Exp Brain Res (2009) 198:339–352
DOI 10.1007/s00221-009-1817-2

RESEARCH ARTICLE

Multisensory integration of drumming actions: musical expertise affects perceived audiovisual asynchrony

Karin Petrini · Sofia Dahl · Davide Rocchesso · Carl Haakon Waadeland · Federico Avanzini · Aina Puce · Frank E. Pollick

Received: 30 September 2008 / Accepted: 11 April 2009 / Published online: 30 April 2009

Abstract We investigated the effect of musical expertise on sensitivity to asynchrony for drumming point-light displays, which varied in their physical characteristics (Experiment 1) or in their degree of audiovisual congruency (Experiment 2). In Experiment 1, 21 repetitions of three tempos × three accents × nine audiovisual delays were presented to four jazz drummers and four novices. In Experiment 2, ten repetitions of two audiovisual incongruency conditions × nine audiovisual delays were presented to 13 drummers and 13 novices. Participants gave forced-choice judgments of audiovisual synchrony. The results of Experiment 1 show an enhancement in experts' ability to detect asynchrony, especially for slower drumming tempos. In Experiment 2 an increase in sensitivity to asynchrony was found for incongruent stimuli; this increase, however, is attributable only to the novice group. Altogether the results indicated that through musical practice we learn to ignore variations in stimulus characteristics that otherwise would affect our multisensory integration processes.

Keywords Synchrony perception · Audiovisual integration · Audiovisual congruency · Drumming actions · Musical expertise

K. Petrini (✉) · F. E. Pollick
Department of Psychology, University of Glasgow,
58 Hillhead Street, Glasgow G12 8QB, Scotland, UK
e-mail: karin@psy.gla.ac.uk

S. Dahl
Department of Media Technology,
Aalborg University Copenhagen, Copenhagen, Denmark

D. Rocchesso
Department of Art and Industrial Design,
IUAV University of Venice, Venice, Italy

C. H. Waadeland
Department of Music,
Norwegian University of Science and Technology,
Trondheim, Norway

F. Avanzini
Department of Information Engineering,
University of Padova, Padova, Italy

A. Puce
Department of Psychological and Brain Sciences,
Indiana University, Bloomington, IN, USA

Introduction

Our understanding of the world comes from a fusion of different senses. In order to socially interact and behave appropriately we need to make sense of biological actions by integrating information from different senses. To us, however, sensory integration may seem effortless. Indeed humans are very good at maintaining accurate judgments of simultaneity for complicated sets of signals, such as vision and audition, which have different processing latencies due to differences in physical and neural transmission (King and Palmer 1985; Fain 2003; Spence and Squire 2003; King 2005). The human brain can tolerate some delay between auditory and visual information with no effect on the quality of audiovisual experience (Dixon and Spitz 1980). In other words, we can tolerate some amount of onset asynchrony between two sources of sensory information and still perceive them as pertaining to a single event. The size of this tolerance provides us with a measure of humans 'Temporal Integration Window' (TIW). The interest in temporal perception of synchrony has developed since Dixon and Spitz (1980) examined and compared the TIW

for audiovisual speech stimuli to that of other kinds of complex non-speech stimuli (a hammer repeatedly hitting a peg). However, since then the investigations concerning audiovisual integration have focused (beside speech) on much more simple and less naturalistic audiovisual events (light flashes, brief noise, clicks, etc.…; Spence et al. 2001; Stone et al. 2001; Sugita and Suzuki 2003; Zampini et al. 2003; Fujisaki and Nishida 2005) than on more complex and environmental non-speech events (music and object actions; Miner and Caudell 1998; Hollier and Rimell 1998; Vatakis and Spence 2006a, b).

The study of complex audiovisual non-speech events would help in better clarifying the extent to which factors influencing sensitivity to asynchrony for simple audiovisual events can account for more complex and natural audiovisual events. However, studies involving object action or music events have been almost completely neglected in favour of those involving complex speech events (Steinmetz 1996; Grant and Greenberg 2001; Grant et al. 2004; Navarra et al. 2005; Vatakis et al. 2007a).

Only recently, Vatakis and Spence (2006a, b) pointed out the necessity of investigating synchrony perception using other non-speech stimuli, such as music and object actions. Specifically, these authors emphasised how equally complex time-varying events (i.e. music) have been ignored, probably because they are not experienced by most people to the same extent as speech. To overcome these limitations Vatakis and Spence (2006a, b) assessed people's sensitivity to audiovisual asynchrony, by using short video clips of speech, object actions and music. They presented the three kinds of video clips at a variety of fixed SOAs (Stimulus Onset Asynchrony), and asked participants to judge which stream, the visual or the auditory, appeared first using a TOJ (Temporal-Order Judgment) task. The results of Vatakis and Spence (2006a, b) showed that people are less sensitive to asynchrony for speech than object actions (in agreement with Dixon and Spitz 1980) but, interestingly, people were found to be less sensitive to asynchrony in musical video clips than for either speech or object action video clips. The authors assumed that the differences in sensitivity to asynchrony for the three kinds of audiovisual events might be a consequence of differences in people's familiarity with the stimuli (i.e. almost all people have high familiarity with speech events, but far fewer are highly familiar with music events). Vatakis and Spence (2006b) addressed this point and found a clear effect of familiarity on sensitivity to asynchrony for musical and object actions events, but not for speech. Furthermore, they also found differences in the level of sensitivity to asynchrony within the musical event category. Whilst for guitar videos the best perceived synchrony was achieved when the visual stream lead the auditory stream, for piano videos the auditory stream had to lead the visual stream.

The complexity of results provided by Vatakis and Spence (2006a, b) leave us with renewed interest in audiovisual temporal perception for complex non-speech events and with a lot of unanswered questions. For example, when demonstrating that familiarity plays a role in sensory integration abilities it may be interesting to test how enhanced levels of expertise (due to long-term training) can affect sensitivity to asynchrony. Addressing such a point would enhance our understanding of how humans' sensitivity to asynchrony is modified by practice and long-term repeated exposure to specific multisensory events; it also would create a background of knowledge for future neuroimaging and electrophysiological studies interested in functional and structural modification of cortical and subcortical regions consequent to musical practice (Bengtsson et al. 2005; Gaser and Schlaug 2003; Hodges et al. 2005; Bermudez and Zatorre 2005).

Music shares a lot of characteristics with speech. As one of the most primitive means of social interaction it is still one of the most effective forms of communication besides speech. Also similarly to speech, it is composed of perceptually discrete elements organised in time-varying sequences. Thus when looking at a musician playing his instrument (as when looking at someone speaking), at a concert or in a video clip, we integrate visual and audio information to create a singular perceptual event. However, the act of making music is not experienced or practiced as much as speech (Vatakis and Spence 2006a, b) and this makes music events a perfect tool to discriminate between experts and non-experts. The playing of music then is not only a useful tool to better understand fundamental issues in speech processing (with which it shares many similarities), but also a special kind of non-speech stimulus requiring extensive study itself. To date only Arrighi et al. (2006) have embraced this exigency by examining the conditions necessary for audiovisual simultaneity using drumming video clips. They showed that the width of the TIW and the point of subjective simultaneity (PSS) varied inversely with drumming tempo (i.e. the faster the drumming tempo the narrower the TIW and the smaller the sound delay necessary to perceive the best synchrony) and that the PSS always occurred when the auditory pattern was delayed with respect to the visual one. These results are in agreement with the findings by Vatakis and Spence (2006b) for guitar videos and object actions, as opposed to speech and piano playing videos. These results provide us with an initial understanding of how vision and sound from different musical events are integrated, and suggest that sensitivity to asynchrony may change considerably with change in stimulus physical characteristics thanks to high flexibility of TIW (Navarra et al. 2005; Vatakis et al. 2007b).

Here, we investigate whether professional musical training affects the human TIW and PSS by exploring any difference

in audiovisual synchrony perception between these groups. To the best of our knowledge no study has attempted to clarify how audiovisual integration processes change with musical expertise. Nevertheless some evidence indicates differences between musicians and non-musicians in detecting asynchrony (Miner and Caudell 1998), while other evidence does not support such a difference (Vatakis and Spence 2006b). Finally, activation differences between musicians and non-musicians have been demonstrated in brain regions thought to be involved in audiovisual integration (Hodges et al. 2005).

