



**NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS**  
DEPARTMENT OF PHILOSOPHY AND HISTORY OF SCIENCE  
DEPARTMENT OF INFORMATICS AND TELECOMMUNICATIONS  
DEPARTMENT OF PSYCHOLOGY  
DEPARTMENT OF PHILOLOGY

**Konstantina Paraskevoudi**

**A.M. 15M12**

**Motion and multisensory processing for the enhancement of rhythm perception**

Thesis submitted  
In partial fulfillment of the requirements for the Masters degree by the  
Interdepartmental Graduate Program in Cognitive Science

Advisory committee:

**Nikolaos Smyrnis, Associate Professor, National and Kapodistrian University of Athens**  
**Argiro Vatakis, Researcher, Cognitive Systems Research Institute**  
**Marios Avraamides, Associate Professor, University of Cyprus**

Athens, Greece  
September 2017

The thesis is approved

Nikolaos Smyrnis.....

Argiro Vatakis.....

Marios Avraamides .....

© 2017 Konstantina Paraskevoudi  
Available online under a Creative Commons Attribution 4.0 Licence (International)  
(CC-BY 4.0) <https://creativecommons.org/licenses/by/4.0/>

## Abstract

Despite the auditory advantage in rhythm perception, rhythm may also be mediated by visual moving (animate or inanimate) and multisensory stimuli. Thus, in the current study, using a perceptual learning paradigm, we, initially, tested whether multisensory training with animate versus inanimate motion enhances processing of metric simple two integer-ratio visual rhythms (Exp. 1). Participants were trained with audiovisual rhythms with auditory tones occurring at the onset of each interval and the visual stimuli being static (i.e., static circles;  $AV_{stat}$ ), moving-inanimate (i.e., moving bar;  $AV_{inan}$ ), or moving-animate (i.e., humanlike point-light figure;  $AV_{an}$ ). A regular beat was also present in all rhythms used. In the pre- and post-training tasks, participants had to respond whether two visual rhythms (i.e., static circles) differed or not. Results showed improved post-training performance for all groups irrespective of the type of audiovisual stimulation. To assess whether this benefit was mainly auditory driven, we introduced visual-only training with a moving stimulus and a regular beat ( $V_{inan}$ ; Exp. 2). Comparison of  $AV_{inan}$  and  $V_{inan}$  showed that visual training also improved post-training performance. Given that simple rhythms were used for Exp. 1 and 2, we tested whether  $AV_{inan}$  training benefits processing of four integer-ratio metric simple and complex visual rhythms (Exp. 3). Results showed no significant improvements in post-training accuracy, a result that could be attributed to increased memory load. Overall, our findings support that training with audiovisual or visual moving stimuli enhances visual rhythm perception, but this enhancement is limited to two integer-ratio rhythms only.

*Keywords:* Rhythm; Multisensory perception; Motion; Perceptual learning.

## Περίληψη

Παρά την κυριαρχία της ακοής στην αντίληψη ρυθμού, ο ρυθμός μπορεί να μεταδοθεί από οπτικά (βιολογική και μη βιολογική κίνηση) και πολυαισθητηριακά ερεθίσματα. Η παρούσα μελέτη εξέτασε αρχικά με ένα έργο αντιληπτικής μάθησης εάν η πολυαισθητηριακή εξάσκηση με βιολογική ή μη-βιολογική κίνηση βελτιώνει την επεξεργασία μετρικών οπτικών ρυθμών δύο διαστημάτων (Πείραμα 1). Στα pre- και post-training έργα, οι συμμετέχοντες έκριναν εάν δύο οπτικοί ρυθμοί (στατικοί κύκλοι) διέφεραν ή όχι. Στην εξάσκηση παρουσιάζονταν πολυαισθητηριακοί ρυθμοί με ακουστικούς τόνους στην αρχή κάθε διαστήματος, ενώ το οπτικό ερέθισμα ήταν στατικό (στατικοί κύκλοι; AVstat), μη-βιολογική (κινούμενη γραμμή; AVinan), ή βιολογική κίνηση (ανθρωπόμορφη φιγούρα; AVan). Σε όλες τις συνεδρίες εμφανιζόταν συστηματικά ένας χτύπος. Οι αναλύσεις έδειξαν βελτιωμένη post-training επίδοση ανεξαρτήτως του είδους της οπτικοακουστικής εξάσκησης. Στο Πείραμα 2 εξετάστηκε η συμβολή της ακοής με μία νέα ομάδα που εξασκήθηκε με οπτικό κινούμενο ερέθισμα και την ακολουθία χτύπων (Vinan). Η σύγκριση των AVinan και Vinan έδειξε ότι και η οπτική εξάσκηση (Vinan) βελτιώνει την post-training επίδοση, κάτι που πιθανόν να οφείλεται στην ευκολία του έργου. Έτσι εξετάστηκε η συμβολή της πολυαισθητηριακής εξάσκησης (AVinan; Πείραμα 3) στην επεξεργασία μετρικών και μη-μετρικών οπτικών ρυθμών τεσσάρων διαστημάτων. Δεν φάνηκε σημαντική post-training βελτίωση, πιθανόν λόγω του αυξημένου μνημονικού φόρτου ως αποτέλεσμα της ύπαρξης τεσσάρων διαστημάτων. Συνολικά, η μελέτη αναδεικνύει τη συμβολή της εξάσκησης με πολυαισθητηριακά και οπτικά κινούμενα ερεθίσματα στην αντίληψη οπτικού ρυθμού, με τα φαινόμενα μάθησης να περιορίζονται όμως σε ρυθμούς δύο διαστημάτων.

*Λέξεις-κλειδιά:* Ρυθμός, Πολυαισθητηριακή αντίληψη, Κίνηση, Αντιληπτική μάθηση

## **Acknowledgements**

To my co-supervisors, Prof. Nikolaos Smyrnis and Prof. Marios Avraamides for their assistance and guidance. My special thanks to Prof. Marios Avraamides and his team in Cyprus for their constant support during the last year of my studies.

To the professors that have always supported and guided me throughout the last two years of the Master's Program: Dr. Irimi Skaliora, Dr. George Giannakopoulos, Dr. John Pezaris, Dr. Thanassis Protopappas, and Dr. Konstantinos Moutoussis.

