Exposure to multisensory and visual static or moving stimuli enhances processing of nonoptimal visual rhythms

Ourania Tachmatzidou[1,2] · Nadia Paraskevoudi[1] · Argiro Vatakis[2]

Abstract

Research has shown that visual moving and multisensory stimuli can efficiently mediate rhythmic information. It is possible, therefore, that the previously reported auditory dominance in rhythm perception is due to the use of nonoptimal visual stimuli. Yet it remains unknown whether exposure to multisensory or visual-moving rhythms would benefit the processing of rhythms consisting of nonoptimal static visual stimuli. Using a perceptual learning paradigm, we tested whether the visual component of the multisensory training pair can affect processing of metric simple two integer-ratio nonoptimal visual rhythms. Participants were trained with static (AVstat), moving-inanimate (AVinan), or moving-animate (AVan) visual stimuli along with auditory tones and a regular beat. In the pre- and posttraining tasks, participants responded whether two static-visual rhythms differed or not. Results showed improved posttraining performance for all training groups irrespective of the type of visual stimulation. To assess whether this benefit was auditory driven, we introduced visual-only training with a moving or static stimulus and a regular beat (Vinan). Comparisons between Vinan and Vstat showed that, even in the absence of auditory information, training with visual-only moving or static stimuli resulted in an enhanced posttraining performance. Overall, our findings suggest that audiovisual and visual static or moving training can benefit processing of nonoptimal visual rhythms.

Keywords Rhythm · Multisensory perception · Motion · Perceptual learning · Audiovisual

Introduction

Rhythm perception is considered by most as tightly associated with the auditory system (e.g., Grahn, 2012; Grahn et al., 2011; Grondin & McAuley, 2009). This seems counterintuitive given that most everyday activities that require efficient processing of temporally structured patterns are inherently multisensory (Ghazanfar, 2013; Grahn & Brett, 2009; Su & Pöppel, 2012). Consider, for example, the rhythmic information contained in the dancing or walking act. We dance by synchronizing our movements to the music and our partner or we maintain rhythmic gait by integrating visual, auditory, tactile, and proprioceptive feedback from the environment. To date, however, most studies on rhythmic processing have focused primarily on auditory rhythms, thereby largely ignoring the contribution of the other senses to rhythm perception.

Recently, a small number of studies have started to investigate the intramodal, as well as the crossmodal differences in rhythm perception and discrimination (Grahn, 2012; Grahn et al., 2011; Hove et al., 2013). Intramodal studies in the auditory domain have shown that the presence of a periodic beat that yields salient physical accents and gives rise to a clear metrical structure enhances auditory rhythm processing as compared with rhythms with irregular temporal structure (Grahn, 2012; Phillips-Silver & Trainor, 2007). Indeed, the beneficial impact of "hearing the beat" of a rhythm (i.e., the regular pulse that serves as a temporal anchor around which events are organized; Iversen et al., 2009) facilitates rhythm processing and encoding (Grahn, 2012; Su, 2014b), as well as motor synchronization (Gan et al., 2015; Grahn, 2012; Grahn & Brett, 2007). Crossmodal studies have also reported modality-dependent effects on rhythm processing. Research has consistently shown an auditory advantage in rhythm perception, which has been attributed to the more fine-grained temporal resolution of the auditory as compared with the visual system (Collier & Logan, 2000; Grahn, 2012; Grahn et al., 2011; Patel et al., 2005). For example, a periodic rhythm

✉ Argiro Vatakis
  argiro.vatakis@gmail.com

1 Department of History and Philosophy of Science, National and Kapodistrian University of Athens, Athens, Greece

2 Multisensory and Temporal Processing Lab (MultiTimeLab), Department of Psychology, Panteion University of Social and Political Sciences, 136 Syngrou Ave, 17671 Athens, Greece

can be efficiently processed by the auditory channel, while the same rhythm cannot be easily recognized when presented in the visual modality (Collier & Logan, 2000; Grahn et al., 2011; Patel et al., 2005). This is further supported by neuro-imaging data that have demonstrated increased activity of timing-related areas (i.e., basal ganglia, putamen) when processing an auditory rhythm as compared with a rhythm mediated by static visual flashes (Grahn et al., 2011; Hove et al., 2013).

Recent data, however, have challenged the currently supported visual inferiority in rhythm processing by demonstrating that rhythm discrimination performance is contingent upon the reliability of the stimulus presented (Gan et al., 2015; Grahn, 2012; Hove et al., 2013). Specifically, it has been suggested that the auditory dominance in rhythm processing may be partly due to the use of nonoptimal visual stimuli such as static flashes (Barakat et al., 2015) that lack spatiotemporal information, while motion—a more optimal visual stimulus (e.g., Ernst & Banks, 2002; Welch & Warren, 1980)—has been found to increase the temporal reliability of visual rhythm encoding (Gan et al., 2015; Grahn, 2012; Hove et al., 2013). The optimality of visual moving stimuli in rhythm perception was first investigated by Grahn (2012). Specifically, she directly compared auditory rhythms with visual rhythms with the latter being formed by a moving line. Three types of rhythmic patterns were used: (a) metric simple (i.e., integer-ratio rhythms with regular temporal accents that provide a clear metrical structure), (b) metric complex (i.e., integer-ratio rhythms with irregular temporal accents), and (c) nonmetric rhythms (i.e., non-integer-ratio rhythms with irregular temporal accents). In each trial, three rhythmic sequences were presented, and participants had to report whether the third sequence differed from the other two or not. The results showed higher accuracy for auditory trials as compared with visual ones, thus supporting the auditory advantage in rhythm processing (Collier & Logan, 2000; Patel et al., 2005). However, performance in the visual trials was also significantly improved, but this was only found for the metric simple rhythms and not the metric complex and nonmetric rhythms, indicating that visual rhythm encoding requires a clear metrical structure. Although these findings support the auditory advantage in rhythm processing, they also demonstrate that visual rhythm processing can also be enhanced when moving stimulation is used.

In addition to the beneficial impact of visual moving stimuli on rhythm perception, multisensory stimulation with visual components consisting of biological movement have also been found to affect the encoding and processing of rhythmic patterns (Su, 2014a, 2014b, 2016; Su & Salazar-López, 2016). Studies using point-light human figures along with auditory rhythmic patterns (Su, 2014a, 2014b, 2016; Su & Salazar-López, 2016) have shown improved discrimination accuracy for audiovisual metric simple (Su, 2014b) and metric complex

rhythms (Su, 2014a) as compared with auditory-only rhythms. This improvement is also in line with several studies reporting enhanced performance in multisensory as compared with unisensory trials (e.g., Alais & Cass, 2010; Roy et al., 2017; Shams et al., 2011). However, no study to-date has directly assessed whether animate and inanimate moving stimuli exert differential influences on rhythm discrimination given that two different mechanisms have been suggested to mediate temporal processing for these types of stimuli (Carrozzo et al., 2010).

Given the increasing evidence suggesting that certain types of visual (e.g., Grahn, 2012; Hove et al., 2013) and multisensory stimulation (e.g., Su, 2014a, 2014b) affect rhythm processing, exposure to such sensory rhythmic stimulation could potentially facilitate subsequent processing of visual rhythms. Facilitation of visual rhythms after training has recently been reported in a perceptual learning study by Barakat et al. (2015). Specifically, after receiving visual, auditory, or audiovisual training, participants in this study had to discriminate between two visual-only rhythmic sequences composed of visual empty intervals (i.e., demarcated by static flashes occurring at the onset and offset of each interval). Results showed that visual training did not contribute to an enhanced posttraining performance, while both the auditory and multisensory training groups were significantly better during the posttraining session. More importantly, these latter two groups did not differ in their posttraining performance, suggesting that multisensory training did not enhance rhythm perception more than the auditory training. One could, thus, argue that the posttraining enhancement observed was auditory-driven, while it remains unanswered whether the absence of posttraining improvement for the visual training group was due to the use of nonoptimal static stimuli (Grahn, 2012; Hove et al., 2013).

As far as we know, no study has yet examined the effects of modality and stimulus attributes such as visual movement or animacy on enhancing rhythm perception in a task consisting of static stimuli. Additionally, no attempts have been made to manipulate the animacy of the training stimulus and directly compare performance following exposure to animate and inanimate movement so as to assess whether the former benefits rhythm processing more than the latter. To address this gap, we examined whether multisensory training with different types of visual stimulation (i.e., static vs. moving and inanimate vs. animate) yield differential learning effects in a subsequent visual rhythm discrimination task consisting of static visual stimuli. We hypothesized that training with audiovisual rhythms, particularly those containing visual movement, would improve the processing of visual static rhythmic patterns due to the visual system's high spatial resolution and enhanced motion processing (e.g., Hove et al., 2013; Welch & Warren, 1980). Furthermore, we reasoned that if biological motion has a beneficial impact on rhythm processing (Su,

2014a, 2014b), then training with auditory rhythms accompanied by animate visual movement would yield better discrimination performance in a subsequent visual-only rhythm discrimination task as compared with training with moving, yet inanimate, visual stimuli.

Experiment 1

Methods

Participants

Fifty-three university students (47 female) aged between 19 and 48 years (mean age = 24 years) took part in the experiment. All participants reported having normal or corrected-to-normal vision and normal hearing. All were naïve as to the purpose of the experiment. The experiment was performed in accordance with the ethical standards laid down in the 2013 Declaration of Helsinki and informed consent was obtained from all participants. To control for potential confounding factors, participants with extensive (over 5 years) musical and/or dance training were removed from further analysis (cf. Grahn & Rowe, 2009; Iannarilli et al., 2013).

