main effect is qualified by a significant interaction with modality, detailed below. Follow-up comparisons between groups revealed that the magnitude of the distractor effect was greatest for musicians ($M = 0.20$, $SE = 0.02$), then controls ($M = 0.16$, $SE = 0.01$), then dancers ($M = 0.13$, $SE = 0.01$; Fig. 3). Note that distractor effects can also be visualized in Fig. 4 as the difference between the highest and lowest mean relative asynchrony within each subplot. Group differences were only significant between musicians and dancers, $t(38) = 2.83$, $p = .007$, $BF = 4.62$,

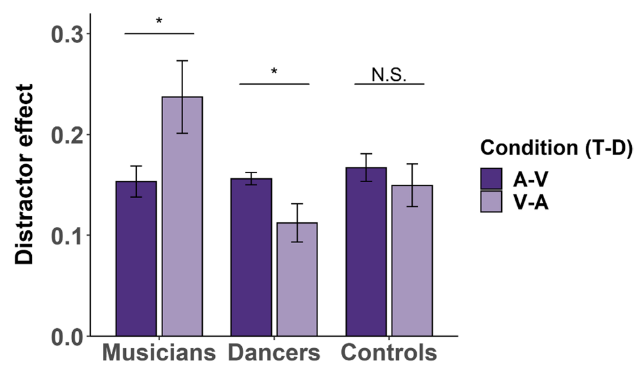

Fig. 3 Distractor effects of group and condition when tapping in the bimodal distractor synchronization task. Distractor effects were quantified as the difference between maximum and minimum relative asynchrony. A-V = auditory target with visual distractor; V-A = visual target with auditory distractor. Error bars indicate standard error of the mean

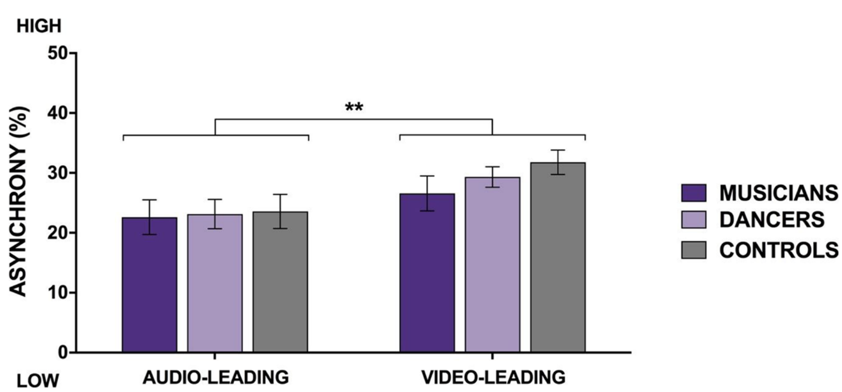

Fig. 2 Effects of group and track type for the "bounce-beep" integration task. Musicians, dancers, and controls performed similarly. For all groups, the window of perceived simultaneity was smaller for clips in which audio led video than when video led audio. The interaction between group and track type was not significant. Error bars indicate standard error of the mean. ** $p < .01$

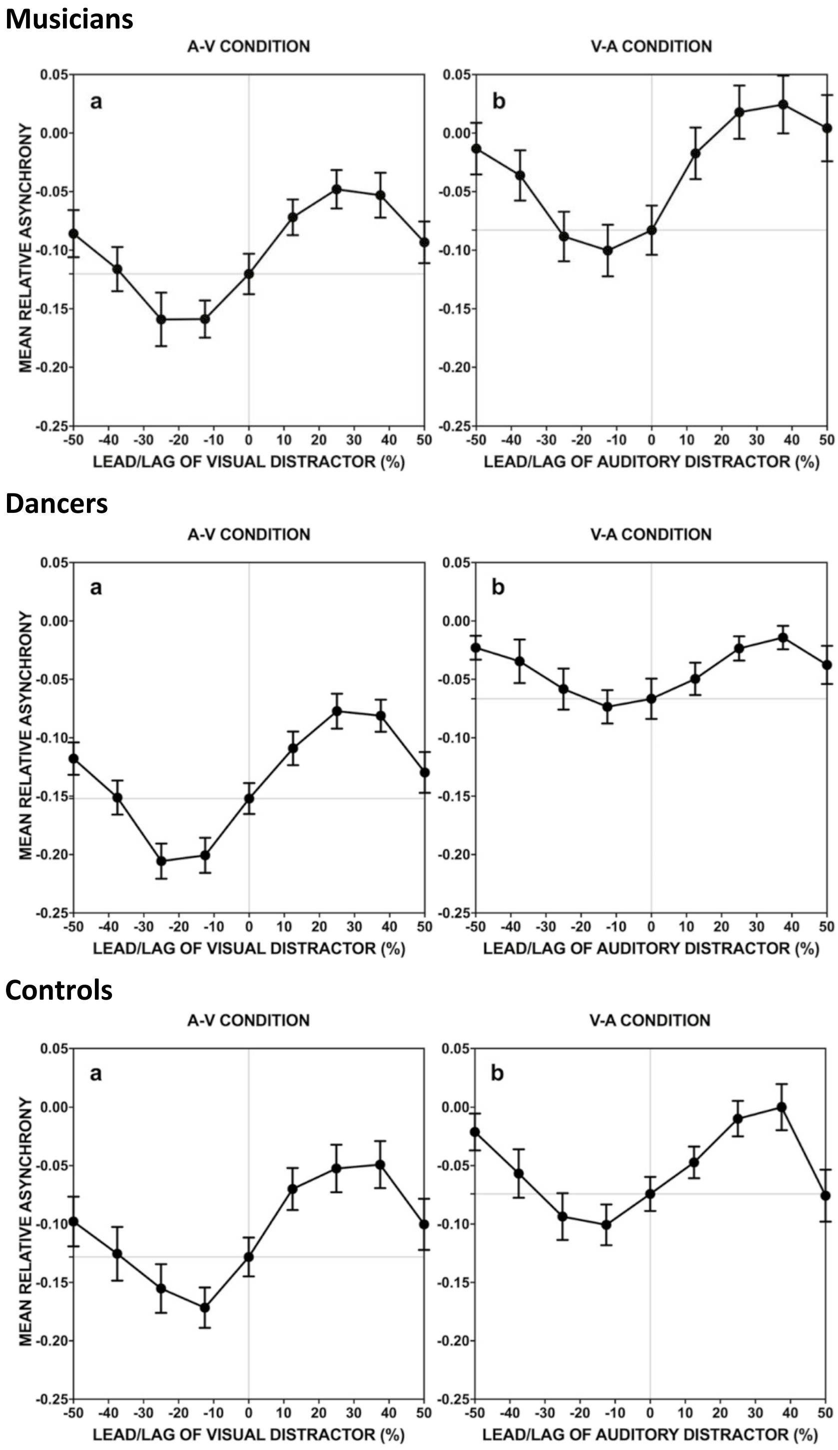

Fig. 4 Mean relative tap asynchrony as a function of distractor lead/lag in the auditory target-visual distractor (A-V) condition (left side) and visual target-auditory distractor (V-A) condition (right side), for musicians (**a**, **b**), dancers (**c**, **d**), and controls (**e**, **f**). The horizontal grey line indicates the mean relative asynchrony at zero lead/lag (vertical grey line). Error bars indicate standard error of the mean

not between controls and musicians, $t(38) = 1.57$, $p = .13$, BF = 0.65, or controls and dancers, $t(38) = 1.45$, $p = .15$, BF = 0.60. There was no main effect of distractor modality, $F(1, 57) = .20$, $p = .66$, $\eta_p^2 = .003$, $BF_{inclu} = 0.21$, but the interaction between group and modality was significant, $F(2, 57) = 5.38$, $p = .007$, $\eta_p^2 = .16$, $BF_{inclu} = 15.91$, confirming that the groups differed in the size of the auditory versus the visual distractor effects. The interaction was driven by a significant difference in the auditory distractor effect for musicians compared to dancers. Musicians were significantly more distracted than dancers by the auditory distractor, t(38) = 3.07, p = .004, BF = 10.18, and dancers, t(38) = 2.10, p = .04, BF= 1.69, while all three groups were similarly distracted by the visual distractor ($ps > .47$, BFs < 0.61). Another way of interpretation, detailed in the sections below, is that musicians showed slightly larger auditory distractor effects than visual distractor effects, whereas dancers showed slightly larger visual than auditory distractor effects (Fig. 3). Controls showed similar distractor effects for both modalities.

