RESEARCH ARTICLE

The benefit of multisensory integration with biological motion signals

**Catarina Mendonça · Jorge A. Santos ·
Joan López-Moliner**

Abstract Assessing intentions, direction, and velocity of others is necessary for most daily tasks, and such information is often made available by both visual and auditory motion cues. Therefore, it is not surprising our great ability to perceive human motion. Here, we explore the multisensory integration of cues of biological motion walking speed. After testing for audiovisual asynchronies (visual signals led auditory ones by 30 ms in simultaneity temporal windows of 76.4 ms), in the main experiment, visual, auditory, and bimodal stimuli were compared to a standard audiovisual walker in a velocity discrimination task. Results in variance reduction conformed to optimal integration of congruent bimodal stimuli across all subjects. Interestingly, the perceptual judgements were still close to optimal for stimuli at the smallest level of incongruence. Comparison of slopes allows us to estimate an integration window of about 60 ms, which is smaller than that reported in audiovisual speech.

Keywords Optimal integration · Biological motion · Velocity perception · Audiovisual integration · Point-light walkers

C. Mendonça · J. A. Santos
School of Psychology, University of Minho, Braga, Portugal

J. López-Moliner
Group of Vision and Control of Action, Institute for Brain, Cognition and Behavior, University of Barcelona, Catalonia, Spain

C. Mendonça (✉)
Laboratory of Visualization and Perception,
Centro de Computação Gráfica, Campus de Azurém,
4800-058 Guimarães, Portugal
e-mail: mendonca.catarina@gmail.com

Introduction

Human perception is well adapted to understanding other people's actions. We are more sensitive and accurate in the processing of biological motion visual signals than of other complex motion patterns (Garcia and Grossman 2007; Hiris et al. 2005; Neri et al. 1998; Shiffrar et al. 1997; Shiffrar and Pinto 2002). Indeed, most of the current research on biological motion processing is devoted to understanding the human expertise in extracting information from simplified visual motion displays, dubbed point-light walkers. Human motion signals, however, are not exclusively visual, and there is evidence that auditory footsteps might be processed in the same brain areas as the visual walkers (Bidet-Caulet et al. 2005). The audiovisual integration of human motion has only recently started to gather due attention. Nevertheless, several audiovisual interaction effects were already found (e.g. Barraclough et al. 2005; Van der Zwan et al. 2009), and differences between the multimodal processing of biological and rigid motion arose. There are direction-selective integration mechanisms (Brooks et al. 2007), which do not occur with rigid motion (Alais and Burr 2004a; Wuerger et al. 2003). The crossmodal temporal matching estimates are consistently better with upright than with inverted or scrambled walkers (Saygin et al. 2008). Also, audiovisual detection thresholds conform to probability summation in rigid motion (Wuerger et al. 2003; Meyer et al. 2005), but the detection of visual biological motion is close to Bayesian fusion if matched with meaningful sounds (Arrighi et al. 2009). Despite these interaction effects, it is still unclear how the visual and auditory cues are integrated, or more specifically, there is yet to unfold the perceptual benefit of multisensory cure integration in biological motion.

Several models have been proposed to account for signal interactions in multimodal perception. Classical hypotheses like modality precision, directed attention and modality appropriateness (for a review, see Welsh and Warren 1980) assumed that, given a discrepancy between two sensory modalities, one modality would bias the other, becoming dominant. Most recently alternative models gained relevance, in which signal integration was conceptualized as a probabilistic process, such as signal detection theory, probability summation, fuzzy logical model of perception, and maximum likelihood (e.g. Massaro and Friedman 1990; Ernst and Bülthoff 2004). The Maximum Likelihood Estimation theory (MLE) has earned growing attention by explaining the traditional crossmodal biases as an imbalance in relative reliability of each sensory cue. According to this model, multisensory integration allows perceptual maximization by decreasing uncertainty, and therefore variability (Clarke and Yuille 1990). The integrated percept is optimal, as it has the maximum reliability, and thus the least possible variability.

Over the last years, a few optimal mechanisms were found in multimodal cue integration, namely in the discrimination of visuo-tactile size (Ernst and Banks 2002), audiovisual position (Alais and Burr 2004b; Battaglia et al. 2003), event number (Andersen et al. 2005; Shams et al. 2005), and arrival time (Wuerger et al. 2010). In velocity discrimination, however, integration effects are not as robust. Vision affects the perceived velocity of sounds in a weighted average fashion (López-Moliner and Soto-Faraco 2007), but there is only a weak tendency for optimal integration of the velocity estimates, which defaults to probability summation when unimodal weights diverge (Bentvelzen et al. 2009).