Apparatus and stimuli

The experiment was conducted in a dimly lit and quiet room. The visual stimuli were presented on a CRT monitor with 60 Hz refresh rate, while the auditory stimuli were presented using two loudspeakers (Creative Inspire 265), placed to the left and right of the monitor. The experiment was programmed using OpenSesame (Version 3.1; Mathôt et al., 2011).

Three types of visual stimuli were utilized to create the visual stream of the audiovisual rhythmic sequences: (a) red and green static circles, (b) a moving bar, and (c) a human point-light figure (PLF). Both the static circles and the moving bar were created using Adobe Illustrator CS6. The moving bar was designed with six different orientations, each one pointing to a different position (separated by approximately 30°) around a central axis of rotation, so that apparent movement could be induced when presented sequentially (cf. Grahn, 2012). The PLF was adopted from the Atkinson et al.'s (2004) stimulus set and was processed in Adobe Premiere Pro CS5. The PLF moved vertically, starting from an upright position, then bending down and, finally, returning to its initial position. We used this kind of movement, since vertical human body movements have been suggested to mediate rhythm more efficiently than horizontal body movements (Nesti et al., 2014; Toiviainen et al., 2010). All stimuli can be accessed online (https://osf.io/pzc2t/).

The auditory stream of the rhythmic sequences utilized was created using Audacity and was composed of two types: (a) a sinewave tone (sampling frequency 44110 Hz) of 43 ms in duration and (b) a pink noise (sampling frequency 44110 Hz) of 50 ms in duration. The former sound was used to create the auditory rhythmic patterns, while the latter the beat sequences. Both the auditory tones and the beat stimuli were presented at 76 dB (as measured from the participant's ear position).

Design

The experiment took place in two separate days (24 to 48 hours apart), with the experimentation of each day lasting approximately 50 minutes. The experiment consisted of four sessions: a) a pretraining session, b) two training sessions, and c) a posttraining session (cf. Barakat et al., 2015; see Fig. 1). For all sessions, the participants completed a two alternative forced choice (2AFC) rhythm discrimination task (i.e., 'same' or 'different'). In each trial, a fixation point was initially presented for 1000 ms followed by the first rhythmic pattern (i.e., "standard"). After 1,100 ms (interstimulus interval; ISI), the second rhythmic sequence (i.e., "comparison") was presented and participants provided a self-paced response. The intertrial interval (ITI) was set at 1,200 ms.

The rhythmic sequences used were metric simple rhythms (cf. Grahn, 2012; Grahn & Brett, 2007) that consisted of six elements of either a short (400 ms) or a long interval (800 ms). The intervals were, thus, related by integer ratios, where 1 = 400 ms and 2 = 800 ms, and had a regular grouping with the beat occurring regularly every two units (cf. Drake, 1993)—that is, every 800 ms (interbeat interval, IBI; cf. Grahn & Brett, 2007). Five rhythmic sequences were used as 'standard' and 'comparison' intervals (i.e., Rhythm A: 111122, B: 112112, C: 112211, D: 211211, and E: 221111), resulting in a factorial 5 × 5 design with 25 rhythm pairs in total (i.e., AA, AB, AC, AD, AE, BA, BB, BC, BD, BE, CA, CB, CC, CD, CE, DA, DB, DC, DD, DE, EA, EB, EC, ED, and EE).

On Day 1 of the experiment, participants completed a set of five practice trials to familiarize themselves with the task. Subsequently, they all performed the pretraining session and the first training session. On Day 2, participants started with the second training session that was followed by the posttraining test (that was identical to the pretraining test).

The pre- and posttraining sessions were composed of visual-only rhythms that consisted of static circles of changing colours (see Fig. 1). During each trial, a circle appeared on the screen and lasted for the whole duration of the respective interval (i.e., 400 or 800 ms). Once the first interval ended, the circle changed colour (green or red, based on the previous circle), which represented the onset of the next element of the rhythmic sequence. Each one of the two rhythms in a given trial consisted of six elements (i.e., six circles). The pre- and posttraining sessions consisted of four repetitions of each rhythm pair, resulting in 100 trials per session in total. Each session lasted approximately 30 minutes.
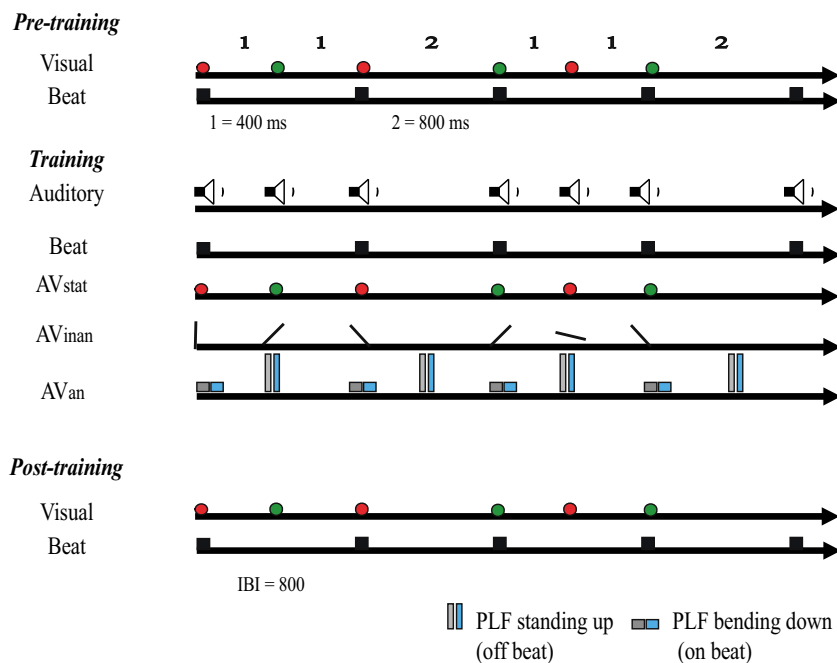
Fig. 1 Schematic of the design and the stimuli used for the rhythmic patterns in Experiment 1. The rhythmic pattern shown here is for the 6-interval 112112 rhythm with 1 = 400 ms and 2 = 800 ms. All participants initially performed a pretraining session consisting of static circles (red and green ellipses) and a regular beat occurring every 800 ms (black square). They were, subsequently, randomly assigned to one of the three training groups and received two training sessions that were separated by a day. The first group (AVstat) was trained with auditory tones and static circles of changing colours (red and green ellipses). The second group (AVinan) was presented with auditory tones and a bar "moving" to different screen locations (black line). The third group (AVan) was trained with auditory tones and a human point-light figure starting in an upright position (grey-blue lines) and then bending down (grey-blue squares). After training, all participants completed the final posttraining session that was the same as the pretraining. (Colour figure online)

For the training phase, all participants were randomly assigned to one of three training groups: audiovisual static circles (AVstat), moving bar (i.e., inanimate stimulus; AVinan), or PLF (i.e., animate stimulus; AVan) group. During the training sessions, participants received feedback for their responses. Each training session included 3 repetitions of each rhythm pair (i.e., 75 trials per session in total) and lasted approximately 20 minutes. The auditory stimulation was the same across the three training groups, with the auditory tones occurring at the onset and offset of each interval, and the beat being presented every 800 ms. The first group (AVstat; N =16, 15 female, age range: 19–38, mean age = 25.2 years) was trained with rhythms consisting of auditory tones and static circles of changing colours. The presentation of the circles was the same as in the pre- and posttraining with the sole exception that, here, the onset of each circle was accompanied by an auditory tone. The second group (AVinan; N = 20, 15 female, age range: 19–48, mean age = 23.5 years) was trained with audiovisual rhythms consisting of a moving bar (cf. Grahn, 2012) that was accompanied by auditory tones. In this case, a line was initially presented in a vertical position and once the rhythmic pattern started, the line changed positions sequentially around a central axis of rotation. The third group (AVan; N = 15, 15

female, age range: 21–39, mean age = 21.9 years) was trained with a human PLF with each PLF cycle lasting 800 ms. Thus, the transition from the upright position to the lowest position of the PLF and the reverse lasted 400 ms each, that is, the beat always occurred at the lowest position of the PLF as suggested by previous studies (cf. Su, 2014a, 2014b).

Procedure

The participants received detailed verbal instructions prior to the start of the experiment, and they were allowed to ask for any clarification. Prior to the start of the experiment, participants completed a practice session to familiarize themselves with the task. They, subsequently, performed the pretest and the first training session (Day 1) that was followed by the second training session and the posttest in Day 2. Participants self-initiated each session. Once both sequences were presented, they were instructed to report as accurately as possible whether the two rhythms differed or not, by pressing the keys "m" and "z" of the keyboard, respectively. Participants were informed that during the training sessions response feedback would be provided, while this would not be

the case for the pre- and posttest. Finally, all participants were allowed to take a break between the experimental sessions.

Results and discussion

Two participants were removed from the analysis due to formal musical and dance training. For all the analyses, Bonferroni-corrected t tests (where $p < .05$ prior to correction) were used for all post hoc comparisons. When sphericity was violated, Greenhouse–Geisser correction was applied. The alpha level was set to 0.05 and the confidence interval to 95%. Moreover, 90% confidence intervals around eta partial square (Steiger, 2004) are reported to facilitate future researchers (Thompson, 2002) and to provide further information on the sufficiency of the sample size (Calin-Jageman, 2018).

