Meaningful sounds enhance visual sensitivity to human gait regardless of synchrony

James Philip Thomas

Department of Psychology, Rutgers University, Newark, NJ

Maggie Shiffrar

Department of Psychology, Rutgers University, Newark, NJ

Previous research demonstrates that meaningfully related sounds enhance visual sensitivity to point-light displays of human movement. Here we report two psychophysical studies that investigated whether, and if so when, this facilitation is modulated by the temporal relationship between auditory and visual stimuli. In Experiment 1, participants detected point-light walkers in masks while listening to footsteps that were either synchronous or out-of-phase with point-light footfalls. The relative timing of auditory and visual walking did not impact performance. Experiment 2 further tested the importance of multisensory timing by disrupting the rhythm of the auditory and visual streams. Participants detected point-light walkers while listening to footstep or tone sounds that were either synchronous or temporally random with regards to point-light footfalls. Heard footsteps improved visual sensitivity over heard tones regardless of timing. Taken together, these results suggest that during the detection of others' actions, the perceptual system makes use of meaningfully related sounds whether or not they are synchronous. These results are discussed in relation to the unity assumption theory as well as recent empirical data that suggest that temporal correspondence is not always a critical factor in multisensory perception and integration.

Introduction

Behaviorally relevant information about people, objects, and events is transmitted to us through multiple sensory modalities, and our phenomenal reality is shaped by the co-occurrence of these sensory events. Consider, for example, a fine dining experience, during which you simultaneously sample the sights, sounds, smells, and tastes of a wonderful meal. In such instances, the experiences that we attribute to one modality, for example the taste of a steak, are strongly influenced by information from other modalities (e.g., Petrini, McAleer, & Pollick, 2010; Spence & Shankar, 2010). Typically, sensory information arriving from common sources is effortlessly integrated in our phenomenal experience.

Given the ubiquity of multisensory information in our daily lives, it is no surprise that the field of multisensory research has expanded in recent years. Fueled by recent technological advances and an increased awareness of the inherently multisensory nature of human perceptual processing, multisensory research has captured the attention of scientists from a variety of perceptual disciplines. Increasingly, investigators interested in a wide variety of perceptual phenomena have turned their efforts towards understanding how information from different sensory modalities interacts to determine perception and behavior.

Multisensory interplay occurs at various stages of processing and is determined by a host of stimulus factors, including spatial location, timing, and semantics, or meaningful associations (Alais, Newell, & Mamassian, 2010; Bedford, 2001; Doehrmann & Naumer, 2008; Radeau & Bertelson, 1977; Welch, 1999; Welch & Warren, 1980). Recently, the role of meaningful associations in multisensory perception has been explored with stimuli ranging in complexity and real-world veridicality, from simple colored circles and color word vocalizations (Laurienti, Kraft, Maldjian, Burdette, & Wallace, 2004), to naturalistic sounds and pictures (e.g., Chen & Spence, 2010; Molholm, Ritter, Javitt, & Foxe, 2004; Iordanescu, Guzman-Martinez, Grabowecky, & Suzuki, 2008), to point-light (e.g., Arrighi, Marini, & Burr, 2009; Thomas & Shiffrar, 2010; van der Zwan et al., 2009) and full light movies (e.g., Schutz & Kubovy, 2009; Vatakis & Spence 2007).

Arguably, the most complex, compelling, and behaviorally relevant stimuli are people, and of course

Citation: Thomas, J. P. & Shiffrar, M. (2013). Meaningful sounds enhance visual sensitivity to human gait regardless of synchrony. *Journal of Vision, 13*(14):8, 1–13, http://www.journalofvision.org/content/13/14/8, doi:10.1167/13.14.8.

doi: 10.1167/13.14.8 Received May 29, 2013; published December 6, 2013 ISSN 1534-7362

we gather information about others through multiple sensory pathways as well. Recent evidence suggests that meaningful congruence impacts the multisensory perception of human speech (e.g., Grant & Seitz, 2000; Vatakis, Ghazanfar, & Spence, 2008; Vatakis & Spence 2007), and also emotions, (e.g., Collignon et al., 2008; Hietanen, Leppanen, Illi, & Surakka, 2004; Pourtois & deGelder, 2002), and actions (e.g., Eramudugolla, Henderson, & Mattingley, 2011; Mitterer & Jesse, 2010; Schutz & Kubovy, 2009).

In the action-perception domain, meaningful congruence has been shown to impact multisensory processing, though it should be noted that not all paradigms provide such evidence. Collectively, evidence from experiments using full-light visual displays has been equivocal, with some reporting that integration is enhanced when crossmodal signals are meaningfully associated (e.g., Eramudugolla, Henderson, & Mattingley, 2011; Mitterer & Jesse, 2010; Schutz & Kubovy, 2009) and others reporting no evidence of enhanced integration (Vatakis, Ghazanfar, & Spence, 2008; Vatakis & Spence, 2007, 2008). A lack of multisensory enhancement for meaningfully related auditory and visual action stimuli in temporal order judgment paradigms has led some investigators to propose that such meaningful congruence effects are in fact specific to the perception of audiovisual speech (e.g., Vatakis, Ghazanfar, & Spence, 2008). However, effects of meaningful associations on multisensory processing have also been observed in recent audiovisual point-light action paradigms. For example, the sounds of female footsteps impact the perception of point-light walker gender (Van der Zwan et al., 2009), and the sounds of tap dancing enhance visual sensitivity to point-light tapping feet (Arrighi et al., 2009). Research from our own lab has shown that visual sensitivity to point-light walking is enhanced when paired with footstep sounds, but not when paired with temporally synchronous but meaningfully unrelated control sounds (Thomas & Shiffrar, 2010).

A critical question that arises is to what degree do action sounds need be synchronous with visually presented actions in order for facilitation to occur? In multisensory research, there is a pervasive and longstanding assumption that whether information from different sensory modalities is integrated depends critically on the temporal relationship between the unisensory streams (e.g., Alais, Newell, & Mamassian, 2010; Spence, 2007; Stein & Meredith, 1993; Welch, 1999). Indeed, evidence from a variety of multisensory perceptual tasks confirms the impact of intersensory timing on multisensory processing. For example, stimulation in one sensory modality influences the perceived location of stimulation in another modality when both are presented synchronously (e.g., Jack & Thurlow, 1973; Radeau & Bertelson, 1977). As well, visual apparent motion impacts perceived auditory motion direction when unisensory streams are presented at the same time (e.g., Soto-Faraco et al., 2002). At threshold, detection of simple visual stimuli improves when paired with synchronous, but not asynchronous, white noise bursts (Bolognini, Frassinetti, Serino, & Làdavas, 2005; Frassinetti, Bolognini, & Làdavas, 2002), and detection of a novel visual target in a series of distracters is enhanced by a novel tone presented synchronously, but not prior to, onset of the visual stimulus (Vroomen & deGelder, 2000). Indeed, evidence suggests that unitary perception breaks down when unisensory temporal offsets exceed what has been termed a temporal window of integration (e.g., Spence & Squire, 2008; Stein & Meredith, 1993; van Wassenhove, Grant, & Poeppel, 2007; Vroomen & Keetles, 2010).