Most importantly, to my supervisor, Dr. Argiro Vatakis. This study would not have been possible without her guidance and unwavering support. Her enthusiasm and innate scientific curiosity have been inspirational, making the last demanding year of the program much easier as it would have been otherwise.

To my lab friends, Lydia, Penny, Vassilis, for sharing all the stress with me, to Vassilis for opening the lab and helping us running our experiments, to my friend, Ilektra, my brother, Kostis, and my parents for their support and patience and, of course, to all the participants for their assistance in completing this study.

To Timos, for always reminding me of the important things after all.

## Contents

<b>Abstract</b> .....	3
<b>Περίληψη</b> .....	4
<b>Acknowledgements</b> .....	5
<b>Contents</b> .....	6
<b>Introduction</b> .....	7
<b>Methods</b> .....	11
Participants .....	11
Apparatus and stimuli.....	11
Design.....	12
Procedure .....	16
<b>Results and discussion</b> .....	16
Training data.....	17
Pre- and post-training data.....	18
<b>Experiment 2</b> .....	21
<b>Methods</b> .....	21
Participants .....	21
Apparatus, stimuli, design, and procedure .....	22
<b>Results and discussion</b> .....	22
Training data.....	22
Pre- and post-training data.....	24
<b>Experiment 3</b> .....	26
Methods.....	26
Participants .....	26
Apparatus and stimuli.....	26
Design and procedure .....	27
<b>Results</b> .....	28
Training data.....	28
Pre- and post-training data.....	28
<b>Discussion</b> .....	30
<b>References</b> .....	34

## Introduction

Rhythm perception is considered by most as tightly associated with the auditory system (e.g., Grahn, 2012; Grahn, Henry, & McAuley, 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 crossmodal, as well as the intramodal differences in rhythm perception and discrimination (Grahn, 2012; Grahn et al., 2011; Hove, Fairhurst, Kotz, & Keller, 2013). Most such studies have reported an auditory advantage in rhythm perception, which has been attributed to the more fine-grained temporal resolution of the auditory as compared to the visual system (Collier & Logan, 2000; Grahn, 2012; Grahn et al., 2011; Patel, Iversen, Chen, & Repp, 2005). For example, a periodic rhythm 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 neuroimaging data that have demonstrated increased activity of timing-related areas (i.e., basal ganglia, putamen) when processing an auditory rhythm

as compared to a rhythm mediated by static visual flashes (Grahn et al., 2011; Hove et al., 2013). Studies have also 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 to 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, Repp, & Patel, 2009) facilitates rhythm processing and encoding (Grahn, 2012; Su, 2014b), as well as motor synchronization (Gan, Huang, Zhou, Qian, & Wu, 2015; Grahn, 2012; Grahn & Brett, 2007).

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 (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, Seitz, & Shams, 2015) that lack spatiotemporal information, while motion – a more optimal visual stimulus (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 moving visual stimuli in rhythm perception was first investigated by Grahn (2012). Specifically, she compared directly auditory rhythms to 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 to visual trials, 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 some types of rhythms can also be processed by vision 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 has 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 to auditory-only rhythms. This improvement is also in line with several studies reporting enhanced performance in multisensory as compared to unisensory trials (e.g., Alais & Cass, 2010; Roy, Lagarde, Dotov, & Dalla Bella, 2016; Shams, Wozny, Kim, & Seitz, 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 animate and inanimate moving stimuli (Carrozzo, Moscatelli, & Lacquaniti, 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 certain types of training has recently been reported in a perceptual learning study (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 post-training performance, while both the auditory and multisensory training groups were significantly better during the post-training session. More importantly, these latter two groups did not differ in their post-training performance, suggesting that multisensory training did not enhance rhythm perception more than the auditory training. One could, thus, argue that the post-training enhancement observed was auditory-driven, while it remains unanswered whether the absence of the post-training improvements 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 as yet examined the effects of modality and stimulus attributes such as visual motion 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 motion so as to assess whether the former benefits rhythm processing more than the latter. In order to address this gap, we, thus, 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 stimuli. We hypothesized that training with audiovisual rhythms, particularly those containing visual motion, would improve the processing of the visual static rhythmic patterns due to the visual system's high spatial resolution and motion processing (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 visual animate movement would yield better discrimination performance in a subsequent visual-only rhythm discrimination task as compared to 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. 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, Vannozzi, Iosa, Pesce, & Capranica, 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, Schreij, & Theeuwes, 2012).

Three types of visual stimuli were utilized to create the visual stream of the 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 implemented in 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 movement, since human vertical body movements have been suggested to mediate rhythm more efficiently than horizontal body movements (Nesti, Barnett-Cowan, MacNeilage, & Bühlhoff, 2014; Toiviainen, Luck, & Thompson, 2010).

The auditory stream of the rhythmic sequences utilized was created using Audacity and was composed of two types: a) a sinewave tone (44110 Hz) of 43 ms in duration and b) a pink noise (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 was divided in two parts that took place in two separate days

(24 to 48 hours apart), with each part lasting approximately 50 minutes. The experiment consisted of four sessions in total: a pre-training session, two training sessions, and a post-training session (cf. Barakat et al., 2015; see Figure 1). For all sessions, the participants completed a two alternative forced choice (2AFC) rhythm discrimination task ('same' or 'different'), with the experimental structure being as follows. In each trial, a fixation point was initially presented for 1000 ms followed by the first rhythmic pattern (i.e., 'standard'). After 1100 ms (inter-stimulus interval; ISI), the second rhythmic sequence (i.e., 'comparison') was presented and participants provided a self-paced response. The inter-trial interval (ITI) was set at 1200 ms.