Here, we were interested in assessing the integration mechanisms of audiovisual biological motion signals at different walking speeds. In walking motion, both velocity and rhythmic cues are present. The translational global motion of the walker provides velocity information, while the frequency of steps in the auditory stimuli and local limb motion in the visual stimuli bring rhythmic information. Such phenomena are interdependent, as when the human walker accelerates or decelerates, so does the frequency of the local periodic events. Therefore, for biological motion, velocity and rhythmic information are coupled and could not be isolated without violating biomechanical constraints and the corresponding perceptual coherence. In rigid motion, as opposed to biological motion, these cues are independent, but it is still a matter of debate whether subjects recover speed from them. In this study, we did not separate both sources of information, as we intended to obtain plausible biomechanical and perceptual walking representations. Therefore, while specifically addressing walking speed, the present work comprised simultaneously rhythm and velocity discriminations of human motion.

Having in mind that auditory delays are frequently required for auditory-visual subjective simultaneity to be reached (e.g. Alais and Carlile 2005; Arrighi et al. 2006; Van Eijk et al. 2008; Vatakis and Spence 2006; Di Luca et al. 2009), a preliminary experiment was conducted. We measured simultaneity judgments at several audiovisual walking asynchronies and obtained a point of subjective simultaneity.

The main experiment consisted of a velocity discrimination task. Visual, auditory, and audiovisual walkers, which either was congruent or incongruent in velocity, were compared to a standard audiovisual stimulus walking at an average rate. We hypothesized that congruent audiovisual stimuli would lower the discrimination thresholds. The discrimination accuracy was compared to optimal predictions and analyzed at different velocity congruency levels. Results are discussed in light of the MLE model and the temporal windows of integration.

Method

Participants

Two subjects took part in the preliminary experiment and seven subjects took part in the main experiment. In the preliminary experiment, one participant was an author (CA) and the other was an untrained naïve subject (SM), one left and the other right handed. In the main experiment, one participant was an author (CA), 3 were trained but naïve to the purpose of the study (BO, LA, RO), and 3 were untrained naïve subjects (CO, JO, RA), 3 left handed and 4 right handed. All underwent visual and auditory standard screening and had normal or corrected-to-normal sensory acuity.

Stimuli and apparatus

The visual stimuli were point-light walkers of 13 white dots, generated in the Laboratory of Visualization and Perception (http://webs.psi.uminho.pt/lvp) by a Vicon motion capture system with 6 cameras MX F20 at 240 Hz and a set of custom LabVIEW implemented routines. All stimuli corresponded to the correct motion coordinates of a 17-year-old male, 1.87 m high, walking at five translational velocities (1.1, 1.3, 1.5, 1.7, and 1.9 m/s) with constant time intervals between steps (590, 560, 530, 500, and 470 ms, respectively, with a maximum variability of ±4 ms). When virtualized as walking at 10.73 m from the observer, the visual angular velocities corresponded to 6, 7, 8, 9, and 10 deg/s. The point-light stimuli were presented through a computer with a graphics card Nvidia Quadro FX 4500, and a 3 chip DLP projector Christie Mirage S + 4 K, with

$1,400 \times 1,050$ pixel resolution at 60 Hz. The image was projected onto a 2.78×2.09 m screen area in a dark room. On the screen, the white dots (54 cd/m²) moved against a black background (0.4 cd/m²) and each dot subtended 0.9 deg of the visual angle.

The auditory stimuli were step sounds from the database of controlled recordings from the College of Charlston (Marcell et al. 2000). They corresponded to the sound of a male walking over a concrete floor. These sounds were auralized by a MATLAB routine with generic head-related transfer functions from the MIT database (http://sound.media.mit.edu/resources/KEMAR.html), as free-field, at the same five velocities and step frequencies as the visual stimuli, at a distance of 10.73 m in space. The auditory stimuli were presented with a Realtec Intel 8280 IBA sound card through a set of in-ear earphones Etymotics ER-4B. The sounds reached the listeners at a maximum level of 35 dB SPL, peaked at 260 Hz, as measured with a Brüel & Kjaer Head and Torso Simulator type 4128-C with Brüel & Kjaer Pulse Analyzer type 3560-C and Pulse CPB Analysis software.

The audiovisual stimuli were programmed and presented through a custom application running on top of Vr Juggler code. The temporal alignment of the sound onset and the displayed frames for each stimulus was checked and adjusted prior to the experiment using a custom-built latency analyzer (Arm7 microprocessor coupled with light and sound sensors). The sensors were targeted at the precise frame of the visual footstep and at the earphones' paired output. Due to limitations of the visualization system, the SOA accuracy was of ± 1 ms and the precision across trials was within a maximum range of ± 6 ms.