The effect of timing on the perception of audiovisual point-light displays has been investigated by several researchers. For instance, both temporal frequency (Saygin, Driver, & de Sa, 2008) and velocity judgments (Mendonca, Santos, & Lopez-Moliner, 2011) are influenced by the degree to which visual and auditory streams are synchronized. As well, in the aforementioned point-light tapping feet study, synchronized yet task-uninformative tap sounds facilitated detection of point-light tapping feet embedded in visual noise, while desynchronized tap sounds did not (Arrighi et al., 2009). These data collectively suggest that audiovisual timing may be critical in multisensory action perception, at least under certain conditions.

Recently, however, the critical importance of timing in multisensory processing has been problematized. Results from recent experiments have challenged the view that all aspects of multisensory interplay are hinged upon temporal synchrony (e.g., Adam & Noppeney, 2010; Arrighi et al., 2009; Chen & Spence, 2010, 2011; Munhall, Gribble, Sacco, & Ward, 1996; Petrini et al., 2010; Raposo, Sheppard, Schracter, & Churchland, 2012; Schneider, Engel, & Debener, 2008; Vroomen & Keetles, 2010). For instance, the emotional content of music affects the perceived emotionality of visual displays that reference the same event (e.g., drumming sounds with drumming motion), regardless of timing (Petrini et al., 2010). Furthermore, the perception of a novel McGurk stimulus persists despite audiovisual lags (e.g., Munhall et al., 1996).

Examples of audiovisual interplay in the face of perceptible temporal asynchrony are not specific to speech processing. For example, enhancements in visual object sensitivity as a function of meaningfully related sounds have been observed when sounds preceded the visual stimulus (e.g., Adam & Noppeney, 2010; Schneider et al., 2008). As well, the Schutz-Lipscomb illusion, wherein the duration of visual marimba gestures biases the perceived duration of

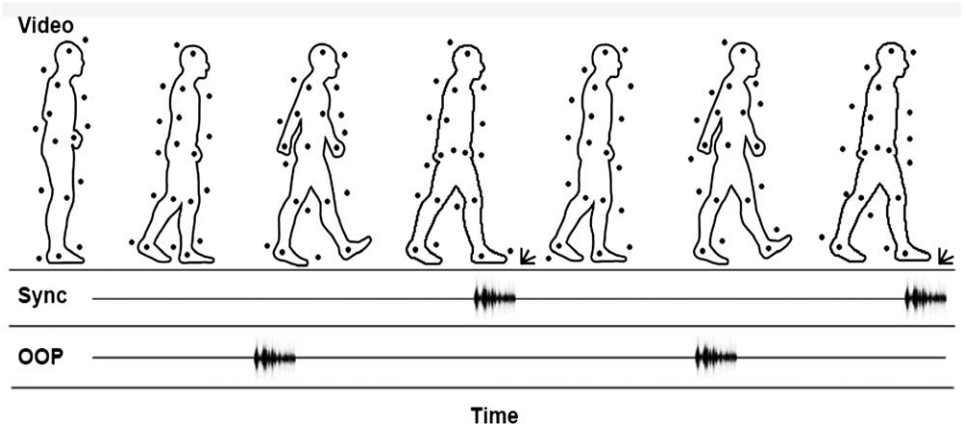

Figure 1. Experiment 1. Diagram depicting the time course of a single gait cycle of a coherent point-light walker in a mask. Time-line shows the onset of sounds occurring either in synchrony with, or out-of-phase with the visual footfalls of the point-light walker. The form of the human body was added here for clarity and did not appear in the experimental trials. Not all point-lights are shown.

heard musical sounds, persists despite auditory delays of up to 700 ms (Schutz & Kubovy, 2009). Whereas there is no doubt that intersensory timing can exert a strong influence on multisensory interplay, it is quite plausible that not all aspects of multisensory interplay are hinged upon temporal synchrony. Thus, the question of whether and how intersensory timing impacts multisensory action processing remains an open question.

Experiment 1

The experiments reported herein were designed to further test the effect of intersensory timing on the perception of audiovisual point-light displays of human movement. In Experiment 1, naïve participants detected masked point-light walkers while listening to meaningfully related sounds (footsteps) that were either synchronous or out-of-phase with the visible gait cycle of the walkers. If intersensory timing is a critical determinant of multisensory action sensitivity, then hearing synchronous footstep sounds should enhance visual sensitivity relative to asynchronous sounds.

Method

Participants

Forty-one undergraduate students at the Newark campus of Rutgers University (mean age 21.2 years) participated in this experiment for course credit. Of these, one was excluded from analyses after it was discovered that he/she had failed to comply with the experimental instructions and chose instead to press one button for every trial. Analyses were conducted on the remaining 40 participants, with 20 each assigned to the synchronous and out-of-phase footstep conditions. All participants reported normal or corrected-to-normal visual and auditory acuity. The Rutgers University Institutional Review Board approved this and the subsequent study and written informed consent was obtained for all participants.

Apparatus

Human motion kinematics were recorded with a ReActor motion capture system (Ascension Technology Corporation). Motion capture data were rendered into point-light movies of human walking motion using Kaydara Motion Builder™ 6.0 software. Sounds were processed on a MacBook computer using a digital audio workstation (Cubase LE 4 by Steinberg). Video and audio were combined using Macintosh iMovie HD 6.0.3 video editing software. The experiment was run on an AMD™ computer with an AMD Athlon™ 64 × 2 dual core processor. Movies were displayed on a 22-in. ViewSonic monitor with a 1280 × 1020 pixel resolution and a refresh rate of 60 Hz. Sounds were played through Bose® Companion 2® speakers which flanked the monitor, facing the participant. The experiment was programmed and controlled using E-Prime® version 2.0.

Stimuli

During motion capture recording, two individuals walked with a neutral emotional gait along a linear path (approximately 6 m) within the ReActor system. This linear walking motion was subsequently rendered into point-light movies. The construction of the point-light movies is described in extensive detail elsewhere (Thomas & Shiffrar, 2010).

In brief, two types of movies were created: "person present" movies, which depicted a spatiotemporally

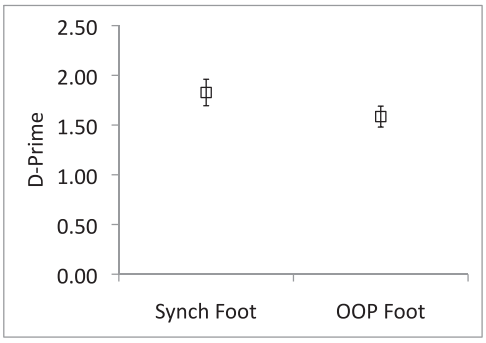

Figure 2. Experiment 1. Detection sensitivity as assessed by D-prime. Mean D-primes as a function of sound timing. Error bars indicate standard errors.