Musicians For both the A-V and V-A conditions, musicians' mean relative asynchronies showed the expected sinusoidal shape (Hove et al., 2013; Repp & Penel, 2002, 2004). The sinusoidal function suggests that the in-phase and anti-phase alignment produced similarly sized asynchronies (Figs. 4a, b). Numerically, relative to the zero lead/lag trials, lagging distractors attracted taps more strongly than leading distractors for both the A-V and V-A conditions. Statistically, the auditory distractor (M = 0.24, SE = 0.04) effect was significantly larger than the visual distractor effect (M = 0.15, SE= 0.02), t(19) = 2.12, p = .048, BF = 1.44. Therefore, musicians were significantly more distracted by the auditory distractors than by the visual distractors.

Dancers For both the A-V and V-A conditions, dancers' mean relative asynchronies also showed the expected sinusoidal shape. However, the sinusoidal amplitude was much shallower for the V-A condition than for the A-V condition (Figs. 4c, d), indicating the auditory distractors were less distracting than the visual distractors. Numerically, relative to the zero lead/lag trials, lagging distractors attracted taps more strongly than leading distractors for both the A-V and the V-A conditions. Statistically, the visual distractor effect (M = 0.16, SE = 0.006) was significantly larger than the auditory distractor effect (M = 0.11, SE= 0.02), t(19) = 2.11, p = .048, BF = 1.44. Thus, in contrast to musicians, dancers were significantly more distracted by the visual distractors than by the auditory distractors.

Controls Like musicians and dancers, controls' mean relative asynchronies showed the expected sinusoidal shape for both the A-V and the V-A conditions (Figs. 4e, f). Numerically, relative to the zero lead/lag trials, the lagging distractors attracted taps more strongly than leading distractors for both the A-V and V-A conditions. Statistically, there was no significant difference between the magnitudes of the visual and auditory distractor effects (visual: M = 0.17, SE= 0.01, auditory: M = 0.15, SE= 0.02), t(19) = .75, p = .46, BF = 0.30. Therefore, controls showed similar distraction for both modalities.

Experiment 1: Discussion

Experiment 1 examined how expertise (in music and dance) and modality (auditory and visual) interact to influence audiovisual integration and synchronization on a "bounce-beep" integration task and a bimodal target-distractor synchronization task, respectively. Performance on the "bounce-beep" integration task did not significantly differ between groups, in contrast to the prediction that musicians and dancers would outperform controls. Although the window of perceived simultaneity was larger for controls than for musicians and dancers, particularly when the video was leading the audio, the difference was not significant. Because audiovisual integration is an automatic process that is important for a range of human behaviours (Adams, 2016; Alais & Burr, 2004; Hartcher-O'Brien et al., 2014), particularly speech (Alsius et al., 2005; Déry et al., 2014), it may be that the experience of audiovisual integration in other domains enabled controls to perform similarly to musicians and dancers.

In contrast, for the bimodal target-distractor synchronization task, expertise did relate to modality differences in distractor effectiveness: musicians' taps were more attracted to auditory distractors than visual distractors, whereas dancers' taps were more attracted to visual than auditory distractors. For controls, we had predicted a greater effect for auditory than visual distractors in line with previous work (Hove et al., 2013; Welch & Warren, 1980). However, controls' taps were similarly attracted to both auditory and visual distractors.

The roughly sinusoidal pattern of relative asynchrony across the different temporal offsets is consistent with previous literature (Hove et al., 2013; Repp & Penel, 2002, 2004), with greater offsets producing larger asynchronies than smaller offsets. However, Hove et al. (2013) found that, for musicians, auditory distractors that were fully anti-phase no longer had an attractor effect, with greater asynchronies observed at offsets between fully in-phase and fully anti-phase. Our study finds something similar, but for all three groups, as anti-phase and in-phase distractors produced similar (although not identical) asynchronies, compared to smaller out-of-phase distractors (Fig. 4). Attractors that are fully anti-phase may become easy to integrate as a point marking an equal subdivision of the target sequence, thus no longer distracting away from the target phase, but becoming incorporated within it.

Previous studies have found that a bouncing ball is more effective than a flashing light for enhancing synchrony and attracting movements, and that this effectiveness as a distractor is enhanced for visual experts (video gamers and ball players) (Hove et al., 2013). Our findings also support that a spatiotemporal visual stimulus can be as effective as an auditory stimulus in attracting taps, but with a bouncing stick-figure as opposed to a ball. In fact, the bouncing stick-figure may be even more effective than a bouncing ball, as asynchrony in controls was no longer better for auditory than visual stimuli, in contrast to previous studies with the bouncing ball.

Experiment 2

The results of Experiment 1 suggested that expertise and modality interacted to affect sensorimotor synchronization. However, sensorimotor synchronization was tested with finger tapping, and it is unclear whether the same interactions would occur for whole-body movements, as dancers are more expert in these types of movements. Thus, the objective of Experiment 2 was to examine whether musicians and dancers rely differentially on auditory and visual modalities when synchronizing with whole body effector movements (i.e., bouncing) as opposed to specific-effector movements (i.e., finger tapping). Participants in this second experiment performed the same audiovisual integration task as in Experiment 1, as well as a modified audiovisual synchronization task, which now involved knee bending, rather than finger-tapping. More specifically, participants bounced in synchrony with an isochronous auditory or visual (bouncing stick figure) target sequence while a distractor sequence was presented in the other modality at one of nine temporal offsets.

For the audiovisual integration task, we expected to replicate the results found in Experiment 1, with no difference between groups. For the audiovisual synchronization task, it was predicted that musicians' auditory experience would lead them to be more influenced by auditory distractors than visual distractors. We predicted that dancers, however, would be more influenced by visual distractors than auditory distractors because of their experience with visual rhythms and the use of whole-body movements. Finally, based on the findings in Experiment 1, it was predicted that controls would not show a bias towards one modality over the other.

Methods

Participants

Three groups of participants were tested: musicians, dancers, and non-musician/non-dancer controls. There were 20 participants in each group, for a total of 60. Age ranged between 18 and 47 years ($M = 22.55$ years, $SD = 4.48$ years). Table 2 summarizes the demographic characteristics of the sample. The criteria for classifying musicians, dancers, and controls were identical to criteria in Experiment 1. All participants reported normal hearing and normal or corrected-to-normal vision. Participants received either two and a half research credits or $25.00 (CAD) for their participation. All participants provided informed consent in accordance with the guidelines approved by the University of Western Ontario Nonmedical Research Ethics Board.

