Expertise with multisensory events eliminates the effect of biological motion rotation on audiovisual synchrony perception

Karin Petrini  
Department of Psychology, University of Glasgow, Glasgow, UK

Samuel Paul Holt  
Department of Psychology, University of Glasgow, Glasgow, UK

Frank Pollick  
Department of Psychology, University of Glasgow, Glasgow, UK

Biological motion, in the form of point-light displays, is usually less recognizable and coherent when shown from a less natural orientation, and evidence of this disruption was recently extended to audiovisual aspects of biological motion perception. In the present study, eight drummers and eight musical novices were required to judge either the audiovisual simultaneity or the temporal order of the movements of a drumming point-light display and the resulting sound. The drumming biological motion was presented either in its upright orientation or rotated by 90, 180, or 270 degrees. Our results support and extend previous findings demonstrating that although the rotation of the point-light display affects the audiovisual aspects of biological motion, this effect disappears when experience with the represented multisensory action is increased through practice.

Keywords: biological motion, audiovisual synchrony, drumming, action expertise


Introduction

Only rarely will our perception rely fully upon a single sense. When engaging in day-to-day activities, it is usually a combination of the information coming from two or more sensory modalities that gives us a complete picture of what is occurring in the world. The temporal co-occurrence of arriving signals is often fundamental to this process of integration and to perceiving the different information as pertaining to the same event (e.g., the Ventriloquism effect). For example, in a recent study, Saygin, Driver, and De Sa (2008) demonstrated that only when the auditory and visual footsteps of a point-light walker (Johansson, 1973) were phase-locked did the inversion of the visual information affect the participants’ ability to detect a temporal frequency mismatch between the signals. They concluded that the visual disruption of multisensory judgments occurred only when the sound was perceived as a consequence of the portrayed action.

The importance of the temporal relation between signals in multisensory integration has stimulated a wide range of studies on synchrony perception. In general, they aim to determine the temporal window, or range of tolerance, of audiovisual asynchrony under which the human brain still integrates the signals (Dixon & Spitz, 1980). Several studies have examined the temporal integration window (TIW) and the point of subject simultaneity (PSS) using rather simple stimuli like clicks and flashes (Fujisaki & Nishida, 2005; Spence, Shore, & Klein, 2001; Stone et al., 2001; Sugita & Suzuki, 2003; Zampini, Shore, & Spence, 2003a, 2003b), while others have examined more complex speech and non-speech stimuli (Hollier & Rimell, 1998; Miner & Caudell, 1998; Petrini, Dahl et al., 2009; Petrini, Russell et al., 2009; Vatakis & Spence, 2006a, 2006b). Furthermore, the effects on sensitivity to asynchrony determined by variation in the physical characteristics of the stimuli (Arrighi, Alais, & Burr, 2006; Petrini, Dahl et al., 2009; Petrini, Russell et al., 2009; van Wassenhove, Grant, & Poeppel, 2007) as well as by top-down factors (Vatakis & Spence, 2007, 2008a) have been well documented.

Prior knowledge acquired through years of experience with a certain audiovisual event appears to be an effective factor in enhancing the ability to detect asynchrony. For example, musical experience has been found on more than one occasion to enhance sensitivity to asynchrony (Hodges, Hairston, & Burdette, 2005; Petrini, Dahl et al., 2009; Petrini, Russell et al., 2009). Petrini et al. demonstrated that the simultaneity judgments of drummers were resilient to disruption of the auditory information (induced by scrambling the impact velocity of the strikes), disruption of the visual information (induced by eliminating the
impact point information from the drumming point-light display), and variation in temporal frequency of the audiovisual information (Pettrini, Dahl et al., 2009; Pettrini, Russell et al., 2009). Vatakis and Spence (2008b) also found that expert pianists’ sensitivity to asynchrony was not affected by the inversion of a piano playing video, while they did find an effect of speech video inversion on the sensitivity to asynchrony of speech experts (almost everyone can be considered an expert in speech perception and production). Similarly, Saygin et al. (2008) found that inversion of point-light walking displays affected the sensitivity to audiovisual mismatch for walking experts (as for speech, almost everyone can be considered an expert in walking perception and production). This posits an interesting question of whether musical expertise is somehow different from other kinds of acquired expertise, and to this end it would be informative to compare musicians’ and non-musicians’ sensitivity to asynchrony as a function of musical display rotation. The use of a side view of point-light drumming displays would combine aspects of the point-light walker displays used by Saygin et al. (2008) with those of musical displays used by Vatakis and Spence (2008b) and would also allow comparisons with previous studies where drummers’ and nonVICES sensitivity to asynchrony was compared using similar displays (Pettrini, Dahl et al., 2009; Pettrini, Russell et al., 2009).