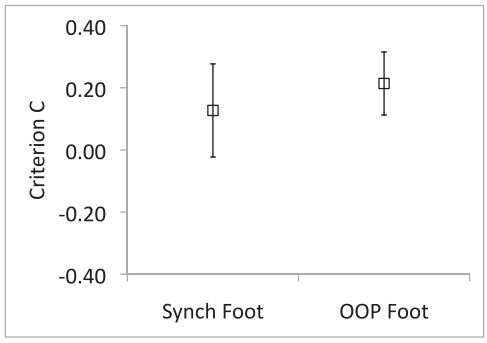

Figure 3. Experiment 1. Bias as assessed by Criterion C. Mean C as a function of sound timing. Error bars indicate standard errors.

coherent point-light walker embedded within a point-light mask, and "person absent" movies, depicting a spatiotemporally scrambled walker embedded within a point-light mask. Both the scrambled walkers and masks were constructed using a classic method (Bertenthal & Pinto, 1994), in which all of the original 13 points comprising each spatiotemporally coherent walker were duplicated and the locations of those points were spatially scrambled. "Person present" movies depicted 13 points defining a coherent point-light walker and 13 points defining the mask. "Person absent" movies depicted 13 points defining the scrambled walker and 13 points defining the mask; thus, "person absent" movies displayed two spatiotemporally scrambled point-light walkers. Both types of movies were rendered from four different observer-centered orientations (walking towards, walking away from, walking rightward, or walking leftward). On average, the duration of each complete step cycle was approximately 660 ms and overall walking speed was approximately 1.5 m/sec, which falls well within normal gait speeds (e.g., Jacobs, Pinto, & Shiffrar, 2004).

Two visually identical sets of movies were rendered, one for each sound condition. For the synchronous sound condition, footstep sounds were inserted into "person present" movies whenever a point-light foot contacted the ground, and in "person absent" movies at the same time points. For the out-of-phase sound condition, footstep sounds were shifted in time so as to be out-of-phase with the footfalls of the point-light walkers by 50%. One half of the out-of-phase movies presented sounds with an auditory lead of approximately 300 ms, whereas the other half presented sounds with an auditory lag of equal duration. Thus, the timing of the walkers' footsteps was consistent with natural human walking, whereas the out-of-phase footstep sounds were shifted in time.

Footstep sounds were downloaded from an online audio archive (www.sounddogs.com). Individual footstep sounds were extracted from a continuous recording of an individual walking in hard-soled shoes on a concrete floor, resulting in 20 unique individual footstep sounds. Multiple individual footstep sounds were used instead of one single, repeated footstep sound in order to increase the ecological validity of the stimuli (Gibson, 1979). Naturally, each footstep sound varied in duration, with a mean average of approximately 245 ms. A 10 ms fade envelope was applied to each footstep sound clip to prevent audible pops at onset and offset. Sounds were rendered in mono, such that sound motion did not accompany the horizontal motion of the point-light stimuli. Additionally, all sounds were equated for subjective loudness, such that sounds did not loom or recede with the in-depth motion of the walkers. Each movie lasted 3000 ms and depicted a smoothly continuous, natural human gait with four to five individual audiovisual footstep events.

The two sets of movies (temporally synchronous and out-of-phase) were visually identical. Each contained the same 56 individual movies (seven scrambled + seven coherent) × (four motion directions). Half of the movies (28) showed a spatiotemporally coherent point-light walker embedded in a scrambled walker mask. The other half showed a scrambled walker in a scrambled walker mask. These movies were visually identical to those used in pervious experiments (see Thomas & Shiffrar, 2010, 2011).

Procedure

To reduce the possibility that participants would infer the experimental hypothesis, and to avoid potential order effects, this and the following experiments utilized a between-subjects design. Participants were randomly assigned to one of two sound conditions: synchronous footsteps or out-of-phase footsteps. The experimental procedures were otherwise identical across participants. Participants sat with their heads in a chinrest facing the monitor at a distance of 90 cm. During instruction, participants were told that they

would view short movies that might or might not contain a point-light person walking within a mask of visually similar point-lights. Participants were instructed to view each movie and report whether they saw a person in each display by pressing the appropriate button (y/n).

The experiment began with eight practice trials, during which participants observed four "person present" and four "person absent" movies. Next, participants completed two experimental blocks each containing 48 movies; 24 "person present" and 24 "person absent" movies, six of each rendered from one of the four walker directions described above. Both experimental blocks were identical, such that each movie was shown twice. Movie presentation was randomized within blocks. No feedback was provided.

Results

The data were analyzed using a signal detection framework (McMillan & Creelman, 1991). Mean accuracy was broken down into proportions of hits and false alarms, which were used to generate two statistics: D-Prime, a measure of sensitivity; and Criterion C, a measure of bias.

A one-way ANOVA on d-prime revealed no effect of sound timing (synchronous versus out-of-phase), $F(1, 28) = 2.053$, $p = 0.160$, $f = 0.23$. It should be noted that a priori power analysis indicated that the experiment had sufficient power to detect a medium-sized effect had one existed. The choice of effect size for the a priori power analysis was based on the effect sizes observed in previous experiments (Thomas & Shiffrar, 2010), which ranged from medium to large. There was also no difference in bias between sound timing conditions, $F(1, 38) = 0.230$, $p = 0.634$, $f = 0.08$.

Repeated-measures ANOVA on motion direction (away, towards, leftwards, and rightwards) revealed no differences in visual sensitivity across groups, $F(3, 117) = 0.464$, $p = 0.708$, $f = 0.11$. Repeated-measures ANOVA with sound timing as the between-subjects factor and motion direction as the within-subjects factor revealed no interaction, $F(3, 114) = 1.475$, $p = 0.225$, $f = 0.20$. As well, there was no main effect of motion direction on observer bias, $F(2.16, 84.17) = 1.271$, $p = 0.287$, $f = 0.18$, and no interaction between sound timing and motion direction, $F(2.17, 82.41) = 1.043$, $p = 0.362$, $f = 0.17$ (Greenhouse-Geisser).

Because temporal asymmetries in multisensory interplay are common (e.g., Morein-Zamir, Soto-Faraco, & Kingston, 2003; van Wassenhove, Grant, & Poeppel, 2007), a paired samples t test was conducted comparing sensitivity to stimuli with an auditory lead to sensitivity to stimuli with a visual lead. This analysis revealed no differences in sensitivity as a function auditory lead versus lag, $t(19) = -0.369$, $p = 0.716$, $d = 0.17$. Visual sensitivity in the out-of-phase footstep condition did not differ as a function of whether the onset of the footstep sounds preceded or followed the onset of the visual footfalls.

Discussion

In Experiment 1, hearing synchronous footsteps conferred no benefit in visual sensitivity over hearing out-of phase-footsteps. This pattern of results is different from that observed by Arrighi and colleagues (2009: experiment 1), who found that visual detection of point light tapping feet was improved by the presentation of synchronous, but not asynchronous, tapping sounds. Rather, these current results appear to suggest that relative timing is not as critical a factor as previously assumed for point-light displays paired with meaningfully related sounds. However, the inherent interpretative difficulties associated with null effects make it difficult to draw a clear conclusion from Experiment 1 alone. This, and the reasons summarized below, motivated the design of Experiment 2.