To investigate the differences between musicians and non-musicians in perceiving temporal synchrony we used point-light displays (Johansson 1973) of drumming actions. The decision to use this kind of display was made after considering the similarity in sensitivity to asynchrony found by Arrighi et al. (2006) when using video clips of drumming actions or biological motion displays that preserved the same motion profile. Also, the use of point-light displays is at the core of current research on audiovisual integration of human actions (Saygin et al. 2008; Brooks et al. 2007; Luck and Sloboda 2007), where the focus of interest concerns the study of biological motion processing from a multisensory point of view. Point-light displays allow us to isolate the effects of perceiving the biological motion from contextual factors, which is key to differentiating between experts and non-experts in the present work. Thus, because our primary interest concerned the effect of expertise on audiovisual integration mechanisms, the stimuli were chosen to enhance differences in familiarity between expert musicians and non-musicians, and to preserve some of the richness of human activity while enabling parametric manipulation of the audio and visual signals. For this purpose we used point-light displays in combination with the synthetic sound of a professional jazz drummer playing a swing groove, where "swing" denotes a jazz style.

The term "swing" is strictly related to rhythmic qualities of music performance so that, when played, the listener is compelled to move fingers, feet, head and so on along with the music. More precisely *swing groove* can be defined as a particular rhythmic ostinato played by jazz drummers (Waadeland 2003, 2006). Rhythmic isochronous sequences have been already widely used in psychoacoustics to study just noticeable differences for timing perturbation of a tone (see Friberg and Sundberg 1995 for a review of the topic), by showing that sensitivity to perturbation changed with changes in IOIs (Inter Onset Intervals), i.e. with changes in tempo. However, whereas Friberg's and Sundberg's (1995) study did not find an effect of expertise on sensitivity to tone perturbation, Drake and Botte (1993) found a significant difference between musicians and non-musicians with regard to temporal discrimination. Hence, examining the effect of expertise on audiovisual synchrony for a rhythmic sequence appears a natural continuation of these studies where only an auditory stream and less naturalistic stimuli were used.

We built on the few studies described above, regarding the playing of music, by assessing musicians' and non-musicians' sensitivity to audiovisual asynchrony in two different experiments. In the first experiment, we investigated whether changes in physical characteristics (i.e. tempo and accent) of a musical pattern differently affect the perception of asynchrony of professional jazz drummers and musical novices. This was done by creating drumming displays for three different tempos (see Arrighi et al. 2006), and three different accents of the swing groove. We also investigated, in a second experiment, whether the elimination of any of the natural correspondence between the time-varying characteristics of the visual and the auditory stimulation (achieved by eliminating the covariation between drummers' movements and relative sound) would affect differently the asynchrony judgments of musicians and non-musicians.

Incongruency between different sensory modalities has been already demonstrated to affect temporal perception (Spence and Walton 2005) and recently a clear effect of audiovisual incongruency has been found on perceived asynchrony when using the 'McGurk effect' (McGurk and MacDonald 1976; MacDonald and McGurk 1978) to disrupt the correspondence between visual and auditory speech information (Munhall et al. 1996). That is, the audiovisual incongruency generated by the McGurk effect was found to facilitate the detection of asynchrony in comparison to the congruent condition (van Wassenhove et al. 2007). Also, this effect of facilitation was found by Vatakis and Spence (2007), using gender mismatch (i.e. hearing the speech of one gender but seeing the face of the other) in their audiovisual temporal order judgments task. Hence, our aim in the second experiment is to examine whether this facilitation can be generalised to other non-speech events and to people with different levels of musical expertise.

Experiment 1

The purpose of this experiment was to investigate differences between expert drummers and novices in perceiving audiovisual synchrony of drumming actions, when different physical characteristics of the played swing groove were manipulated. Specifically, we aim to examine whether changes in characteristics of the musical pattern such as tempo and accent would differently affect drummers and novices.

Methods

Participants

Four expert jazz drummers (all males with eight to 13 years of drumming training and with a mean age of 20.5) and four novices (all males with no experience in drumming or any other music instrument and with a mean age of 19.5), with normal hearing and normal or corrected-to-normal visual acuity, participated in the experiment. The study received IRB approval and all participants gave informed consent to participate. Participants received cash or course credits for their participation.

Apparatus and stimuli

All the visual stimuli were presented on a Sony Trinitron screen, by a Macintosh G4 OS9, with resolution 1,024 × 768 pixels, 32-bit colour depth, and refresh rate of 60 Hz. Auditory stimuli were digitized at a rate of 44,100 Hz and presented through two high-quality loudspeakers (Harmon/Kardon) flanking the computer screen and lying in the same plane 1 m from the participant. Speaker separation was 55 cm and stimuli range intensities at the sound source varied from 65 to 92 dB.

Point-light displays

Point-light displays of a professional jazz drummer playing a swing groove (Fig. 1a) were generated using 3D motion capture data (240 Hz), previously obtained by Waadeland (2003, 2006) to analyse jazz drummers' movements during performance. The data coordinates representing the shoulder, elbow, wrist, hand joints and two points on the drumstick (one at the level of the grip and the other of the drumstick head) of the drummer's right side were used to reproduce the video and the audio for the drumming displays. For the video we converted the 3D motion coordinates of the drummer performance into point-light displays, by importing them into a Matlab script and using the Psychtoolbox routines (Brainard 1997; Pelli 1997). The points were represented in the display by white discs (luminance: 85 cd/m²; diameter: 2 mm) on a black background (luminance: 0.12 cd/m²), while a thick white line, oriented 25° from horizontal, represented the drumhead (width: 2.2 cm; height: 2 mm; luminance: 85 cd/m²). As the drummer performed the swing groove at three different tempos (60, 90 and 120 beats per minute) and with three different accents (on the first, second or third beat), nine sets of movement data (three tempos × three accents) were converted into point-light displays. The videos were produced using OSX and OpenGL functions of Matlab toolbox version 4, and saved as AVI files of different duration (25, 20 and 15 s) in accord with the drumming tempo (60, 90, and 120 BPM).

Sound

The matching synthetic sounds were obtained by a simulation of the first 25 modes of a circular membrane (Fontanaet al. 2004) that takes as input the time and impact velocity of a strike and output the resulting audio signal. Both time and impact velocity were derived by plotting the displacement and velocity of the drumstick head marker versus time and selecting, for each impact, the frame at which the drumstick head velocity changed from negative to positive (Fig. 2) representing the strike event (Dahl 2004, p 765). To use only the vertical displacement and vertical velocity of the drumstick marker versus time (Dahl 2004, p 765), motion data were rotated to a coordinate frame where horizontal was parallel to the drumhead and vertical perpendicular to the drumhead, before collecting impact times and relative velocities (Fig. 2). The resulting nine sounds were saved as WAV files and the duration of the sound files, as well as for the videos, varied (25, 20 and 15 s) in accord with the drumming tempo. The range of intensities at the sound source was 65–90, 71–91 dB, and 76–92, respectively, for 60, 90 and 120 BPM.

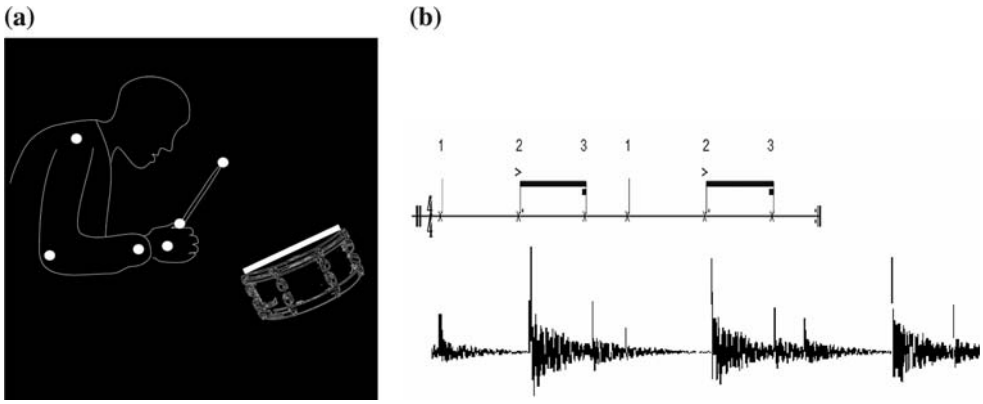

Fig. 1 **a** Frame sample of the jazz drummer point-light displays; **b** waveform sample of the 9 impacts sound selection, with relative 3-beat cyclic pattern for accent on 2nd beat in the swing groove (see Waadeland 2006 for more details). The numbers 1, 2 and 3 at the top of the figure indicate, respectively, the 1st, 2nd, and 3rd beat in the pattern, while >indicates the accented beat