Procedure

Measuring perceived asynchronies

We first wanted to know the magnitude of the perceived asynchrony between the visual and auditory cues to motion and the width of the integration window to help us interpret the data from the main experiment. Audiovisual congruent stimuli moving at 6, 7, 8, 9, and 10 deg/s, with step frequencies of 1.69, 1.78, 1.89, 2.0, and 2.13 Hz, respectively, were presented with a duration of 1.6 s. Stimuli delays consisted of the time differences between the frame when the visual foot touched the ground and the moment when the auditory step sound reached the subject's ear. Each SOA was randomly varied across all stimulus velocities, from −120 ms (auditory lead) to 120 ms (visual lead) in 30 ms intervals. All stimuli were pseudo-randomly presented, 20 trials per asynchrony level for each velocity (5×20), with an interstimulus interval of 0.8 s. Participants had to judge if the visual and the auditory signals were synchronized or not.

Main experiment

In each trial of the main experiment, there was a standard stimulus and a test stimulus. The standard stimulus was always an audiovisual congruent walker, moving at 8 deg/s, with a step frequency of 1.89 Hz and the duration of 1.6 s. The test stimuli could either be unimodal (visual or auditory) or bimodal. There were 5 possible stimulus velocities of the visual and auditory stimuli: 6, 7, 8, 9, or 10 deg/s, with the respective 5 step frequencies of 1.69, 1.78, 1.89, 2.0, and 2.13 Hz. The bimodal stimuli corresponded to all visual and auditory velocities factorially paired (5×5), which were therefore congruent or incongruent. Incongruence varied from 1 deg/s of velocity difference to 4 deg/s. There were three possible test stimulus durations: 1.5, 1.6, and 1.7 s. This duration variation was too small to introduce differences in the number of visual and auditory steps, but it was large enough to assure that the starting and ending points as well as the first and last frame of the point-light stimuli were not informative. As such, for example, slower stimuli did sometimes travel longer distances than faster stimuli. Both visual and auditory signals moved from left to right and crossed hemifields at half the stimulus duration. There were always 3 steps, and stimuli were arranged in such way that the second step occurred precisely when crossing the visual mid-line. Thus, in the bimodal incongruent stimuli, the second step would always be perceptually aligned and the mismatch was equally distributed by both the first and the third step (see Fig. 1). Also and as a result from the preliminary experiment, where we found that a delay was required for simultaneity to be reached (see Fig. 2), all sounds in this experiment lagged an additional amount of 30 ms.

There were 20 trials per stimulus, distributed in 4 experimental blocks. Each block presented all stimulus types (visual, auditory, and bimodal) pseudo-randomly. In a 2-IFC task, participants were instructed to respond in which interval (standard or test) was the fastest walker.

Results and discussion

Perceived asynchronies

In the preliminary experiment, several auditory-visual SOAs were tested in a simultaneity judgment task. The relative frequency of "synchronized" estimates as a function of SOA was well fitted by a Gaussian curve (Fig. 2).