Because meaningful congruency has been shown to influence audiovisual temporal perception (e.g., Mitterer & Jesse, 2010; Vatakis & Spence, 2007), we considered the possibility that participants might have actually perceived the temporally offset footstep sounds as synchronous with the visual footfalls of the walker, amounting to a point-light walker temporal ventriloquism effect (e.g., Aschersleben & Bertelson, 2003; Bertelson & Aschersleben, 2003; Morein-Zamir, Soto-Faraco, & Kingston, 2003). However, in the current experiment, the temporal offset between visual footfalls and the out of synch footstep sounds was approximately 300 ms, a value which far exceeds the temporal window of integration for walking motion estimated by Mendonca and colleagues (2011). We also considered the possibility that the rhythm of the out-of-phase footstep streams could have aided participants in detecting point-light walkers. In other words, it is possible that participants were able to use the rhythm of the sound stream, which was preserved in the out-of-phase footsteps condition, to anticipate the timing of the gait cycle of the point-light walkers. Experiment 2 was conducted in order to test this alternative hypothesis, and further probe the importance of timing in audiovisual action detection.

Experiment 2

In Experiment 2, participants detected coherent point-light walkers in a mask while listening to sounds

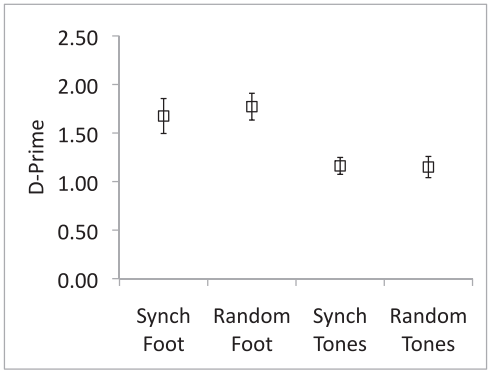

Figure 4. Experiment 2. Detection sensitivity as assessed by D-prime. Mean D-primes as a function of sound type and sound timing. Error bars indicate standard errors.

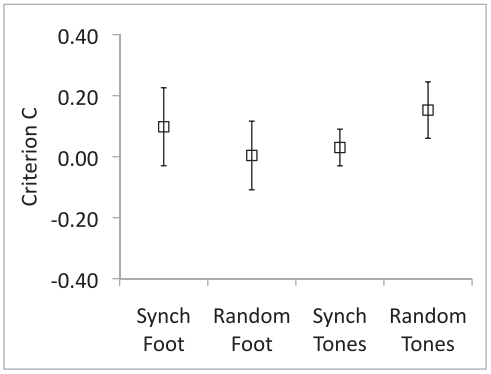

Figure 5. Experiment 2. Bias as assessed by Criterion C. Mean C as a function of sound type and sound timing. Error bars indicate standard errors.

that were either meaningfully related (footsteps) or not (tones), as well as temporally synchronous or temporally random with regards to the gait cycle of the visual walkers. We hypothesized that hearing synchronous footsteps would result in greater walker detection sensitivity than would hearing synchronous tones, in agreement with previous results (Thomas & Shiffrar, 2010). If intersensory timing is a critical determinant of multisensory action sensitivity, then hearing synchronous footstep sounds should also enhance visual sensitivity relative to asynchronous footsteps.

Method

Participants

Sixty-five undergraduate students at Rutgers University–Newark (mean age 19.9 years) participated in this experiment for course credit. Of these, five were excluded from analyses after a one-sample t test indicated their accuracy was not different from what would be expected if they had responded randomly (50%). Analyses were conducted on data from the remaining 60 participants, with 15 each assigned to one of four sound-synchrony conditions. All participants reported normal or corrected-to-normal visual and auditory acuity and provided informed consent.

Stimuli and procedure

Two different types of sounds were used in this experiment: the footstep sounds from Experiment 1 and a pure tone obtained from an audio archive (www.audiosparx.com). The 1000 Hz tone was matched in duration to the average duration of the footstep sounds, and was treated to the same 10 ms fade envelope. As in Experiment 1, all sounds were rendered in mono such that no apparent sound motion was present. Sounds were equated for subjective loudness.

Four visually identical sets of movies were rendered, one for each sound condition. For the temporally synchronous sound conditions, footstep or tone sounds were inserted into "person present" movies whenever a point-light foot contacted the ground, and in "person absent" movies at the same time points, as in Experiment 1. For the temporally random sound conditions, footstep or tone sounds were inserted into randomly selected frames of each movie. Each movie consisted of 90 total frames, and sounds may have onset on any given frame, with the constraint that no two sounds could overlap in time. This manipulation had the effect of breaking the constant phase relationship between the auditory and visual components that was present in Experiment 1. As the temporally random condition movies played, the timing of the sounds was completely decoupled from the footfalls of the visual walker, and in no way resembled the temporal pattern of a natural walking gait. The visual stimuli and experimental procedures were otherwise identical to those used in Experiment 1.

Results

A two-way ANOVA on d-prime revealed a main effect of sound type (footsteps vs. tones), $F(1, 56) = 18.324$, $p < 0.001$, $f = 0.57$, no effect of sound timing (synchronous vs. random), $F(1, 56) = 0.101$, $p = 0.752$, $f = 0.04$, and no interaction, $F(1, 56) = 0.163$, $p = 0.688$ $f = 0.05$. Post-hoc tests (Bonferroni corrected) indicated that visual sensitivity to coherent point-light walkers in the presence of synchronous footsteps was significantly greater than visual sensitivity with either synchronous ($p = 0.049$) or temporally random tones ($p = 0.042$). As well, visual sensitivity to point light walkers in the presence of temporally random footsteps was significantly greater than with synchronous ($p = 0.012$) or temporally random tones ($p = 0.010$). In other words,

participants in both footstep conditions exhibited greater visual sensitivity than participants in both tone conditions. Visual sensitivity with synchronous and temporally random footsteps did not differ ($p = 1.00$). As well, sensitivity with synchronous and random tones was equivalent ($p = 1.00$). Repeated-measures ANOVA revealed no effect of motion direction on Criterion C, $F(2.14, 125.94) = 0.982$, $p = 0.382$, $f = 0.13$, and all interactions between motion direction, sound type, and sound synchrony were not significant, all $p > 0.05$.

Repeated-measures ANOVA again revealed no main effect of motion direction across groups, $F(3, 177) = 0.769$, $p = 0.536$, $f = 0.11$. As well, there was no interaction between motion direction and sound type (footsteps vs. tones), $F(3, 168) = 2.221$, $p = 0.088$, $f = 0.20$, and no interaction between motion direction and sound timing (synchronous vs. random), $F(3, 168) = 0.105$, $p = 0.957$, $f = 0.04$. The three-way interaction was also not significant, $F(3, 168) = 0.162$, $p = 0.922$, $f = 0.05$.