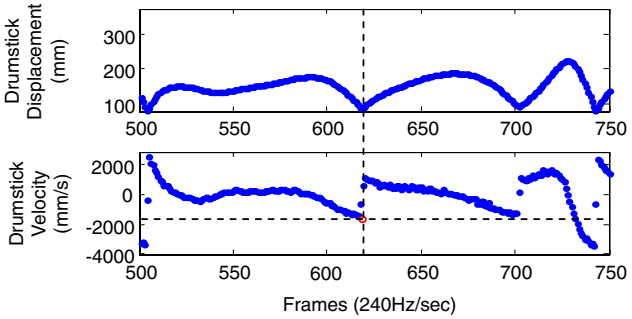

Fig. 2 Vertical displacement of drumstick marker versus time (*top diagram*) for 250 frames (240 Hz/s), and corresponding vertical drumstick marker velocity (*bottom diagram*) for a swing groove played at 120 BPM/1st accent. The *vertical dashed line* shows the correspondence of impact displacement and change in velocity direction. The last point before the change in velocity direction (*red open circle*) was taken as the impact frame and the corresponding velocity derived (*horizontal dashed line*)

Movies

To produce the audiovisual QuickTime[1] movies (60 Hz) used as stimuli (Movie 1: example of movie with the drummer playing the swing groove at 120 BPM and with accent on 2nd beat), the video (AVI) and audio (WAVE) files were imported in Adobe Premiere 1.5. Hence, they were combined by delaying the video with respect to the audio (AV) by 0, 4, 8, 12 and 16 frames (corresponding to AV lags: 0, –66.67, –133.33, –200, and –266.67 ms) and the audio with respect to the video (VA) of the same frames' numbers (corresponding to VA lags: 0, 66.67, 133.33, 200, and 266.67 ms) for a total of nine audiovisual lags. After discarding the initial 5-s of the audio and video files (to assure a stabilization of the drummer performance) a sound selection of nine impacts (Fig. 1b) was made always starting from two frames before the first impact and ending at one frame before the tenth impact. The sound selection was kept constant for the nine different lags within each tempo/accent condition, while the video was selected each time accordingly. Hence, the video sequence was shifted along the Adobe Premiere timeline to be delayed (AV) or advanced (VA) with respect to selected audio to create, respectively, the four negative and four positive audiovisual lags. Since the total amount of audiovisual display available was much longer than that needed for the nine impacts there was always audio and video at the beginning and at the end of each experimental display.

The resulting QuickTime movies were compressed using QuickTime Pro 6, and were shown to participants through Showtime (Watson and Hu 1999), a component of Psychophysics Toolbox (Brainard 1997; Pelli 1997) extensions to Matlab (see also Zhou et al. 2007, for Showtime used in multi-sensory integration). As the number of impacts was kept constant at nine for all the tempo and accent conditions, the duration of the final movies varied from 5.21 s for the slower tempo to 3 s for the faster one.

Procedure

Observers sat in a darkened room at a distance of approximately 1 m from the screen where the stimulus was displayed. The experimenter, after briefly explaining the task to the participant, asked him to read the instructions written on the screen. The experiment consisted of 21 blocks of the 81 movies (for a total of 1,701 presentations) run in three, daily-separated sessions of seven blocks each (for a total of 567 presentations), to control for effects of fatigue. Participants had to press "1" on the keypad if for them the drummer's movements were in synchrony with the sound, or to press "3" if they were perceived as asynchronous. Before starting the experiment participants performed a brief set of practice trials to familiarise themselves with the task and to give them also the possibility to ask for further clarifications in case the task was not completely understood. After the short training the experimenter left the participant alone and the participant could start the experiment by clicking the mouse. The 81 movies within each block were presented randomly, and at the end of each block participants could take a few minutes' break and start the next one by clicking the mouse. After each movie, a short text message was displayed at the centre of the black screen to remind participants of the task and of the keys to use. When either response key was pressed the next movie was automatically displayed and so on until the end of the block.

Results and discussion

In this experiment, we delayed by a variable amount either the audio or the video of point-light movies of a jazz drummer playing a swing groove and measured the frequency with which four expert jazz drummers and four novices judged the video and audio of the movies to be in synchrony. The results for the three different tempo × three accent conditions are shown in Fig. 3. The primary analysis involved finding the best-fitting Gaussian to synchrony response rates, from which the TIW and the PSS could be

[1] Using a series of static frames sampled at 30 Hz like Arrighi et al. (2006) would increase the possibility that the frame showing the actual point of contact between the percussor and the resonator is not actually shown (see Arrighi et al. 2006, p 266); we used QuickTime movies sampled at 60 Hz to maintain as much as possible the original information of the drummer's performance.

derived. The Gaussians were taken as an estimate[2] of participants' TIW width and the time of the peaks of the Gaussians as their PSS. This analysis is identical to that used by Arrighi et al. (2006) and Petrini et al. (2009) so as to allow comparisons between the data.

In Fig. 3 lines represent the best-fitting Gaussian curves (all with R^2 values ranging from .84 to .96 for novices and from .93 to .99 for drummers[3]) for both experts' and novices' averaged data (top diagrams and bottom diagrams, respectively). The three colours and styles—blue-solid, green-dashed, red-dotted—correspond, respectively, to the three different drumming tempos (60, 90, and 120 BPM), plotted separately for each accent condition (from left to right). The inter-subject variability within each group was found to be quite low as indicated by the error bars (representing the standard error) in Fig. 3. The PSS or the perceived best alignment between the visual and auditory stimuli occurred, with no exception, when the audio was delayed with respect to the video. However, there is a general effect of expertise on the width of TIW and on the size of the best auditory delay to perceive synchrony; indeed experts' PSSs and TIW widths are much smaller than those of novices.

Figure 4 plots the PSSs and TIW widths against drumming tempo after collapsing the data across the three accent conditions. Accent had very little effect. The only apparent effect occurs at the faster tempo for novices' PSS, which is closer to the physical point of synchrony for the 1st accent condition with respect to the other two. Novices' PSSs for 60–120 BPM occurred when the auditory stream was delayed from about 80 to 40 ms, while drummers' PSS occurred when the auditory stream was delayed from about 40–20 ms (Fig. 4). Furthermore, novices' TIW widths decreased from approximately 200 to 150 ms with increasing tempo, while drummers' TIW widths from 130 to 100 ms. A linear relationship between delay and tempo (Arrighi et al. 2006; Petrini et al. 2009) can well describe the trend in the data for both PSS and TIW width as shown by the slopes of the best linear fits in Fig. 4. Hence, for both measures of sensitivity to asynchrony the effect of tempo is more accentuated for the novices' group when compared to the drummers' group.

An analysis of variance with a 2 (expertise) × 3 (tempo) × 3 (accent) mixed factorial design was conducted on the PSS and TIW width data to test our observations. The between-subjects factor "expertise" was found to have a significant effect on both PSS ($F(1, 6) = 6.345$, $p = .045$) and TIW width ($F(1, 6) = 8.919$, $p = .024$). Also, the within factor "tempo" significantly affected the PSS ($F(2, 12) = 5.366$, $p = .021$) and TIW width ($F(2, 12) = 9.701$, $p = .003$), while no effect of accent was found on either PSS ($F(2, 12) = 1.137$, $p = .35$) or TIW width ($F(2, 12) = 1.850$, $p = .19$); likewise a tempo × expertise interaction was not found for either PSS ($F(2, 12) = 1.592$, $p = .24$) or TIW width ($F(2, 12) = 1.670$, $p = .22$). However, Fig. 4 reveals that while drummers do not differ much in their judgment of asynchrony when presented with swing groove stimuli played at 60, 90 and 120 BPM, novices do differ consistently in their sensitivity to asynchrony especially when it comes to the slower tempo (60 BPM). In other words, only the performance of the novices' group is significantly worsened by the slower tempo condition. A series of repeated contrast measures were run on the PSS and TIW width data to compare the effect of each level of tempo, after collapsing the data across the different accent conditions (Fig. 4). A significant interaction between tempo and expertise was found when comparing the TIW width for swing groove played at 60 BPM to that played at 90 ($F(1, 22) = 5.558$, $p = .028$), while no significant effect of interaction was found for the two faster tempos ($F(1, 22) = .012$, $p = .915$). Also a similar result was found when comparing the PSSs for 60 and 90 BPM, though for this measure the interaction was only marginally significant ($F(1, 22) = 4.220$, $p = .052$).