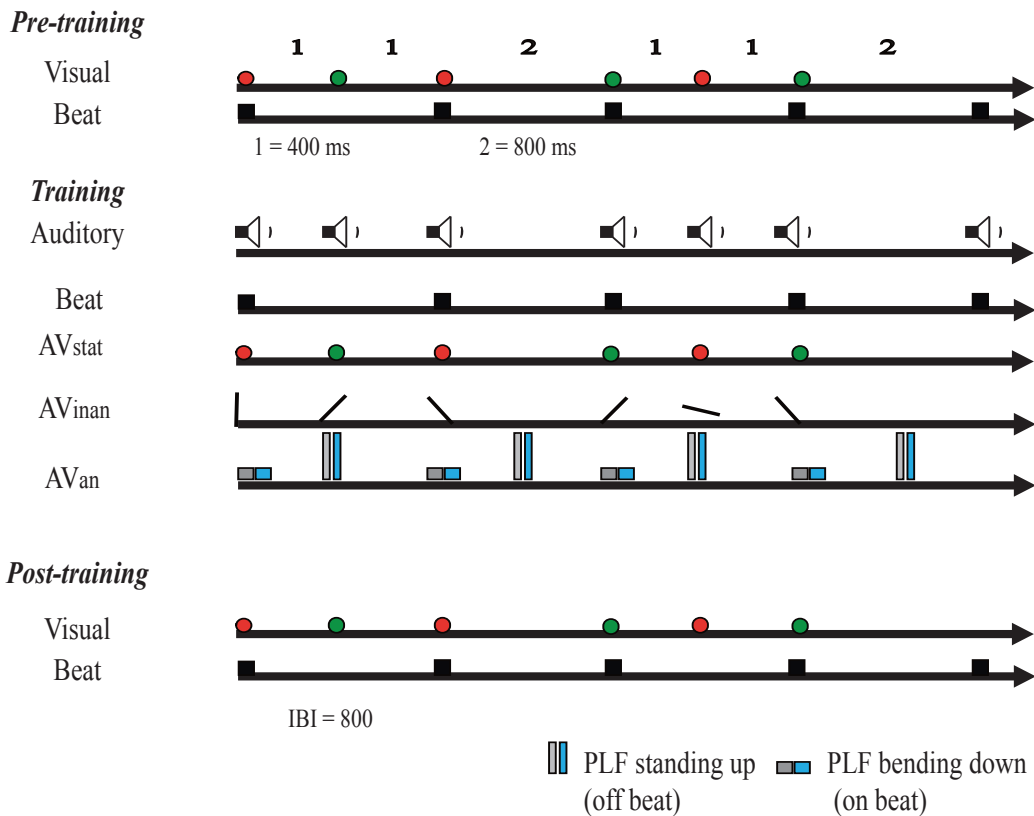
The rhythmic sequences used were metric simple rhythms (cf. Grahn, 2012; Grahn & Brett, 2007) that consisted of six elements of either a short (400 ms) or a long interval (800 ms). The intervals were, thus, related by integer ratios, where 1 = 400 ms and 2 = 800 ms, and had a regular grouping with the beat occurring regularly every 2 units (cf. Drake, 1993), that is every 800 ms (inter-beat 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 5x5 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, EE).

At the start of the first part of the experiment (Day 1), participants completed a set of five practice trials in order to familiarize themselves with the task. Subsequently, they all performed the pre-training session and the first training session. On Day 2, participants started with the second training session that was followed by the post-training test (which was identical to the pre-training test).

The pre- and post-training sessions were composed of visual-only rhythms that

consisted of static circles of changing colours (see Figure 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 post-training sessions consisted of 4 repetitions of each rhythm pair, resulting in 100 trials per session in total. Each session lasted approximately 30 minutes.

For the training phase, all participants were randomly assigned to one of the three training groups: audiovisual static circles (AV<sub>stat</sub>), moving bar (i.e., inanimate stimulus; AV<sub>inan</sub>), or PLF (i.e., animate stimulus; AV<sub>an</sub>) 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 (AV<sub>stat</sub>;  $N = 16$ , 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 pre- and post-training with the sole exception that, here, the onset of each circle was accompanied by an auditory tone. The second group (AV<sub>inan</sub>;  $N = 20$ , 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.



*Figure 1.* Schematic illustration 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 pre-training 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 one day. The first group (AV<sub>stat</sub>) was trained with auditory tones and static circles of changing colours (red and green ellipses). The second group (AV<sub>inan</sub>) was presented with auditory tones and a bar ‘moving’ to different screen locations (black line). The third group (AV<sub>an</sub>) 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 post-training session that was exactly the same as the pre-training.

The third group (AV<sub>an</sub>;  $N = 15$ , 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 in order to familiarize themselves with the task. They, subsequently, performed the pre-test and the first training session (Day 1). The second part started with the second training session followed by the post-test (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 buttons 'm' and 'z', 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 post-test. 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 reported here, Bonferroni-corrected t-tests (where  $p < .05$  prior to correction) were used for all post-hoc comparisons.



### *Training data*

The training data (i.e., percent 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: AV<sub>stat</sub>, AV<sub>inan</sub>, AV<sub>an</sub>) as the between-participants factor. The analysis showed a significant main effect of Group [ $F(2,48) = 5.54, p < .01, \eta^2 = .19$ ], with the AV<sub>stat</sub> group performing significantly better ( $M = .916$ ) as compared to both the AV<sub>inan</sub> ( $M = .823$ ) and the AV<sub>an</sub> ( $M = .820$ ) group (see Figure 2). The higher performance of the AV<sub>stat</sub> group could be attributed to the prior exposure to the pre-training session (i.e., identical visual stimulation), however, it should be noted that both AV<sub>inan</sub> and AV<sub>an</sub> groups also reached high performance accuracy with a mean accuracy over 80%. A significant main effect of Training Session was also obtained [ $F(1,48) = 12.25, p < .001, \eta^2 = .203$ ], with all groups having higher accuracy scores during the second training session ( $M = .871$ ) as compared to the first ( $M = .835$ ). 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 [ $F(10.72,514.68) = 11.3, p < .001, \eta^2 = .191$ ], with certain pairs having systematically lower accuracy ( $M_{BC} = .703, M_{CB} = .692, M_{CD} = .740, M_{DD} = .774$ ) as compared to others that were significantly easier to discriminate ( $M_{AE} = .925, M_{BE} = .939, M_{CA} = .933, M_{DA} = .948, M_{EA} = .957, M_{EB} = .908$ ). 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 [ $F(21.45,514.68) = 1.84, p < .05, \eta^2 = .07$ ] was obtained, with the AV<sub>stat</sub> 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 to the AV<sub>an</sub> ( $M_{AE} = .900$ ,  $M_{BA} = .844$ ,  $M_{DA} = .889$ ,  $M_{EA} = .922$ ,  $M_{EB} = .844$ ) and AV<sub>inan</sub> ( $M_{AE} = .875$ ,  $M_{CE} = .867$ ,  $M_{DD} = .692$ ) group, while AV<sub>an</sub> performed significantly worse than the other two groups when the rhythm pair was BC (AV<sub>an</sub> = .500, AV<sub>inan</sub> = .858, AV<sub>stat</sub> = .750). The interactions between Training Session and Group [ $F(2,48) = .65$ ,  $p > .05$ ], Training Session and Rhythm Pair [ $F(13.58,651.74) = 1.54$ ,  $p > .05$ ], and Group, Training Session, and Rhythm Pair [ $F(27.16,651.74) = 1.27$ ,  $p > .05$ ] did not reach significance.