Training data

The training data (i.e., percentage correct detections of "same" or "different" rhythmic pairs) were analyzed via a mixed analysis of variance (ANOVA) with Training Session (2 levels: Session 1 vs. Session 2) and Rhythm Pair (25 levels) as the within-participant factors, and Group (3 levels: AVstat, AVinan, AVan) as the between-participants factor. The analysis showed a significant main effect of Group, $F(2, 48) = 5.94$, $p = .005$, $\eta_p^2 = .20$, CI [.04, .334], with the AVstat group

Training Experiment 1

Fig. 2 Mean discrimination accuracy during the two training sessions for the three training groups (AVstat, AVinan, AVan) in Experiment 1. Significant differences between the groups ($p < .05$) are indicated by the asterisk. The error bars represent the standard error of the mean

performing significantly better ($M = .916$) as compared with both the AVinan ($M = .819$) and the AVan ($M = .820$) group (see Fig. 2). The higher performance of the AVstat group could be attributed to the prior exposure to the pretraining session (i.e., identical visual stimulation); however, it should be noted that both AVinan and AVan groups also reached high levels of performance accuracy with a mean accuracy over 80%. A significant main effect of Training Session was also obtained, $F(1, 48) = 12.39$, $p < .001$, $\eta_p^2 = .21$, CI [.058, .335], with all groups having higher accuracy scores during the second training session ($M = .868$) as compared with the first ($M = .832$). Thus, showing that even one training session was sufficient to yield higher discrimination accuracy for all three groups. We also obtained a significant main effect of Rhythm Pair, $F(10.97, 526.55) = 11.61$, $p < .001$, $\eta_p^2 = .19$, CI [.132, .228], with certain pairs having systematically lower accuracy ($M_{AA} = .777$, $M_{BB} = .791$, $M_{BC} = .725$, $M_{CB} = .693$, $M_{CC} = .771$, $M_{CD} = .722$, $M_{DD} = .761$, $M_{DE} = .794$) as compared with others that were significantly easier to discriminate ($M_{AC} = .905$, $M_{AD} = .902$, $M_{AE} = .918$, $M_{BA} = .915$, $M_{BE} = .938$, $M_{CA} = .928$, $M_{CE} = .918$, $M_{DA} = .951$, $M_{EA} = .961$, $M_{EB} = .908$). The data showed that the rhythms B, C, and D (i.e., 112112, 112211, and 211211, respectively) were particularly difficult to discriminate in certain types of pairing, yet the performance was still above chance level.

A significant interaction between Rhythm Pair and Group was obtained, $F(21.94, 526.55) = 2.10$, $p = .003$, $\eta_p^2 = .08$, CI [.014, .080], with the AVstat group having significantly higher accuracy scores in some rhythm pairs ($M_{AE} = 1$, $M_{BA} = .969$, $M_{CE} = .990$, $M_{DA} = .990$, $M_{DD} = .885$, $M_{EA} = 1$, $M_{EB} = .979$) as compared with the AVan ($M_{AE} = .900$, $M_{BA} = .844$, $M_{DA} = .889$, $M_{EA} = .922$, $M_{EB} = .844$) and AVinan ($M_{AE} = .867$, $M_{CE} = .867$, $M_{DD} = .675$) groups, while AVan performed significantly worse than the other two groups when the rhythm pair was BC (AVan = .500, AVinan = .875, AVstat = .750). Furthermore, a significant interaction between Training Sessions and Rhythm Pair was also obtained, $F(13.56, 651.04) = 1.74$, $p = .046$, $\eta_p^2 = .04$, CI [.002, .040]. The accuracy scores of numerous rhythm pairs were significantly lower during the first Training Session ($M_{AD} = .837$, $M_{AE} = .856$, $M_{BA} = .869$, $M_{CA} = .876$) in comparison to the second ($M_{AD} = .967$, $M_{AE} = .980$, $M_{BA} = .961$, $M_{CA} = .980$). The interactions between Training Session and Group, $F(2, 48) = 0.63$, $p = .539$, $\eta_p^2 = .03$, and Group, Training Session, and Rhythm Pair, $F(27.13, 651.04) = 1.27$, $p = .162$, $\eta_p^2 = .05$, did not reach significance.

Pre- and posttraining data

For the main analysis, we compared the pre- and posttraining performance to test for potential learning effects following training. The pre- and posttest responses were analyzed via a
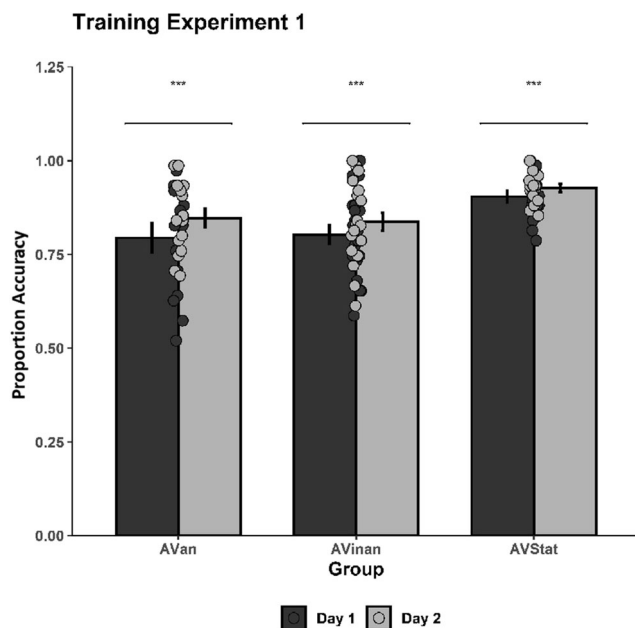
mixed ANOVA with Session (2 levels: Pretraining vs. Posttraining) and Rhythm Pair (25 levels) as within-participant factors, and Group (3 levels: AVstat, AVinan, AVan) as between-participants factor. A significant main effect of Session was obtained, $F(1, 48) = 59.23$, $p < .001$, $\eta_p^2 = .55$, CI [.383, .655], with all groups performing better ($M = .742$) during posttraining as compared with the pretraining session ($M = .639$; see Fig. 3). A significant main effect of Rhythm Pair was also obtained, $F(13.44, 645.08) = 13.21$, $p < .001$, $\eta_p^2 = .22$, CI [.156, .245], with certain rhythm pairs being more accurately discriminated (i.e., $M_{AE} = .819$, $M_{EA} = .833$, $M_{EC} = .821$, $M_{EE} = .836$) as compared with others (i.e., $M_{BC} = .537$, $M_{CB} = .480$, $M_{CD} = .615$, $M_{DC} = .620$, $M_{DE} = .561$).

Further examination of this effect showed that Rhythm E (i.e., 221111) was easier to discriminate from other rhythms suggesting that this rhythmic pattern was more efficiently processed and maintained in memory as compared with the other rhythmic sequences. This was not the case for pairs including the Rhythms B (i.e., 112112), C (i.e., 112211), and D (i.e., 211211) that lead the participants to lower discrimination accuracy. No main effect of Group was obtained, $F(2, 48) = 0.66$, $p = .521$, $\eta_p^2 = .03$], while the interactions between Group and Session and between Group, Session, and Rhythm Pair did not reach significance, $F(2, 48) = 0.01$, $p = .988$, $\eta_p^2 = .01$, and $F(26.98, 647.50) = 1.27$, $p = .162$, $\eta_p^2 =$

Fig. 3 Mean discrimination accuracy during the two main sessions (pre- and posttraining) for the three training groups (AVstat, AVinan, AVan) in Experiment 1. Significant differences between the pre- and posttest sessions ($p < .001$) are indicated by two asterisks. The error bars represent the standard error of the mean

.05, respectively. These findings suggest that the type of the visual component presented during training did not modulate posttraining performance. However, we found a significant interaction between Group and Rhythm Pair, $F(26.88, 645.08) = 1.68$, $p = 0.17$, $\eta_p^2 = .07$, CI [.005, .058]. Specifically, although performance was equal between groups, the AVstat and AVinan groups differed significantly in their discrimination accuracy for the rhythm CC ($M = .820$ and .650, respectively). We also found a significant interaction between Session and Rhythm Pair, $F(13.49, 647.50) = 4.46$, $p < .001$, $\eta_p^2 = .09$, CI [.037, .102], with 18 out of 25 rhythm pairs being more accurately discriminated in the posttest as compared with the pretest. These findings demonstrate that training had a beneficial impact on discrimination performance for most rhythmic patterns.

Overall, the results of Experiment 1 showed that the different training stimulation utilized resulted in similar posttest performance for all training groups, despite the main effect of group during training. That is, irrespective of the training stimulus type (i.e., static or moving, animate or inanimate), the multisensory perceptual training implemented enhanced the processing of subsequently presented visual-only, static rhythms. The absence of group differences could be attributed to the fact that the auditory stimulation was identical for all training groups. Thus, it could be the case that the benefit obtained after training was driven solely or mainly due to the contribution of audition, thereby providing support for the modality appropriateness hypothesis (i.e., theory supporting that the most reliable modality will dominate the final percept depending on the task utilized; Welch & Warren, 1980). An alternative explanation of our findings could be the ease of the task. The results showed that, even during the pretraining session, most participants exhibited high discrimination accuracy, which could be due to low task difficulty. This ease of rhythm discrimination could be attributed either to the presence of the explicit beat (cf. Su, 2014a), the low complexity of the rhythms presented (that consisted of only two interval types; i.e., 400 and 800 ms; cf. Barakat et al., 2015; Drake, 1993), or the rhythm type utilized that had a clear metrical structure (i.e., metric simple rhythms; cf. Grahn, 2012; Su, 2014b).