In summary, musical expertise was found to affect sensitivity to asynchrony by reducing the extent of both TIW and of PSS. Also, consistent with the results of Arrighi et al. (2006), we found that sound must be delayed with respect to sight to produce the perception of best audiovisual synchrony, and that the size of the sound delay (PSS) is negatively correlated with drumming tempo. The effect of tempo on sensitivity to asynchrony was found to be more evident for the novices group, which performed poorly in detecting asynchrony at the slower drumming tempo (Petrini et al. 2009). This is an interesting result that indicates how musical practice enhances sensibility to asynchrony by mitigating the influence of other signals' characteristics such as tempo.

However, would this difference between drummers' and novices' sensitivity to asynchrony still be evident when decreasing the coherence between auditory and visual stimulation? In other words, will drummers show a higher

[2] The estimate of TIW width was derived by calculating the SD of the normal distribution, which is known to be equal to FWHM/2.3548. The FWHM or full width at half maximum is a simple measure of the width of a distribution, and is easily obtained from empirical distributions. The FWHM for a distribution described by the probability density $f(x)$ is defined by the absolute difference between one point on the left x_2 and one on the right x_1 of the mode x_m (defined by $f(x_m) = \max$), with $f(x_1) = f(x_2) = f(x_m)/2$.

[3] R^2 measures how successful the fit is in explaining the variation of the data. That is, R^2 is the square of the correlation between the response values and the predicted response values. R^2 is defined as the ratio of the sum of squares of the regression (SSR) and the total sum of squares (SST). R^2 can take on any value between 0 and 1, with a value closer to 1 indicating that a greater proportion of variance is accounted for by the model.

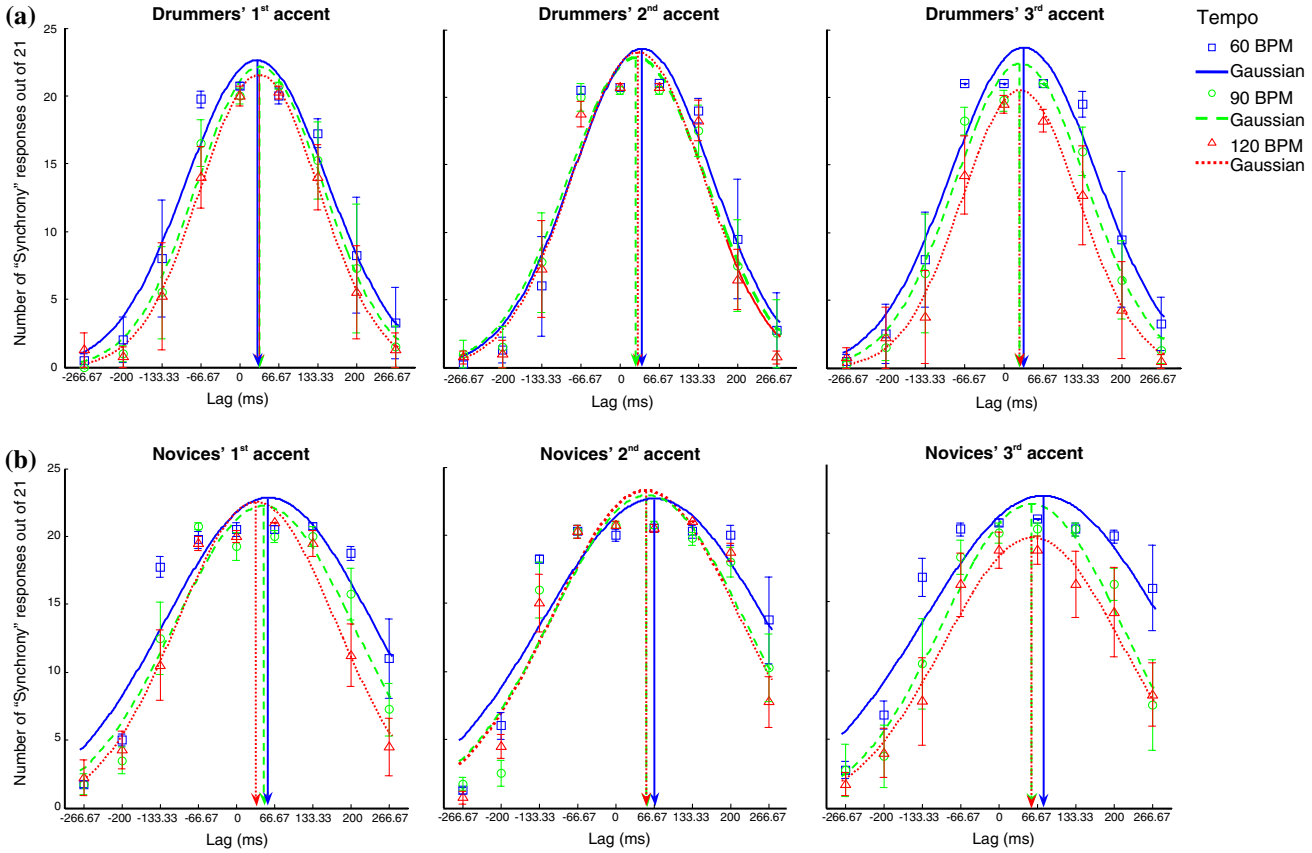

Fig. 3 Results for Experiment 1: Number of "Synchrony" responses out of 21 as a function of audiovisual delay for drummers (*top diagrams*) and novices (*bottom diagrams*) showed separately for the three accent conditions. Negative delays indicate AV lags (visual stream was delayed), while positive delays VA lags (visual stream was advanced). *Blue-solid, green-dashed, red-dotted lines* represent the best-fitting Gaussian curves, respectively, for 60, 90 and 120 BPM, while symbols of the same colours represent the data. The peaks of the Gaussian curves provide an estimate of the PSS (point of subjective simultaneity). The *error bars* represent the standard errors of the mean

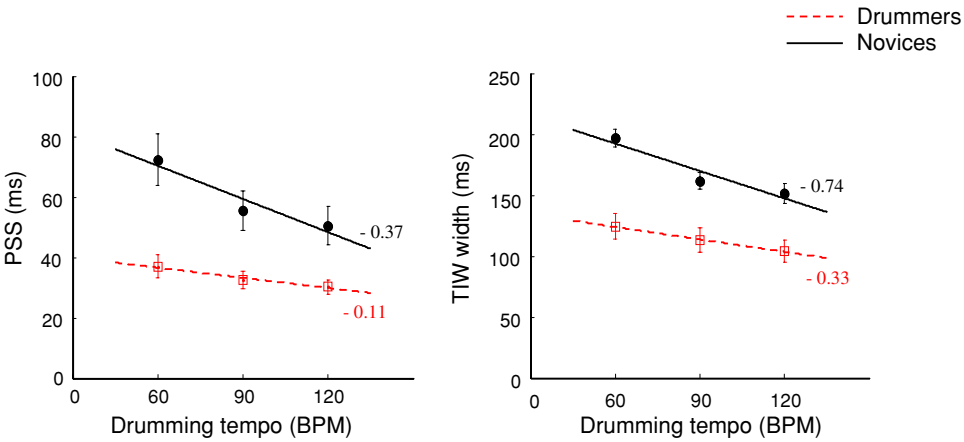

Fig. 4 Results for Experiment 1: PSS and TIW width for drummers and novices plotted as a function of drumming tempos, after collapsing the data across the different accent conditions. The *solid* and *dashed lines* are the best linear fits to the data, with their slopes. The *error bars* represent the standard errors of the mean

sensitivity to asynchrony even when their expectancy is contradicted? To answer to this question we investigated the effect of expertise on perceived audiovisual synchrony in a further experiment, where the effect of audiovisual incongruency was examined.