The overall results show that as the absolute SOA increased the average simultaneity judgments decreased. However, the distribution revealed a positive bias, as the highest proportion of synchronized answers occurred not for the stimuli with the 0 ms SOA (relative frequency of
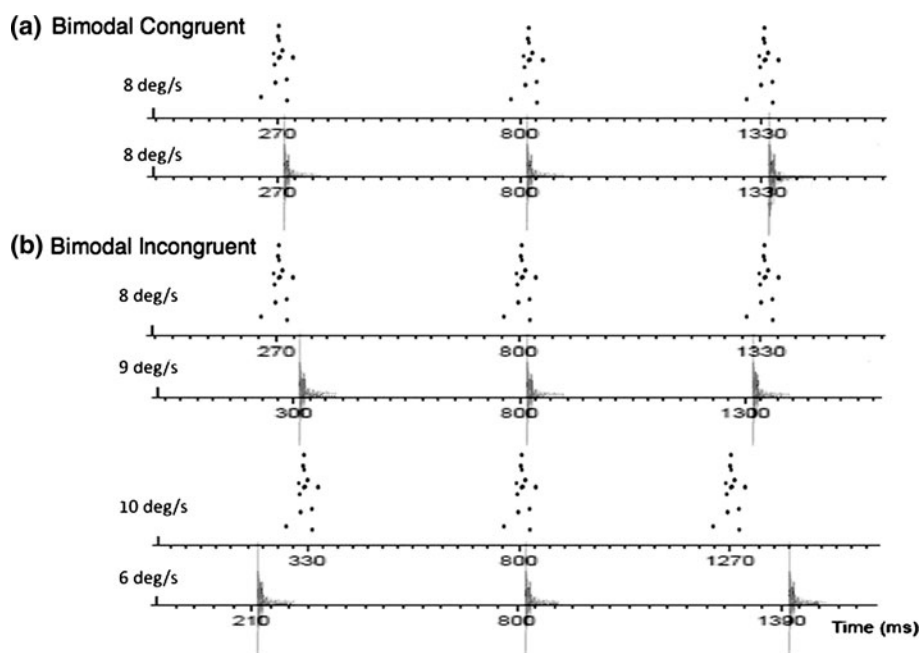
Fig. 1 Examples of bimodal stimuli. **a** Bimodal Congruent. Both visual (*upper line*) and auditory (*lower line*) signals moved at 8 deg/s (standard stimulus). The first step occurred 270 ms after the trial start, the second step occurred at the 800th ms (half of the 1,600 ms long stimulus), and the third step occurred at the 1,330th ms. There was an actual delay of 30 ms between the moment when the visual foot reached the ground and the moment when the sound reached the subject's inner ears. **b** Bimodal incongruent. The *upper* example is a stim-

ulus with 8 deg/s of visual velocity 9 deg/s auditory velocity. The second step co-occurred after 800 ms, but the first and last steps were mismatched; the sound arrived 30 ms later than the image in the first step and 30 ms earlier in the third step (0 ms later in the first step and 60 ms later in the last one, considering the 30 ms audio lag). The *bottom* example is the most incongruent bimodal stimulus. Here, the maximum time mismatch was 120 ms (150 ms in the last step, with the audio lag)
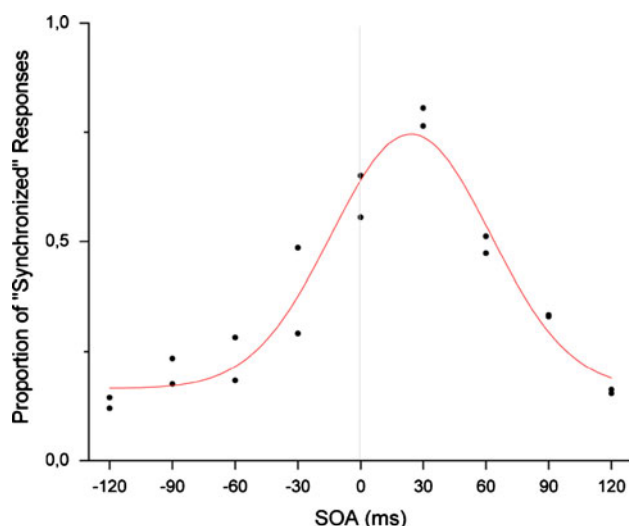

Fig. 2 Proportion of the pooled synchronized responses as a function of each auditory-visual delay, from −120 ms (auditory first) to 120 ms (visual first). Each dot corresponds to the responses of each of the two participants (100 trials/dot). Gaussian fit with mean = 25 ms and $\sigma = 38.2$ ms

shifted to the right, with a point of subjective simultaneity (PSS) of 25 ms (95% CI of 5.27) and a discrimination threshold (1σ) of 38.2 ms.

These results support the broad understanding that for subjective simultaneity to be reached visual lead is required (e.g. Arrighi et al. 2006).

Results in this experiment were taken into account in the subsequent experimental design. Since our stimuli were displayed with several realistic depth cues such as perspective, relative change of the walkers' size, and dot dimension, as well as sound attenuation effects, we chose to keep a delay of the auditory signal for all stimuli, consistent with our distribution peak and signal propagation differences. Therefore, in the main experiment, congruent bimodal walkers (proximal stimuli) were always mismatched by 30 ms and incongruent walkers followed the same rule, with varying delays.

