# Meaningful auditory information enhances perception of visual biological motion

**Roberto Arrighi**

Istituto Nazionale di Ottica Applicata (INOA), Firenze, Italy, & Facoltà di Psicologia, Università degli Studi di Firenze, Firenze, Italy    ✉

**Francesco Marini**

Facoltà di Psicologia, Università degli Studi di Firenze, Firenze, Italy    ✉

**David Burr**

Facoltà di Psicologia, Università degli Studi di Firenze, Firenze, Italy, & Istituto di Neuroscienze del CNR, Pisa, Italy    ✉

Robust perception requires efficient integration of information from our various senses. Much recent electrophysiology points to neural areas responsive to multisensory stimulation, particularly audiovisual stimulation. However, psychophysical evidence for functional integration of audiovisual motion has been ambiguous. In this study we measure perception of an audiovisual form of biological motion, tap dancing. The results show that the audio tap information interacts with visual motion information, but only when in synchrony, demonstrating a functional combination of audiovisual information in a natural task. The advantage of multimodal combination was better than the optimal maximum likelihood prediction.

Keywords: biological motion, audiovisual integration, cross-modal perception

## Introduction

Robust perception requires integration between various senses (Ernst & Bülthoff, 2004; Stein & Meredith, 1993), and much physiological evidence points to the existence of neurons that respond to stimuli in more than one modality. For example Meredith and Stein (1986, 1996) and Stein, Meredith, and Wallace (1993) found in cat superior colliculus multimodal neurons that respond to visual as well as auditory stimuli and integrate (non-linearly) signals of these two modalities. Indeed they claim that many neurons with receptive fields spatially aligned across modalities show a super-additive response to coincident and colocalized multimodal stimulation while they show a reduction in activity, or no change, for spatially and temporally incoherent stimuli. Other studies revealed neurons with similar functional properties in the anterior ectosylvian sulcus AES (Wallace, Meredith, & Stein, 1992).

Recent physiological studies on monkey have revealed neurons in the ventral premotor cortex (area F5) that show the characteristics of *mirror neurons* along with specific sensitivity to audiovisual stimuli. Mirror neurons respond both when a monkey observes an action as well as when he makes a similar action (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Rizzolatti, Fadiga, Gallese, & Fogassi,

1996). In addition, audiovisual mirror neurons respond even when only the sound of an action is perceived, without any significant activation for auditory stimuli not related to action (Kohler et al., 2002). Auditory and visual signals interact in defining the response profile of these neurons. For example, around one third of them show strongest responses when visual and auditory patterns of an action they are sensitive to are presented together relative to unimodal presentations (Keysers et al., 2003). These neurophysiological results reinforced behavioral studies in cat (Stein, Huneycutt, & Meredith, 1988) and also humans (Alais & Burr, 2004b; Ernst & Banks, 2002; Frassinetti, Bolognini, & Làdavas, 2002), indicating that mammalian perceptual systems are indeed capable of exploiting multiple sensory stimulation to facilitate the detection of external signals.

One area of interest is auditory visual integration of moving stimuli. Neurophysiological data show that auditory and visual motion are integrated in cat superior colliculus neurons (Wallace & Stein, 1997). Human fMRI data show that many cortical areas are activated by both visual and auditory moving stimuli. For example, Howard et al. (1996) showed that three different kinds of visual motion (optic flow, coherent and biological motion) yield foci of activation within stimulus-specific regions of the left superior temporal gyrus (STG), regions that they were able to localize unambiguously in the auditory cortex.

Lewis, Beauchamp, and De Yoe (2000) investigated with fMRI cross-modal perception in a speed comparison task and found neurons that respond to stimuli of both modalities in lateral parietal cortex, lateral frontal cortex, anterior midline as well as anterior insular cortex.

Surprisingly, the psychophysical results are less clear. Both Alais and Burr (2004a) and Wuerger, Hofbauer, and Meyer (2003) reported facilitation for audiovisual motion perception, but the effects were small and, importantly, unspecific for direction: rightward auditory motion facilitated leftward visual motion as much as did leftward auditory motion. This small audiovisual advantage was consistent with a statistical combination of information, rather than of a functional neural integration of visual and auditory motion. On the other hand Meyer, Wuerger, Röhrbein, and Zetzsche (2005) reported that audiovisual integration was stronger and direction specific when visual stimuli comprised spatially curtailed objects in motion (rather than whole-field motion) and auditory stimuli a physical signal source providing high-quality localization. Recently, Brooks et al. (2007) demonstrated audiovisual integration for point-source biological motion (Johansson, 1973), showing that reaction times for detecting visual biological motion embedded in noise was enhanced by the presence of congruent auditory "motion" (advancing or retreating steps). In this study we pursue further audiovisual integration for a form of biological motion where both sight and sound provide useful information: tap dancing. We show that with these natural-like stimuli, integration does occur, at a level greater than predicted by mere statistical advantage.