Experiment 2

This experiment aimed to investigate the effect of audiovisual incongruency on drummers' and novices' judgments of audiovisual asynchrony. Thus, we aimed to examine

whether audiovisual incongruency would facilitate the detection of asynchrony even for our complex non-speech stimuli and for groups with different levels of musical expertise.

To investigate the effect of audiovisual congruency, only one tempo and one accent condition were selected. This was done to reduce as much as possible any confounding effect of tempo or accent on drummers' and novices' sensitivity to asynchrony for the congruent condition (for which the original auditory and visual stimulation was used), so that any effect of expertise found on PSS and TIW width could be explained only by the elimination of audiovisual congruency. From the results of the first experiment we knew that the faster tempo (120 BPM) was the one for which both groups were more sensitive to asynchrony as demonstrated by the PSS and TIW width data. Moreover, the second accent condition of swing groove appeared to be the one for which both drummers' and novices' TIW width was very consistent across all the different tempos (Fig. 3: central top and bottom diagrams).

Before running the experiment we ran a pilot with a new group of three expert jazz drummers and a new group of three novices using a similar apparatus to that of Experiment 1. Once the 120 BPM/2nd condition was selected we manipulated the covariation between sound and sight to study the effect of audiovisual incongruency. In the pilot three different levels of incongruency were used (*Congruent condition*: where the covariation between the auditory and visual information was maintained; *Averaged/Incongruent condition*: where the covariation between auditory and visual information was eliminated by giving as input to the sound model the averaged impact velocity; *scrambled/incongruent condition*: where the covariation between auditory and visual information was eliminated by giving as input to the sound model the list of impact velocities in randomized order). The procedure and task of the pilot was the same as in Experiment 1, but the stimuli were presented to the participants through two different sound sources (headphones and loudspeakers). Results showed a significant effect of audiovisual incongruency on sensitivity to asynchrony and no effect of the different sound sources. Thus, we selected, for the main experiment, only two levels of incongruency (*congruent-scrambled/incongruent*) along with the headphones as sound source.

Methods

Participants

Thirteen new drummer experts (all males with a mean of 23 years of drumming training and with a mean age of 36.7) and 13 new novices (all males with no experience in drumming or any other music instrument and with a mean age of 36.6), were recruited for this experiment. All participants had normal or corrected-to-normal visual acuity and hearing. All of them gave informed consent to participate in the study which was approved by the Institutional Review Board at West Virginia University and received cash for their participation.

Apparatus and stimuli

All the visual stimuli were presented on a Macintosh PowerBook G3, running OS9, with resolution 1,024 × 768 pixels, and a refresh rate of 60 Hz. Auditory stimuli were presented through headphones (Beyer Dynamic DT Headphones). Stimuli range intensities at the sound source varied 50–70 dB. The tempo and accent condition of all stimuli, as mentioned above, was 120 BPM/2nd accent. The two levels of incongruency used were:

Congruent condition

The congruency between auditory and visual stimulation was maintained by using a point-light display for the condition 120 BPM/2nd accent, previously used in Experiment 1, where input to the sound model (Fontana et al. 2004) was the original impact velocity (Fig. 5a). The wave file used to create the congruent drumming stimulus had a mean intensity of 71 dB and duration of 15 s.

Scrambled/incongruent condition

To eliminate both audiovisual covariation and regularity in the sound the scrambled list of impact velocities (Fig. 5b) was given as input to the sound model. That is, the same original impact values of the congruent condition were used but they were given in a scrambled order to the sound model. As previously, the impact time remained unchanged. The WAV file used to create the scrambled/incongruent drumming stimulus had mean intensity of 71 dB and duration of 15 s.

Procedure

The procedure was identical to that of Experiment 1, except that it consisted of 10 blocks of 18 movies each.

Results and discussion

The averaged results are shown in Fig. 6a where the black-solid, and black-dashed lines represent the best-fitting Gaussian curves (all with R^2 values ranging from .77 to .97 for novices and from .81 to .99 for drummers), respectively, for the Congruent (C), and scrambled/incongruent condition

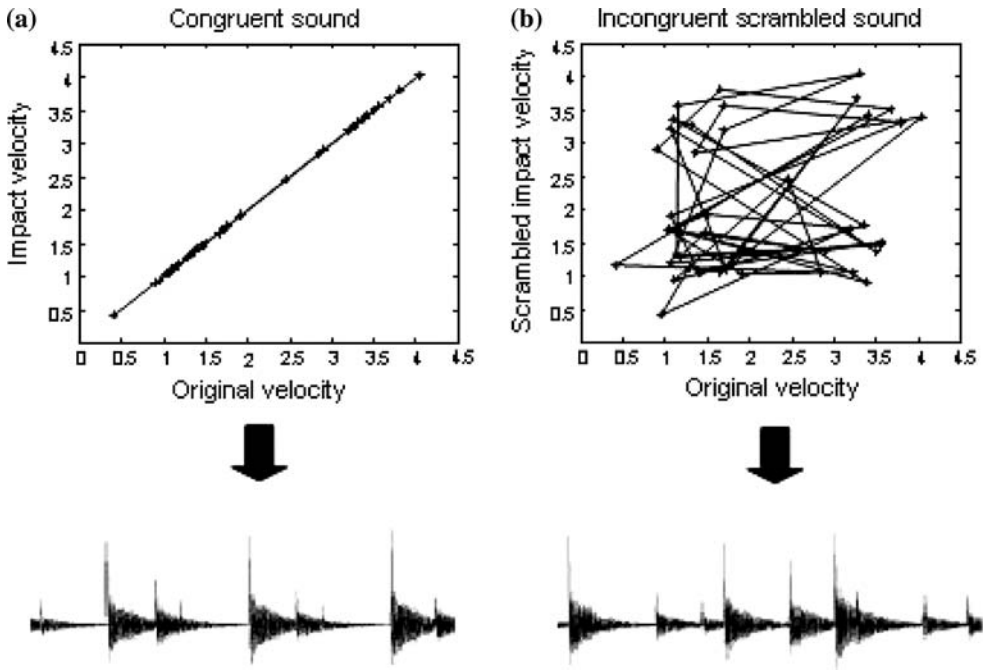

Fig. 5 Example of the mapping from original velocities measured to the impact velocities used by the algorithm to generate the sounds. Conditions used include (a) the original and (b) the scrambled velocities. The two audio files resulting from this manipulation were used to create audiovisual drumming displays with different level of visual and auditory incongruency (congruent and scrambled/incongruent)

(S/I). Overall, drummers have much smaller TIWs width similarly to what we found in Experiment 1. Also an apparent difference in the effect of incongruency on drummers' and novices' TIW width emerged (Fig. 6). Indeed, it is easy to see from Fig. 6b that the effect of incongruency on the TIW width is only obviously present for the novices group, which have an evident improvement in detecting asynchrony when auditory and visual stimuli do not covary together. This was not true for experts, who had the same kind of performance with both congruent and incongruent audiovisual stimuli.

The inter-subject variability is once again quite low as indicated by the error bars in Fig. 6a, and though the PSS values for both groups in this experiment are higher than Experiment 1, this is probably due to the reduced number of stimulus presentations.

An analysis of variance 2 (expertise) × 2 (congruency) for a mixed factorial design was conducted on the PSS and TIW width data to test our observations. The between-subjects factor "expertise" (PSS: $F(1, 24) = 0.036$, $p = .851$; TIW width: $F(1, 24) = 14.037$, $p = .001$), the within-subjects factor "congruency" (PSS: $F(1, 24) = 2.134$, $p = .157$; TIW width: $F(1, 24) = 32.730$, $p < .001$) and the interaction between the two factors (PSS: $F(1, 24) = 1.209$, $p = .282$; TIW width: $F(1, 24) = 11.926$, $p = .002$) were found to be highly significant *only* for the TIW width.

Figure 6b plots PSS and TIW width against incongruency. A significant effect of congruency is evident only for novices, but not for drummers, which is why we found a significant interaction between expertise and incongruency. This result is made evident by the steeper slope (−34.82 ms/I) in Fig. 6b for the novices group when going from congruent to incongruent condition.