A two-way ANOVA on Criterion C revealed no effect of sound type, $F(1, 56) = 0.158$, $p = 0.693$, $f = 0.05$; no effect of sound timing, $F(1, 56) = 0.020$, $p = 0.888$, $f = 0.00$; and no interaction, $F(1, 56) = 1.138$, $p = 0.291$, $f = 0.13$. Thus, the differences in visual sensitivity observed were not accompanied by shifts in bias. In other words, participants were not more likely to report seeing a person in the mask simply because they heard footsteps.

Discussion

In agreement with previous research, meaningfully related sounds enhanced visual sensitivity to point-light walkers relative to meaningfully unrelated sounds (Thomas & Shiffrar, 2010). The goal of the current set of experiments was to examine the impact of meaningful audiovisual associations relative to the impact of audiovisual timing. Interestingly, audiovisual timing did not significantly impact performance in the point-light walker detection task. This result is surprising given the pivotal role that intersensory timing is presumed to play in multisensory processing. Yet, previous research has revealed multisensory facilitation effects in the face of perceptible temporal mismatches (e.g., Adam & Noppeney, 2010; Chen & Spence, 2011; Petrini et al., 2010; Schneider et al., 2008). For example, enhancements in visual object sensitivity as a function of meaningfully related sounds have been observed when sounds preceded the visual stimulus (e.g., Adam & Noppeney, 2010; Schneider et al., 2008).

However, an important issue to be addressed is how these results contrast with those from previous point-light investigations in which timing was a defining factor (Mendonca et al., 2011; Saygin et al., 2008). In a study by Saygin and colleagues (2008), participants were required to compare the temporal frequency (number of cycles over time) of a point-light stimulus and a sound stream of pure tones. The visual gestalt of an upright and coherent (vs. a scrambled or inverted) point-light walker improved auditory-visual frequency judgments, but only when the visual walker and sound stream were closely coupled in time. When the visual walker and sound stream were presented out-of-phase, accuracy was equivalent across visual stimuli. The authors concluded that the visual gestalt of the point-light walker aided audiovisual comparisons, and that this facilitation depended upon whether temporal evidence suggested that sounds could conceivably have been produced by the visual stimulus. In another recent study by Mendonca and colleagues (2011), participants made velocity discriminations with unimodal auditory, visual, and audiovisual point-light walkers. Synchronous audiovisual point-light walkers improved velocity judgments (relative to unimodal visual walkers or footstep sounds) in an optimal fashion (Mendonca et al., 2011). Slightly asynchronous audiovisual point-light walkers (temporal misalignment not exceeding 60 ms) also improved performance relative to unimodal trials, while audiovisual trials with greater degrees of asynchrony did not.

In both of the above studies, timing had a critical impact on performance. However, though similar to the current experiments in regards to stimuli (point-light walkers and sound streams), comparing results across experiments employing very different tasks (and thus presumably processes) is difficult. In the frequency judgment task, participants were required to actively compare two co-occurring rhythms (Saygin et al., 2008) and report whether they were the same or different. Similarly, Mendonca et al. (2011) employed a velocity judgment task in which participants compared the velocity of visual, auditory, and audiovisual walking to a standard audiovisual walker. Both experiments required participants to focus on, and specifically compare the timing of, the stimuli (note that for human walking, rhythm and velocity are coupled, so judgments of velocity and rhythm cannot be dissociated). The current experiments employed a masked point-light walker detection task, and thus utilized a response dimension (do you see a person: yes/no) that is orthogonal to timing. Therefore, it is quite plausible that the divergence of our results simply reflects differences in processing demands.

What is more intriguing is how our results diverged from another point-light action detection task, in which no benefit in detection sensitivity was found when asynchronous tap sounds were presented during masked coherent and scrambled movies of tap dancing feet (Arrighi et al., 2009). The cause of this divergence

of results is not entirely clear, but several explanations are plausible.

The movements of the point-light walkers used in the current experiment were rhythmic and cyclical, in contrast to the arrhythmic tap dancing feet utilized by Arrighi and colleagues (2009). It is therefore possible that audiovisual timing is less critical for stimuli with a regular, cyclical rhythm. Indeed, it has been previously suggested that the temporal correspondence of auditory and visual signals is a strong cue to unity for erratic, arrhythmic pattern (e.g., Noesselt et al., 2007). Another difference between these stimuli was their average temporal frequency. The cyclical walking motion used here progressed at approximately 4.5 Hz, in contrast to the relatively higher-frequency motion of the tap dancing feet. Differences in temporal frequency could also account for this divergence of results (e.g., Noesselt et al., 2005).

It could also be the case that audiovisual timing has a greater impact on visual sensitivity to categories of actions with which the observer has less experience. Most individuals have extensive experience both viewing and performing walking, while very few have similarly extensive experience viewing and performing tap dancing. Indeed, action expertise modulates the perception of audiovisual point-light actions (Petrini, et al., 2009; Petrini, Holt, & Pollick, 2010; Petrini, Russell, & Pollick, 2009).

Another difference between the current experiments and the point-light tap study was the density of the visual mask. The current experiments utilized a 13-point visual mask for all point-light stimuli, while the point-light tap dancing feet were masked by a significantly greater number of mask dots (determined individually by an adaptive QUEST procedure targeting 75% correct performance). It is possible that the timing of the auditory stimulus is more critical for detecting more heavily masked stimuli.

A related issue is that of intersubject variability. Indeed, recent evidence suggests that individual observers vary significantly in their ability to process point-light biological motion in a mask (Miller & Saygin, 2013). It is therefore possible that utilization of a fixed number of masking dots across participants reduced the power of the current experiments to observe a subtle effect of interstimulus timing on the visual detection of point-light walking motion.

Alternatively, the divergence in results could be related to other differences in the specific stimuli used. In the point-light tap study, participants detected point-light tapping feet, each defined by three point-light markers placed on the shoes of the tap dancer during motion capture. Only the foot motion of the actor was present in displays; the action was represented as the coordinated motion of two sets of three rigidly connected dots. The current point-light walker detection experiments utilized whole-body point-light walkers, which ambulated across the screen relative to the observer. It is therefore possible that the increased veridicality of the full-bodied point-light walker, coupled with its linear locomotion, activated different and/or additional processing mechanisms than those recruited for the processing of individual point-light feet. Indeed, evidence suggests that multisensory whole and partial biological motion signals are processed differently (Saygin et al., 2008).

General discussion

Two psychophysical studies were conducted to examine the role of timing in the perception of audiovisual point-light displays. Hearing footstep sounds enhanced visual sensitivity to walking motion regardless of whether they were synchronous with the walker's footfalls. In fact, hearing temporally random footstep sounds, whose timing was completely decoupled from the motion of the visual stimulus, enhanced sensitivity relative to hearing simple tones that were synchronous with the visual walkers' footfalls. These results suggest that, at least under some conditions, simply hearing the sounds of human footsteps can increase visual sensitivity to human walking motion. Meaningful associations may drive multisensory interplay across wide temporal windows.