The potential contribution of audition to the posttraining enhancements observed in Experiment 1 led us to a follow-up experiment to further examine the contribution of audition during training. We reasoned that if the posttraining improvement in Experiment 1 resulted from the presence of auditory information, then training with a visual-only moving stimulus would not be sufficient to yield this enhancement in posttraining performance when compared with the multisensory case of Experiment 1 (cf. Barakat et al., 2015). If, however, visual motion can mediate the rhythmic information needed for increasing discrimination accuracy, then the posttraining performance
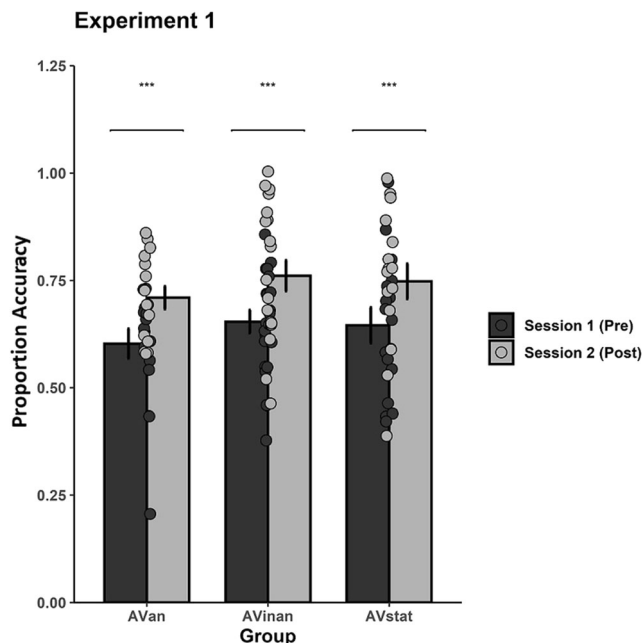
could be enhanced following training with visual-only moving stimuli. In Experiment 2, therefore, we kept the experimental structure and design of Experiment 1 with the sole difference of the training stimulation, which was now composed of visual-only rhythmic patterns. Specifically, we trained participants with the moving bar and static circles utilized in Experiment 1 in the absence of auditory tones (i.e., Vinan and Vstat group).

Experiment 2

Methods

Participants

Twenty-two new university students (18 female) aged between 19 and 20 years old (mean age = 19.5 years) took part in this experiment.

Apparatus, stimuli, design, and procedure

These were the same as for Experiment 1 with the sole exception that instead of being trained with multisensory rhythms, participants received a unisensory training with a moving bar (Vinan) or static circles (Vstat), where the visual stimulus was presented without the auditory tones at the onset and offset of each interval (see "Stimuli" for Experiment 1). The beat sequence was maintained (i.e., IBI = 800 ms). We used the moving bar as one of our training stimuli due to Grahn's (2012) findings of visual moving stimuli mediating rhythmic information. We also used the static circles to further support our results from Experiment 1 by showing that the outcomes are, indeed, not driven by audio.

Results and discussion

For all the analyses, Bonferroni-corrected t-tests (where $p < .05$ prior to correction) were used for all post hoc comparisons. When sphericity was violated, Greenhouse–Geisser correction was applied. The alpha level was set to 0.05 and the confidence interval to 95%.

Training data

A mixed ANOVA with Training Session (2 levels: Session 1 vs. Session 2) and Rhythm Pair (25 levels) as the within-participant factors, and Group (Vstat, Vinan) as the between-participants factor was conducted. A significant main effect of Group, $F(1, 19) = 1.894$, $p < .001$, $\eta_p^2 = .091$, CI [NA, .350] was obtained, with the static group performing significantly better ($M = .746$) as compared with the moving-bar group ($M = .663$; see Fig. 4). A main effect of Rhythm Pair was also
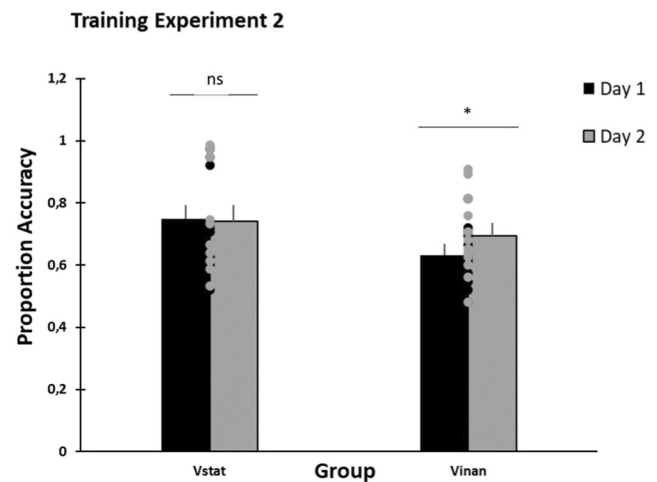

Training Experiment 2

Fig. 4 Mean discrimination accuracy during the two training sessions for the static circles (Vstat) and the moving bars (Vinan) training groups. Significant differences between the groups and the training sessions ($p < .001$) are indicated by two asterisks. The error bars represent the standard error of the mean

obtained, $F(24, 456) = 1.889$, $p = .007$, $\eta_p^2 = .090$, CI [.007, 094], with some rhythmic pairs being particularly difficult to discriminate ($M_{BA} = .630$, $M_{BB} = .616$, $M_{BC} = .597$, $M_{BD} = .674$, $M_{BE} = .657$, $M_{CB} = .653$, $M_{CC} = .664$). We also obtained an interaction between Group and Rhythm Pair, $F(24, 456) = 1.921$, $p = .006$, $\eta_p^2 = .092$, CI [.008, .096], with some rhythm pairs being more accurately discriminated by the static training group (i.e., AB, AE, CA, DA, DE, EA, EB) as compared with the bar training group. We also obtained an interaction between Group and Training Session, $F(1, 19) = 6.186$, $p = .022$, $\eta_p^2 = .246$, CI [.002, .499]). The interactions between Training Session and Rhythm, $F(24, 456) = 1.234$, $p = .206$, $\eta_p^2 = .061$, and Training Session, Group, and Rhythm Pair, $F(24, 456) = .809$, $p = .727$, $\eta_p^2 = .041$, did not reach significance.

Pre- and posttraining data

A mixed ANOVA with Session (2 levels: Pretraining vs. Posttraining) and Rhythm Pair (25 levels) as within-participant factors, and Group (2 levels: Vstat, Vinan) as the between-participants factor was conducted. A significant main effect of Session was obtained, $F(1, 19) = 31.439$, $p < .001$, $\eta_p^2 = .623$, CI [.289, .762], with both groups exhibiting a significant improvement in posttraining performance ($M = .718$) as compared with the pretraining ($M = .691$; see Fig. 5). Thus, despite the absence of the auditory tone stimulation in Experiment 2, training with visual-only moving stimuli continued to enhance posttraining performance in a task where the rhythms consisted of static visual stimuli.

We also obtained a significant interaction between Session and Rhythm Pair, $F(24, 456) = 4.186$, $p < .001$, $\eta_p^2 = .181$, CI [.081, 200], with 16 out of the 25 rhythm pairs being
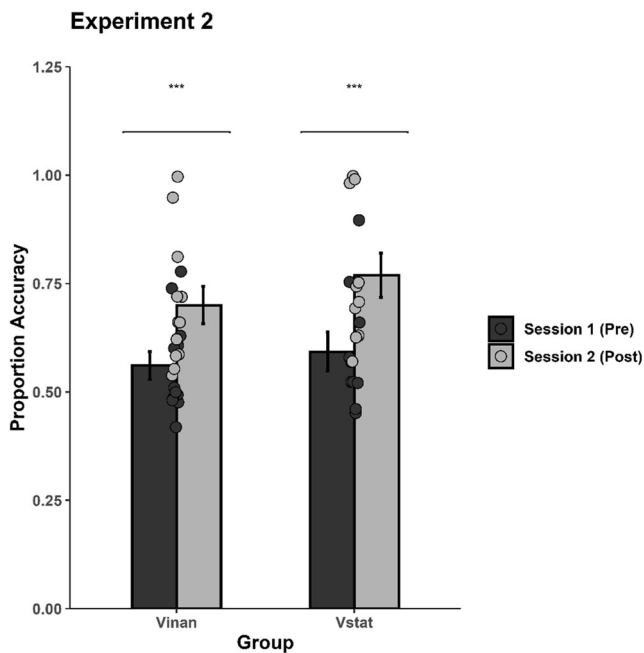
Experiment 2

Fig. 5 Mean discrimination accuracy during the pre- and posttraining sessions for the static circles (Vstat) and the moving bars (Vinan) training groups. Significant differences between the pre- and posttest sessions (p < .001) are indicated by two asterisks. The error bars represent the standard error of the mean

significantly more accurately discriminated during the posttraining as compared with the pretraining (i.e., AB, AD, BA, BC, BD, BE, CA, CB, CE, DA, DB, DC, EA, EB, EC, ED). The main effect of Group, $F(1, 19) = 1.370, p = .256, \eta_p^2 = .067$, Rhythm, $F(24, 456) = 1.467, p = .061, \eta_p^2 = .073$, and the interactions between Session and Group, $F(1, 19) = .284, p = .60, \eta_p^2 = .015$, Rhythm Pair and Group, $F(24, 456) = 1.467, p = .072, \eta_p^2 = .072$, and Session, Rhythm Pair, and Group, $F(24, 456) = 1.267, p = .180, \eta_p^2 = .063$, did not reach significance.

Overall, the results of Experiment 2 demonstrated that visual-only training with either a static or moving stimulus can enhance rhythm perception even in the absence of auditory rhythmic stimulation. In particular, training with a visual-only moving (i.e., a moving bar) or static stimulus (i.e., static circles) improved the participants' processing and discrimination ability of metric simple visual rhythms consisting of static stimuli. More importantly, we did not find any enhancement differences between visual static and visual moving training, suggesting that both visual stimuli are sufficient to improve discrimination accuracy of the two integer-ratio visual rhythms we used in the pre- and posttraining sessions. One potential explanation for the enhancement after the visual moving training could also be that the intervals between targets could have helped participants to accurately predict their location (Pfeuffer et al., 2020; Wagener & Hoffmann, 2010). Moreover, participants are able of learning such complex spatiotemporal patterns and this process most likely is implicit

(Kirkham et al., 2007). However, until now, training with nonoptimal rhythms has not been found sufficient to improve rhythm discrimination (Barakat et al., 2015; Zerr et al., 2019). Thus, it is possible that the similarity of visual static stimuli with the main task stimuli influenced the participants' performance.