Main experiment

In the main experiment, visual, auditory, and audiovisual walkers at different speeds and congruency levels were compared to a standard bimodal walker at 8 deg/s. As we used a speed discrimination task instead of a detection one

0.6) but for the stimuli with 30 ms SOA, vision had to be presented first (a 0.78 relative frequency). The best-fit Gaussian, adjusted with MLE fitting procedure, was thus

(as it has often been the case) in which the standard was always bimodal some caveats need to be taken into account in the analysis. We will compare auditory, visual, and audiovisual psychometric curves, which have been obtained against the same (bimodal) reference. When uncertainty is high, potential biases might appear in the unimodal conditions in the form of choosing the bimodal standard more often. If subjects did so, we would be able to detect it by observing fluctuations in the mean (PSE) of the curve. Nevertheless, to further clarify the sensory modality impact of the standard stimulus, an experiment not reported here was run with a similar methodology as in the main experiment, where visual, auditory, and audiovisual standards were used with 3 all-naïve participants. The results from that experiment revealed no differences in the discrimination slopes. Once these cautions were taken, MLE predictions in variance reduction were tested for all subjects individually and for the pooled data results.

As there were no differences between results for each stimulus duration, nor starting or ending points, all data were pooled together for analysis purposes. All results in the main experiment are presented with respect to stimulus velocity (not stimulus step frequency). This was a matter of choice, mainly due to the fact that participants were indeed instructed to make a velocity judgment. We first obtained the proportion of faster responses against the velocity for the different conditions and fitted cumulative Gaussians to these data distributions. This was done for individual subjects and pooled data. The parameters of the fit, mean and SD, defined the PSE and the JND (=1SD), respectively. The individual and group variances for each condition were obtained from the estimated SD. The group and individual variability in terms of the variance is presented in Fig. 3.

According to MLE, the benefit of multiple sensory cue integration lays in the reduced variance and therefore increased precision of the perceptual judgments. Overall, unimodal precision varied greatly across subjects. Variance was lower in the visual than in the auditory discriminations of 3 participants, and 4 participants had overlapping unimodal variances. Nonetheless, optimal integration predictions were within the 95% confidence interval of the audiovisual congruent condition for all subjects. The discrimination data pooled over subjects are displayed in Fig. 4 together with the best cumulative Gaussian fits.

Congruent audiovisual stimuli yielded the steepest slopes (smallest deviation, $\sigma = 0.6$), better than those of visual ($\sigma = 0.8$) and auditory ($\sigma = 0.92$) stimuli alone, and significantly different from best unimodal (bootstrap $P = 0.043$). Interestingly, bimodal stimuli with a velocity incongruence of only 1 deg/s also revealed a steep function ($\sigma = 0.78$) with a slope which did not differ from bimodal congruent nor best unimodal (bootstrap $P = 0.554$ and $P = 0.32$, respectively). The integration disruption was
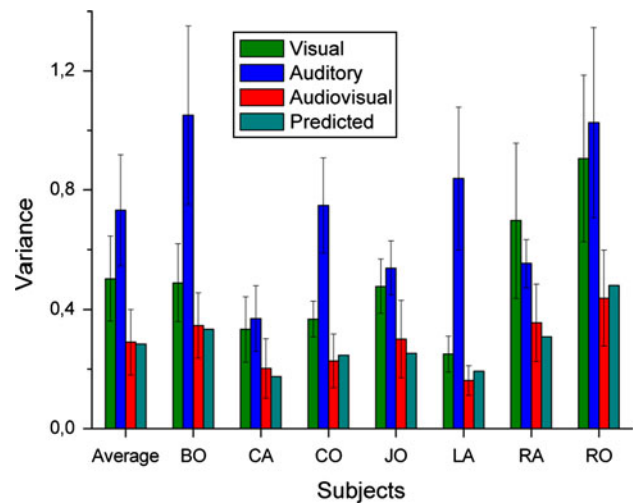

Fig. 3 Velocity discrimination variances in the visual, auditory, and audiovisual congruent stimuli and the optimal integration predictions for each subject. Variance values were obtained by the squared standard deviations of the cumulative Gaussians fitted for each participant. From those values, the predicted optimal integration variances were calculated as $\sigma^2 AV = \sigma^2 A \sigma^2 V / \sigma^2 A + \sigma^2 V$. *Error bars* display 95 percent confidence intervals obtained from each Gaussian's standard error
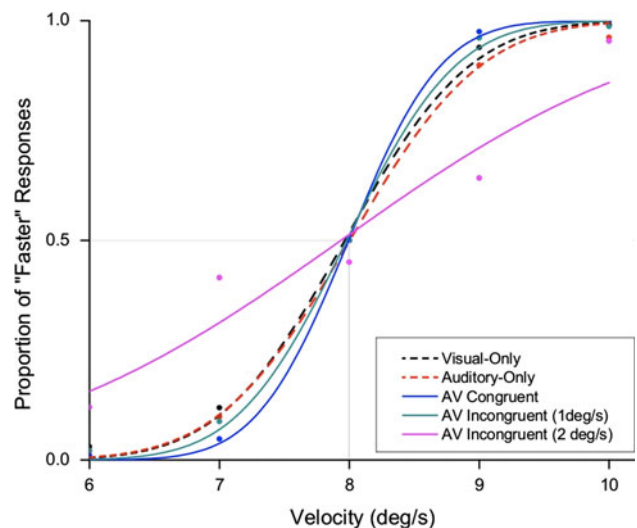