Whereas there is no doubt that timing can be an important factor in multisensory processing, and whereas evidence suggests that the integration of auditory and visual information into a unified percept depends upon some degree of temporal alignment (e.g., Spence & Squire, 2008), the current results are consistent with increasing evidence that all multisensory interactions do not hinge on temporal synchrony (e.g., Driver & Noesselt, 2008). In fact, multisensory facilitation effects in the face of temporal misalignments are not uncommon (e.g., Adam & Noppeney, 2010; Arrighi, Marini, & Burr, 2009; Chen & Spence, 2011; Petrini et al., 2010; Raposo, Sheppard, Schracter, & Churchland, 2012; Schneider et al., 2008). For instance, in one recent multisensory decision-making paradigm, both humans and rodents were able to effectively integrate auditory and visual rate information despite audiovisual asynchronies (Raposo et al., 2012). Prior evidence also exists for multisensory enhancements as a function of meaningful associations between asynchronous audiovisual stimuli. As aforementioned, emotions conveyed through music affect the perception of emotions conveyed visually when auditory and visual streams reference the same event, regardless of audiovisual timing (Petrini et al., 2010). In the domain of multisensory object priming, meaningful

congruency effects are also observed despite asynchrony, for example, when meaningfully related images speed the identification of subsequently heard sounds (Schneider et al., 2008).

What's more, evidence suggests that the *meaningful content* of stimuli can change observers' perception of the *timing* of events depicted in another modality (e.g., Mitterer & Jesse, 2010; Schutz & Kubovy, 2009; Vatakis & Spence, 2007). For instance, meaningfully related visual gestures bias the perceived duration of action sounds (Schutz & Kubovy, 2009). As well, research utilizing temporal frequency judgments has revealed that viewing a human actor playing the keys of a piano influences the perceived timing of the piano's sound (Mitterer & Jesse, 2010). Thus, meaningful associations can impact the perception of what have been traditionally conceptualized as "bottom-up" stimulus attributes such as timing.

According to the unity assumption model, whether or not unisensory stimuli are integrated depends upon many stimulus factors, including both timing and meaningful associations. Stimuli that are not closely coupled in time are unlikely to reference the same event, and are therefore unlikely to be integrated into a coherent percept (e.g., Welch & Warren, 1980). One striking quality of the unity assumption model is the rather strict predictions it makes about the relationship between the observer's unity assumption and their subsequent phenomenal perceptual experiences. Accordingly, if perceptual cues are reasonably consonant, then perceptual fusion should be experienced, and intersensory discrepancies should not be perceived. Conversely, if intersensory discrepancies are significant enough in quantity and/or magnitude such that that the unity assumption is not met, there is no impetus for multisensory interplay.

This stringent formulation of the unity assumption may not adequately describe all aspects of multisensory perception. For one thing, it cannot account for multisensory priming effects, in which multisensory interactions are observed despite perceptible mismatches in "bottom-up" stimulus properties like timing (e.g., Chen & Spence, 2011). Indeed, evidence is accumulating to suggest that multisensory interactions may be more flexible and dynamic than the unity assumption model allows for. For instance, in an experiment utilizing the McGurk-MacDonald illusion, a clear dissociation between phonological and temporal perception was observed (Soto-Faraco & Alsius, 2009). At some auditory-visual ISIs, participants perceived illusory McGurk syllables, while at the same time they were consciously aware of the asynchrony of their unimodal components. McGurk syllables are also perceived with gender-incongruent stimuli, as when a male voice is dubbed onto a female face. Such results present a challenge to the unity assumption model, which predicts that intersensory discrepancies should not be perceived if perceptual fusion occurs. It is quite possible that multisensory integration is more flexible, and perhaps more task-dependent, than originally thought. Auditory and visual information may not need to be integrated into a unified percept in order for multisensory interplay to occur (Driver & Noesselt, 2008). Rather than a unified multisensory integration system with an all-or-nothing unity assumption component, multisensory processing may graded, such that different stimulus attributes may interact through different mechanisms, potentially resulting in various dimensions of perceptual fusion and segregation (Soto-Faraco & Alsius, 2009).

Whereas the studies reported herein were solely psychophysical in nature, speculation regarding the neural mechanisms undergirding the observed effects is warranted. Potential mechanisms likely include a functional circuit involving the Superior Temporal Sulcus (STS) and Premotor Cortex (PMC). Both regions have been implicated in the visual processing of human actions (e.g., Saygin, 2007), and both are selectively modulated by action-related sounds (e.g., Bidet-Caulet, Voisin, Bertrand, & Fonlupt, 2005; Engel, Frum, Puce, Walker, & Lewis, 2009; Pizzamiglio et al., 2005; Saarela & Hari, 2008). STS has long been regarded as integral in the perception of point-light displays of human movement (e.g., Grossman & Blake, 2002), and researchers have posited a role for STS in the integration of auditory and visual cues at an action-specific level of representation (e.g., Barraclough, Xiao, Baker, Oram, & Perrett, 2005; James, VanDerKlok, Stevenson, & James, 2011). The PMC also responds to action-related stimuli in multiple modalities (e.g., Kaplan & Iacoboni, 2007; Kohler et al., 2002). That the PMC supports action perception agrees with behavioral (e.g., Loula, Prasad, Harber, & Shiffrar, 2005; Shiffrar & Freyd, 1990, 1993), electrophysiological and neuroimaging evidence (e.g., Calvo-Merino, Glaser, Grezes, Passingham, & Haggard, 2005; Calvo-Merino, Grezes, Glaser, Passingham, & Haggard, 2006; Gazzola, Aziz-Zadeh, & Keysers, 2006; Saygin, 2007; Saygin, Wilson, Hagler, Bates, & Sereno, 2004; Stevens, Fonlupt, Shiffrar, & Decety, 2000), as well as theories that implicate the motor system in the perception of human actions (e.g., Decety & Grezes, 1999; Hommel, Musseler, Aschersleben, & Prinz, 2001; Prinz, 1997; Wilson & Knoblich, 2005).

The rapid detection of conspecifics is paramount for survival and social success, and sensory information from multiple modalities can aid in this process. Indeed, actions are often heard before they are seen, and action sounds can serve an important role in signaling socially relevant events (Aglioti & Pazzaglia, 2010). This and previous data suggest that hearing footstep sounds can increase visual sensitivity to

walking motion. This increase in visual sensitivity could potentially be due to enhanced activation of action-specific representations, though more work is required to adequately test this hypothesis. It is logically plausible that such a mechanism could have evolved, given its potential behavioral utility. Indeed, such a mechanism could potentially service not only negative (enhanced threat detection leading to appropriate fight or flight behavior) but also positive (enhanced person detection in situations where affiliation is crucial, i.e., being stranded alone in the wilderness) interpersonal interactions, and in doing so optimize positive behavioral outcomes. Additionally, such a mechanism could have practical implications for certain individuals, such as security guards, who must remain vigilant to the presence of others despite a myriad of perceptual obstacles. Indeed, if sounds can enhance visual sensitivity to certain movement patterns, auditory information may be a critical addition to traditional visual surveillance in security situations where the rapid detection of others is paramount.