Table 2 Participant characteristics for Experiment 2

	Musicians	Dancers	Controls
Sex	11 females 9 males	19 females 1 male	12 females 8 males
Age range, y	19 to 34	19 to 25	18 to 47
Age, y (M ± SD)	22.90 ± 3.86	21.30 ± 2.11	23.45 ± 6.35
Music training, y (M ± SD)	11.25 ± 3.45	0.55 ± 1.23	0.32 ± 0.65
Dance training, y (M ± SD)	0.22 ± 0.57	11.45 ± 5.20	0.05 ± 0.22
Starting age, y (M ± SD)	8.50 ± 3.65	9.85 ± 5.49	NA
Weekly practice, h (M ± SD)	2.60 ± 2.80	4.80 ± 5.86	NA

Procedure

Audiovisual integration was tested using the same "bounce-beep" integration task used in Experiment 1, while audiovisual synchronization was tested using a variant of the bimodal target-distractor synchronization task (Chen et al., 2002; Repp & Penel, 2002, 2004) assessed with whole-body movements recorded by a motion capture system. All participants completed the audiovisual integration task followed by the audiovisual synchronization task in one session. The entire session took approximately 2.5 h. All participants were fully debriefed after the study.

For the bimodal target distractor synchronization task, participants bounced in synchrony with a target isochronous auditory sequence (500 Hz, 10 ms tones, with onset and offset ramps of 5 ms) or a target visual sequence (the bouncing stick figure), while a distractor sequence was presented in the other modality at one of nine temporal offsets. Identical to Experiment 1, both target and distractor sequences consisted of 32 events with an IOI of 925 ms. The nine temporal phase displacements between the target and distractor sequences ranged from –50% and +50% of the IOI: 0%, ± 12.5%, ± 25%, ± 37.5%, ± 50%. The task consisted of 108 trials; each of the nine temporal offsets occurred with the auditory sequence as the distractor and the visual sequence as the distractor, and repeated six times each. The order of the 108 trials was randomized for each participant, and broken into 12 blocks of nine trials to prevent fatigue. The auditory sequences were played over Dell A215 speakers at a comfortable volume, while the visual sequences were rear projected from a NEC LT260 projector onto a 3 × 4 ft screen

(Fig. 5). The projected image was approximately 35 × 25 in. Participants were instructed to face the projection screen while standing approximately 55 in. away as they bounced in synchrony with the target sequence by bending their knees, ensuring that the bottom of their bounces synchronized with the target sequence, starting at the third event of each target sequence, and to continue until the end of the trial, while ignoring the distractor sequence.

Participants' movements were recorded by a three-camera optoelectronic recording system (Optotrak, Northern Digital Inc., Waterloo, Canada). The system captured the three-dimensional (3-D) positions of infrared emitting diodes (IREDs) attached to black foam knee pads (two IREDs on each knee) worn by the participant. Using custom in-house software (OTCollect, programmed by Haitao Yang), the 3-D positions of each IRED were recorded at 250 Hz as the participant bounced, and used to calculate the spatial displacement of the knees. Bounce times were derived from the lowest point. The motion capture was time locked to the start of the trial. Each trial was recorded for 30 s.

Similar to Experiment 1, an attention check was used to ensure participants were paying attention to the visual distractors. For the A-V condition, participants were required to report at the end of each trial whether the joints of the bouncing stick figure briefly changed colour (from black to red). For the V-A condition, participants reported whether one of the auditory tones changed in pitch (from 500 Hz to 700 Hz). The attention check occurred on half of the trials and was randomized to occur between event 13 and 21. Responses regarding colour or pitch changes were made verbally by the participant and recorded by the experimenter on a PC desktop.

Statistical analyses

Identical analyses to Experiment 1 were conducted for group demographics related to training, performance on the "bounce-beep" integration task and attention check accuracy for the bimodal target-distractor synchronization task. Custom in-house software (OTDisplay, programmed by Haitao Yang) was used to calculate the bounce times. Bounce times were calculated for each of the four IREDs, and then averaged to get bounce times for each trial. Distractor effects were calculated as in Experiment 1. To quantify the strength of the distractor effect, the range of values for the mean relative asynchrony scores (the maximum value of the nine mean relative asynchrony scores minus the minimum value of the nine mean relative asynchrony scores) for each participant was calculated (Hove et al., 2013). Differences in distractor effects were then analysed using a 3 × 2 mixed ANOVA with group as a between-subject variable and distractor modality as the within-subject variable. Follow-up pairwise t-tests were conducted where appropriate to determine the nature of any interactions and p-values were corrected using the Holm correction for multiplicity. All hypothesis tests used $\alpha = .05$ for significance. Data were analyzed with the same software as in Experiment 1.

Results

Group demographics

One-way ANOVAs conducted on years of music and dance training revealed significant group differences for both music, $F(2, 57) = 169.08$, $p < .001$, $\eta_p^2 = .86$, and dance

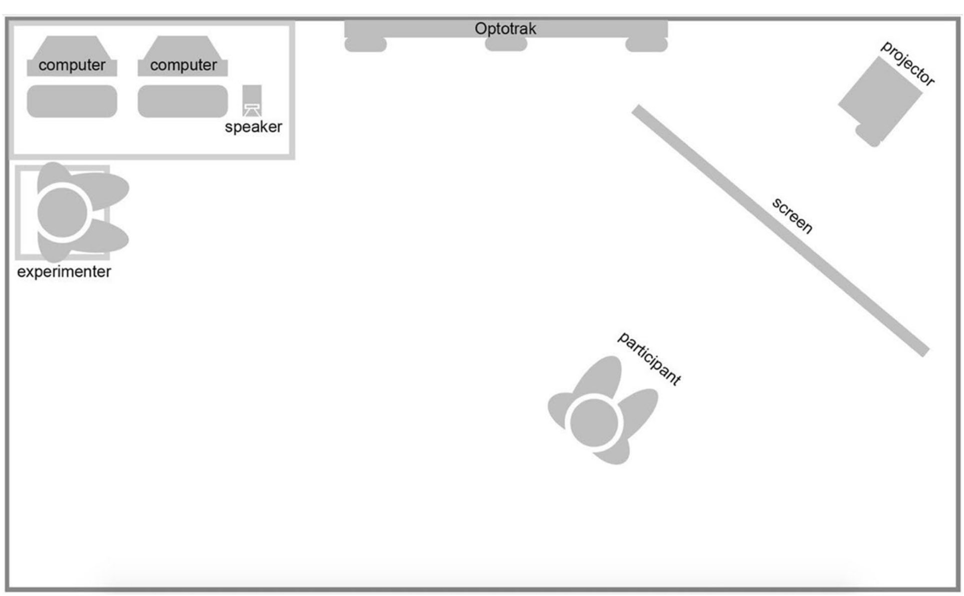

Fig. 5 Schematic diagram of the experimental setup

training, $F(2, 57) = 93.51$, $p < .001$, $\eta_p^2 = .77$. Post hoc comparisons confirmed that musicians and controls significantly differed in years of music training, $t(38) = 13.93$, $p < .001$. Musicians and dancers also significantly differed, $t(38) = 13.07$, $p < .001$, but there were no significant differences in years of music training between controls and dancers, $t(38) = .72$, $p = .48$. Likewise, for years of dance training, post hoc comparisons confirmed that dancers and controls significantly differed in years of dance training, $t(38) = 9.80$, $p < .001$. Dancers and musicians also significantly differed, $t(38) = 9.60$, $p < .001$. However, there were no significant differences in years of dance training between controls and musicians, $t(38) = 1.27$, $p = .21$.

Independent-samples t-tests between musicians and dancers were conducted on years of training, starting age of training, and hours of practice per week in their respective expertise. Musicians and dancers did not significantly differ in the years of training in their respective expertise, $t(38) = .14$, $p = .89$, $\eta_p^2 = .001$, or starting age, $t(38) = .92$, $p = .37$, $\eta_p^2 = .02$. Similarly, musicians and dancers did not significantly differ on the number of hours they practiced per week, $t(38) = 1.51$, $p = .14$, $\eta_p^2 = .06$.