In summary, the results of this further experiment both confirm and extend the results of Experiment 1. Indeed, a general effect of expertise on the TIW width was found as well as in Experiment 1, confirming the enhancement in sensibility to asynchrony for the drummers group. However, the audiovisual incongruency was found to narrow only the TIW of the novices group, suggesting that only this group is facilitated by a mismatch between auditory and visual information when detecting asynchrony between the two signals.

General discussion

The results of the two experiments described in this study support previous findings on synchrony perception of audiovisual actions, and also add further evidence to help us understand how musical expertise affects perceived sensitivity to audiovisual asynchrony.

Effect of tempo and accent on sensitivity to asynchrony

In Experiment 1, in agreement with drumming results of Arrighi et al. (2006), we found that sound must be delayed

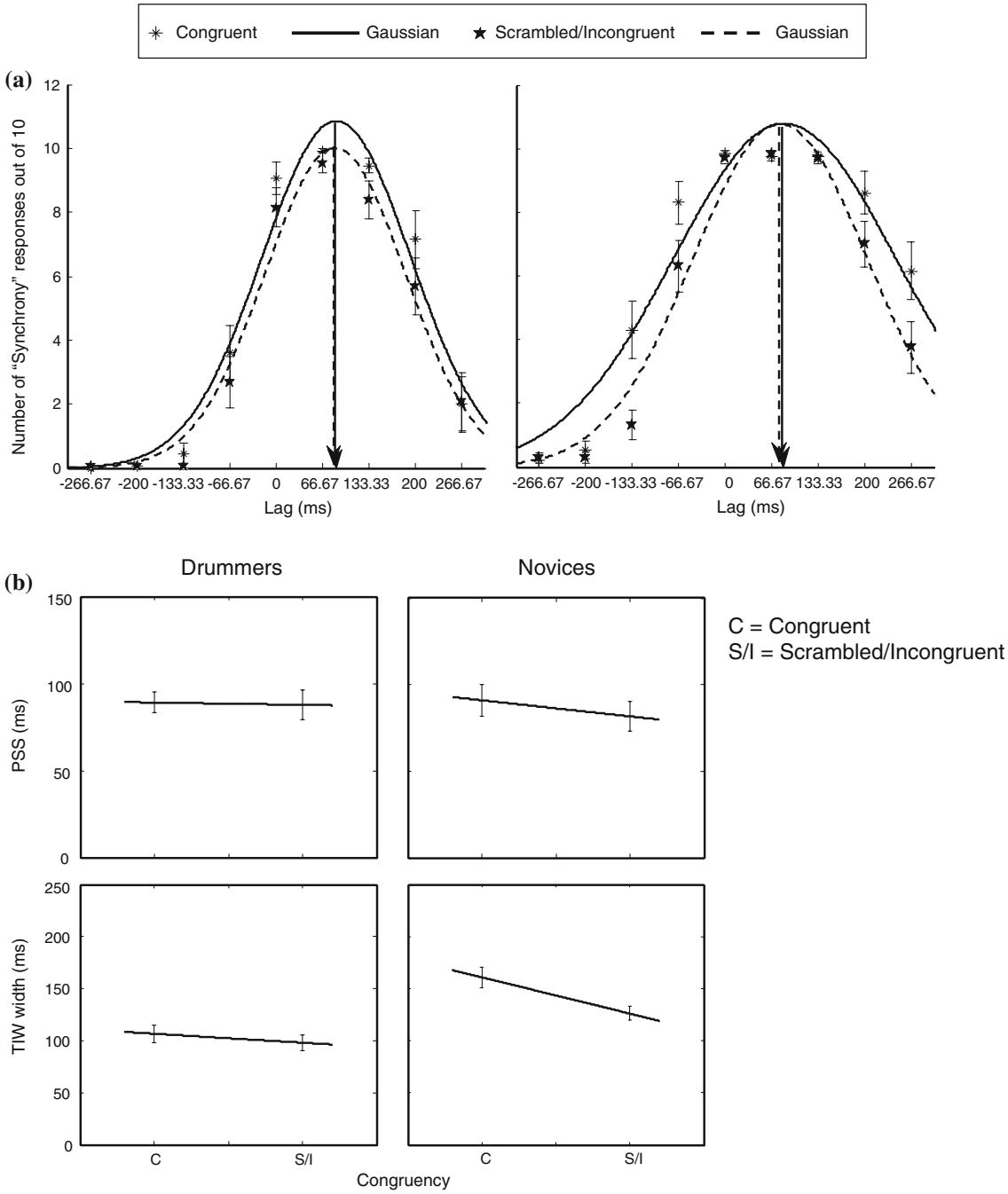

Fig. 6 Results for Experiment 2: **a** Number of "Synchrony" responses out of ten as a function of audiovisual delay for drummers (on the *left*) and novices (on the *right*), where negative delays indicate AV lags (visual stream was delayed), while positive delays VA lags (visual steam was advanced). *Black-solid* and *black-dashed lines* represent the best-fitting Gaussian curves, respectively, for congruent (C) and scrambled/incongruent (S/I) condition. **b** PSS and TIW width plotted as a function of audiovisual incongruency. The *solid lines* represent the linear fits to the data, with their slopes. The *error bars* represent the standard errors of the mean

with respect to sight to produce the perception of best audiovisual synchrony, and that the size of the sound delay (PSS) is negatively correlated with drumming tempo. That is, in order to perceive synchrony the sound has to be delayed less for a faster tempo of swing groove than for a slower tempo. This effect of tempo increase on synchronisation has also been found recently by Luck and Sloboda (2007) in their study of synchronization to three-beat conductors' gestures indicating that faster tempos consistently improve perception of synchrony no matter what kind of

task participants are required to do (e.g. judge simultaneity vs. tap in synchrony with the beat). Also, in agreement with Arrighi et al. (2006) we found that the range of delays supporting audiovisual synchrony (TIW width) could be very broad, especially at slow tempo, and that the increased range with drumming tempo depends on an asymmetric lowering of the upper bound (Novices: −25 to 170 ms for 60 BPM; −25 to 136 ms for 90 BPM; −25 to 126 ms for 120 BPM; Drummers: −25 to 99 ms for 60 BPM; −24 to 89 ms for 90 BPM; −21 to 82 ms for 120 BPM), especially for novices. Finally, the width of the TIW was found to be inversely related to drumming tempo, with the width decreasing as the drumming tempo increased.

Whereas the lagging of auditory information with respect to visual information can be accounted for by adaptation of the central nervous system to differences in the speed of light and sound (Massaro et al. 1996), we do not have a clear idea of why PSS and TIW width should decrease with drumming tempo. Arrighi et al. (2006) explained this result by indicating that when people are judging simultaneity between two signals (in a phase-based task), rather than simultaneity between changes in the two signals, frequency matters (Clifford et al. 2003) and sensitivity to asynchrony increases for higher frequency. Thus they argued that their task was necessarily a phase alignment task because in their movies the frame that showed the actual point of contact between hand and drum could be missing. However, unlike Arrighi et al. (2006) our movies were sampled at 60 Hz instead of 30 Hz, enhancing the possibility that the frame showing the contact point between drumstick and drumhead was presented, thus reducing the possibility for the task to be performed by phase alignment.

An alternative explanation of the effect of different tempos (different frequencies in other terms) on sensitivity to asynchrony could refer to the notion of humans' spontaneous tempo. Spontaneous tempo, as determined from subjects freely tapping out a rhythm with their finger, was found to average ∼2 Hz (Kay et al. 1987; Collyer et al. 1994; Moelants 2002), and also was found to correlate with the cadence of walking in locomotion experiments (Murray et al. 1964) as well as in extended periods of natural activity (MacDougall and Moore 2005). Furthermore, a preference towards 2 Hz frequency of movement has been observed in music (Moelants 2002), where a clear preference for rhythm of 2 Hz (120 BPM) was found sampling a wide selection of Western music and asking participants to tap along with it. Therefore, if we possess a preferred tempo at around 120 BPM, we should expect that we would be better in detecting audiovisual asynchrony for swing groove approaching this tempo. Our results are consistent with this assumption.