Biological motion, in the form of point-light displays, is usually less recognizable and coherent when shown from a less natural orientation than upright (Pavlova & Sokolov, 2000; Shipley, 2003; Sumi, 1984; Troje & Westhoff, 2006), and this disruption appears to interfere with the audiovisual aspects of biological motion perception (Saygin et al., 2008). This phenomenon, however, is not limited to biological motion. The inversion of face stimuli also affects the audiovisual aspects of speech perception (Green, 1994; Jordan & Bevan, 1997; Vatakis & Spence, 2008b). Interestingly, the inversion of the visual information in other kinds of stimuli such as audiovisual music has been found not to affect the perceived synchrony between the two sensory signals (Vatakis & Spence, 2008b). Vatakis and Spence (2008b) accounted for this finding by suggesting that the loss of configural information caused by inversion is more detrimental to the processing of face stimuli than to non-face stimuli. Indeed, the authors underlined how face/speech stimuli are more unnatural when presented in an inverted orientation than other kind of non-face stimuli, and how this could lead to different brain processes that would cause an apparent delay in the processing of the visual signal only for the face stimuli. As an alternative explanation for their results, Vatakis and Spence (2008b) also indicated the greater social relevance of the face stimuli when compared with other non-face stimuli. Comparing musical experts and musical novices by using musical point-light displays would further clarify whether the effect of biological motion inversion on multisensory judgments (Saygin et al., 2008) extends to other kind of stimuli like musical displays, or whether it is limited to more day-to-day activities like walking or face-to-face speech interaction, in accord with Vatakis and Spence’s (2008b) findings with full videos.

Another factor important for interpreting the results of Vatakis and Spence (2008b) is that they used the temporal order judgment (TOJ) task to derive the best perceived alignment between the auditory and visual information for upright and inverted displays. However, it is becoming more and more evident that if one wants an appropriate estimate of PSS, an SJ task should be preferred to a TOJ (Schneider & Bavelier, 2003; Van Eijk, Kohlrausch, Juola, & Van De Par, 2008; Vatakis, Navarra, Soto-Baraclo, & Spence, 2007a; Zampini, Guest, Shore, & Spence, 2005). A problem with the PSS obtained by using the TOJ task is that participants do not directly judge simultaneity, and thus the found effects, or lack of such, of visual inversion might be caused by a different process (Mittrani, Shekerdiiiski, & Yakimoff, 1986; Van Eijk et al., 2008). Additionally, Vatakis and Spence (2008b) indicated that the use of the TOJ task might have been the reason why participants in their study did not differ in their sensitivity to signals’ temporal order for upright and inverted stimuli. That is, the orientation of the visual information should not really matter when judging whether the visual stream or the auditory was presented first. On the other hand, the SJ task may be more sensitive to this kind of manipulation, thus uncovering differences in sensitivity to asynchrony between audiovisual upright and inverted displays.

In the experiment reported here, we tried to overcome these task-related limitations by using both SJ and TOJ tasks to investigate how experience with a certain multisensory event affects audiovisual synchrony perception of upright and rotated biological motions. To this end, a point-light display representing a professional drummer playing a simple swing groove (Pettrini, Dahl et al., 2009; Pettrini, Russell et al., 2009; Waadeland, 2003, 2006) was shown, together with the reproduced sound, to a group of expert drummers and musical novices. Percussionists’ movements have previously been successfully used to study audiovisual integration (Armontrout, Schultz, & Kubovy, 2009; Arrighi et al., 2006; Pettrini, Dahl et al., 2009; Pettrini, Russell et al., 2009; Schutz & Kubovy, 2009; Schutz & Lipscomb, 2007). The percussionist movement is very evident to everyone (as are most impact actions) in that it involves the whole arm, not only part of it or just the fingers as is the case with most other musical instruments. Nevertheless, the specific rhythmic pattern used (swing groove in our case) makes a perfect simple stimulus to differentiate between novices and experts.

### Methods

#### Participants

A total of 16 participants were recruited, all males between the ages of 21 and 36. All had normal hearing
One point was located at the shoulder, elbow, wrist, hand joints, and drumstick coordinates of the point-light displays represented the drumhead configuration represented the movements of a professional jazz drummer playing a swing groove as viewed from the right-hand side (Figure 1a). The points were represented in the display by white disks (luminance: 85 cd/m²; diameter: 2 mm) on a black background (luminance: 0.12 cd/m²), while a thick white line, oriented 25 degrees anticlockwise from horizontal, represented the drumhead (width: 2.2 cm; height: 2 mm; luminance: 85 cd/m²). Stimuli were presented on a Sony Trinitron screen placed 1 m from where the participant was seated. Screen resolution was 1024 × 768 pixels with a screen refresh rate of 60 Hz. Audio stimuli were digitized at a rate of 44,100 Hz and presented through high-quality headphones (Beyer Dynamic DT Headphones). The experiment was run on a Macintosh OS9 platform and the displays were presented to participants by using Showtime (Watson & Hu, 1999), a component of Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) extensions to Matlab.