"Bounce-beep" integration task

Performance on the "bounce-beep" integration task did not significantly differ between groups, $F(2, 57) = 2.05$, $p = .14$, $\eta_p^2 = .07$, $BF_{inclu} = 0.36$. Musicians ($M = 23.09\%$, $SE = 2.06\%$), dancers ($M = 26.27\%$, $SE = 1.57\%$), and controls ($M = 28.25\%$, $SE = 2.12\%$) all performed similarly on the task. However, performance did significantly differ for track type, $F(1, 57) = 7.27$, $p = .009$, $\eta_p^2 = .11$, $BF_{inclu} = 9.62$. For all groups, the window of perceived simultaneity was smaller when the audio led the video ($M = 22.86\%$, $SE = 1.78\%$) compared to when the video led the audio ($M = 28.88\%$, $SE = 1.28\%$), suggesting that there is a larger integration window when vision leads audition. The interaction between group and track type was not significant, $F(2, 57) = 1.02$, $p = .37$, $\eta_p^2 = .03$, $BF_{inclu} = 0.36$ (Fig. 6).

Bimodal target-distractor synchronization task

Attention check accuracy No participants were excluded as all scored at least 85%. Average response accuracy was 96.8%. Accuracy did not significantly differ between groups, $F(2, 57) = 1.38$, $p = .26$, $\eta_p^2 = .05$ (musicians: $M = 96\%$, $SE = 9.26\%$; dancers: $M = 98\%$, $SE = 7.41\%$; controls: $M = 96\%$, $SE = 9.64\%$). Thus, participants were attending to the distractor sequences while synchronizing to the target sequences.

Relative asynchronies

Comparison between groups The 3 × 2 mixed ANOVA with group as the between-subjects variable and distractor modality as the within-subject variable on the magnitude of the distractor effect did not produce a significant main effect of group, $F(2, 57) = 0.92$, $p = .41$, $\eta_p^2 = .03$, $BF_{inclu} = 0.23$ (Fig. 7). Thus, the magnitude of the distractor effect was similar for musicians ($M = 0.17$, $SE = 0.2$), dancers ($M = 0.18$, $SE = 0.02$), and controls ($M = 0.20$, $SE = 0.02$). However, there was a main effect of distractor modality, $F(1, 57) = 178.59$, $p < .001$, $\eta_p^2 = .76$, $BF_{inclu} = 1.20 \times 10^{22}$. For all three groups, visual distractors ($M = 0.27$, $SE = 0.01$) were significantly more distracting than auditory distractors ($M = 0.09$, $SE = 0.01$). The interaction between group and distractor modality was not significant, $F(2, 57) = .28$, $p = .75$, $\eta_p^2 = .01$, $BF_{inclu} = 0.16$.

For the A-V conditions, all three groups' mean relative asynchronies showed the expected sinusoidal shape consistent with previous literature (Hove et al., 2013; Repp & Penel,

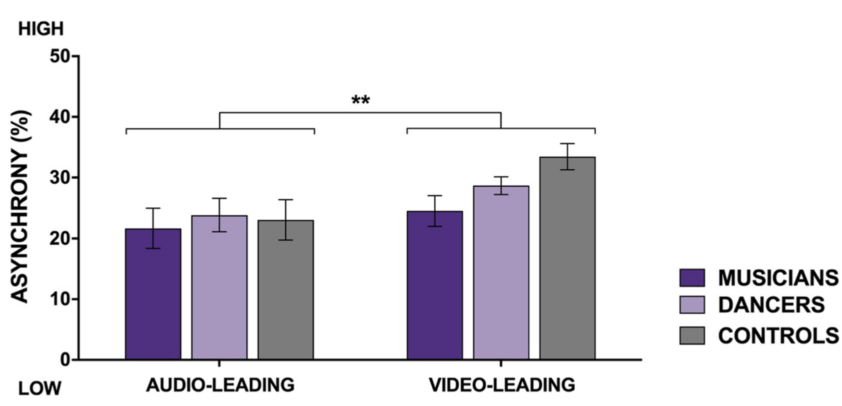

Fig. 6 Effects of group and track type for the "bounce-beep" integration task. Musicians, dancers, and controls performed similarly. For all three groups, the window of perceived simultaneity was smaller for clips which audio led video than when video led audio. The interaction between group and track type was not significant. Error bars indicate standard error of the mean. ** $p < .01$

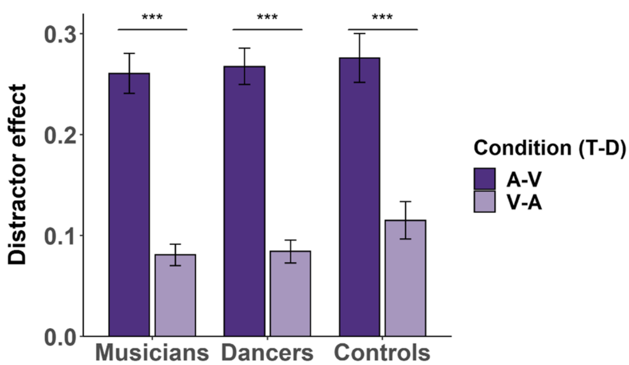

Fig. 7 Distractor effects of group and condition when bouncing in the bimodal distractor synchronization task. Distractor effects were quantified as the difference between maximum and minimum relative asynchrony. A-V = auditory target with visual distractor; V-A = visual target with auditory distractor. Error bars indicate standard error of the mean

2002, 2004). Numerically, relative to the zero lead/lag trials, leading visual distractors attracted bounces more strongly than lagging distractors (Figure 8a, c, e). However, for the V-A conditions, mean relative asynchronies showed only a small effect of auditory distractors on bounces (Figs. 8b, d, f).

Experiment 2: Discussion

Experiment 2 examined how expertise (in music and dance) and modality (auditory and visual) interacted to influence audiovisual integration and synchronization using a "bounce-beep" integration task and a bimodal target-distractor synchronization task assessed with whole-body movements, respectively. As predicted, the bounce-beep results replicated Experiment 1. The window of perceived simultaneity was larger for visual-leading stimuli than auditory-leading stimuli, and there were no differences between groups. For the bimodal target-distractor synchronization task, it was predicted that musicians would be more distracted by auditory than visual distractors, whereas dancers would be more distracted by visual than auditory distractors and would show a larger visual effect here than in Experiment 1, because of their experience with visual rhythms and the use of whole-body movements. However, the magnitude of the distractor effect was similar for musicians, dancers, and controls. Moreover, the effect of the visual distractor was increased in Experiment 2 relative to Experiment 1, and the effect of the auditory distractors was reduced in Experiment 2 relative to Experiment 1, numerically speaking. This reduction seems most likely caused by the use of bouncing as the synchronization action, compared to tapping, which is usually used (Chen et al., 2002; Hove et al., 2013; Repp & Penel, 2002, 2004). It may be that, by mirroring the movements of the bouncing stick figure, participants' focus on the visual modality was significantly increased relative to tapping to the same stimulus, in order to match the bouncing movements. Participants would be more easily able to compare their body position to the visual stimulus position throughout the bouncing trajectory, as there are more points of similarity between bouncing movements and the bouncing stimulus, compared to tapping movements and the bouncing stimulus. Evidence for bias to imitate observed movements is prevalent (Bonda et al., 1996; Downing et al., 2001, 2006; Grossman et al., 2000; Vaina et al., 2001). Thus, a desire to synchronize throughout the trajectory rather than just at the low point may have enhanced processing of the visual target over the auditory distractor, relative to Experiment 1.