To compare task performance after unisensory and multi-sensory training, we performed a combined analysis of the data from Experiment 1 (AVstat, AVinan) and Experiment 2 (Vstat, Vinan). For the training data, we used a mixed ANOVA with Training Session (2 levels: Session 1 vs. Session 2) and Rhythm Pair (25 levels) as within-participant factors, and Group (four levels: AVstat, AVinan, Vstat, Vinan) as the between-participants factor. We obtained a main effect of Training Session, $F(1, 53) = 11.751, p = .001, \eta_p^2 = .181$, CI [.032, .353] with the participants' performance being significantly better in Session 2 ($M = .800$) than in Session 1 ($M = 772$). We also obtained a main effect of Rhythm Pair, $F(12.876, 682.429) = 6.527, p < .001, \eta_p^2 = .110$, CI [.055, .139], with some rhythms having systematically lower accuracy ($M_{CB} = .681, M_{CD} = .688, M_{DD} = .679$) compared with others ($M_{AD} = .871, M_{AE} = .856, M_{EB} = .880, M_{EC} = .858$). A significant main effect of Training group was observed, $F(1, 53) = 14.179, p < .001, \eta_p^2 = .445$, CI [.048, .382], with the groups AVstat ($M = .916$) and AVinan ($M = .819$) performing better than Vstat ($M = .746$) and Vinan ($M = .663$). The interaction between Rhythm pair and Training group was also found significant, $F(38.628, 682.429) = 2.201, p < .001, \eta_p^2 = .111$, CI [.025, .105], with the training groups discriminating more accurately specific rhythms. The interactions between Training Session and Group, $F(3, 53) = 2.461, p = .073, \eta_p^2 = .122$, Training Session and Rhythm pair, $F(1.809, 59.119) = 1.622, p = .063, \eta_p^2 = .030$, and Training Session, Rhythm Pair, and Group, $F(3.869, 59.119) = 44.604, p = .228, \eta_p^2 = .061$, were not found significant.

For the pre- and posttraining data, we conducted a mixed ANOVA with Session (2 levels: Pretraining vs. Posttraining) and Rhythm Pair (25 levels) as within-participant factors, and Group (4 levels: AVstat, AVinan, Vstat, Vinan) as the between-participants factor. A significant main effect of Session was obtained, $F(1, 54) = 77.224, p < .001, \eta_p^2 = .590$, CI [.407, .696], with all groups performing better posttraining ($M = .745$) than pretraining ($M = .613$). A main effect of Rhythm Pair was also obtained, $F(10.571, 77.682) = 6.527, p < .001, \eta_p^2 = .120$, CI [.230, .535], with some rhythm pairs being particularly difficult to discriminate ($M_{BC} = .579, M_{BB} = .616, M_{CB} = .522, M_{DE} = .585$). Additionally, an interaction between Session and Rhythm Pair, $F(7.960, 66.459) = 6.468, p < .001, \eta_p^2 = .107$, CI [.189, .520], and Rhythm Pair and Training Group, $F(7.221, 77.682) = 1.673, p = .006, \eta_p^2 = .085$, CI [NA, .207], was also obtained. The main effect of Training group and the interactions between Session and Training Group, $F(3, 54) = 1.198, p = .319, \eta_p^2$

= .062, and Session, Rhythm Pair, and Training Group, $F(4.806, 66.459) = 1.302$, $p = .100$, $\eta_p^2 = .067$, did not reach significance. The results of this combined analysis further support our previous findings that unisensory training with moving or static stimuli can indeed improve rhythm discrimination accuracy.

So far, it is not clear whether the results obtained in the two experiments described were due to perceptual learning or due to the mere exposure to the experimental conditions. Perceptual learning can be defined as the performance enhancement on a task, emerging from prior perceptual experience (Watanabe & Sasaki, 2015). This enhanced performance comes, commonly, because of a training with feedback period (Dosher & Lu, 2017; Hammer et al., 2015). However, it is possible that improvement can also occur after a period of mere exposure to numerous stimulus features (Liu et al., 2010; Watanabe et al., 2001). To address this potential "mere exposure" confound, we conducted a control experiment to further examine whether an effect of session (pre- vs. posttraining) will be obtained in the absence of feedback during training. If this effect is present in the control experiment, the results obtained in Experiment 1 and Experiment 2 may simply reflect mere exposure effects rather than learning processes. However, if no posttraining enhancements occur in the absence of feedback, our effects can be attributed to perceptual learning processes.

Twenty-three university students (21 females) aged between 20 and 37 years old (mean age = 22.5 years) took part in this control experiment. We kept stimuli, design, and procedure identical to Experiment 1 with only two exceptions: we removed feedback during the training sessions, and we only utilized the AVinan condition in order to reduce experimentation time. The training data were analyzed using a repeated-measures ANOVA between Session (two levels: Session 1, Session 2) and Rhythm Pair (25 levels). The alpha level was set to 0.05 and the confidence interval to 95%. The analyses showed no statistically significant main effect of Session, $F(1, 22) = 2.49$, $p = .129$, $\eta_p^2 = .10$, with performance during Session 2 ($M = .769$) remaining at the same level with Session 1 ($M = .746$; see Fig. 6). A main effect of Rhythm Pair was obtained, $F(24, 528) = 7.06$, $p < .001$, $\eta_p^2 = .24$, CI [.159, .260], with some rhythmic pairs being particularly difficult to discriminate (i.e., AC, BC, CD, DC, DE, ED). The interaction between Training Session and Rhythm Pair, $F(24, 528) = 0.96$, $p = .524$, $\eta_p^2 = .04$, did not reach significance.

The pre- and posttraining data were analyzed by using a repeated-measures ANOVA between Session (two levels: Pretraining vs. Posttraining) and Rhythm Pair (25 levels). The main effect of Session did not reach significance, $F(1, 22) = 0.52$, $p = .480$, $\eta_p^2 = .02$ (see Fig. 7), so in the absence of feedback during training we did not obtain a significant improvement in performance. However, a main effect of Rhythm Pair was revealed, $F(9.14, 201.12) = 10.40$, $p <$
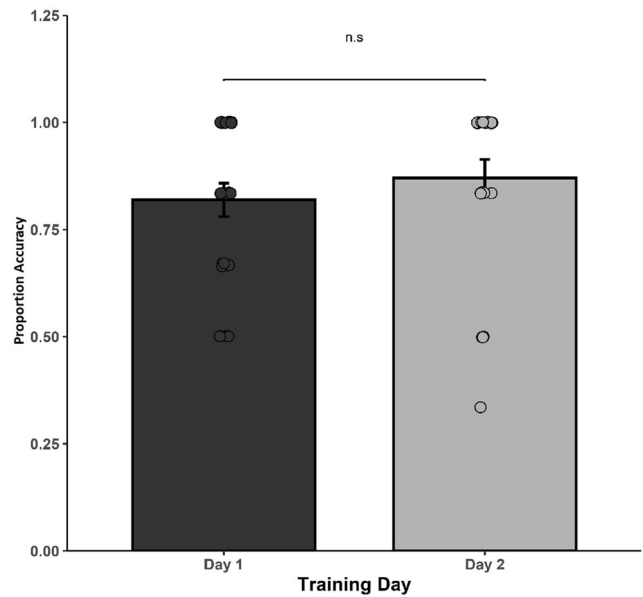

Fig. 6 Mean discrimination accuracy during the two training sessions for the control group. Significant differences between the training sessions ($p < .001$) are indicated by two asterisks. The error bars represent the standard error of the mean

.001, $\eta_p^2 = .32$, CI [.207, .372], with some rhythm pairs having systematically lower accuracy scores (i.e., AB, BA, BC, BD, CB, CD, DB, DC, DE, ED; see also Fig. 8 for a summation on rhythms mean accuracy). The interaction between Rhythm Pair and Session, $F(10.33, 227.24) = 1.26$, $p =$
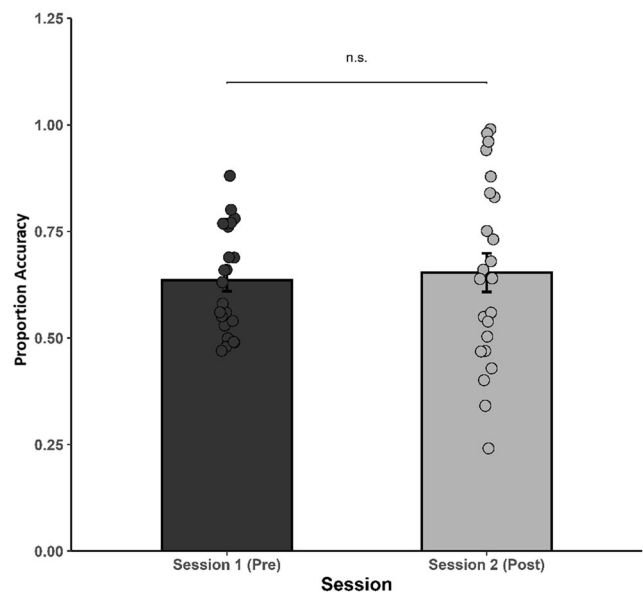