Fig. 4 Best-fit cumulative Gaussians, with unconstrained parameters, obtained from the faster than standard judgments of all participants (420 trials per dot) for the unimodal (*dashed lines*) and audiovisual (*full lines*) stimuli

observed in the stimuli with a velocity incongruence of 2 deg/s ($\sigma = 1.08$), and bimodal stimuli with larger velocity differences did not fit Gaussians. Importantly, PSE values remained close to standard in all conditions.

The unexpected steepness of the slopes at the smallest level of incongruence led us to wonder about a window of audiovisual integration, which might stretch to such velocity mismatch. While little is known about windows of velocity integration, there is a growing body of information on the
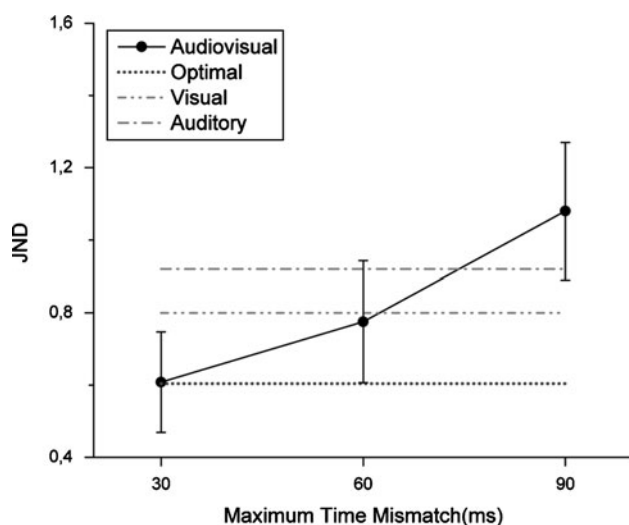
Fig. 5 Just noticeable differences (JND) as a function of the temporal alignment. *Error bars* are confidence intervals calculated with the bootstrap technique in 1,000 simulations (Efron et al. 1993)

temporal windows of such processes (e.g. van Wassenhove et al. 2007). Taking a new look over Fig. 1, we might notice that indeed our incongruent stimuli do vary in temporal alignment. Summing up the 30 ms auditory delay we added to all bimodal stimuli (see preliminary experiment), we have that congruent stimuli were mismatched in 30 ms, 1 deg/s incongruent stimuli were maximally mismatched in 60 ms and 2 deg/s incongruent stimuli in 90 ms. In Fig. 5, the discrimination thresholds are displayed as a function of the temporal mismatch of the bimodal stimuli.

The JND of the congruent stimuli, which were mismatched in 30 ms, fell well within optimal integration prediction and, as said above, was significantly lower than the best unimodal. The sensitivity of stimuli with the 60 ms mismatch was better than the best unimodal (visual) and worse than audiovisual congruent, but did not differ from either. Stimuli mismatched in 90 ms had the highest discrimination thresholds. These results are consistent with the optimal mechanisms in the processing of biological motion. They further support the conceptualization of a window of integration within which multisensory biological signals might be fused.

Discussion

The experiments we report here were devoted to understanding the multisensory integration of biological motion at several walking speeds. In a preliminary experiment, we assessed the audiovisual PSS of the point-light steps and the footstep sounds to accurately produce perceptually aligned and congruent bimodal stimuli. We found that stimuli

with a 30 ms auditory lag were the most perceived as synchronized. This result is consistent with the assumption that visual lead is necessary for perceived simultaneity of audiovisual stimuli, and it is in line with several PSS values recently reported in literature (e.g. Arrighi et al. 2006; Vatakis et al. 2006; see Van Eijk et al. 2008 for a review). Also, this finding supports the assumption that the necessary asynchrony might follow physical rules (Alais and Carlile 2005; Sugita and Suzuki 2003), since for our stimuli's virtual localization an actual 30 ms auditory delay was to be expected. The width of the Gaussian fit for the synchronized judgments might also be informative, as it revealed a JND of ± 38.2 ms. Such value, which reflects a temporal window of perceived audiovisual simultaneity for human walkers, is consistent with the integration window found in the main experiment (stimuli lagging from -13.2 to 63.2 ms would be perceived as synchronous). This window is smaller than that reported for audiovisual speech, object action, and music play (Van Wassenhove et al. 2007; Vatakis et al. 2006), but larger than the window for flashes-clicks, and other non-biological motion signals (e.g. Arrighi et al. 2006; Van Eijk et al. 2008). However, for a more accurate understanding of the simultaneity perception in audiovisual biological motion, a broader study is needed.