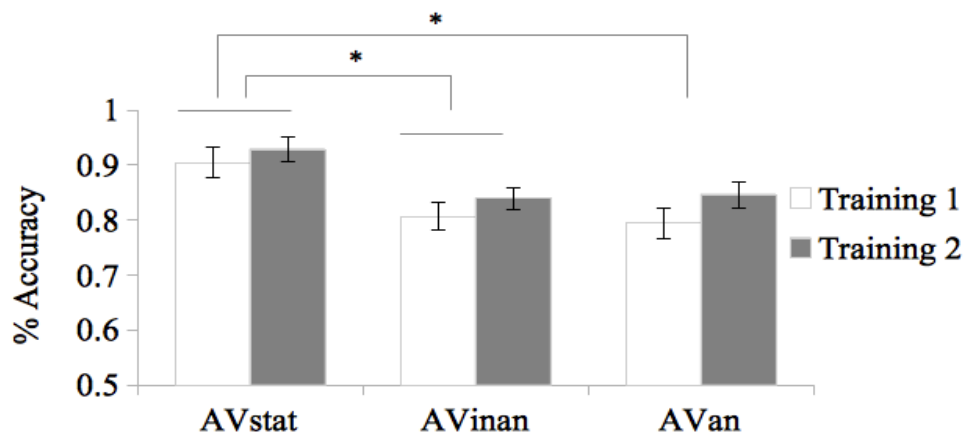


Figure 1. Mean discrimination accuracy during the two training sessions for the three training groups (AV<sub>stat</sub>, AV<sub>inan</sub>, AV<sub>an</sub>) in Experiment 1. Significant differences between the groups ( $p < .05$ ) are indicated by the asterisk. The error bars represent the standard error of the means.

#### *Pre- and post-training data*

For the main analysis, we compared the pre- and post-training performance in order to test for potential learning effects following training. The pre- and post-test responses were analyzed via a mixed ANOVA with Session (2 levels: Pre-training vs. Post-training) and Rhythm Pair (25 levels) as within-participant factors, and Group (3 levels: AV<sub>stat</sub>, AV<sub>inan</sub>, AV<sub>an</sub>) as between-participants factor. A significant main effect of

Session was obtained [ $F(1,48) = 58.26, p < .001, \eta^2 = .55$ ], with all groups performing better ( $M = .740$ ) during post-training as compared to the pre-training session ( $M = .634$ ; see Figure 3). A significant main effect of Rhythm Pair was also obtained [ $F(13.44,645.08) = 13.45, p < .001, \eta^2 = .22$ ], with certain rhythm pairs being more accurately discriminated (i.e.,  $M_{AE} = .815, M_{EA} = .831, M_{EC} = .818, M_{EE} = .841$ ) as compared to others (i.e.,  $M_{BC} = .531, M_{CB} = .478, M_{CD} = .613, M_{DC} = .617, M_{DE} = .551$ ).

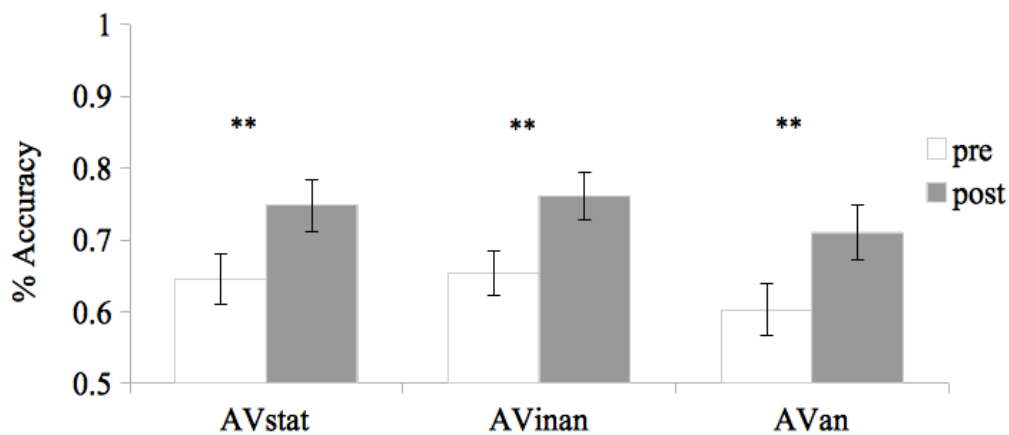


Figure 2. Mean discrimination accuracy during the two main sessions (pre- and post-training) for the three training groups (AVstat, AVinan, AVan) in Experiment 1. Significant differences between the pre- and post-test sessions ( $p < .001$ ) are indicated by two asterisks. The error bars represent the standard error of the means.