The selected tempo of the swing groove was 120 beats per minute (BPM) while the accent was positioned on the second beat. This condition was found to produce the smallest difference between drummers and novices and was therefore appropriate for this experiment (Petrini, Dahl et al., 2009). The swing groove video and audio footage were combined using Adobe Premier 1.5 to create the different asynchronous conditions by delaying either the video with respect to the audio or the audio with respect to the video. Whereas the negative lags (−66.67 ms, −133.33 ms, −200 ms, and −266.67 ms) indicated the conditions for which the auditory information preceded the visual, the positive lags (+66.67 ms, +133.33 ms, +200 ms, and +266.67 ms) indicated the conditions for which the visual information preceded the auditory. Thus, a total of 9 audiovisual lags were created, for which the timings were selected as being similar to those used in previous multisensory drumming studies (Arrighi et al., 2006; Petrini, Dahl et al., 2009; Petrini, Russell et al., 2009). Each clip was edited to be 3 s long and contained a total of 9 drum strikes. The movies were created from a much longer clip, thus ensuring that there was always video and audio footage at the beginning and end of each clip. The temporal profiles of all nine stimuli are presented in Figure 2. The auditory information was kept constant while the visual information was changed accordingly by shifting this information along the timeline. Figure 2 shows that this manipulation allowed us to keep constant to three the number of strikes per second for both the auditory and visual information. However, while the auditory signal was held constant, the visual signal was advanced or delayed, and this resulted in the density of audio events remaining constant while the density of the visual signal changed for the different lags. This density is reflected in the average interval between events and varied with lag for the visual displays (i.e., visual average event intervals of 329, 329, 350, 350, 350, 310, 310, 310, and 310 ms for lags of −267, −200, −133, −67, 0, 67, 133, 200, and 267 ms, respectively). Because this variation might affect the identification of single strikes and inter-strike intervals (Fujisaki & Nishida, 2007), we examined whether the performance on the different experimental tasks could be explained by differences in density of the visual signal. Finally, the nine produced audiovisual drumming displays (Figure 1a, Movie 1) were rotated by 90° (Figure 1b, Movie 2), 180° (Figure 1c, Movie 3), and 270° (Figure 1d, Movie 4) degrees to obtain corresponding drumming displays at

Figure 1. The four orientation conditions as viewed by participants. Each one of the nine (a) original time lag conditions was rotated by (b) 90°, (c) 180°, and (d) 270° to create a further 27 displays. The outlines of the drum and drummer are for illustrative purposes only and were not present in the actual displays.

and normal or corrected-to-normal vision and received a monetary incentive for their participation. Eight participants were expert drummers with drumming experience varying from 8 to 19 years and a mean age of 27 (±4.79). The remaining eight participants were musical novices with no musical experience and a mean age of 26.5 (±5.73).

Apparatus and stimuli

The apparatus and basic stimuli used in the present experiment were the same as those used in Petrini, Dahl et al. (2009) and Petrini, Russell et al. (2009). The data coordinates of the point-light displays represented the shoulder, elbow, wrist, hand joints, and drumstick (Waadeland, 2003, 2006). One point was located at the hand grip level and another at the drumstick head level. The configuration represented the movements of a professional jazz drummer playing a swing groove as viewed from the right-hand side (Figure 1a). The points were represented in the display by white disks (luminance: 85 cd/m²; diameter: 2 mm) on a black background (luminance: 0.12 cd/m²), while a thick white line, oriented 25 degrees anticlockwise from horizontal, represented the drumhead (width: 2.2 cm; height: 2 mm; luminance: 85 cd/m²). Stimuli were presented on a Sony Trinitron screen placed 1 m from where the participant was seated. Screen resolution was 1024 × 768 pixels with a screen refresh rate of 60 Hz. Audio stimuli were digitized at a rate of 44,100 Hz and presented through high-quality headphones (Beyer Dynamic DT Headphones). The experiment was run on a Macintosh OS9 platform and the displays were presented to participants by using Showtime (Watson & Hu, 1999), a component of Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) extensions to Matlab.