Fig. 7 Mean discrimination accuracy during the pre- and posttraining sessions for the control group. Significant differences between the pre- and posttest sessions ($p < .001$) are indicated by two asterisks. The error bars represent the standard error of the mean
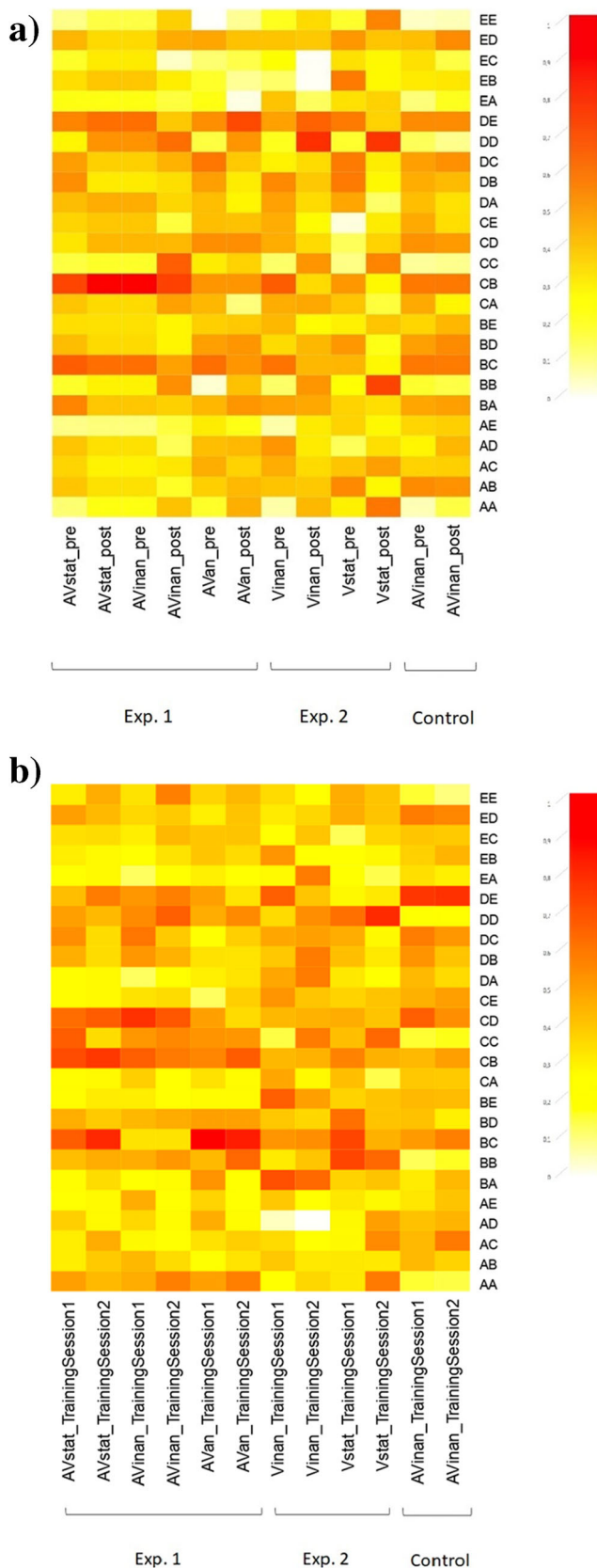
◄ Fig. 8 The above heatmaps depict (**a**) the mean accuracies of the 25 rhythm pairs used in the pre- and posttraining sessions, grouped by training group (AVstat, AVinan, Avan, Vinan, Vstat) and session (Pre-, Posttraining) for Experiments 1 and 2 and the Control experiment, (**b**) the mean accuracies of the 25 rhythm pairs used in each training group, grouped by training group (AVstat, AVinan, Avan, Vinan, Vstat) and session (Training Session 1, Training Session 2) for Experiments 1 and 2 and the Control experiment. The lower mean accuracies are associated with lighter colorings (i.e., yellow) while the higher mean accuracies are associated with darker colorings (i.e., red). (Colour figure online)

.253, $\eta_p^2 = .05$, did not reach significance. Overall, the results suggest that the findings we obtained in Experiments 1 and 2 can be attributed to perceptual learning processes, since the same design, but without feedback (supported as essential to learning; Goldhacker et al., 2013), did not yield significant results.

General discussion

In the present study, we used a rhythm perceptual learning paradigm, where we manipulated the type of the visual stimulus (i.e., moving vs. static and animate vs. inanimate; Experiment 1) and the training modality (Experiment 2) so as to investigate the potential of posttraining enhancement of visual rhythm processing. Our results showed that visual rhythm perception can be enhanced when using both moving and static and/or animate and inanimate stimuli in training, when the rhythmic information is mediated by audiovisual (Experiment 1) or visual only stimuli (Experiment 2), but only if trial-by-trial feedback is provided during training (Control experiment).

The main aim of Experiment 1 was to assess the effects of stimulus attributes during perceptual training (e.g., visual movement or animacy) on the discrimination of subsequently presented rhythms consisting of nonoptimal static stimuli. Previous studies employing discrimination tasks have reported that compared with visual rhythms consisting of static stimuli, exposure to visual stimuli with spatiotemporal information (e.g., visual movement) may increase the temporal reliability of visual rhythm encoding (Gan et al., 2015; Grahn, 2012; Hove et al., 2013), and, thus, facilitate rhythm processing (Grahn, 2012; Hove et al., 2013). Given that no studies have tested the effects of visual movement on rhythm perceptual learning, we reasoned that the latter findings may explain why perceptual learning effects are not observed following training with visual-only, nonoptimal, static rhythms (Barakat et al., 2015). To address this possibility, we manipulated the visual component of the multisensory rhythms during training. However, contrary to our predictions, the results obtained in Experiment 1 showed that all three different forms of training led to a significantly better posttraining performance,

irrespective of the type of the visual stimulation presented. Consistent with previous evidence reporting an auditory advantage in rhythm processing (Collier & Logan, 2000; Patel et al., 2005) and auditory-driven training effects in perceptual learning tasks (Barakat et al., 2015), one possible explanation of our findings is that the similar performance between the different training groups could be driven by the multisensory benefit during training, and, in particular, from the contribution of audition to the multisensory stimulus presented, in agreement with the optimal integration hypothesis (Ernst & Banks, 2002).

A secondary aim of Experiment 1 was to investigate whether animate or inanimate moving stimuli exert differential influences on subsequent visual rhythm processing. However, contrary to previous findings supporting that biological motion affects time estimates (Blake & Shiffrar, 2007; Carrozzo et al., 2010; Lacquaniti et al., 2014; Mendonça et al., 2011; Orgs et al., 2011), facilitates temporal prediction of actions as compared with inanimate moving stimuli (Stadler et al., 2012), and improves synchronization and rhythm discrimination accuracy (Su, 2014a, 2014b, 2016; Su & Salazar-López, 2016; Wöllner et al., 2012), we did not observe any animacy-related enhancements in Experiment 1. This could potentially be attributed to the stimulus design adopted (i.e., Su, 2014b). Specifically, as in Su's study, the PLF's movement in Experiment 1 had a fixed timing (i.e., the repetitive movement lasted 500 ms in Su's study and 800 ms in our study), thus not mediating the duration of the accompanying auditory rhythm, which consisted of intervals of different durations (i.e., 250, 500, 750, or 1000 ms and 400 or 800 ms in Su's and our study, respectively). This may have minimized the potential of observing an animacy-driven benefit (AVan) when comparing the performance between the different multisensory training groups. The absence of animacy-related enhancements in our study raises the possibility that the beneficial impact of biological motion reported by Su (2014b) could be due to modality differences that resulted from comparing performance between multisensory (auditory tones and PLF) and auditory-only rhythms. We, therefore, suggest that the enhancements reported by Su in the multisensory as compared with the auditory-only trials were not due to the beneficial impact of biological motion on rhythm processing but instead due to the behavioral benefits associated with multisensory as compared with unisensory stimulation (e.g., Huang et al., 2012; Stein & Stanford, 2008). Future studies need to account for the temporal aspects of auditory and visual rhythms that are mediated by biological motion to gain a better understanding of the potential animacy-related enhancements in rhythm processing.

In Experiment 2, we eliminated the potential effects of auditory dominance in Experiment 1 since posttraining enhancement was also obtained despite the absence of auditory information. While, indeed, visual-only moving stimuli have been found to improve rhythm perception (Grahn, 2012; Hove et al., 2013; Repp & Su, 2013), our study extends that body of literature by being the first to investigate whether the processing of two integer-ratio metric simple visual rhythms consisting of untrained static stimuli can be enhanced after training with rhythms containing motion information. Contrary to the findings reported by Zerr et al. (2019) and Barakat et al. (2015), we found that visual-only training can be as efficient as multisensory training in improving posttraining discrimination performance. The absence of significant posttraining enhancements for the visual-only training group in the above-mentioned studies might, thus, simply reflect the inefficiency of training with visual-only static stimuli in yielding learning effects, since the visual system rarely processes temporal information that lacks a spatial translation (Hove et al., 2013). Overall, while exposure to auditory stimulation have been found to facilitate subsequent visual rhythm processing (Barakat et al., 2015; Collier & Logan, 2000; Grahn et al., 2011; McAuley & Henry, 2010), our results are the first to extend these findings by showing that discrimination of two integer-ratio visual rhythms consisting of static stimuli can be facilitated by prior exposure to both audiovisual (Experiment 1; cf. Barakat et al., 2015; Zerr et al., 2019) and visual (Experiment 2) stimuli. The ability to predict the timing and the allocation pattern of an event can enhance information processing and this can be achieved when this event appears rhythmically (i.e., speech, music, biological motion; Breska & Deouell, 2017). Johndro et al. (2019) further investigated auditory rhythms and found that their temporal characteristics are able of directing attention and, moreover, enhancing the encoding of visual stimuli into memory. Our results extend previous work by showing that the spatiotemporal characteristics of visual dynamic rhythms enhance the processing of static visual rhythms but only in the presence of feedback (Control experiment). The use of such predictive stimuli during training trials failed to enhance posttraining performance in the absence of feedback.