Further examination of this effect showed that rhythm E (i.e., 221111) seemed to be easier to discriminate from other rhythms suggesting that this rhythmic pattern was more efficiently processed and maintained in memory as compared to 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 [ $F(2,48) = .66, p > .05$ ], while the interactions between Group and Session and between Group, Session, and

Rhythm Pair did not reach significance ( $[F(2,48) = .012, p > .05]$  and  $[F(26.98,647.5) = 1.27, p > .05]$ , respectively). These findings suggest that the type of the visual component during training did not modulate post-training performance. However, we found a significant interaction between Group and Rhythm Pair  $[F(26.88,645.08) = 1.68, p < .05, \eta^2 = .07]$ , with AV<sub>stat</sub> being significantly better in their discrimination of rhythm CC ( $M = .820$ ) as compared to the AV<sub>inan</sub> group ( $M = .650$ ). We also found a significant interaction between Session and Rhythm Pair  $[F(13.49,647.5) = 4.27, p < .001, \eta^2 = .08]$ , with 17 out of 25 rhythm pairs being more accurately discriminated in the post-test as compared to the pre-test. 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 post-test performance for all training groups, despite the main effect of group during training. That is, irrespective of the training stimulus type (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 from the contribution of audition, thereby providing support for the modality appropriateness hypothesis (i.e., the 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 pre-training 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 post-training enhancements observed in Experiment 1 led us to a second experiment focusing on the contribution of audition during training. We reasoned that if the post-training improvement in Exp. 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 post-training performance when compared to the multisensory case of Exp. 1 (cf. Barakat et al., 2015). If, however, visual motion is able to mediate the rhythmic information needed for increasing discrimination accuracy, then the post-training performance could be enhanced following training with visual-only moving stimuli. In Experiment 2, therefore, we kept the experimental structure and design of Exp. 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 utilized in Exp. 1 in the absence of the auditory tone stimulation (i.e., *Vinan* group).

## **Experiment 2**

### ***Methods***

#### *Participants*

Twenty-two new university students (20 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 exactly the same as for Exp. 1 with the sole exception that instead of being trained with multisensory rhythms, participants received a unisensory training with a moving bar ( $V_{\text{inan}}$ ), 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 the sole training stimulus for two reasons: to save experimental time and, most importantly, due to Grahn's (2012) findings of visual moving stimuli mediating rhythmic information.

### ***Results and discussion***

To assess the effects of visual-only training and its post-training differences from the multisensory training, we performed a combined analysis of the data from Experiments 1 ( $AV_{\text{inan}}$ ) and 2 ( $V_{\text{inan}}$ ). Two participants from Experiment 2 were removed from the analysis due to formal musical and dance training.

#### *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 ( $AV_{\text{inan}}$ ,  $V_{\text{inan}}$ ) as the between-participants factor was conducted. A significant main effect of Training Session [ $F(1,38) = 24.21$ ,  $p < .001$ ,  $\eta^2 = .39$ ] was obtained, with both groups having higher accuracy scores during the second training session ( $M = .778$ ) as compared to the first one ( $M = .729$ ; see Figure 4). We also obtained a main effect of Group [ $F(1,38) = 22.65$ ,  $p < .001$ ,  $\eta^2 = .37$ ], with the multisensory group ( $AV_{\text{inan}}$ ) performing significantly

better ( $M = .823$ ) as compared to the visual-only group ( $V_{\text{in}}$ ;  $M = .684$ ). This is in line with findings reporting enhanced performance during audiovisual as compared to visual-only trials (Su, 2014a, 2014b). A main effect of Rhythm Pair was also obtained [ $F(10.26,389.75) = 6.598, p < .001, \eta^2 = .15$ ], with some rhythmic pairs being particularly difficult to discriminate ( $M_{AA} = .646, M_{CB} = .633, M_{CD} = .621, M_{DD} = .6, M_{DE} = .663$ ). We also obtained an interaction between Training Session and Rhythm Pair [ $F(13.75,522.46) = 2.7, p < .01, \eta^2 = .07$ ], with some rhythm pairs being more accurately discriminated during the second training session (i.e., AC, AD, AE, BA, BC, CA, DC, EB) as compared to the first session. The interactions between Training Session and Group [ $F(1,38) = 2.84, p > .05$ ], Rhythm and Group [ $F(10.26,389.75) = .940, p > .05$ ], and Training Session, Group, and Rhythm Pair [ $F(13.75,522.46) = .783, p > .05$ ] did not reach significance.

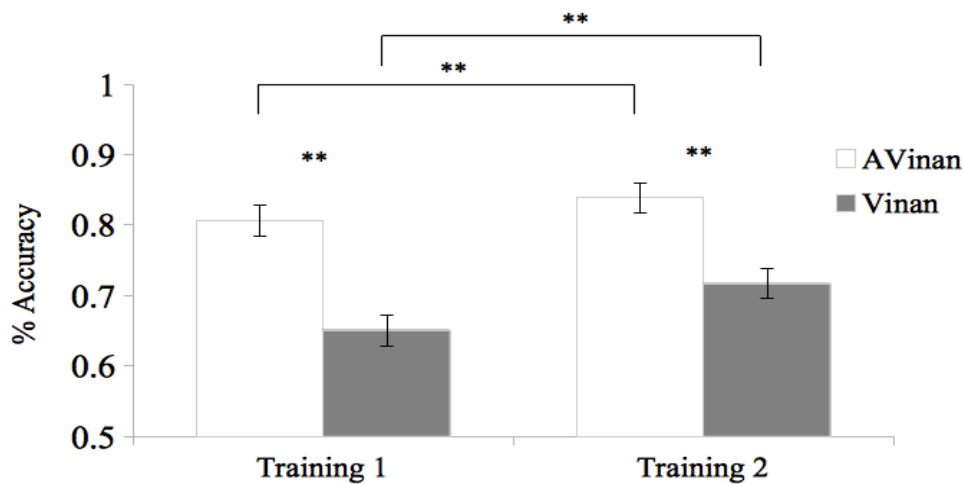
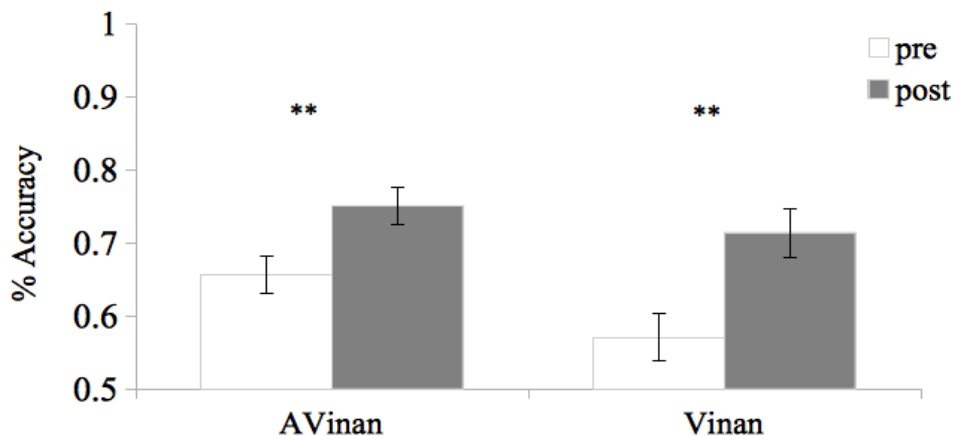