# Methods

## Subjects

Two of the authors (RA and FM) and one naive female (mean age 26 years), all with normal hearing and normal or corrected visual acuity, served as subjects. All gave informed consent to participate in the study that was conducted in accordance with the guidelines of the University of Florence. The tasks were performed in a dimly lit, sound attenuated room.

## Apparatus and stimuli
### Stimuli

Several movies of a professional tap dancer performing standard routines were recorded with a tripod-mounted Sony DCR-HC19E digital camera equipped with an Advanced HAD CCD 800,000 pixel sensor, 25 Hz, in a rehearsal room of the dance school "Dance Connection" of Signa (Florence). The dancer wore dark clothes with 6 white plastic markers (3 markers per foot: ankle, heel, and tiptoe) applied on the dark tap dancing shoes. We recorded a total of 600 s, then selected three sequences of 3-s duration (see example in Movie 1).

Video tracks from the selected clips were imported to the computer and converted into sequences of bitmap images and subsequently imported into a Matlab script. The script automatically identified frame by frame the position of the six foot markers (allowing for manual correction of any errors by mean of a graphic interface). The markers were replaced with 1° diameter white disks, creating the light-point tap dancing movies (Movie 2). The auditory soundtrack from the original movies was also imported into a Matlab script and edited to optimize the signal level (of the taps) by filtering out background noise. For the second experiment (summation), auditory stimuli were then standardized by substituting the taps with a single template tap sound (Movie 3).

## Experimental procedures

In the experimental sessions all visual stimuli were displayed with a Cambridge VSG 2/5 framestore on a Sony Trinitron CRT monitor with a resolution of $800 \times 600$ pixels, 100 Hz, and mean luminance of 25 cd/m$^2$. The video screen subtended $90 \times 57°$ from the subjects viewing distance of 57 cm. Light-point tap dance stimuli consisted of six black disks of 1° diameter displayed on a mid gray rectangle, the same size as the original movie frames, $320 \times 240$ pixels ($9.4° \times 7°$), centered within the monitor screen. Noise motion sequences were created from the point-light tap sequences, by sampling short sequences of 75 frames of a single dot, and displaying it in random positions and orientations. Thus the noise had very similar spatial and temporal characteristics as the tap dance signal and was very effective in reducing its visibility (see, for example, Movie 4).

Auditory stimuli were digitized at 80 kHz and presented through two high-quality loud speakers (Yamaha MSP5) flanking the computer screen and lying in the same plane 60 cm from the subject. Speaker separation was 90 cm, and stimuli intensity was 85 dB at the sound source. Auditory noise consisted of auditory templates randomly scattered along the 3-s duration to yield a sound without any predefined beat (example in Movie 5).

We measured sensitivity for detecting tap dance sequences under various conditions. In all cases subjects were presented with two sequences, either visual or auditory or both, each of 3-s duration. One sequence comprised only noise, the other the point-light tap dance sequence embedded in noise (with total dots matched). Subjects were required to identify in 2AFC which sequence contained the tap dance sequence (no feedback was given). The amount of noise varied from trial to trial, following the adaptive QUEST procedure (Watson & Pelli, 1983). Psychophysical functions were calculated and fitted with a raised cumulative Gaussian curve (with

asymptotes at 0.5 and 1), to yield an estimate of sensitivity (defined as 75% correct response), with standard error (calculated by bootstrap). A total of 500 trials per condition was collected for each subject.

# Results

## Experiment 1: Facilitation

We measured the interaction of visual and auditory information in two ways: facilitation and summation. The facilitation experiment was designed to see if auditory signals, not informative on their own, could increase visual performance. There were three separate experimental conditions:

1. visual stimuli displayed with no sound (2AFC with one interval containing the tap dance sequence);
2. the visual sequence with the taps of the dance sequence added to *both* sequences (so as to be uninformative on its own);
3. the visual sequence with the auditory soundtrack added out of synchrony, again presented in both sequences.