Keywords: multisensory perception, action perception, timing, semantics, biological motion, point-light walkers

Acknowledgments

This research was supported by the NSF (grant EXP-SA 0730985) and the Simons Foundation (grant 94915). The authors thank Ashley Blanchard, Adam Doerrfeld, John Franchak Sr., Jen Gomes, Steve Ivory, Christina Joseph, and Shabeer Wali.

Commercial relationships: none.
Corresponding author: James Philip Thomas.
Email: jpthomas@psychology.rutgers.edu.
Address: Department of Psychology, Rutgers University, Newark, NJ.

References

Adam, R., & Noppeney, U. (2010). Prior auditory information shapes visual category-selectivity in ventral occipito-temporal cortex. *NeuroImage, 52,* 1592–1602.

Aglioti, S. M., & Pazzaglia, M. (2010). Representing actions through their sound. *Experimental Brain Research, 206*(2), 141–151.

Alais, D., Newell, F. N., & Mamassian, P. (2010). Multisensory processing in review: From physiology to behavior. *Seeing and Perceiving, 23,* 3–38.

Arrighi, R., Marini, F., & Burr, D. (2009). Meaningful auditory information enhances perception of visual biological motion. *Journal of Vision, 9*(4):25, 1–7, http://www.journalofvision.org/content/9/4/25, doi:10.1167/9.4.25. [PubMed] [Article]

Aschersleben, G., & Bertelson, P. (2003). Temporal ventriloquism: Crossmodal interaction on the time dimension 2. Evidence from sensorimotor synchronization. *International Journal of Psychophysiology, 50,* 157–163.

Barraclough, N. E., Xiao, D., Baker, C. I., Oram, M. W., & Perrett, D. I. (2005). Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *Journal of Cognitive Neuroscience, 17*(3), 377–391.

Bedford, F. L. (2001). Towards a general law of numerical/object identity. *Current Psychology of Cognition, 20*(3), 113–175.

Bertelson, P., & Aschersleben, G. (2003). Temporal ventriloquism: Crossmodal interaction on the time dimension 1. Evidence from auditory-visual temporal order judgment. *International Journal of Psychophysiology, 50,* 147–155.

Bertenthal, B. I., & Pinto, J. (1994). Global processing of biological motion. *Psychological Science, 5*(4), 221–225.

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.

Bolognini, N., Frassinetti, F., Serino, A., & Làdavas, E. (2005). "Acoustical vision" of below threshold stimuli: Interaction among spatially converging audiovisual inputs. *Experimental Brain Research, 160,* 273–282.

Calvo-Merino, B., Glaser, D. E., Grezes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: An fMRI study with expert dancers. *Cerebral Cortex, 15,* 1243–1249.

Calvo-Merino, B., Grezes, J., Glaser, D. E., Passingham, R. E., & Haggard, P. (2006). Seeing or doing? Influence of motor familiarity in action observation. *Current Biology, 16,* 1905–1910.

Chen, Y.-C., & Spence, C. (2010). When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked pictures. *Cognition, 114,* 389–404.

Chen, Y.-C., & Spence, C. (2011). Crossmodal semantic priming by naturalistic sounds and spoken words enhances visual sensitivity. *Journal of Experimental Psychology: Human Perception and Performance, 37*(5), 1554–1568.

Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., & Lepore, F. (2008). Audio-visual integration of emotion expression. *Brain Research, 1242,* 126–135.

Decety, J., & Grezes, J. (1999). Neural mechanisms subserving the perception of human actions. *Trends in Cognitive Sciences, 3*(5), 172–178.

Doehrmann, O., & Naumer, M. J. (2008). Semantics and the multisensory brain: How meaning modulates processes of audio-visual integration. *Brain Research, 1242,* 136–150.

Driver, J., & Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on "sensory-specific" brain regions, neural responses, and judgments. *Neuron, 57,* 11–23.

Engel, L. R., Frum, C., Puce, A., Walker, N. A., & Lewis, J. W. (2009). Different categories of living and non-living sound-sources activate distinct cortical networks. *NeuroImage, 47*(4), 1778–1791.

Eramudugolla, R., Henderson, R., & Mattingley, J. B. (2011). Effects of audio-visual integration on the detection of masked speech and non-speech sounds. *Brain and Cognition, 75,* 60–66.

Frassinetti, F., Bolognini, N., & Làdavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research, 147,* 332–323.

Gazzola, V., Aziz-Zadeh, L., & Keysers, C. (2006). Empathy and the somatotopic auditory mirror system in humans. *Current Biology, 16,* 1824–1829.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Grant, K. W., & Seitz, P. F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *Journal of the Acoustical Society of America, 108*(3), 1197–1208.

Grossman, E. D., & Blake, R. (2002). Brain areas active during visual perception of biological motion. *Neuron, 35,* 1167–1175.

Hietanen, J. K., Leppanen, J. M., Illi, M., & Surakka, V. (2004). Evidence for the integration of audiovisual emotional information at the perceptual level of processing. *European Journal of Cognitive Psychology, 16*(6), 769–790.

Hommel, B., Musseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences, 24*(5), 849–937.

Iordanescu, L., Guzman-Martinez, E., Grabowecky, M., & Suzuki, S. (2008). Characteristic sounds facilitate visual search. *Psychonomic Bulletin & Review, 15*(3), 548–554.

Jack, C. E., & Thurlow, W. R. (1973). Effects of degree of visual association and angle of displacement on the "ventriloquism" effect. *Perceptual and Motor Skills, 37,* 967–979.

Jacobs, A., Pinto, J., & Shiffrar, M. (2004). Experience, context, and the visual perception of human movement. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 822–835.

James, T. W., VanDerKlok, R. M., Stevenson, R. A., & James, K. H. (2011). Multisensory perception of action in posterior temporal and parietal cortices. *Neuropsychologia, 49,* 108–114.

Kaplan, J. T., & Iacoboni, M. (2007). Multimodal action representation in human left ventral premotor cortex. *Cognitive Processes, 8,* 103–113.

Kohler, E., Keysers, C., Umiltà, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science, 297,* 846–848.

Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., & Wallace, M.T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Experimental Brain Research, 158,* 405–414.

Loula, F., Prasad, S., Harber, K., & Shiffrar, M. (2005). Recognizing people from their movement. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 210–220.

McMillan, N., & Creelman, C. (1991). *Detection theory: A user's guide*. Cambridge, UK: Cambridge University Press.

Mendonca, C., Santos, J. A., & Lopez-Moliner, J. (2011). The benefit of multisensory integration with biological motion signals. *Experimental Brain Research, 213,* 185–192.

Miller, L. E., & Saygin, A. P. (2013). Individual differences in the perception of biological motion: Links to social cognition and motor imagery. *Cognition, 128,* 140–148.

Mitterer, H., & Jesse, A. (2010). Correlation versus causation in multisensory perception. *Psychonomic Bulletin & Review, 17*(3), 329–334.

Molholm, S., Ritter, W., Javitt, D. C., & Foxe, J. J. (2004). Multisensory visual–auditory object recognition in humans: A high-density electrical mapping study. *Cerebral Cortex, 14,* 452–465.