Figure 3. Mean discrimination accuracy during the two training sessions for the multisensory (AVin; Exp. 1) and the unisensory (Vin; Exp. 2) training groups. Significant differences between the groups and the training sessions ( $p < .001$ ) are indicated by two asterisks. The error bars represent the standard error of the mean.

### *Pre- and post-training data*

A mixed ANOVA with Session (2 levels: Pre-training vs. Post-training) and Rhythm Pair (25 levels) as within-participant factors, and Group (2 levels: AVinan, Vinan) as the between-participants factor was conducted. A significant main effect of Session was obtained [ $F(1,38) = 69.03, p < .001, \eta^2 = .65$ ], with both groups exhibiting a significant improvement in post-training performance ( $M = .738$ ) as compared to the pre-training ( $M = .613$ ; see Figure 5). Thus, despite the absence of auditory tone stimulation in Exp. 2 and despite the main effect of Group we found in the training data, training with visual-only moving stimuli continued to enhance post-training performance in a task where the rhythms consisted of static visual stimuli.



*Figure 4.* Mean discrimination accuracy during the pre- and post-training sessions for the multisensory (AVinan; Exp. 1) and the unisensory (Vinan; Exp. 2) training groups. Significant differences between the pre- and post-test sessions ( $p < .001$ ) are indicated by two asterisks. The error bars represent the standard error of the means.

The analysis also revealed a main effect of Rhythm Pair [ $F(12.63, 479.93) = 8.88, p < .001, \eta^2 = .19$ ], with some rhythm pairs having systematically lower accuracy scores ( $M_{CB} = .490, M_{CD} = .580, M_{DE} = .560$ ) as compared to others ( $M_{EA} = .820, M_{EC} = .820$ ).

Similar to our first experiment, the pairs with higher accuracy scores have the rhythm E



as one of the pairs. We also obtained a significant interaction between Session and Rhythm Pair [ $F(12.62,479.43) = 5.07, p < .001, \eta^2 = .12$ ], with 16 out of 25 rhythm pairs being significantly more accurately discriminated during the post-training as compared to the pre-training (i.e., AB, AC, AD, AE, BA, BC, BD, BE, CA, CD, CE, DA, DB, DC, DE, EC). The main effect of Group [ $F(1,38) = 3.022, p > .05$ ] and the interactions between Session and Group [ $F(1,38) = 1.431, p > .05$ ], Rhythm Pair and Group [ $F(12.63,479.93) = .832, p > .05$ ], and Session, Rhythm Pair, and Group [ $F(12.62,479.43) = 1.038, p > .05$ ] did not reach significance.

Overall, the results of Experiment 2 demonstrated that visual-only training with a moving stimulus can enhance rhythm perception even in the absence of auditory rhythmic stimulation. In particular, training with a visual-only moving stimulus (i.e., a moving bar) improved processing and discrimination ability of metric simple visual rhythms consisting of static stimuli. More importantly, we did not find any enhancement differences between multisensory and visual training, suggesting that visual moving stimuli are sufficient to improve discrimination accuracy of the two integer-ratio visual rhythms we used in the pre- and post-training sessions.

Although both unimodal and multisensory training sessions lead to discrimination accuracy enhancement, it still remains unclear whether or not multimodality can in some form lead to a higher rhythm discrimination advantage as compared to unimodal stimulation. In Experiments 1 and 2, we observed high accuracy scores even during pre-training, which might suggest that the task could have been too easy for the participants. This could have minimized the potential of observing a multisensory advantage. Thus, in Experiment 3, we manipulated the complexity (i.e., by manipulating the number of integer ratios) and metricality of the rhythmic patterns

presented by introducing four integer-ratio metric simple and metric complex rhythms (cf. Grahn, 2012; Grahn & Brett, 2007) contrary to the two integer-ratio metric simple rhythms used in Exps. 1 and 2. This would allow us to examine whether perceptual learning with multisensory stimulation (as shown in Exp. 1) would result in enhanced post-training performance in a more demanding visual-only rhythm discrimination task with four integer-ratio metric simple and metric complex rhythms consisting of static stimuli.

### **Experiment 3**

#### *Methods*

##### *Participants*

Twenty new university students (16 female) aged between 19 and 43 years old (mean age = 21.2 years) took part in this experiment.

##### *Apparatus and stimuli*

The apparatus and stimuli were exactly the same as for Experiment 2 with the sole exception that multisensory training was implemented and different types of rhythms were utilized. Additionally, instead of using empty auditory intervals we utilized filled ones, given that auditory filled intervals were more analogous to the ('filled') moving bar. The duration of the tones depended on the intervals within a rhythmic pattern with each tone occurring at the onset of each interval and ending 35 ms before the offset of the interval in order to create a silent gap between the intervals (cf. Grahn & Brett, 2007).

### *Design and procedure*

These were exactly the same as for Experiments 1 and 2 with the sole exception that instead of using two integer-ratio metric simple rhythms (cf. Drake, 1993), we utilized metric simple and metric complex rhythms of multiple intervals (adopted from Grahn, 2012, Table 1). Both rhythm types were composed of 5 elements and 4 interval durations that were related by integer ratios with the shortest being 200 ms. The rest of the intervals were multiples of the shortest interval (i.e., 1 = 200 ms, 2 = 400 ms, 3 = 600 ms, and 4 = 800 ms). For the metric simple rhythms, the intervals were arranged in groups of four units (e.g., 22413) with a regular beat occurring every 4 units (IBI = 800 ms). In contrast, for the metric complex rhythms, the intervals did not have any regular grouping (e.g., 14232), thus the beat did not co-occur with the onset and offset of 4-unit grouping.