Finally, no effect of accent was found in the first experiment. This result suggests that striking velocity and preparatory height (Dahl 2004; Schutz and Lipscomb 2007) are not salient information when judging synchrony between visual and auditory streams, at least as long as coherent audiovisual information is presented. Thus, though the kinematic information contained in the accented gestures are probably very important for musicians (Luck and Sloboda 2007; Luck and Nte 2008) when performing along with conducting gestures, they might become irrelevant when no motor synchronization is required. However, because in our experiment we compared three different accented conditions (i.e. the accent was always present but in different positions of the cyclic drumming sequence), instead of comparing a condition with an accent to one without an accent, we cannot exclude the possibility that the presence of an accent in itself would make a difference in the way we judge audiovisual asynchrony. Nevertheless, because accents can be defined as events that attract attention (Jones 1987; Palmer and Krumhansl 1990), in order to compare a condition with accents with a condition without accents the attention demand should be carefully balanced.

Effect of audiovisual incongruency on sensitivity to asynchrony

In Experiment 2 we additionally found, in agreement with van Wassenhove et al. (2007) and Vatakis and Spence (2007), a consistent effect of incongruency on TIW width, with the tolerance to asynchrony (TIW width) decreasing when the audiovisual incongruency was introduced. This result is similar to what van Wassenhove et al. (2007) found for incongruent and congruent audiovisual speech stimuli (syllables). The congruent condition was found by the authors to elicit a broader tolerance to asynchrony than the incongruent in complete accord with our results. However, on average the size of the integration window for audiovisual drumming events (∼112 ms) was found to be much smaller than that found by van Wassenhove et al. (2007; ∼200 ms) for audiovisual speech events. This difference in the extent of TIW for drumming and speech stimuli might be a result of differences in duration between drumming impacts and syllables. That is, because it is known that a TIW for audiovisual speech of ∼200 effectively corresponds to the average of syllable duration across languages (Arai and Greenberg 1997), the narrowing of TIW for audiovisual drumming could depend, at least in part, on the general short duration of different kinds of impact (usually in the order of few milliseconds: Wagner 2006, p 30). Hence, the smaller width of the TIW for drumming stimuli might ensue from a compromise between the duration of contact time and the tolerance to asynchrony necessary to successfully integrate sound and sight in a unique percept

despite differences in arrival and time processing of these two signals. Yet, it is more important to underline the evidence that incongruency between auditory and visual streams appears to enhance people's sensibility to asynchrony for speech syllables as well as drumming stimuli. That is, for both drumming and speech the visual action of motion and the resulting sound are integrated to a greater degree when the two kinds of information covary together.

Effect of expertise on sensitivity to asynchrony

In both Experiment 1 and 2, we showed that novices and drummers highly differ in their performance. Our results indicate a clear effect of expertise on the sensitivity to asynchrony, showing that professional drummers were less tolerant to asynchrony as compared to novices, in that TIW width and PSS values were found to be much smaller for the former group than for the latter. These findings are in line with those of previous research that have shown differences between musicians and non-musicians in temporal integration ability (Miner and Caudell 1998; Hodges et al. 2005), in tapping synchronization to gestures (Luck and Sloboda 2007; Luck and Nte 2008), and in brain areas involved in audiovisual integration (Hodges et al. 2005).

Most importantly in Experiment 1 we found that the effect of tempo on sensitivity to asynchrony (Arrighi et al. 2006; Petrini et al. 2009) was more evident for the novices group (Petrini et al. 2009). This is an interesting difference between drummers' and novices' sensitivity that focuses upon the slower tempo. Novices were clearly much more tolerant than drummers to audiovisual asynchrony for the swing groove played at 60 BPM than for those played at the faster tempos. This difference between experts and novices for very slow tempo might be accounted for by a higher ability of musicians to tap at slower rates than non-musicians, as demonstrated by results of forced tapping tasks. That is, the musicians can produce a wider range of tempos, especially slower tempos, than non-musicians (Drake et al. 2000). If musicians are able to produce a wider range of slow tempos, while novices cannot, then this might explain why only novices are very poor in detecting audiovisual asynchrony at low frequency. In other words, drummers might learn through practice to process the two signals of the audiovisual event without any interference from additional physical characteristics of the music stimulus, and thus give unbiased judgments on whether those signals temporally belong to a unique event.

In line with this view are the results of Experiment 2 showing that only novices' sensitivity to asynchrony appears significantly enhanced by audiovisual incongruency between the auditory and visual stream. Indeed, from the results of Experiment 2 it emerged that incongruency narrows the TIW width only for the novices group, while there was no such effect on drummers' TIW. This result might indicate that drummers, due to practice, are not further facilitated in their detection of audiovisual asynchrony by a mismatch between the two signals. That is, drummers might be able to treat as separate the two signals in both cases of audiovisual congruency or incongruency, and consequently be better in evaluating whether or not the two are in synchrony. This hypothesis needs to be tested by future research by investigating whether the effect of audiovisual incongruency on sensitivity to asynchrony decreases after a long musical training period.

Based on our findings we can speculate that, after a long period of musical training, some characteristics of the audiovisual stimulation (such as tempo and congruency), that would normally be used by our neural system to recalibrate our tolerance to signals asynchrony (Vroomen et al. 2004; Navarra et al. 2005), are ignored. In other words, with musical practice the binding process changes in such a way that additional factors are no longer used by our neural system to integrate the multisensory information, because the system reaches a very high and unbiased level of precision itself. If this is true, it means that with practice the cross-modal binding of stimuli might not be modulated anymore by variations in the low-level stimulus properties, even if could still be modulated by top-down factors.

One top-down factor which has been proposed to be important for multisensory processing is the "unity assumption" (Vatakis and Spence 2007), which holds that the binding of different sources of sensory information will be influenced by whether or not they are held to correspond to the same category. For example, Vatakis and Spence (2007) showed that whether the gender of a face and voice matched or mismatched had an effect on audiovisual binding even when the temporal alignment and correlation between streams was high. While such a finding is important, it is not directly relevant to our results in Experiment 2 since we manipulated the congruency of audiovisual information at a low-level, and always within the category of drumming. Cross-category presentation of music stimuli is possible and has been used by Schutz and Kubovy (2008) with presentation of visual stimulation from one music category with auditory stimulation from a different music category. Therefore, it would be interesting to examine whether the difference between drummers and novices would stand also when presenting, for example, a drumming display together with a guitar sound and vice versa. However, the lack of an effect of the "unity assumption" on temporal order judgments shown by Vatakis and Spence (2008) for musical stimuli suggests that this kind of top-down factor might be specific to speech (Vatakis et al. 2008).

Finally, in view of all the behavioural differences that have been found between musical experts and novices in this and previous studies (Miner and Caudell 1998; Hodges

et al. 2005; Petrini et al. 2009) it would be very important to run further brain imaging studies (Bengtsson et al. 2005; Gaser and Schlaug 2003; Hodges et al. 2005; Bermudez and Zatorre 2005) to better understand which of these differences are reflected in structural changes rather than functional rearrangement of the brain areas involved in multisensory integration processes.

Acknowledgments This work was supported by a grant from the British Academy (LRG-42455). We would like to thank Melanie Russell for her contribution during the pilot phase of this study and Gerry Rossi of Strathclyde University for his help in recruiting the jazz drummers. We also would like to thank Jim Kay and Michael Kubovy for their valuable suggestions on data analysis and statistical methods.

References

Arai T, Greenberg S (1997). The temporal properties of spoken Japanese are similar to those of English. In: Proceedings of Eurospeech, Rhodes, Greece, pp 1011–1014

Arrighi R, Alais D, Burr D (2006) Perceptual synchrony of audiovisual streams for natural and artificial motion sequences. J Vision 6:260–268

Bengtsson SL, Nagy Z, Skare S, Forsman L, Forssberg H, Ullén F (2005) Extensive piano practicing has regionally specific effects on white matter development. Nature Neurosci 8:1148–1150

Bermudez P, Zatorre RJ (2005) Differences in gray matter between musicians and nonmusicians. Ann N Y Acad Sci 1060:395–399

Brainard DH (1997) The psychophysics toolbox. Spatial Vision 10:433–436

Brooks A, van der Zwan R, Billard A, Petreska B, Clarke S, Blanke O (2007) Auditory motion affects visual biological motion processing. Neuropsychologia 45:523–530

Clifford CWG, Arnold DH, Pearson J (2003) A paradox of temporal perception revealed by a stimulus oscillating in colour and orientation. Vision Res 43:2245–2253

Collyer CE, Broadbent HA, Church RM (1994) Preferred rates of repetitive tapping and categorical time production. Percept Psychophys 55:443–453