The selected tempo of the swing groove was 120 beats per minute (BPM) while the accent was positioned on the second beat. This condition was found to produce the smallest difference between drummers and novices and was therefore appropriate for this experiment (Petrini, Dahl et al., 2009). The swing groove video and audio footage were combined using Adobe Premier 1.5 to create the different asynchronous conditions by delaying either the video with respect to the audio or the audio with respect to the video. Whereas the negative lags (−66.67 ms, −133.33 ms, −200 ms, and −266.67 ms) indicated the conditions for which the auditory information preceded the visual, the positive lags (+66.67 ms, +133.33 ms, +200 ms, and +266.67 ms) indicated the conditions for which the visual information preceded the auditory. Thus, a total of 9 audiovisual lags were created, for which the timings were selected as being similar to those used in previous multisensory drumming studies (Arrighi et al., 2006; Petrini, Dahl et al., 2009; Petrini, Russell et al., 2009). Each clip was edited to be 3 s long and contained a total of 9 drum strikes. The movies were created from a much longer clip, thus ensuring that there was always video and audio footage at the beginning and end of each clip. The temporal profiles of all nine stimuli are presented in Figure 2. The auditory information was kept constant while the visual information was changed accordingly by shifting this information along the timeline. Figure 2 shows that this manipulation allowed us to keep constant to three the number of strikes per second for both the auditory and visual information. However, while the auditory signal was held constant, the visual signal was advanced or delayed, and this resulted in the density of audio events remaining constant while the density of the visual signal changed for the different lags. This density is reflected in the average interval between events and varied with lag for the visual displays (i.e., visual average event intervals of 329, 329, 350, 350, 350, 310, 310, 310, and 310 ms for lags of −267, −200, −133, −67, 0, 67, 133, 200, and 267 ms, respectively). Because this variation might affect the identification of single strikes and inter-strike intervals (Fujisaki & Nishida, 2007), we examined whether the performance on the different experimental tasks could be explained by differences in density of the visual signal. Finally, the nine produced audiovisual drumming displays (Figure 1a, Movie 1) were rotated by 90° (Figure 1b, Movie 2), 180° (Figure 1c, Movie 3), and 270° (Figure 1d, Movie 4) degrees to obtain corresponding drumming displays at
three less natural orientations. This gave us a total of 36 clips (4 orientations × 9 audiovisual delays).

Procedure

Participants were seated in a darkened room approximately 1 m from the screen. They received a written explanation of what the drumming point-light display looked like and where it would appear on the screen, along with what the sound stimuli would be. Participants were asked to make temporal order judgments (TOJs) or synchrony judgments (SJs) in two separate sessions, and the task order was counterbalanced across participants. Hence, half the participants (i.e., four drummers and four novices) completed the SJ task first, while the remaining half completed the TOJ task first. The two task sessions were carried out on different days.

Before starting the experiment, three practice trials were presented in order to familiarize participants with the tasks and nature of the stimuli. At the end of each trial (and throughout the experiment), a reminder of the response keys was given. Responses were made by pressing one of two keys on the keyboard. For the TOJ task, participants were asked which information came first, the visual or the auditory. If participants were unsure, they were asked to guess. For the SJ task, subjects were asked to judge whether the visual and the auditory information were synchronous or asynchronous. Each trial consisted of a randomly presented clip with a certain orientation and audiovisual lag. Each one of these conditions was presented once in a block, meaning that one block consisted of a randomized presentation of the 36 displays. Between each block, participants were encouraged to have a brief break. In total, participants completed 20 blocks, meaning that each participant completed 720 trials for each of the TOJ and SJ tasks.

Results

The best-fitting graphs to the synchrony and visual first judgments, respectively, are shown in Figure 3 for the four different levels of the display rotation (0, 90, 180, and 270 degrees). The standard deviations of the Gaussian curves (all with $R^2$ values > 0.75) were taken as an estimate of participants' Temporal Integration Window (TIW) for the SJ (Simultaneity Judgments) task, while the standard deviation of each Cumulative Gaussian curve was taken as an estimate of TIW for the TOJ (Temporal Order Judgments) task. The point of subjective simultaneity (PSS) for the SJ task was derived as an estimate of the Gaussian's peak, while the PSS for the TOJ task was derived as an estimate of the Cumulative Gaussian's intercept at the 50% point (corresponding to 10 on the $Y$-axis). The estimates for the TOJ task were calculated from the data of four out of eight novices and five out of eight drummers, because the data of the remaining participants produced PSS values that were beyond the range of tested audiovisual delays, and exploration of the data of these participants revealed that they had not been able to discriminate between the “auditory first” and “visual first” conditions (Figure 4). This evidence, together with the larger variability found for the TOJ estimates (bottom diagrams in Figures 6 and 7), suggests that this task was much more challenging than the SJ task for both groups.

Figure 3 shows that drummers have narrower TIW and smaller PSS than novices, in agreement with previous findings (Petrini, Dahl et al., 2009; Petrini, Russell et al., 2009), and most importantly that the PSSs estimated by using the two different tasks (SJ and TOJ) are much closer for the drummers than for the novices group. The results for the PSS and TIW estimates are presented in Figures 6 and 7. In neither the SJ nor the TOJ task was there an evident change in PSS for either the novices or drummers. However, novices showed an increase in TIW when going...
from 0° rotation to 90° rotation for the SJ task but not for the TOJ task.