Six rhythmic patterns were used as 'standard' for both the metric simple (i.e., A: 22413, B: 31413, C: 31422, D: 41331, E: 43113, and F: 43122) and metric complex rhythms (i.e., A: 33141, B: 41133, C: 41232, D: 11343, E: 13242, and F: 23241). Every rhythmic pattern had a corresponding deviant sequence that served as the 'comparison' stimulus in half of the trials. As in Grahn's (2012) study, the deviant sequences were in the same category as the standard sequences (i.e., for the metric simple standards, metric simple deviants were used; A: 22431, B: 31431, C: 13422, D: 43131, E: 41313, F: 41322, while for the metric complex standards, the deviants were also metric complex patterns; i.e., A: 31341, B: 14133, C: 14232, D: 13143, E: 31242, F: 23214). Half of the trials contained a deviant sequence as the 'comparison' stimulus (i.e., 'different' trials), while in the other half of the trials the second 'comparison' sequence was the same as

the first one (i.e., 'same' trials). This led to 12 trials with metric simple rhythms (i.e., 6 'same' pairs and 6 'different' pairs) and 12 trials with metric complex rhythms, thereby resulting in 24 rhythm pairs in total. The pre- and post-training sessions included 5 repetitions of each rhythm pair, resulting in 120 trials in total. Each training session consisted of 3 repetitions for each rhythm pair, resulting in 144 training trials in total. The presentation of each rhythm pair was randomized.

## **Results**

Five participants were removed from the analysis due to formal musical and dance training.

### *Training data*

A repeated-measures ANOVA with the factors of Training Session (2 levels: Session 1 vs. Session 2) and Rhythm Type (2 levels: Metric complex vs. Metric simple) was conducted. A significant main effect of Training Session [ $F(1,179) = 12.28, p < .01, \eta^2 = .06$ ] was obtained, with participants having higher discrimination accuracy during the second training session ( $M = .831$ ) as compared to the first ( $M = .744$ ). The main effect of Rhythm Type did not reach significance [ $F(1,179) = 1.37, p > .05$ ], with performance for metric complex rhythms ( $M = .790$ ) being similar to that for metric simple rhythms ( $M = .816$ ). The interaction between Training Session and Rhythm Type did not reach significance [ $F(1,179) = .000, p > .05$ ].

### *Pre- and post-training data*

A repeated-measures ANOVA with Session (2 levels: Pre- vs. Post-training) and Rhythm Type (2 levels: Metric complex vs. Metric simple) was conducted. This analysis

did not yield any significant effects of Session [ $F(1,179) = 1.95, p > .05$ ] or Rhythm Type [ $F(1,179) = .665, p > .05$ ] and the interaction between Session and Rhythm Type was also not significant [ $F(1,179) = 2.156, p > .05$ ]. It should be noted here that for both rhythm types (i.e., metric complex and simple), high pre-training accuracy scores (approximately 70%) were obtained. Surprisingly, despite the increased number of integer ratios in the rhythms used in Exp. 3, we noted a trend for slightly better accuracy scores for the metric simple rhythms in pre-training ( $M = .707$ ) as compared to the accuracy scores for the two integer-ratio metric simple rhythms in Exp. 1 ( $AV_{stat} = .646, AV_{inan} = .654, AV_{an} = .603$ ). Further inspection of the data revealed that the post-training accuracy scores in all three experiments reached a plateau at approximately 75%. These observations suggest that participants in Exp. 1 initially performed worse as compared to those of Exp. 3, but the training they received led to enhanced post-training performance, while this was not the case in Exp. 3. This might suggest that the factor responsible for the absence of learning effects in Exp. 3 was not task difficulty, but rather the increased memory load due to the increased number of intervals.

Contrary to Experiments 1 and 2, Experiment 3 showed that the number of integer ratios of the rhythmic patterns determines whether learning effects will be manifested behaviorally in a task consisting of static stimuli. Specifically, we noted a trend toward increased post-training accuracy for the metric complex as compared to metric simple rhythms, a finding that seems to be consistent with recent evidence suggesting that training on more difficult judgments may improve temporal acuity in a temporal perceptual learning paradigm (De Nier, Koo, & Wallace, 2016). Overall, the absence of post-training enhancement could be due to the use of four integer-ratio rhythms in this experiment as compared to the two integer-ratio rhythms used in Exps. 1 and 2 (cf.

Drake, 1993; Collier & Logan, 2000; Patel et al., 2005), which has probably prevented the improvement in post-training discrimination ability. It should be noted, however, that the null post-training effects could also be due to the small sample size used in this experiment.

## **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) and training modality (Experiments 1 and 2), as well as the complexity (i.e., number of integer ratios; cf. Drake, 1993) and metricality (i.e., metric simple vs. metric complex; cf. Grahn, 2012) of the rhythmic patterns (Experiment 3) so as to investigate the potential of post-training 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 stimulus types in training, when the rhythmic information is mediated by audiovisual or visual moving stimuli (i.e., Exps. 1 and 2). However, this enhancement is contingent on the number of integer ratios used within the rhythmic sequence (cf. Drake, 1993), with increased number of integers leading to a potential memory load and, thus, the absence of any processing benefits (i.e., Exp.3).