Overall, we expected whole-body movements to engage dancers' movements to a greater degree than musicians' and controls' movements, consistent with previous work (Gardner et al., 2015; Shimada, 2010; Vogt et al., 2007). For example, in musicians, brain areas involved in both action observation and execution respond when musicians are observing musically familiar actions (Bangert & Schlaug, 2006; Pau et al., 2013; Proverbio et al., 2014). Likewise, in dancers, overlaps in activation are observed when dancers are observing dance movements within their motor repertoire (Calvo-Merino et al., 2005; Cross et al., 2006; Pilgramm et al., 2010). Although the choice to use a bouncing stick figure was because the knee bending motion was familiar for dancers, the goal was also to have a movement all groups could perform easily. Therefore, as knee bending is somewhat familiar to musicians and controls, the movement may not have specifically engaged dancers' expertise.

General discussion

The current studies examined how expertise in music and dance interacted with modality (auditory and visual) to affect performance on an audiovisual integration task and a bimodal target-distractor synchronization task. Musicians and dancers, as well as non-musician/non-dancer controls, completed two tasks measuring audiovisual integration (perception) and synchronization (production) with finger tapping and knee bending. It was predicted that musicians and dancers would not significantly differ in performance on the audiovisual integration task but would outperform controls (Karpati et al., 2016). However, performance for the audiovisual integration task in Experiments 1 and 2 did not significantly differ between musicians, dancers, and controls. Although the window of perceived simultaneity was numerically larger for controls than for both musicians and dancers, the difference was not significant. The window of audiovisual integration appeared to be around 25% and 30% of the inter-onset interval for all three groups. Therefore,

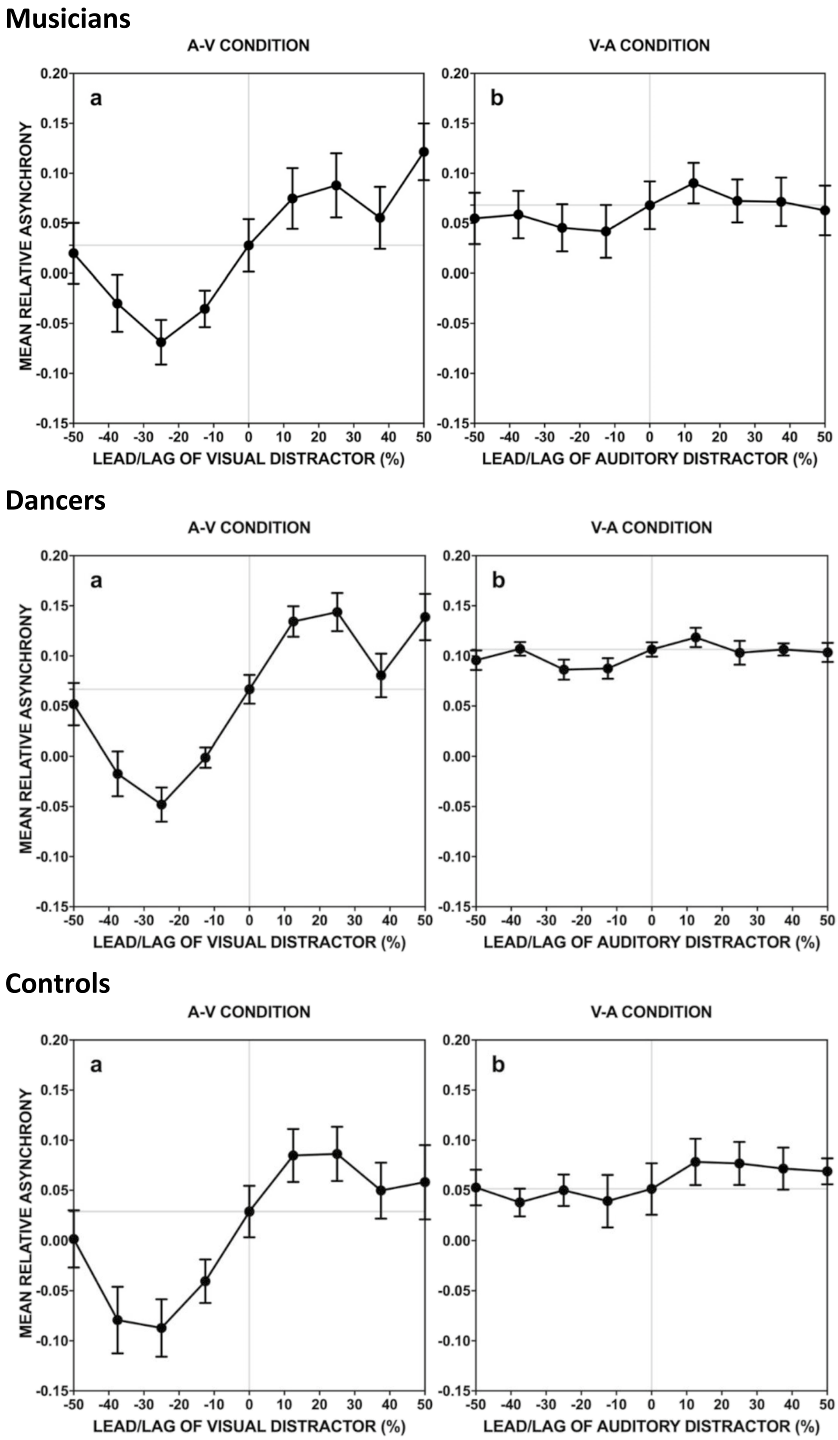

Fig. 8 Mean relative bounce asynchrony as a function of distractor lead/lag in the auditory target-visual distractor (A-V) condition (left side) and visual target-auditory distractor (V-A) condition (right side), for musicians (**a**, **b**), dancers (**c**, **d**), and controls (**e**, **f**). The horizontal grey line indicates the mean relative asynchrony at zero lead/lag (vertical grey line). Error bars indicate standard error of the mean

despite specialized training, we did not find evidence that musicians and dancers have an advantage on the audiovisual integration over controls. This may be consistent with the automaticity of audiovisual integration as a part of human behaviour (Adams, 2016; Alais & Burr, 2004; Hartcher-O'Brien et al., 2014).

The window of perceived simultaneity for the "bounce-beep" integration task was larger for visual-leading than auditory-leading stimuli. That is, the audio and video were more likely to be perceived as simultaneous when vision led audition. This aligns with previous studies also showing that audiovisual asynchrony scores are larger if video leads audio rather than if the audio leads video (Dixon & Spitz, 1980; Kayser et al., 2008; Keetels & Vroomen, 2012; Vatakis & Spence, 2006; Zampini et al., 2005). One speculation is that auditory stimuli are processed faster than visual stimuli (Keetels & Vroomen, 2012), therefore when vision leads audition the two stimuli are more likely to be perceived as occurring at the same time. One important difference between the visual and auditory stimuli is that the visual stimuli were continuous while the auditory stimuli were discrete (onsets occurred at one precise point in time). Thus, the moment the bouncing figure was perceived to be at the bottom might have been imprecise compared to the auditory stimulus. This might have affected the window of perceived simultaneity. A future study could investigate whether the same effects and similar window sizes are obtained when a continuous auditory sinusoidal stream is used along with a continuous visual stream.