Dahl S (2004) Playing the accent—comparing striking velocity and timing in an ostinato rhythm performed by four drummers. Acta Acustica United Acustica 90:762–776

Dixon NF, Spitz L (1980) The detection of auditory visual desynchrony. Perception 9:719–721

Drake C, Botte MC (1993) Tempo sensitivity in auditory sequences: evidence for a multiple-look model. Percept Psychophys 54:277–286

Drake C, Jones MR, Baruch C (2000) The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. Cognition 77:251–288

Fain GL (2003) Sensory transduction. Sinauer Associates, Sunderland, MA

Fontana F, Rocchesso D, Avanzini F (2004) Computation of nonlinear filter networks containing delay-free paths. In Proceedings of the 7th international conference on digital audio effects (DAFX-04), Naples, Italy, pp 113–118

Friberg A, Sundberg J (1995) Time discrimination in a monotonic, isochronous sequence. J Acoust Soc Am 98:2524–2531

Fujisaki W, Nishida S (2005) Temporal frequency characteristics of synchrony–asynchrony discrimination of audio–visual signals. Exp Brain Res 166:455–464

Gaser C, Schlaug G (2003) Brain structures differ between musicians and non-musicians. J Neurosci 23:9240–9245

Grant KW, Greenberg S (2001) Speech intelligibility derived from asynchronous processing of auditory–visual speech information. In: Proceedings of AVSP 2001 international conference of auditory–visual speech processing, Scheelsminde, Denmark, pp 132–137

Grant KW, van Wassenhove V, Poeppel D (2004) Detection of auditory (cross-spectral) and auditory–visual (cross-modal) synchrony. J Acoust Soc Am 108:1197–1208

Hodges DA, Hairston WD, Burdette JH (2005) Aspects of multisensory perception: the integration of visual and auditory information in musical experiences. Ann N Y Acad Sci 1060:175–185

Hollier MP, Rimell AN (1998) An experimental investigation into multi-modal synchronisation sensitivity for perceptual model development. 105th AES Convention Preprint No. 4790

Johansson G (1973) Visual perception of biological motion and model for its analysis. Percept Psychophys 14:201–211

Jones MR (1987) Dynamic pattern structure in music: recent theory and research. Percept Psychophys 41:621–634

Kay BA, Kelso JAS, Saltzman EL, Schöner G (1987) Space–time behavior of single and bimanual rhythmical movements: data and limit cycle model. J Exp Psychol: Human Percept Perform 13:178–192

King AJ (2005) Multisensory integration: strategies for synchronization. Curr Biol 15:R339–R341

King AJ, Palmer AR (1985) Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus. Exp Brain Res 60:492–500

Luck G, Nte S (2008) An investigation of conductors' temporal gestures and conductor— musician synchronization, and a first experiment. Psychol Music 36:81–99

Luck G, Sloboda JA (2007) An investigation of musicians' synchronization with traditional conducting beat patterns. Music Perform Res 1:26–46

MacDonald J, McGurk H (1978) Visual influences on speech perception processes. Percept Psychophys 24:253–257

MacDougall HG, Moore ST (2005) Marching to the beat of the same drummer: the spontaneous tempo of human locomotion. J Appl Physiol 99:1164–1173

Massaro DW, Cohen MM, Smeele PM (1996) Perception of asynchronous and conflicting visual and auditory speech. J Acoust Soc Am 100:1777–1786

McGurk H, MacDonald J (1976) Hearing lips and seeing voices. Nature 264:746–748

Miner N, Caudell T (1998) Computational requirements and synchronization issues for virtual acoustic displays. Presence-Teleoperat Virtual Environ 7(4):396–409

Moelants D (2002) Preferred tempo reconsidered. In: Proceedings of the 7th international conference on music perception and cognition, Sydney, 2002, Stevens C, Burnham D, McPherson G, Schubert E, Renwick J (eds). Causal productions, Adelaide, pp 580–583

Munhall KG, Gribble P, Sacco L, Ward M (1996) Temporal constraints on the McGurk effect. Percept Psychophys 58:351–362

Murray M, Drought AB, Kory RC (1964) Walking patterns of normal men. J Bone Joint Surgery 46:335–360

Navarra J, Vatakis A, Zampini M, Soto-Faraco S, Humphreys W, Spence C (2005) Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration. Cognitive Brain Res 25(2):499–507

Palmer C, Krumhansl CL (1990) Mental representations for musical meter. J Exp Psychol Hum Percept Perform 16:728–741

Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spatial Vision 10:437–442

Petrini K, Russell M, Pollick F (2009) When knowing can replace seeing in audiovisual integration of actions. Cognition 110:432–439

Saygin AP, Driver J, de Sa VR (2008) In the footsteps of biological motion and multisensory perception: judgements of audio-visual temporal relations are enhanced for upright walkers. Psychol Sci 19:469–475

Schutz M, Kubovy M (2008) The effect of tone envelope on sensory integration: support for the 'unity assumption'. J Acoust Soc Am 123:3412

Schutz M, Lipscomb S (2007) Hearing gestures, seeing music: vision influences perceived tone duration. Perception 36:888–897

Spence C, Squire S (2003) Multisensory integration: maintaining the perception of synchrony. Curr Biol 13:519–521

Spence C, Walton M (2005) On the inability to ignore touch when responding to vision in the crossmodal congruency task. Acta Psychol 118:47–70

Spence C, Shore DI, Klein RM (2001) Multisensory prior entry. J Exp Psychol: Gen 130:799–832

Steinmetz R (1996) Human perception of jitter and media synchronization. IEEE 14:61–72

Stone JV, Hunkin NM, Porrill J, Wood R, Keeler V, Beanland M, Port M, Porter NR (2001) When is now? Perception of simultaneity. Proc R Soc, B 268:31–38

Sugita Y, Suzuki Y (2003) Audiovisual perception: implicit estimation of sound-arrival time. Nature 421:911

van Wassenhove V, Grant KW, Poeppel D (2007) Temporal window of integration in auditory-visual speech perception. Neuropsychologia 45:598–607

Vatakis A, Spence C (2006a) Audiovisual synchrony perception for speech and music using a temporal order judgment task. Neurosci Lett 393:40–44

Vatakis A, Spence C (2006b) Audiovisual synchrony perception for music, speech, and object actions. Brain Res 1111:134–142

Vatakis A, Spence C (2007) Crossmodal binding: evaluating the "unity assumption" using audiovisual speech stimuli. Percept Psychophys 69:744–756

Vatakis A, Spence C (2008) Evaluating the influence of the 'unity assumption' on the temporal perception of realistic audiovisual stimuli. Acta Psychol 127:12–23

Vatakis A, Navarra J, Soto-Faraco S, Spence C (2007a) Audiovisual temporal adaptation of speech: temporal order versus simultaneity judgments. Exp Brain Res 185:521–529

Vatakis A, Navarra J, Soto-Faraco S, Spence C (2007b) Temporal recalibration during asynchronous audiovisual speech perception. Exp Brain Res 181:173–181

Vatakis A, Ghazanfar AA, Spence C (2008) Facilitation of multisensory integration by the "unity effect" reveals that speech is special. J Vision 8:1–11

Vroomen J, Keetels M, de Gelder B, Bertelson P (2004) Recalibration of temporal order perception by exposure to audio-visual asynchrony. Cognitive Brain Res 22:32–35

Waadeland CH (2003) Analysis of jazz drummers' movements in performance of swing grooves-a preliminary report. In: Bresin R (ed) Proceedings of Stockholm music acoustic conference, volume II, Stockholm, pp 573–576

Waadeland CH (2006) Strategies in empirical studies of swing groove. Studia Musicol Norvegica 32:169–191

Wagner A (2006) Analysis of drumbeats—interaction between Drummer, Drumstick and Instrument. Master's Thesis at the Department of Speech, Music and Hearing (TMH)

Watson AB, Hu J (1999) ShowTime: a QuickTime-based infrastructure for vision research displays. Perception 28 ECVP Abstract Supplement, 45b

Zampini M, Shore DI, Spence C (2003) Audiovisual temporal order judgments. Exp Brain Res 152:198–210

Zhou F, Wong V, Sekuler R (2007) Multi-sensory integration of spatio-temporal segmentation cues: one plus one does not always equal two. Exp Brain Res 180:641–654