In the main experiment, we focused on the multisensory integration of biological motion at different walking speeds. There were visual, auditory, and bimodal human walkers, congruent and incongruent, which were compared to a standard audiovisual walker in a velocity discrimination task. All participants discriminated more accurately the bimodal congruent than the unimodal stimuli, and their accuracy with such stimuli conformed to optimal integration mechanisms.

Optimal mechanisms had been found before with audiovisual motion in a time-to-arrival experiment (Wuerger et al. 2010). However, those findings were obtained in an experiment where participants had received a large amount of training, with feedback, before the experimental sessions. Furthermore, during those procedures, subjects were only exposed to congruent stimuli, and therefore they would be able to develop optimal strategies. Conversely, during our experiments, participants remained untrained, since most of the randomly displayed bimodal stimuli were incongruent (20 out of 25), and thus integration learning was impaired. Another study directly assessed the audiovisual integration of velocity signals and looked for optimal integration mechanisms (Bentvelzen et al. 2009). In that study, the authors had found only a weak tendency to follow MLE predictions, as just some of the subjects did discriminate velocity in an optimal fashion, whereas the others had performed worse than with the best unimodal stimuli. The authors had interpreted this as separate integration

mechanisms derived from different unimodal weights: only those subjects who had similar visual and auditory accuracies would follow MLE. Here, in contrast, optimal integration was a robust effect across all subjects, despite different unimodal discrimination performances, previous training, or experimental goal awareness.

The integration effects that we find are consistent with previous neuroimaging data. The posterior superior temporal sulcus (STSp), a brain region selectively activated for the sight of biological motion stimuli, is also hypothesized to be a structure where multisensory integration might occur (Calvert 2001). Single cell-recordings showed significant interactions between visual and auditory processing of human actions in this area (Barraclough et al. 2005), and fMRI studies revealed that STSp responds to the auditory stimulation of human steps (Bidet-Caulet et al. 2005). Our findings might be read as further evidence for a possible neural specialization in the processing of human motion-related signals.

An unexpected outcome of our findings was the discrimination slope of the stimuli at the smallest level of incongruence. The JND of these stimuli was still lower than in the best unimodal and remained close to optimal. A window of integration might explain this result. According to this hypothesis, optimal mechanisms are not exclusive to a given pair of audiovisual stimuli, but might occur along a small continuum of slightly mismatched stimuli, which still allows for unity assumption. Such windows of integration have long been studied and carry an extensive body of experimental support in audiovisual speech (e.g. Yabe et al. 1997; Van Wassenhove et al. 2007). In our study, stimuli at the smallest level of incongruence were mismatched in 1 deg/s of angular velocity and were also temporally mismatched in a maximum of ±60 ms. Therefore, we propose the existence of a temporal window of integration in biological motion signals, which might be about ±60 ms wide. Interestingly, this value is close to the −13.2/+63.2 ms synchrony interval we report in the preliminary experiment. On the other hand, this window is much smaller than the ±200 ms audiovisual integration window found in audiovisual speech (Van Wassenhove et al. 2007). A narrower window in speed perception would have a higher adaptive value as it would allow to ascertain small discrepancies in situations when judging the speed accurately is crucial.

In sum, firstly, we argue that audiovisual congruent walkers benefit from cue integration in an optimal fashion. We cannot fully claim for MLE predictions, as we did not address PSE weighting by directly manipulating cue reliability. Nevertheless, we should stress that we found an optimal benefit of multisensory integration in lowering discrimination thresholds. Secondly, we argue that these integration processes might occur within a well-defined temporal interval of stimulus alignment. Further research should approach the spatiotemporal constraints of biological motion multisensory cue fusion.