In Experiment 1, a variant of a previously used bimodal target-distractor synchronization paradigm was used for the audiovisual synchronization task (Chen et al., 2002; Repp & Penel, 2002; Repp & Penel, 2004), pitting a visual bouncing stick figure against an auditory metronome. When tapping, musicians were more distracted by auditory than visual distractors, whereas dancers were more distracted by visual than auditory distractors. Controls showed no difference between modalities. These results are broadly consistent with another study that compared musicians and video gamers/ball players on a similar task (Hove et al., 2013). For musicians, auditory was more distracting than visual distractors, whereas the opposite was observed for video gamers and ball players (Hove et al., 2013).

In Experiment 2, participants bounced rather than tapped to the target sequence, and the stick figure was projected life-size rather than watched on a computer screen. Again, it was predicted that musicians would have a strong bias for the auditory modality, whereas dancers would have a strong bias for the visual modality, but the magnitude of the visual distractor effect for dancers would be larger than observed in Experiment 1 because of the use of whole-body movements. In contrast to Experiment 1, however, all three groups were biased to the visual modality. The changes in mean relative asynchrony scores for the A-V conditions produced a prominent sinusoidal pattern, consistent with previous literature, while the changes in mean relative asynchrony scores for the V-A conditions produced a flattened pattern, suggesting that there was little effect of the auditory distractors (Hove et al., 2013; Repp & Penel, 2002, 2004). It is unclear why the visual dominance was so much stronger for knee bending than finger tapping, though it likely relates to the increased similarity between the observed and performed movements (Bonda et al., 1996; Downing et al., 2001, 2006; Grossman et al., 2000; Vaina et al., 2001). Additionally, the visual stimulus (the bouncing stick figure) was continuous, which may have provided more temporal information than the discrete auditory beeps. Although this continuous aspect of the stimuli did not seem to affect finger tapping synchronization in Experiment 1, the combination of the similarity to the performed movement and continuous nature of the visual distractor in Experiment 2 may have amplified its effect in the bimodal synchronization task, resulting in a similar distractor effect across all three groups. Therefore, similar to the recommendation for the bounce-beep integration task, future studies could also evaluate how continuous auditory and visual stimuli would affect performance in the bimodal target-distractor synchronization task. Finally, differences in tempo associated with each type of movement could also be a reason for different results between Experiments 1 and 2. More specifically, effector-specific movements are often done at a faster tempo than whole body movements (Burger et al., 2013). Therefore, it is possible that the slow tempo for the effector specific movements led to larger distractor effects in Experiment 1.

One thing that is clear, however, is that results from both Experiment 1 and Experiment 2 provide further evidence against the modality appropriateness hypothesis, if it is interprted as auditory information being more appropriate than visual information for performance in temporal tasks. More specifically, our results show that this interpretation of the hypothesis does not hold when participants have different expertise (Experiment 1) and when using different movements to synchronize (Experiment 2). Conversely, one might argue that the hypothesis remains true if it is interpreted as people will use the modality that is best suited for the task at hand. However, this makes it hard to differentiate from other hypotheses, such as the Bayesian optimal integration (BOI) hypothesis, which may be equally or more suited to explain these results. The BOI hypothesis stipulates that the bias caused by a stimulus depends on the relative precision of the information it provides (Hove et al., 2013; Körding & Wolpert, 2004). For example, in the case of Experiment 2, the bouncing stick figure provides more (precise) information that is relevant to the task (bouncing) compared to the beep, resulting in bias towards to visual distractor and less bias towards the auditory distractor. Similarly, the BOI hypothesis may explain the different biases seen in Experiment 1. Even though musicians and dancers are receiving the same information (i.e., stimuli), the type of information (auditory or visual) is processed more precisely depending on the expertise, resulting in bias for the auditory modality for musicians and the visual modality for dancers. Of course, these are *a posteriori* interpretations and the BOI hypothesis was not specifically tested in the current study:

quantifying the "precision" of the information presented in each modality is beyond the scope of this study. However, it remains that the current evidence suggests the revision of the modality appropriateness hypothesis, at least in the context of bimodal target-distractor synchronisation tasks, to specify the determining factors of the "appropriate modality".

Overall, the results also suggest that dynamic motion may optimize visual rhythm processing and synchronization (Hove & Keller, 2010; Hove et al., 2010; Su, 2014). Several previous studies showing that visual stimuli are less effective at engaging movements often lack dynamic motion, which may provide rich spatiotemporal information (Chen et al., 2002; Repp & Penel, 2002, 2004). Like a bouncing ball (Hove et al., 2013, Iversen et al., 2015), the bouncing stick figure had a continuous motion that consisted of a repetitive knee-bending motion generated from a dancer's bouncing trajectory. The choice to use a bouncing stick figure rather than a bouncing ball was to find a movement that was both ecologically valid and relevant for dancers. It remains to be seen whether a bouncing stick figure is more effective at enhancing rhythm synchronization than other dynamic visual stimuli, such as a bouncing ball, that do not mirror the synchronized body movement.

Conclusion

Overall, we did not find differences in audiovisual integration between groups on the "bounce-beep" integration task. This finding was robust across two experiments. However, when tapping during the bimodal target-distractor task, musicians were more distracted by the auditory than visual distractors, whereas dancers were more distracted by the visual than the auditory distractors. When bouncing, all groups were more biased toward the visual than auditory modality: the visual bouncing stick figure was significantly more distracting than the auditory tones. Thus, we demonstrate that using a dynamic visual stimulus and a bouncing movement influences audiovisual integration, as measured by sensorimotor synchronization performance. In addition, these results provide further evidence for rejecting the modality appropriateness hypothesis, which predicts bias for the auditory modality in tasks requiring temporal processing.

Open Practices Statement The study's design and analyses were not preregistered. Data were collected in 2014 and 2015 and cannot be shared because participants were not asked whether they consented to sharing their data. Study materials can be provided on demand.

Authors' note We have no known conflicts of interest to disclose.

Funding NSERC Discovery Grant to JAG (RGPIN-2016-05834), James S. McDonnell Foundation Understand Human Cognition Scholar Award to JAG (DOI:https://doi.org/10.37717/220020403).

References

Adams, W. J. (2016). The development of audio-visual integration for temporal judgements. *PLoS Computational Biology, 12*(4), e1004865. https://doi.org/10.1371/journal.pcbi.1004865

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14*(3), 257–262. https://doi.org/10.1016/j.cub.2004.01.029

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology, 15*(9), 839–843. https://doi.org/10.1016/j.cub.2005.03.046

Bangert, M., & Schlaug, G. (2006). Specialization of the specialized in features of external human brain morphology. *European Journal of Neuroscience, 24*, 1832–1834. https://doi.org/10.1111/j.1460-9568.2006.05031.x

Bläsing, B., Calvo-Merino, B., Cross, E. S., Jola, C., Honisch, J., & Stevens, C. J. (2012). Neurocognitive control in dance perception and performance. *Acta Psychologica, 139*(2), 300–308. https://doi.org/10.1016/j.actpsy.2011.12.005

Bonda, E., Petrides, M., Ostry, D., & Evans, A. (1996). Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *Journal of Neuroscience, 16*(11), 3737–3744.