Before analyzing the PSS and TIW width of drummers and novices, we examined whether the variation in visual density affected the number of correct judgments of “synchrony” and “asynchrony”. This analysis was possible only for the SJ task, as for the TOJ task, there is no real correct answer for the audiovisual synchrony condition.
Figure 5 shows the results for the SJ task and indicates that this variation did not result in a decrease in correct judgments (novices $R^2 = 0.07$ and drummers $R^2 = 0.02$). An analysis of variance with a 2 (expertise) $\times$ 4 (display orientation) mixed factorial design was conducted on the PSS and TIW width data separately for the SJ and TOJ tasks. In addition, a Tukey’s test for multiple comparisons was carried out, when appropriate, to examine in detail the effect of display rotation on the estimate values.

**Simultaneity judgments**

The between-subjects factor “expertise” was found to have a significant effect on both PSS ($F(1, 14) = 4.952, p = 0.043$) and TIW width ($F(1, 14) = 20.193, p < 0.001$). In addition, the within factor “display orientation” was found to significantly affect the TIW ($F(3, 12) = 5.407, p = 0.003$), while no effect was found for the PSS ($F(3, 12) = 0.927, p = 0.436$). A marginally significant interaction was also found between display orientation and expertise for the TIW estimates ($F(3, 12) = 2.823, p = 0.050$) but not for the PSS ($F(3, 12) = 0.872, p = 0.463$). The Tukey’s test carried out on the novices’ TIW estimates showed a significant difference in sensitivity to asynchrony when rotating the display from 0° to 90° (Tukey’s test: $q = 5.76, p < 0.01$) but not when further rotating from 90° to 180° (Tukey’s test: $q = 1.66, p > 0.05$) or from 180° to 270° (Tukey’s test: $q = 1.78, p > 0.05$). The same test carried out on the drummers’ TIW estimates showed no significant difference in sensitivity to asynchrony when rotating the display from 0° to 90° (Tukey’s test: $q = 0.97, p > 0.05$).

Figure 4. Example of a drummer and a novice who could not discriminate between “auditory first” and “visual first” but could discriminate between “synchrony” and “asynchrony”. The drummer’s data are presented on the left-hand side, while those of the novice are on the right-hand side. Different colored lines and symbols indicate different display orientations. For the SJ judgments, the Gaussian curves are also reported.
Thus, novices became more tolerant to asynchrony when the display was initially rotated from the 0° orientation (M = 158 ms) to the 90° orientation (M = 174 ms), while the level of tolerance to asynchrony for the drummers group did not vary with the display rotation (0°: M = 54; 90°: M = 58).

To examine for possible effects of adaptation to the presented audiovisual time lags and consequent recalibration of the point of perceived simultaneity (Fujisaki, Shimojo, Kashino, & Nishida, 2004; Vatakis, Navarra, Soto-Faraco, & Spence, 2007b; Vroomen, Keetels, de Gelder, & Bertelson, 2004), we compared the PSS obtained by fitting the synchrony data for the first ten trials with those obtained for the second ten trials. Hence, an analysis of variance with a 2 (expertise) × 2 (trials) × 4 (display orientation) mixed factorial design was conducted on the PSS. No significant effect of trials (F(1, 14) = 0.790, p = 0.389), of interaction between trials and expertise (F(1, 14) = 0.674, p = 0.425), or of interaction between trials and orientation (F(3, 12) = 1.566, p = 0.249) was found. Therefore, no evident effect of adaptation to the audiovisual lag timings used was found.

Temporal order judgments

The between-subjects factor “expertise” was found to have a significant effect on PSS (F(1, 7) = 39.076, p < 0.001) but not on TIW width (F(1, 7) = 0.013, p = 0.912). The within factor “display orientation” was not significant for either PSS (F(3, 5) = 0.162, p = 0.920) or TIW (F(3, 5) = 0.825, p = 0.494). In addition, no significant interaction was found between display orientation and expertise for the PSS estimates (F(3, 5) = 0.378, p = 0.770) or for the TIW (F(3, 5) = 0.547, p = 0.655). In other words, for the TOJ task no effect of visual rotation was found on either the TIW or PSS for both groups (bottom diagrams in Figures 6 and 7).

As in the SJ task, we also examined the data for possible effects of adaptation by comparing the PSS data for the first ten trials with those obtained for the second ten trials. Hence, an analysis of variance with a 2 (expertise) × 2 (trials) × 4 (display orientation) mixed factorial design was conducted on the PSS. Once again no significant effect of trials (F(1, 7) = 1.330, p = 0.287), of interaction between trials and expertise (F(1, 7) = 0.118, p = 0.741), or of interaction between trials and orientation (F(3, 5) = 1.042, p = 0.450) was found. Thus similarly to the SJ task, no effect of adaptation to the presented audiovisual lag timings was found.

Discussion

In the present study, a group of drummers and a group of musical novices were required to judge either the simultaneity or the temporal order between the movements of a
drummer (in the form of a point-light display) and the resulting sound. Our results support and extend previous findings demonstrating that though the rotation of the point-light display affects the audiovisual aspects of biological motion, this effect disappears with high levels of experience in the represented multisensory events.