An alternative explanation of our findings would be that they may reflect enhancements due to mere exposure to the rhythms during training rather than learning processes. Given that feedback is indeed important for learning to occur (Dosher & Lu, 2017), we addressed this possibility by performing a control experiment without feedback during the training sessions. Through this control experiment, we showed that the posttraining enhancements obtained depend on the presence of feedback during training. These results are also in line with evidence showing that training with trial-by-trial feedback enhanced temporal acuity for audiovisual stimuli unlike simple exposure to the stimuli (De Niear et al., 2017). Most perceptual learning studies use trial-by-trial feedback as it is correlated with performance improvement, yet learning can be the result of block feedback as well as no feedback at all (Liu et al., 2014). Moreover, while feedback

was mainly seen as a way to make perceptual learning easier rather than produce it, Choi and Watanabe (2012) support that feedback can induce learning in an orientation discrimination task by increasing participants' sensitivity even for trials where the actual stimuli were replaced by noise.

In the experiments described in this paper, we utilized metrical and low complexity rhythms that consisted of only two types of intervals (i.e., two-integer-ratio metric simple rhythms). This might explain the high accuracy scores we observed even during pretraining. It remains unknown, whether this training-driven enhancement would be evident in more complex rhythms (e.g., three- or four-integer-ratio rhythms). Although feedback may be considered as an important factor of perceptual learning (Powers et al., 2009; Seitz et al., 2006), training task difficulty seems to affect performance and maybe interact with feedback. Indeed, De Niear et al. (2016) suggested that harder training procedures may lead to a noticeable performance increment, while Goldhacker et al. (2013) claimed that feedback might be helpful for easier tasks, but it can prevent learning when it comes to more challenging ones. On the other hand, Gabay et al. (2017) claimed that an overall highly intense training procedure may not be as productive as an easier one, while Sürig et al. (2018) suggested that adapting training task difficulty to each participant's abilities can lead to the optimal learning outcomes. Future experimentation may allow a clearer picture in terms of role of task difficulty in training and perceptual learning.

Increasing task difficult by increasing the number of intervals would potentially result in increased memory load, thereby rendering it unlikely to efficiently store and process the rhythmic patterns presented, which could, in turn, affect the transfer of learning (Teki & Griffiths, 2014). This is in line with the Scalar Expectancy Theory (SET), an internal clock model presented by Gibbon (1977). The SET model shares some common elements with previous internal clock models (cf. Creelman, 1962; Treisman, 1963) like the pacemaker, the counter/accumulator, and the decision process, but with the addition of a mechanism consisting of two memory stores, the working memory (short term) and the reference memory (long term) store. When timing a stimulus, pulses are produced by the pacemaker, which are then collected at the level of the accumulator. Those pulses are stored in working memory and are compared with those that were already stored in the reference memory. Then, a decision-making process allows for the time estimation requested by the participant. What is interesting about this system is that it can be initiated, paused, and reset with the aim of providing time estimations for individual or multiple events that take place at the same time (Allman et al., 2014). Moreover, while the mean duration of a time interval increases, the standard deviation of the duration estimate increases as well, thus attributing the naming scalar to this model (Rhodes, 2018). Indeed, considering the SET beyond the context of a single interval, it has been hypothesized that the reference memory gets overloaded with increasing number of

intervals, thereby resulting in worse memory performance (Teki & Griffiths, 2014). This is in line with evidence supporting that rhythm discrimination tasks require working memory resources so as to compare the standard rhythms to the comparison stimuli (Leow & Grahn, 2014), while studies have also shown that rhythms of more integer ratios are less efficiently processed as compared with two integer-ratio rhythms (i.e., as those used in Experiment 1; Drake, 1993). This is probably why multisensory training can yield enhanced posttraining performance in a visual-only rhythm discrimination task with two integer-ratio rhythms (cf. Experiment 1; Barakat et al., 2015), while four integer-ratio visual rhythms of static stimuli cannot be accurately discriminated (Collier & Logan, 2000). The effects of integer-ratio on rhythm processing are further supported by neuroimaging data showing that when compared with simple isochronous rhythmic sequences, the processing of four integer-ratio metric rhythms results in increased activation in the superior prefrontal cortex, an area that has been suggested to be responsible for the memory representation of more complex rhythm sequences (Bengtsson et al., 2009). Taken together, these findings suggest that studies that test one or two different interval lengths, as the ones used in this study, cannot necessarily be generalized to timing of four different interval lengths (Grahn, 2012). Future work is needed to investigate the modulatory effects of integer-ratio on rhythm perceptual learning, by testing whether learning effects can be obtained across different levels of rhythm complexity (e.g., rhythms consisting of two or more integer ratios).

In conclusion, utilizing a perceptual learning paradigm, we showed that the processing of nonoptimal visual rhythms benefits from training with multisensory and visual moving or static stimuli. Moreover, we showed that these benefits are unlikely to be the result of a mere effect of exposure since no enhancements were found in the absence of feedback. However, given that we used low complexity metric simple rhythms, we suggest that the role of task difficulty in rhythm perceptual learning should be further investigated. Future work should also aim to highlight the rhythmic structure of visual sequences by using more naturalistic and complex body movements (e.g., dancing) that might be more efficient in communicating the rhythmic information through different body parts so as to further optimize visual rhythm perception.

References

Alais, D., & Cass, J. (2010). Multisensory perceptual learning of temporal order: Audiovisual learning transfers to vision but not audition. *PLOS ONE, 5*(6), e11283. https://doi.org/10.1371/journal.pone.0011283

Allman, M. J., Teki, S., Griffiths, T. D., & Meck, W. H. (2014). Properties of the internal clock: First- and second-order principles of subjective time. *Annual Review of Psychology, 65*, 743–771. https://doi.org/10.1146/annurev-psych-010213-115117

Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., & Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception, 33*(6), 717–746. https://doi.org/10.1068/p5096

Barakat, B., Seitz, A. R., & Shams, L. (2015). Visual rhythm perception improves through auditory but not visual training. *Current Biology, 25*(2), R60–R61. https://doi.org/10.1016/j.cub.2014.12.011

Bengtsson, S. L., Ullén, F., Ehrsson, H. H., Hashimoto, T., Kito, T., Naito, E., Forssberg, H., & Sadato, N. (2009). Listening to rhythms activates motor and premotor cortices. *Cortex, 45*(1), 62–71. https://doi.org/10.1016/j.cortex.2008.07.002

Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology, 58*(1), 47–73. https://doi.org/10.1146/annurev.psych.57.102904.190152

Breska, A., & Deouell, L. Y. (2017). Neural mechanisms of rhythm-based temporal prediction: Delta phase-locking reflects temporal predictability but not rhythmic entrainment. *PLOS Biology, 15*(2), e2001665. https://doi.org/10.1371/journal.pbio.2001665

Calin-Jageman, R. J. (2018). The new statistics for neuroscience majors: Thinking in effect sizes. *Journal of Undergraduate Neuroscience Education, 16*(2), E21–E25.

Carrozzo, M., Moscatelli, A., & Lacquaniti, F. (2010). Tempo Rubato: Animacy speeds up time in the brain. *PLOS ONE, 5*(12), e15638. https://doi.org/10.1371/journal.pone.0015638

Choi, H., & Watanabe, T. (2012). Perceptual learning solely induced by feedback. *Vision Research, 61*, 77–82. https://doi.org/10.1016/j.visres.2012.01.006

Collier, G. L., & Logan, G. (2000). Modality differences in short-term memory for rhythms. *Memory & Cognition, 28*(4), 529–538. https://doi.org/10.3758/bf03201243

Creelman, C. D. (1962). Human discrimination of auditory duration. *Journal of the Acoustical Society of America, 34*, 582–593.

De Niear, M. A., Gupta, P. B., Baum, S. H., & Wallace, M. T. (2017). Perceptual training enhances temporal acuity for multisensory speech. *Neurobiology of Learning and Memory, 147*, 9–17. https://doi.org/10.1016/j.nlm.2017.10.016

De Niear, M. A., Koo, B., & Wallace, M. (2016). Multisensory perceptual learning is dependent upon task difficulty. *Experimental Brain Research, 234*(11), 3269–3277. https://doi.org/10.1007/s00221-016-4724-3

Dosher, B., & Lu, Z. (2017). Visual perceptual learning and models. *Annual Review of Vision Science, 3*, 343–363. https://doi.org/10.1146/annurev-vision-102016-061249

Drake, C. (1993). Reproduction of musical rhythms by children, adult musicians, and adult nonmusicians. *Perception & Psychophysics, 53*(1), 25–33. https://doi.org/10.3758/bf03211712

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*(6870), 429–433. https://doi.org/10.1038/415429a

Gabay, Y., Karni, A., & Banai, K. (2017). The perceptual learning of time-compressed speech: A comparison of training protocols with different levels of difficulty. *PLOS ONE, 12*(5), e0176488. https://doi.org/10.1371/journal.pone.0176488

Gan, L., Huang, Y., Zhou, L., Qian, C., & Wu, X. (2015). Synchronization to a bouncing ball with a realistic motion trajectory. *Scientific Reports, 5*(1). https://doi.org/10.1038/srep11974

Ghazanfar, A. A. (2013). Multisensory vocal communication in primates and the evolution of rhythmic speech. *Behavioral Ecology and Sociobiology, 67*(9), 1441–1448. https://doi.org/10.1007/s00265-013-1491-z

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review, 84*, 279–325.