Buchanan, J. J., Zihlman, K., Ryu, Y. U., & Wright, D. L. (2007). Learning and transfer of a relative phase pattern and a joint amplitude ratio in a rhythmic multijoint arm movement. *Journal of Motor Behavior, 39*(1), 49–67. https://doi.org/10.3200/JMBR.39.1.49-67

Burger, B., Thompson, M. R., Luck, G., Saarikallio, S., Toiviainen, P. (2013). Influences of rhythm- and timbre-related musical features on characteristics of music-induced movement. *Frontiers in Psychology*, *4*, https://doi.org/10.3389/fpsyg.2013.00183

Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired skills: an fRMI study with expert dancers. *Cerebral Cortex, 15*(8), 1243–1249. https://doi.org/10.1093/cercor/bhi007

Chen, Y., Repp, B. H., & Patel, A. D. (2002). Spectral decomposition of variability in synchronization and continuation tapping: comparisons between auditory and visual pacing feedback conditions. *Human Movement Science, 21*, 515–532. https://doi.org/10.1016/s0167-9457(02)00138-0

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

Cross, E. S., de C Hamilton, A. F., & Grafton, S. T. (2006). Building a motor simulation de novo: observation of dance by dancers. *Neuroimage, 31*(3), 1257–1267. https://doi.org/10.1016/j.neuroimage.2006.01.003

de Boer-Schellekens, L., Eussen, M., & Vroomen, J. (2013). Diminished sensitivity of audiovisual temporal order in autism spectrum disorder. *Frontiers in Integrative Neuroscience, 7*, 8. https://doi.org/10.3389/fnint.2013.00008

Déry, C., Campbell, N. K. J., Lifshitz, M., & Raz, A. (2014). Suggestion overrides automatic audiovisual integration. *Consciousness and Cognition, 24*, 33–37. https://doi.org/10.1016/j.concog.2013.12.010

Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception, 9*(6), 719–721. https://doi.org/10.1068/p090719

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Social Neuroscience, 293*(5539), 2470–2473. https://doi.org/10.1126/science.1063414

Downing, P. E., Peelen, M. V., Wiggett, A. J., & Tew, B. D. (2006). The role of the extrastraite body area in action perception. *Science, 1*(1), 52–62. https://doi.org/10.1080/17470910600668854
Drake, C. (1998). Psychological processes involved in the temporal organization of complex auditory sequences: universal and acquired processes. *Music Perception, 16*(1), 11–26. https://doi.org/10.2307/40285774
E-Prime (Version 2.0) [Computer software]. Pennsylvania: Psychology Software Tools.
Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191.
Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160.
Faulkenberry, T. J., Ly, A., & Wagenmakers, E.-J. (2020). Bayesian inference in numerical cognition: a tutorial using JASP. *Journal of Numerical Cognition, 6*(2), 231–259. https://doi.org/10.5964/jnc.v6i2.288
Fernandes, L. F., & de Barros, R. M. (2012). Grip pattern and finger coordination differences between pianists and non-pianists. *Journal of Electromyography and Kinesiology, 22*, 412–418. https://doi.org/10.1016/j.jelekin.2012.02.007
Fiedler, A., O'Sullivan, J. L., Schröter, H., Miller, J., & Ulrich, R. (2011). Illusory double flashes can speed up responses like physical ones: evidence from the sound-induced flash illusion. *Experimental Brain Research, 214*(1), 113. https://doi.org/10.1007/s00221-011-2811-z
Gardner, T., Goulden, N., & Cross, E. S. (2015). Dynamic modulation of the action observation network by movement familiarity. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 35*(4), 1561–1572. https://doi.org/10.1523/JNEUROSCI.2942-14.2015
Gentilucci, M., & Cattaneo, L. (2005). Automatic audiovisual integration in speech perception. *Experimental Brain Research, 167*(1), 66–75. https://doi.org/10.1007/s00221-005-0008-z
Grahn, J. A. (2012). See what I hear? Beat perception in auditory and visual rhythms. *Experimental Brain Research, 220*(1), 51–61. https://doi.org/10.1007/s00221-012-3114-8
Grahn, J. A., Henry, M. J., & McAuley, J. D. (2011). FMRI investigation of cross-modal interactions in beat perception: audition primes vision, but not vice versa. *NeuroImage, 54*, 1231–1243. https://doi.org/10.1016/j.neuroimage.2010.09.033
Grondin, S., & McAuley, J. D. (2009). Duration discrimination in crossmodal sequences. *Perception, 38*(10), 1542–1559. https://doi.org/10.1068/p6359
Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., & Blake, R. (2000). Brain areas involve in perception of biological motion. *Journal of Cognitive Neuroscience, 12*(5), 711–720. https://doi.org/10.1162/089892900562417
Guttman, S. E., Gilroy, L. A., & Blake, R. (2005). Hearing what the eyes see: auditory encoding of visual temporal sequences. *Psychological Science, 16*(3), 228–235. https://doi.org/10.1111/j.0956-7976.2005.00808.x
Hartcher-O'Brien, J., Di Luca, M., & Ernst, M. O. (2014). The duration of uncertain times: audiovisual information about intervals is integrated in a statistically optimal fashion. *PLoS ONE, 9*(4), e96134. https://doi.org/10.1371/journal.pone.0089339
Honing, H., Merchant, H., Háden, G. P., Prado, L., & Bartolo, R. (2012). Rhesus monkeys (*Macaca mulatta*) detect rhythmic groups in music, but not the beat. *PLoS ONE, 7*(12), e51369. https://doi.org/10.1371/journal.pone.0051369
Hove, M. J., & Keller, P. E. (2010). Spatiotemporal relations and movement trajectories in visuomotor synchronization. *Music Perception, 28*(1), 15–26. https://doi.org/10.1525/mp.2010.28.1.15
Hove, M. J., Spivey, M. J., & Krumhansl, C. L. (2010). Compatibility of motion facilitates visuomotor synchronization. *Journal of Experimental Psychology: Human Perception and Performance, 36*(6), 1525–1534. https://doi.org/10.1037/a0019059
Hove, M. J., Iversen, J. R., Zhang, A., & Repp, B. H. (2013). Synchronization with competing visual and auditory rhythms: bouncing ball meets metronome. *Psychological Research, 77*(4), 388–398. https://doi.org/10.1007/s00426-012-0441-0
Innes-Brown, H., Barutchu, A., Shivdasani, M. N., Crewther, D. P., Grayden, D. B., & Paolini, A. G. (2011). Susceptibility to the flash-beep illusion is increased in children compared to adults. *Developmental Science, 14*(5), 1089–1099. https://doi.org/10.1111/j.1467-7687.2011.01059.x
Inui, N., & Ichihara, T. (2001). Comparison of the relation between timing and force control during finger-tapping sequences by pianists and non pianists. *Motor Control, 5*, 385–398. https://doi.org/10.1123/mcj.5.4.385
Iversen, J. R., Patel, A. D., Nicodemus, B., & Emmorey, K. (2015). Synchronization to auditory and visual rhythms in hearing and deaf individuals. *Cognition, 134*, 232–244. https://doi.org/10.1016/j.cognition.2014.10.018
Jäncke, L., Loose, R., Lutz, K., Specht, K., & Shah, N. J. (2000). Cortical activations during paced finger-tapping applying visual and auditory pacing stimuli. *Cognitive Brain Research, 10*(1–2), 51–66. https://doi.org/10.1016/S0926-6410(00)00022-7
Karpati, F. J., Giacosa, C., Foster, N. E. V., Penhune, V. B., & Hyde, K. L. (2016). Sensorimotor integration is enhanced in dancers and musicians. *Experimental Brain Research, 234*, 893–903. https://doi.org/10.1007/s00221-015-4524-1
Kato, M., & Konishi, Y. (2006). Auditory dominance in the error correction process: a synchronized tapping study. *Brain Research, 1084*(1), 115–122. https://doi.org/10.1016/j.brainres.2006.02.019
Kayser, C., Petkov, C. I., & Logothetis, N. K. (2008). Visual modulation of neurons in auditory cortex. *Cerebral Cortex, 18*(7), 1560–1574. https://doi.org/10.1093/cercor/bhm187
Keetels, M., & Vroomen, J. (2012). Perception of synchrony. In M. M. Murray & M. T. Wallace (Eds.), *The Neural Bases of Multisensory Processes* (9th ed.). CRC Press/Taylor & Francis.
Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception, 36*(14), 1–16.
Körding, K., & Wolpert, D. (2004). Bayesian integration in sensorimotor learning. *Nature, 427*, 244–247. https://doi.org/10.1038/nature02169
Ladda, A. M., Wallwork, S. B., & Lotze, M. (2020). Multimodal Sensory-Spatial Integration and Retrieval of Trained Motor Patterns for Body Coordination in Musicians and Dancers. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.576120
Lakens, D. (2022). Sample Size Justification. Collabra. *Psychology, 8*(1), 33267. https://doi.org/10.1525/collabra.33267
Large, E. W., & Palmer, C. (2002). Perceiving temporal regularity in music. *Cognitive Science, 26*, 1–37. https://doi.org/10.1207/s15516709cog2601_1
Lederman, S. J., & Klatzky, R. L. (2004). Multisensory texture perception. *Handbook of multisensory processes*, (pp. 107-122)
Lorås, H., Sigmundsson, H., Talcott, J. B., Öhberg, F., & Stensdotter, A. K. (2012). Timing continuous or discontinuous movements across effectors specified by different pacing modalities and intervals. *Experimental Brain Research, 220*(3–4), 335–347. https://doi.org/10.1007/s00221-012-3142-4
McAuley, J. D., & Henry, M. J. (2010). Modality effects in rhythm processing: auditory encoding of visual rhythms is neither obligatory nor automatic. *Attention, Perception, & Psychophysics, 72*(5), 1377–1389. https://doi.org/10.3758/app.72.5.1377