One of the main aims of Experiment 1 was to investigate whether animate or inanimate moving stimuli exert differential influences on subsequent processing of visual rhythms. Contrary to previous findings that biological motion affects time estimates (Blake & Shiffrar, 2007; Carrozzo et al., 2010; Lacquaniti, Carrozzo, D'Avella, Scaleia, Moscatelli, & Zago, 2014; Mendonça, Santos, & López-Moliner, 2011; Orgs, Bestmann, Schuur, & Haggard, 2011), facilitates temporal prediction of actions as

compared to inanimate moving stimuli (Stadler, Springer, Parkinson, & Prinz, 2012), and improves synchronization and rhythm discrimination accuracy (Su, 2014a, 2014b, Su, 2016; Su & Salazar-López, 2016; Wöllner, Deconinck, Parkinson, Hove, & Keller, 2012), we did not observe any animacy-related enhancements. This is probably due to the stimulus design we adopted (i.e., Su, 2014b), where the auditory stream consisted of intervals of different durations (i.e., 250, 500, 750, or 1000 ms), while the visual stimulation had a fixed timing (i.e., repetitive bouncing movement consisting of a downward and an upward movement phase, where each phase lasted 250 ms, thus corresponding to an inter-bounce interval of 500 ms), thus potentially any effect obtained could be driven by the auditory modality. We, therefore, speculate that the timing of the PLF's movement was the factor that might have minimized the potential of observing an animacy-driven benefit ( $AV_{an}$ ) in Exp. 1.

In Experiment 2, we eliminated the potential auditory dominance in Experiment 1, since post-training enhancement was also obtained when visual only stimulation was presented (cf. Grahn, 2012; Hove et al., 2013; Repp & Su, 2013). Our study adds to the current body of literature by being the first to show that training with two integer-ratio metric simple visual rhythms consisting of untrained static stimuli can enhanced processing of visual rhythms of static stimuli. This was in contrast to Barakat et al.'s (2015) findings where no significant post-training enhancements for the visual-only training group was observed. This conflict could 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).

The beneficial impact of multisensory training exhibited in Experiment 1 was not

replicated in Experiment 3. Given that the crucial difference between Exp. 1 and 3 was the number of integer ratios within the rhythm, it is possible that increasing the number of intervals might have in turn increased the memory load, thereby rendering it unlikely to efficiently store and process the rhythmic patterns and, thus, affecting the transfer of learning (Teki & Griffiths, 2014). Indeed, considering the Scalar Expectancy Theory (SET; the theory that posits that the standard deviation of time estimates increases linearly as a constant fraction of the mean; Gibbon, Church, & Meck, 1984) 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 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 to two integer-ratio rhythms (i.e., as those used in Exp. 1; Drake, 1993). This is further supported by neuroimaging data showing that when compared to 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 (i.e., Exp. 1) cannot necessarily be generalized to timing of four different interval lengths (i.e., Exp. 3; Grahn, 2012).

In conclusion, utilizing a perceptual learning paradigm, we showed visual rhythm processing can be benefited from training with multisensory and visual moving stimuli, this enhancement is modulated by the memory demands placed by the task. The



specific role of memory and task difficulty in rhythm perceptual learning requires further investigation. Future work on rhythm processing and perceptual learning should also focus in the use of more naturalistic and complex body movements (e.g., dancing), which are more efficient in communicating rhythmic information in multiple modalities.

## 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. doi:10.1371/journal.pone.0011283
- 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. doi: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. doi:10.1016/j.cortex.2008.07.002
- Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology*, *58*(1), 47-73. doi:10.1146/annurev.psych.57.102904.190152
- Carrozzo, M., Moscatelli, A., & Lacquaniti, F. (2010). Tempo Rubato: Animacy speeds up time in the brain. *PLoS ONE*, *5*(12), e15638. doi:10.1371/journal.pone.0015638
- Collier, G. L., & Logan, G. (2000). Modality differences in short-term memory for rhythms. *Memory & Cognition*, *28*(4), 529-538. doi:10.3758/bf03201243
- De Niar, M. A., Koo, B., & Wallace, M. T. (2016). Multisensory perceptual learning is dependent upon task difficulty. *Experimental Brain Research*, *234*(11), 3269-3277. doi:10.1007/s00221-016-4724-3
- Drake, C. (1993). Reproduction of musical rhythms by children, adult musicians, and adult nonmusicians. *Perception & Psychophysics*, *53*(1), 25-33.

doi: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. doi:10.1038/415429a

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).

doi: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. <http://dx.doi.org/10.1007/s00265-013-1491-z>.

Gibbon J., Church R. M., & Meck, W. H. (1984). Scalar timing in memory. *Annals of the New York Academy of Sciences*. *423*, 52-77.

doi:10.1111/j.1749-6632.1984.tb23417.x

Grahn, J. A. (2012). See what I hear? Beat perception in auditory and visual rhythms. *Experimental Brain Research*, *220*(1), 51-61. doi: 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. doi: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. doi: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 : The Official Journal of the Society for Neuroscience*, 29(23), 7540-7548. <http://doi.org/10.1523/JNEUROSCI.2018-08.2009>
- Grondin, S., & McAuley, J. D. (2009). Duration discrimination in crossmodal sequences. *Perception*, 38(10), 1542-1559. doi:10.1068/p6359
- 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). doi: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. doi: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. doi:10.1111/j.1749-6632.2009.04579.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 Neurobotics*, 8. doi: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*, 325-338. doi:10.1007/978-1-4939-1782-2\_17
- 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. doi: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. doi: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. doi: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. <http://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. doi: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. doi:10.1007/s00221-004-2159-8
- 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>
- 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>
- Roy, C., Lagarde, J., Dotov, D., & Bella, S. (2017). Walking to a multisensory beat. *Brain*

- and Cognition*, 113, 172-183. doi:10.1016/j.bandc.2017.02.00
- Shams, L., Wozny, D., Kim, R., & Seitz, A. (2011). Influences of multisensory experience on subsequent unisensory processing. *Frontiers in psychology*, 2, 264.  
doi: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. doi:10.1007/s00426-012-0431-2
- 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.  
doi: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. doi: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.  
doi:10.1155/2016/1678390

- Teki, S., & Griffiths, T. D. (2014). Working memory for time intervals in auditory rhythmic sequences. *Frontiers in Psychology, 5*, 1329.  
doi:10.3389/fpsyg.2014.01329
- Toiviainen, P., Luck, G., & Thompson, M. R. (2010). Embodied meter: Hierarchical eigenmodes in music-induced movement. *Music Perception, 28*(1), 59-70.  
doi:10.1525/mp.2010.28.1.59
- Welch, R., & Warren, D. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 88*, 638-667.  
<http://dx.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.  
doi:10.1037/a0028130