References

Alais D, Burr D (2004a) No direction-specific bimodal facilitation for audiovisual motion detection. Cog Brain Res 19:195–196

Alais D, Burr D (2004b) The ventriloquist effect results from near-optimal bimodal integration. Curr Biol 14:257–262

Alais D, Carlile S (2005) Synchronizing to real events: subjective audiovisual alignment scales with perceived auditory depth and speed of sound. Proc Natl Acad Sci 102:2244–2247

Andersen TS, Tiippana K, Sams M (2005) Maximum likelihood integration of rapid flashes and beeps. Neurosci Lett 380:155–160

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

Arrighi R, Marini F, Burr D (2009) Meaningful auditory information enhances perception of visual biological motion. J Vision 9:1–7

Barraclough NE, Xiao D, Baker CI, Oram MW, Perrett DI (2005) Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. J Cog Neurosci 17:377–391

Battaglia PW, Jacobs RA, Aslin RN (2003) Bayesian integration of visual and auditory signals for spatial localization. J Opt Soc Am 20:1391–1397

Bentvelzen A, Leung J, Alais D (2009) Discriminating audiovisual speed: Optimal integration defaults to probability summation when companent reliabilities diverge. Perception 28:966–987

Bidet-Caulet A, Voisin J, Bertrand O, Fonlupt P (2005) Listening to a walking human activates the temporal biological motion area. NeuroImage 28:132–139

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

Calvert GA (2001) Crossmodal processing in the human brain: insights from functional neuroimaging studies. Cereb Cortex 11:1110–1123

Clarke JJ, Yuille AL (1990) Data fusion for sensory information processing. Kluwer Academic, Boston

Di Luca M, Machulla TK, Ernst MO (2009) Recalibration of multisensory simultaneity: Cross-modal transfer coincides with change in perceptual latency. J Vision 9:1–16

Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. Chapman and Hall, New York

Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415:429–433

Ernst MO, Bülthoff HH (2004) Merging the senses into a robust percept. Tends Cog Sci 8:162–169

Garcia JO, Grossman ED (2007) Necessary but not sufficient: Motion perception is required for perceiving biological motion. Vision Res 48:1144–1149

Hiris E, Krebeck A, Edmonds J, Stout A (2005) What learning to see arbitrary motion tells us about biological motion perception. J Exp Psychol 31:1096–1106

López-Moliner J, Soto-Faraco S (2007) Vision affects how fast we hear sounds move. J Vision 6:1–7

Marcell MM, Borella D, Greene M, Kerr E, Rogers S (2000) Confrontation naming of environmental sounds. J Clin Exp Neuropsychol 22:830–864

Massaro DW, Friedman D (1990) Models of integration given multiple sources of information given multiple sources of information. Psychol Rev 97:229–252

Meyer GF, Wuerger SM, Rohrbein F, Zetzsche C (2005) Low-level integration of auditory and visual motion signals requires spatial co-localization. Exp Brain Res 166:538–547

Neri P, Morrone MC, Burr DC (1998) Seeing biological motion. Nature 395:894–896

Saygin AP, Driver J, de Sa VR (2008) In the footsteps of biological motion and multisensory perception. Psychol Sci 19:469–475

Shams L, Ma WJ, Beierholm M (2005) Sound-induced flash illusion as an optimal percept. NeuroReport 16:1923–1927

Shiffrar M, Pinto J (2002) The visual analysis of bodily motion. In: Prinz W, Hommel B (eds) Common Mechanisms in Perception and Action–Attention and Performance XIX. Oxford University Press, Oxford, pp 381–399

Shiffrar M, Lichtey L, Chatterjee SH (1997) The perception of biological motion across apertures. Percept Psychophys 59:51–59

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

Van der Zwan R, MacHatch C, Kozlowski D, Troje NF, Blanke O, Brooks A (2009) Gender bending: Auditory cues affect visual judgments of gender in biological motion displays. Exp Brain Res 198:373–382

Van Eijk RLJ, Kohlrausch A, Juola JF, Van de Par S (2008) Audiovisual synchrony and temporal order judgments: Effects of experimental method and stimulus type. Percept Psychophys 70:955–968

Van Wassenhove V, Grant KW, Poeppel D (2007) Temporal window of integration in audiovisual speech perception. Neurophychologia 45:598–607

Vatakis A, Spence C (2006) Audiovisual synchrony perception for music, speech, and object actions. Brain Res 111:134–142

Welsh RB, Warren DH (1980) Immediate perceptual response to intersensory discrepancy. Psychol Bull 88:638–667

Wuerger SM, Hofbauer M, Meyer GF (2003) The integration of auditory and visual motion signals at threshold. Percept Psychophys 65:1188–1196

Wuerger S, Meyer G, Hofbauer M, Zetzsche C, Schill K (2010) Motion extrapolation of auditory-visual targets. Information Fusion 11:45–50

Yabe H, Tervaniemi M, Reinikainen K, Näätänen R (1997) Temporal window of integration revealed by MMN to sound omission. Neuroreport 8:1971–1974