Merker, B., Madison, G., & Eckerdal, P. (2009). On the role and origin of isochrony in human rhythmic entrainment. *Cortex, 45*(1), 4–17. https://doi.org/10.1016/j.cortex.2008.06.011

Neumann, O., & Niepel, M. (2004). Timing of perception and perception of time. In C. Kaernbach, E. Schroger, & H. Müller (Eds.), *Psychophysics Beyond Sensation: Laws and Invariants of Human Cognition* (pp. 245–269). Lawrence Erlbaum Associates.

Nguyen, T., Sidhu, R. K., Everling, J. C., Wickett, M. C., Gibbings, A., & Grahn, J. A. (2022). Beat Perception and Production in Musicians and Dancers. *Music Perception, 39*(3), 229–248. https://doi.org/10.1525/mp.2022.39.3.229

Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception, 11*(4), 409–464. https://doi.org/10.2307/40285633

Patel, A. D., Iversen, J. R., Chen, Y., & Repp, B. H. (2005). The influence of metricality and modality on synchronization with a beat. *Experimental Brain Research, 163*, 226–238. https://doi.org/10.1007/s00221-004-2159-8

Pau, S., Jahn, G., Sakreida, K., Domin, M., & Lotze, M. (2013). Encoding and recall of a finger sequence in experienced pianists compared with musically naïve controls: a combined behavioral and functional imaging study. *Neuroimage, 64*, 379–387. https://doi.org/10.1016/j.neuroimage.2012.09.012

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

Proverbio, A. M., Calbi, M., Manfredi, M., & Zani, A. (2014). Audiovisuomotor processing in the musician's brain: an ERP study on professional violinists and clarinetists. *Scientific Reports, 4*, 5866. https://doi.org/10.1038/srep05866

Repp, B. H. (2003b). 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*(4), 355–370. https://doi.org/10.1080/00222890309603156

Repp, B. H. (2004). On the nature of phase attraction in sensorimotor synchronization with interleaved auditory sequences. *Human Movement Science, 23*(3–4), 389–413.

Repp, B. H. (2005). Sensorimotor synchronization: a review of the tapping literature. *Psychonomic Bulletin and Review, 12*(6), 969–992. https://doi.org/10.3758/BF03206433

Repp, B. H. (2010). Sensorimotor synchronization and perception of timing: effects of music training and task experience. *Human Movement Science, 29*(2), 200–213. https://doi.org/10.1016/j.humov.2009.08.002

Repp, B. H., & Penel, A. (2002). Auditory dominance in temporal processing: new evidence from synchronization with simultaneous visual and auditory sequences. *Journal of Experimental Psychology: Human Perception and Performance, 28*(5), 1085–1099. https://doi.org/10.1037/0096-1523.28.5.1085

Repp, B. H., & Penel, A. (2004). Rhythmic movement is attracted more strongly to auditory than to visual rhythms. *Psychological Research, 68*(4), 252–270. https://doi.org/10.1007/s00426-003-0143-8

Repp, B. H. (2003a). Phase attraction in sensorimotor synchronization with auditory sequences: effects of single and periodic distractors on synchronization accuracy. *Journal of Experimental Psychology, 29*(2). https://doi.org/10.1037/0096-1523.29.2.290

Shimada, S. (2010). Deactivation in the sensorimotor area during observation of a human agent performing robotic actions. *Brain and Cognition, 72*, 394–399. https://doi.org/10.1016/j.bandc.2009.11.005

Sofianidis, G., Hatzitaki, V., Grouios, G., Johannsen, L., & Wing, A. (2012). Somatosensory driven interpersonal synchrony during rhythmic sway. *Human Movement Science, 31*, 553–566. https://doi.org/10.1016/j.humov.2011.07.007

Su, Y.-H. (2014). Audiovisual beat induction in complex auditory rhythms: point-light figure movement as an effective visual beat. *Acta Psychologica, 151*, 40–50. https://doi.org/10.1016/j.actpsy.2014.05.016

Su, Y.-H., & Pöppel, E. (2012). Body movement enhances the extraction of temporal structures in auditory sequences. *Psychological Research, 76*(3), 373–382. https://doi.org/10.1007/s00426-011-0346-3

Thullier, F., & Moufti, H. (2004). Multi-joint coordination in ballet dancers. *Neuroscience Letters, 369*(1), 80–84. https://doi.org/10.1016/j.neulet.2004.08.011

Tillmann, B. (2008). Music cognition: learning, perception, expectations. In R. Kronland-Martinet, S. Ystad, & K. Jensen (Eds.), *Computer Music Modeling and Retrieval* (pp. 11–33). Springer.

Vaina, L. M., Solomon, J., Chowdhury, S., Sinha, P., & Belliveau, J. W. (2001). Functional neuroanatomy of biological motion perception in humans. *Proceedings of the National Academy of Sciences of the United States of America, 98*(20), 11656–11661. https://doi.org/10.1073/pnas.191374198

Vatakis, A., & Spence, C. (2006). Audiovisual synchrony perception for music, speech, and object actions. *Brain Research, 1111*(1), 134–142. https://doi.org/10.1016/j.brainres.2006.05.078

Verheul, M. H., & Geuze, R. H. (2004). Bimanual coordination and musical experience: the role of intrinsic dynamics and behavioral information. *Motor Control, 8*, 270–291. https://doi.org/10.1123/mcj.8.3.270

Vogt, S., Buccino, G., Wohlschläger, A. M., Canessa, N., Shah, N. J., Zilles, K., Eickhoff, S. B., Freund, H. J., Rizzolatti, G., & Fink, G. R. (2007). Prefrontal involvement in imitation learning of hand actions: effects of practice and expertise. *Neuroimage, 37*, 1371–1383. https://doi.org/10.1016/j.neuroimage.2007.07.005

Welch, R., & Warren, D. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 88*, 638–667. https://doi.org/10.1037/0033-2909.88.3.638

Zampini, M., Guest, S., Shore, D. I., & Spence, C. (2005). Audio-visual simultaneity judgments. *Perception & Psychophysics, 67*(3), 531–544. https://doi.org/10.3758/BF03193329