The results of both the SJ and TOJ tasks showed much higher sensitivity to asynchrony in the group of drummers, which is in accord with previous findings (Petrini, Dahl et al., 2009; Petrini, Russell et al., 2009). Indeed, the PSSs obtained from both tasks were significantly closer to the zero physical point of synchrony for the drummers than for the novices. In addition, the TIW width was narrower for the drummers in the SJ task, though this was found not to be the case for the TOJ task. The comparison between TOJ and SJ data can help us understand whether the effect of expertise results from an enhancement in cross-modal processing or from the ability to process one modality with particular accuracy. If drummers have an enhanced ability to process information from one modality, we should expect their higher sensitivity to be more evident in the TOJ than the SJ task, in that the first task requires participants to judge explicitly the occurrence of one modality with respect to the other. However, this is not supported by the results, which suggest that the main process influenced by musical expertise is that of integration. Future research should further investigate this point by comparing drummers’ and non-musicians’ sensitivity to asynchrony for audiovisual displays in which the reliability of one type of information is decreased by, for example, adding noise (Ernst & Banks, 2002; Collignon et al., 2008; Landy, Maloney, Johnston, & Young, 1995). A lack of difference between groups’ performance in this condition would reinforce the idea that the use of both modalities in a cross-modal process is the key to differentiating between drummers and novices.

The results for the SJ task indicated that musical novices became more tolerant to asynchrony when the point-light drumming display was rotated to a less natural view, while the drummers did not. However, this difference between the groups’ TIW width did not extend to the TOJ task, for which no effect of display orientation was found. Hence, in agreement with Vatakis and Spence (2008b), we found no effect of display rotation when using a TOJ task to measure the sensitivity to asynchrony for an audiovisual musical event. On the contrary, we did find an effect of display orientation on the sensitivity to asynchrony when using an SJ task. This result offers further evidence that the SJ and TOJ tasks probably measure very different processes (Mitrani et al., 1986; Van Eijk et al., 2008), an assumption further supported by the inability of some of the participants to perform the TOJ task (Spence...
et al., 2001; Vatakis & Spence, 2006b; Zampini et al., 2003a, 2003b). One possible explanation for the difference in results for the SJ and TOJ tasks is that the SJ might depend primarily on “synchrony” judgments, with the TOJ relying more upon “successiveness” judgments. As previously suggested by Shore, Gray, Spry, and Spence (2005), the TOJ may be a more demanding task and reveal more subtle effects than the SJ task. Our results not only support the idea that the TOJ is more demanding than the SJ but also confirm that the occurrence of negative PSS (in which auditory information precedes the visual) is increased by temporal order judgments (e.g., Vatakis & Spence, 2006a, 2006b) when compared to simultaneity judgments (Van Eijk et al., 2008).

An alternative explanation for the root of the SJ and TOJ differences may rely on different kinds of response bias. For example, PSS derived from the TOJ task can be modulated by the nature of the response requested from participants (“Which stimulus came first?” vs. “Which stimulus came second?”, Shore, Spence, & Klein, 2001) as well as by the absence of a correct answer for the synchronous condition of the stimulus (Van Eijk et al., 2008), while PSS derived from the SJ task can be altered by the assumption that the auditory and visual information should be bound together (Vatakis & Spence, 2007).

However, despite all these differences between the SJ and TOJ tasks, the PSS results of the present study tell us quite similar things about the effects of display orientation and expertise. That is, we did not find an effect of display orientation on the PSS for either tasks, and we found a similar shift of the drummers’ PSS toward the zero point of physical synchrony for both tasks. This allows us to assume that even if the two tasks measure different temporal aspects of audiovisual synchrony perception, they are nevertheless linked. Thus, although there seems to be an agreement that the SJ may provide a more sensitive measure of PSS (Van Eijk et al., 2008; Vatakis et al., 2007a), the TOJ may be better suited to examine sensitivity to temporal order. In some instances, however, the use of an alternative method may be required. Using a similarity judgment task to indirectly measure temporal discrimination would, for example, eliminate all the response biases reported for the SJ and TOJ. Participants, for example, could be asked to judge whether a comparison stimulus is the “same” or “different” from a standard in a pair of stimuli separated by a visual mask, in which the first audiovisual stimulus (the standard) of the pair is always synchronous and the second (the comparison) is any one out of all the possible stimuli (i.e., synchronous or asynchronous). The proportion of “same” judgments
obtained could then be presented as a function of audiovisual lag timings and the PSS and TIW derived by the fitting of the observed data. Alternatively, a direct measure of asynchrony could be derived by rating of the mismatch extent. That is, participants could be asked after each presentation of the display how far apart the audiovisual signals were, on a scale from, for example, 0 to 100. The collected rating could be then used as a function of audiovisual lag timings to find the PSS and TIW width.