Morein-Zamir, S., Soto-Faraco, S., & Kingston, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research, 17,* 154–163.

Munhall, K. G., Gribble, P., Sacco, L., & Ward, M.

(1996). Temporal constraints on the McGurk effect. *Perception & Psychophysics, 58*(3), 351–362.

Noesselt, T., Fendrich, R., Bonath, B., Tyll, S., & Heinze, H.-J. (2005). Close in time when farther in space – Spatial factors in audiovisual temporal integration. *Cognitive Brain Research, 25,* 443–458.

Noesselt, T., Reiger, J. W., Schoenfeld, M. A., Kanowski, M., Hinrichs, H., Heinze, H.-J., & Driver, J. (2007). Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *The Journal of Neuroscience, 27*(42), 11431–11441.

Petrini, K., Dahl, S., Rocchesso, D., Waadeland, C. H., Avanzini, F., Puce, A., & Pollick, F. (2009). Multisensory integration of drumming actions: Musical experience affects perceived audiovisual asynchrony. *Experimental Brain Research, 198,* 339–352.

Petrini, K., Holt, S. P., & Pollick, F. (2010). Expertise with multisensory events eliminates the effect of biological motion rotation on audiovisual synchrony perception. *Journal of Vision, 10*(5):2, 1–14, http://www.journalofvision.org/content/10/5/2, doi:10.1167/10.5.2. [PubMed] [Article]

Petrini, K., McAleer, P., & Pollick, F. (2010). Audiovisual integration of emotion signals from music improvisation does not depend on temporal correspondence. *Brain Research, 1323,* 139–148.

Petrini, K., Russell, M., & Pollick, F. (2009). When knowing can replace seeing in audiovisual integration of actions. *Cognition, 111*(3), 432–439.

Pizzamiglio, L., Aprile, T., Spitoni, G., Pitzalis, S., Bates, E., D'Amico, S., & Di Russo, F. (2005). Separate neural systems for processing action- or non-action related sounds. *NeuroImage, 24,* 852–861.

Pourtois, G., & deGelder, B. (2002). Semantic factors influence multisensory pairing: a transcranial magnetic stimulation study. *NeuroReport, 13*(12), 1567–1573.

Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology, 9,* 129–154.

Radeau, M., & Bertelson, P. (1977). Adaptation to auditory-visual discordance and ventriloquism in semirealistic situations. *Perception & Psychophysics, 22*(2), 137–146.

Raposo, D., Sheppard, J. P., Schracter, P. R., & Churchland, A. K. (2012). Multisensory decision-making in rats and humans. *Journal of Neuroscience, 32*(11), 3726–3735.

Saarela, M. V., & Hari, R. (2008). Listening to humans walking together activates the social brain circuitry. *Social Neuroscience, 3*(3-4), 401–409.

Saygin, A. P. (2007). Superior temporal and premotor brain areas necessary for biological motion perception. *Brain, 130,* 2452–2461.

Saygin, A. P., Driver, J., & de Sa, V. R. (2008). In the footsteps of biological motion and multisensory perception: Judgments of audiovisual temporal relations are enhanced for upright walkers. *Psychological Science, 19*(5), 469–475.

Saygin, A. P., Wilson, S. M., Hagler, D. J., Bates, E., & Sereno, M. I. (2004). Point-light biological motion perception activates human premotor cortex. *Journal of Neuroscience, 24*(27), 6181–6188.

Schneider, T. R., Engel, A. K., & Debener, S. (2008). Multisensory identification of natural objects in a two-way crossmodal priming paradigm. *Experimental Psychology, 55*(2), 121–132.

Schutz, M., & Kubovy, M. (2009). Causality and cross-modal integration. *Journal of Experimental Psychology: Human Perception and Performance, 35*(6), 1791–1810.

Shiffrar, M., & Freyd, J. J. (1990). Apparent motion of the human body. *Psychological Science, 1,* 257–264.

Shiffrar, M., & Freyd, J. J. (1993). Timing and apparent motion path choice with human body photographs. *Psychological Science, 4,* 379–384.

Soto-Faraco, S., & Alsius, A. (2009). Deconstructing the McGurk-MacDonald Illusion. *Journal of Experimental Psychology: Human Perception and Performance, 35*(2), 580–587.

Spence, C. (2007). Audiovisual multisensory integration. *Acoustical Science and Technology, 28,* 61–70.

Spence, C., & Shankar, M. U. (2010). The influence of auditory cues on the perception of, and responses to, food and drink. *Journal of Sensory Studies, 25*(3), 406–430.

Spence, C., & Squire, S. (2008). Multisensory integration: Maintaining the perception of synchrony. *Current Biology, 13,* R519–R521.

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge, MA: MIT Press.

Stevens, J. A., Fonlupt, P., Shiffrar, M., & Decety, J. (2000). New aspects of motion perception: Selective neural encoding of apparent human movements. *Neuroreport, 11,* 109–115.

Thomas, J. P., & Shiffrar, M. (2010). I can see you better if I can hear you coming: Action-consistent sounds facilitate the visual detection of human gait. *Journal of Vision, 10*(12):14, 1–11, http://www.journalofvision.org/content/10/12/14, doi:10.1167/10.12.14. [PubMed] [Article]

Thomas, J. P., & Shiffrar, M. (2011). *Footstep sounds increase sensitivity to point-light walking when visual cues are weak*. Poster presented at the Vision Sciences Society Annual Meeting, Naples, FL.

van der Zwan, R., MacHatch, C., Kozlowski, D., Troje, N. F., Blanke, O., & Brooks, O. (2009). Gender bending: Auditory cues affect visual judgments of gender in biological motion displays. *Experimental Brain Research, 198,* 373–382.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech. *Neuropsychologia, 45,* 598–607.

Vatakis, A., Ghazanfar, A. A., & Spence, C. (2008). Facilitation of multisensory integration by the "unity effect" reveals that speech is special. *Journal of Vision, 8*(9):14, 1–11, http://www.journalofvision.org/content/8/9/14, doi:10.1167/8.9.14. [PubMed] [Article]

Vatakis, A., & Spence, C. (2007). Crossmodal binding: Evaluating the "unity assumption" using audiovisual speech stimuli. *Perception & Psychophysics, 69*(5), 744–756.

Vatakis, A., & Spence, C. (2008). Evaluating the influence of the "unity assumption" on the temporal perception of realistic audiovisual stimuli. *Acta Psychologica, 127,* 12–23.

Vroomen, J., & deGelder, B. (2000). Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Experimental Psychology: Human Perception and Performance, 26,* 1583–1590.

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: A tutorial view. *Attention, Perception, & Psychophysics, 72*(4), 871–884.

Welch, R. B. (1999). Meaning, attention, and the "unity assumption" in the intersensory bias of spatial and temporal perceptions. In G. Aschersleben, T. Bachmann, & J. Musseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 363–369). Amsterdam, The Netherlands: Elsevier Science, BV.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 88,* 638–667.

Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin, 133*(3), 460–473.