Goldhacker, M., Rosengarth, K., Plank, T., & Greenlee, M. (2013). The effect of feedback on performance and brain activation during perceptual learning. *Vision Research, 99*, 99–110. https://doi.org/10.1016/j.visres.2013.11.010

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., & Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. *Journal of Cognitive Neuroscience, 19*(5), 893–906. https://doi.org/10.1162/jocn.2007.19.5.893

Grahn, J. A., & Brett, M. (2009). Impairment of beat-based rhythm discrimination in Parkinson's disease. *Cortex, 45*(1), 54–61. https://doi.org/10.1016/j.cortex.2008.01.005

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*(2), 1231–1243. https://doi.org/10.1016/j.neuroimage.2010.09.033

Grahn, J. A., & Rowe, J. B. (2009). Feeling the beat: Premotor and striatal interactions in musicians and non-musicians during beat perception. *The Journal of Neuroscience, 29*(23), 7540–7548. https://doi.org/10.1523/JNEUROSCI.2018-08.2009

Grondin, S., & McAuley, J. D. (2009). Duration discrimination in crossmodal sequences. *Perception, 38*(10), 1542–1559. https://doi.org/10.1068/p6359

Hammer, R., Sloutsky, V., & Grill-Spector, K. (2015). Feature saliency and feedback information interactively impact visual category learning. *Frontiers in Psychology, 6*, 74. https://doi.org/10.3389/fpsyg.2015.00074

Hove, M. J., Fairhurst, M. T., Kotz, S. A., & Keller, P. E. (2013). Synchronizing with auditory and visual rhythms: An fMRI assessment of modality differences and modality appropriateness. *NeuroImage, 67*, 313–321. https://doi.org/10.1016/j.neuroimage.2012.11.032

Huang, J., Gamble, D., Sarnlertsophon, K., Wang, X., & Hsiao, S. (2012). Feeling music: Integration of auditory and tactile inputs in musical meter perception. *PLOS ONE, 7*(10). https://doi.org/10.1371/journal.pone.0048496

Iannarilli, F., Vannozzi, G., Iosa, M., Pesce, C., & Capranica, L. (2013). Effects of task complexity on rhythmic reproduction performance in adults. *Human Movement Science, 32*(1), 203–213. https://doi.org/10.1016/j.humov.2012.12.004

Iversen, J., Repp, B., & Patel, A. (2009). Top-down control of rhythm perception modulates early auditory responses. *Annals of the New York Academy of Sciences, 1169*(1), 58–73. https://doi.org/10.1111/j.1749-6632.2009.04579.x

Johndro, H., Jacobs, L., Patel, A. D., & Race, E. (2019). Temporal predictions provided by musical rhythm influence visual memory encoding. *Acta Psychologica, 200*, Article 102923. https://doi.org/10.1016/j.actpsy.2019.102923

Kirkham, N. Z., Slemmer, J. A., Richardson, D. C., & Johnson, S. P. (2007). Location, location, location: development of spatiotemporal

sequence learning in infancy. *Child Development, 78*(5), 1559–1571. https://doi.org/10.1111/j.1467-8624.2007.01083.x

Lacquaniti, F., Carrozzo, M., D'Avella, A., Scaleia, B. L., Moscatelli, A., & Zago, M. (2014). How long did it last? You would better ask a human. *Frontiers in Neurorobotics, 8.* https://doi.org/10.3389/fnbot.2014.00002

Leow, L.-A., & Grahn, J. A. (2014). Neural mechanisms of rhythm perception: Present findings and future directions. *Advances in Experimental Medicine and Biology Neurobiology of Interval Timing, 829*, 325–338. https://doi.org/10.1007/978-1-4939-1782-2_17

Liu, J., Dosher, B., & Lu, ZL. (2014). Modeling trial by trial and block feedback in perceptual learning. *Vision Research*, *99*, 46–56. https://doi.org/10.1016/j.visres.2014.01.001

Liu, J., Lu, Z. L., & Dosher, B. A. (2010). Augmented Hebbian reweighting: Interactions between feedback and training accuracy in perceptual learning. *Journal of Vision, 10*(10), 29. https://doi.org/10.1167/10.10.29

Mathôt, S., Schreij, D., & Theeuwes, J. (2011). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

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

Mendonça, C., Santos, J. A., & López-Moliner, J. (2011). The benefit of multisensory integration with biological motion signals. *Experimental Brain Research, 213*(2/3), 185–192. https://doi.org/10.1007/s00221-011-2620-4

Nesti, A., Barnett-Cowan, M., MacNeilage, P. R., & Büthoff, H. H. (2014). Human sensitivity to vertical self-motion. *Experimental Brain Research, 232*(1), 303–314. https://doi.org/10.1007/s00221-013-3741-8

Orgs, G., Bestmann, S., Schuur, F., & Haggard, P. (2011). From body form to biological motion the apparent velocity of human movement biases subjective time. *Psychological Science, 22*(6), 712–717. https://doi.org/10.1177/0956797611406446

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*(2), 226–238. https://doi.org/10.1007/s00221-004-2159-8

Pfeuffer, C. U., Aufschnaiter, S., Thomaschke, R., & Kiesel, A. (2020). Only time will tell the future: Anticipatory saccades reveal the temporal dynamics of time-based location and task expectancy. *Journal of Experimental Psychology: Human Perception and Performance, 46*(10), 1183–1200. https://doi.org/10.1037/xhp0000850

Phillips-Silver, J., & Trainor, L. J. (2007). Hearing what the body feels: Auditory encoding of rhythmic movement. *Cognition, 105*(3), 533–546. https://doi.org/10.1016/j.cognition.2006.11.006

Powers 3rd, A. R., Hillock, A. R., & Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. *The Journal of Neuroscience, 29*(39), 12265–12274. https://doi.org/10.1523/JNEUROSCI.3501-09.2009

Repp, B. H., & Su, Y.-H. (2013). Sensorimotor synchronization: A review of recent research (2006-2012). *Psychonomic Bulletin & Review, 20*(3), 403–452. https://doi.org/10.3758/s13423-012-0371-2

Rhodes, D. (2018). On the distinction between perceived duration and event timing: towards a unified model of time perception. *Timing & Time Perception, 6*(1), 90–123 ISSN 2213-445X.

Roy, C., Lagarde, J., Dotov, D., & Bella, S. (2017). Walking to a multi-sensory beat. *Brain and Cognition, 113*, 172–183. https://doi.org/10.1016/j.bandc.2017.02.00

Seitz, A., Nanez, J., Holloway, S., Tsushima, Y., & Watanabe, T. (2006). Two cases requiring external reinforcement in perceptual learning. *Journal of Vision, 6*, 966–973. https://doi.org/10.1167/6.9.9

Shams, L., Wozny, D., Kim, R., & Seitz, A. (2011). Influences of multisensory experience on subsequent unisensory processing. *Frontiers in Psychology, 2*, 264. https://doi.org/10.3389/fpsyg.2011.00264

Stadler, W., Springer, A., Parkinson, J., & Prinz, W. (2012). Movement kinematics affect action prediction: Comparing human to non-human point-light actions. *Psychological Research, 76*(4), 395–406. https://doi.org/10.1007/s00426-012-0431-2

Steiger, J. H. (2004). Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods, 9*, 164–182.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience, 9*(5), 255–266. https://doi.org/10.1038/nrn2377

Su, Y.-H. (2014a). 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. (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. (2016). Visual tuning and metrical perception of realistic point-light dance movements. *Scientific Reports, 6*(1), 22774. https://doi.org/10.1038/srep22774

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

Su, Y.-H., & Salazar-López, E. (2016). Visual timing of structured dance movements resembles auditory rhythm perception. *Neural Plasticity, 2016*, 1678390. https://doi.org/10.1155/2016/1678390

Sürig, R., Bottari, D., & Röder, B. (2018). Transfer of audio-visual temporal training to temporal and spatial audio-visual tasks. *Multisensory Research, 31*(6), 556–578. https://doi.org/10.1163/22134808-00002611

Teki, S., & Griffiths, T. D. (2014). Working memory for time intervals in auditory rhythmic sequences. *Frontiers in Psychology, 5*, 1329. https://doi.org/10.3389/fpsyg.2014.01329

Thompson, B. (2002). What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher, 31*(3), 25–32. https://doi.org/10.3102/0013189X031003025

Toiviainen, P., Luck, G., & Thompson, M. R. (2010). Embodied meter: Hierarchical eigenmodes in music-induced movement. *Music Perception, 28*(1), 59–70. https://doi.org/10.1525/mp.2010.28.1.59

Treisman, M. (1963). Temporal discrimination and the indifference interval. Implications for a model of the "internal clock". *Psychological Monographs, 77*, 1–31.

Wagener, A., & Hoffmann, J. (2010). Temporal cueing of target-identity and target-location. *Experimental Psychology, 57*(6), 436–445. https://doi.org/10.1027/1618-3169/a000054

Watanabe, T., Náñez, J., & Sasaki, Y. (2001). Perceptual learning without perception. *Nature, 413*, 844–848. https://doi.org/10.1038/35101601

Watanabe, T., & Sasaki, Y. (2015). Perceptual learning: Toward a comprehensive theory. *Annual Review of Psychology, 66*, 197–221. https://doi.org/10.1146/annurev-psych-010814-015214

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

Wöllner, C., Deconinck, F. J., Parkinson, J., Hove, M. J., & Keller, P. E. (2012). The perception of prototypical motion: Synchronization is enhanced with quantitatively morphed gestures of musical conductors. *Journal of Experimental Psychology: Human Perception and Performance, 38*(6), 1390–1403. https://doi.org/10.1037/a0028130

Zerr, M., Freihorst, C., Schütz, H., Sinke, C., Müller, A., Bleich, S., Münte, T. F., & Szycik, G. R. (2019). Brief sensory training narrows the temporal binding window and enhances long-term multimodal speech perception. *Frontiers in Psychology, 10*, 2489. https://doi.org/10.3389/fpsyg.2019.02489

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.