The found effect of biological motion rotation on novices’ audiovisual synchrony judgments extends the findings of Saygin et al. (2008) to another kind of audiovisual biological motion event. The biological foot-steps stimuli used by Saygin et al. are very similar to the biological drumming stimuli we used here, in that participants could predict the sound and the impact occurrence in both cases. However, our results indicate that expertise with a certain action eliminates the effect of biological motion rotation on audiovisual synchrony perception. In contrast, the results of Saygin et al. (2008) indicated that participants’ audiovisual simultaneity judgments were affected by the biological motion rotation, despite the fact that all the participants can be considered experts in walking. We think there are two possible explanations for this apparent inconsistency between our results and those of Saygin et al. (2008). First of all, although the two movements are similar in the sense that they depict an impact between the articulation of the body and a surface, they are also quite different regarding the specific kind of movement. Walking requires the whole body to move from one point to another, in contrast with a drumming arm movement for which the body is stable. Furthermore, the flexibility allowed in walking movements is quite constrained when compared with arm impact movements similar to the ones we used in the present study. These impact movements can be quite easily performed from different and less natural body positions (for example, when lying down on the floor), while the same cannot be said for walking. Thus, it might be the case that the acquired model for the drumming arm movements can be mentally rotated by experts because they know that these movements are not impossible to perform, but novices who do not possess such a specific model might find this more difficult. On the contrary, experts in walking might not be able to easily mentally rotate the acquired model for the actions since they are impossible to perform. An alternative explanation would imply a special status of musical stimuli in synchrony perception. That music is not similar to any other audiovisual event has been demonstrated by studies comparing sensitivity to asynchrony in speech, music, and object action stimuli (Vatakis & Spence, 2006a, 2006b). These authors found a difference in sensitivity to asynchrony for these three kinds of audiovisual events and assumed that this difference might be a consequence of differences in participants’ familiarity with the stimuli. Vatakis and Spence (2006b) addressed this point and found a clear effect of familiarity on sensitivity to asynchrony for musical and object action events but not for speech. Since almost everyone has a high familiarity with speech as well as with walking audiovisual events, but fewer are highly familiar with music events, this might suggest that an expertise in music should be considered unusual and not achievable by everyone.

The need to study whether expertise influences the effect of biological motion rotation on audiovisual judgments was underlined by Saygin et al. (2008) as being important to understanding the mechanisms at play. Our findings indicate that expertise eliminates such an effect of visual disruption, in line with Vatakis and Spence’s (2008b) results with expert pianists and full videos. We think that this effect of expertise is probably due to the musicians’ ability to use an internal model (Kawato, 1999; Wolpert, Ghahramani, & Jordan, 1995) of that action to decide on the co-occurrence of the auditory and visual signals even when the visual information is disrupted (Petriani, Russell et al., 2009). Hence, as Saygin et al. suggested, the perception of body movements can engage the observers’ motor system (Saygin, 2007; Saygin, Wilson, Hagler, Bates, & Sereno, 2004), where the acquired model of action might be encoded. Drummers, in contrast with novices, probably encoded through practice a model of drumming swing groove movements. Due to this acquired model, expert drummers might overcome the biological motion disruption in the drumming display representation.

An alternative explanation might be that drummers, through musical practice, enhance their ability to determine the co-occurrence of the auditory and visual information for any kind of multisensory event. However, Petriani, Russell et al. (2009) demonstrated that this cannot be the whole story. Indeed, when drummers judged the simultaneity between the drumming biological motion and the sound of a display from which the impact point information was eliminated, their results were reminiscent of tapping tasks (Aschersleben & Prinz, 1995; Miyake, Onishi, & Pöppel, 2004), indicating that the acquired information for that specific action was used. In other words, when presented with only the point-light arm information (from shoulder to wrist) of the drumming display, not only were drummers still able to discriminate between synchronous and asynchronous displays, while novices were not (Petriani, Russell et al., 2009), but also their point of subjective simultaneity occurred in some instances when the sound was leading the sight, showing the same anticipatory effect as that found in tapping tasks (Aschersleben & Prinz, 1995; Miyake et al., 2004). This interpretation suggests that drummers can use their motor experience to perceive the simultaneity of sight and sound. Along the same lines, the enhanced ability of percussionists (when compared to non-percussionist musicians and non-musicians) to use haptic information to adjust their striking movements in a multisensory setting suggests that acquired experience...
with a specific multisensory action matters more than general musical expertise when a representation of that action is required (Giordano, Avanzini, Wanderley, & McAdams, 2009). The idea that musical training with different instruments gives rise to specialization is also supported by studies revealing differences in brain structures and functions when comparing different kinds of musicians (see Tervaniemi, 2009 for a review).

Still, in the present study, the impact point between drumstick and drumhead was always evident in the drumming point-light display, despite the biological motion rotation. Thus, the possibility that both drummers and novices used only that information and performed a phase-base task (Arrighi et al., 2006) cannot be excluded. This might explain why the biological motion rotation did not affect the drummers’ judgments while it did affect the novices’ performance, if drummers are better able to ignore any kind of discrepancy between the auditory and visual stimulations (Petrini, Dahl et al., 2009). However, Petrini, Dahl et al. (2009) showed that when a discrepancy between the two signals was introduced, novices became more sensitive to asynchrony, in agreement with previous findings (van Wassenhove et al., 2007; Vatakis & Spence, 2007), and not less sensitive as we found here. That is, when presented with drumming displays in which the covariation between the drummer’s arm movements and the resulting sound was eliminated (e.g., a higher preparatory height and strike velocity of the drummer’s arm did not result in a more intense sound), novices showed an enhanced ability to detect asynchrony in comparison with their judgment of displays where this covariation was maintained (Petrini, Dahl et al., 2009). Thus, the effect of biological rotation on the novices’ audiovisual judgments cannot be a result of the same process. Furthermore, the finding by Saygin et al. (2008) that the effect of biological motion rotation on multisensory judgments disappeared when the information was limited to the point-light feet information (where the impact occurred) seems to support an alternative explanation linked to the entire action. This does not exclude, however, the possibility that practice with a specific set of multisensory events might also change the effect of biological motion rotation of other kinds of multisensory actions or of reduced representation of them. Further research is necessary to establish the extent of this expertise effect on audiovisual biological motion perception by using similar paradigms to ours and to that of Saygin et al. (2008), but comparing different sets of natural actions and kinds/levels of expertise.

Finally, the effect of experience and orientation on sensitivity to asynchrony might depend on the improved ability of drummers to judge the temporal relationship between the audiovisual features and/or to correctly select physically corresponding feature pairs (Fujisaki & Nishida, 2007, 2008). In other words, drummers could be more sensitive to asynchrony because they are more able to match the produced sound with a certain arm movement, preparatory height, or strike velocity. However, as mentioned above, Petrini, Dahl et al. (2009) gathered evidence against this possibility, showing that drummers’ synchrony judgments were not affected by the elimination of the audiovisual covariation, while novices’ judgments were affected. These findings indicate that the enhanced drummers’ sensitivity to asynchrony probably depends on their ability to judge the temporal relationship between the auditory and visual signals, while for novices sensitivity may be based on feature matching processes (Fujisaki & Nishida, 2007, 2008). That is, when presented with biological displays rotated to a more unnatural orientation, novices would lose the ability to match the corresponding audiovisual features and consequently become less sensitive to asynchrony. However, the decrease in physical covariation between the signals previously helped novices in detecting asynchrony (Petrini, Dahl et al., 2009), while here the novices’ performance worsened. Hence, a feature-based process of audiovisual synchrony perception cannot be applied to both results. If we accept that the loss of correspondence between the signals, driven by a mismatch in the signals’ physical characteristics, helped novices in their performance in Petrini, Dahl et al.’s (2009) study by decreasing the participants’ assumption that the two signals went together or belonged to the same event as dictated by the “Unity assumption” theory (Schutz & Kubovy, 2008; Vatakis & Spence, 2007, 2008a), here we have to claim a different kind of process. Indeed, despite the rotation of the drummer’s movements, the physical correspondence between the features of the two signals is untouched; thus it is probable that either the low-level disruption of biological motion processing (Saygin et al., 2008) or the low- and/or high-level changes in attention processing (Fujisaki & Nishida, 2008) decreased the novices’ sensitivity to asynchrony. To understand better whether the present effect of expertise and display rotation on novices’ performance is due to bottom-up or top-down factors, an experiment similar to that of Vatakis and Spence (2008a), where they examined asynchrony perception using congruent (e.g., video of a piano with sound of a piano) and incongruent (e.g., video of piano with guitar sound) musical displays, should be replicated by using both full videos and point-light displays of musical events in natural and less natural orientations and also using participants with different levels of musical expertise.

In conclusion, our experiment demonstrated that the gestalt of upright point-light drumming enhances the detection of audiovisual asynchrony for musical novices but not for expert drummers. Hence, the nature of the visual stimulation can affect the perceived synchrony between the two sensory signals, but the extent of this effect is constrained by the level of experience with a particular multisensory event. Future neuroimaging (e.g., fMRI or TMS) studies could contribute to our understanding of whether the network subtending the audiovisual synchrony perception of drumming actions is the same or different between drummers and novices, and
whether primary motor cortex (M1) or motor planning and control areas (e.g., PMA, SMA, and cerebellum) participate together with integration areas (e.g., MTG, STG) in this process or whether the process relies on motor areas only when the visual information is degraded.

Acknowledgments

This work was supported by a grant from the ESRC (RES-060-25-0010).

Commercial relationships: none.
Corresponding author: Dr. Karin Petrini.
Email: karin@psy.gla.ac.uk.
Address: Department of Psychology, University of Glasgow, 58 Hillhead Street, Glasgow G12 8QB, United Kingdom.

References