The results of observer FM are shown in Figure 1. The psychometric function for the purely visual condition (blue lines) virtually overlaps that for the condition with desynchronized sound, showing that desynchronized auditory information does not affect visual performance.
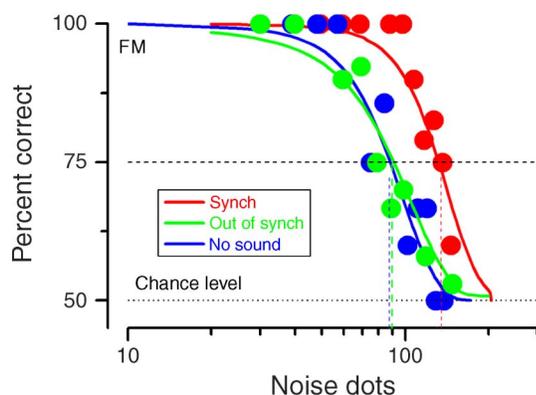
Figure 1. Audio facilitation of visual sensitivity to perceive tap dance motion. Data show the percentage of correct responses as a function of noise level (number of noise dots added to the stimulus). Blue, red, and green symbols indicate, respectively, experimental conditions of vision alone, audio soundtrack added in synchrony, and audio soundtrack added out of synchrony. The continuous lines are the best-fitting cumulative Gaussian curves to the data (50 trials/data point). Sensitivity, defined as 75% of correct response, was found to be highest when auditory information were in phase relative to the visual tap sequences.
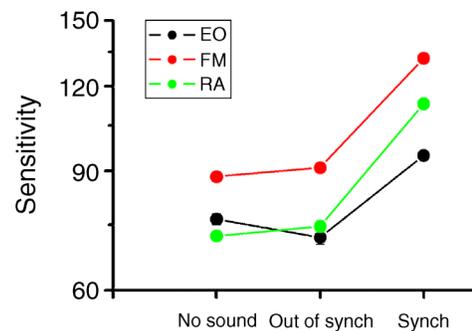
Figure 2. Sensitivity for perceiving point-light tap dancers embedded in noise for three different conditions: vision only, desynchronized audio, and synchronized audio. Only when auditory tap sounds were synchronized to visual streams was there an improvement in sensitivity.

However, when the visual and auditory information were synchronized, there was a small but consistent improvement, suggesting that the perceptual system could use coincident auditory information to help disentangle the visual stimulus from the noise.

Figure 2 plots sensitivities for the three subjects in each condition. For all subjects sensitivity is highest when the auditory information was present and in phase with the visual sequence. The increase in sensitivity was on average of factor of 1.4.

## Experiment 2: Summation

The previous experiment showed that auditory information, not informative on its own, can facilitate visual discrimination of tap dancing. This second experiment was designed to investigate summation between auditory and visual signals when both are informative, and at threshold. In order to do this, we first measured discrimination thresholds for visual and auditory signals presented on their own. The visual condition was the same as that described for the facilitation experiment, with no sound. The auditory version was similar, in that two sequences were presented, one comprising the soundtrack of the tap dance (with standardized taps) with added noise taps inserted at random times, the other comprising only noise taps.

Figure 3 shows psychometric functions for subject FM for the auditory and visual conditions. There was a large difference (around a factor of 10) in absolute sensitivity between auditory and visual thresholds, but the slope of the two functions were very similar for all three subjects. For the summation task the visual and auditory stimuli were normalized to be equally detectable. In the summation conditions, the two presentations were bimodal: one sequence comprised both the visual and the auditory tap dance sequences, and the other only noise (both visual and auditory). In one condition the visual and auditory signals
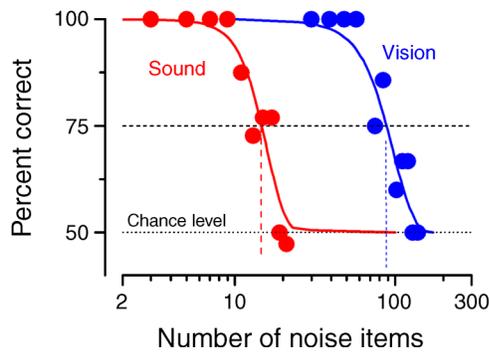
Figure 3. Sensitivity (of subject FM) for detecting the interval containing the tap dance signal, either visually (blue symbols) or auditorially (red symbols). The psychometric functions are similar in shape and width, but the auditory functions are about 1 log unit less sensitive than the visual ones.

were synchronized, in another they were presented out of phase. In every trial, the strength of auditory and visual noise was yoked together, with $X$ times more noise dots than noise taps, where $X$ is the ratio of visual to auditory sensitivity for that subject (about 10 for all observers).

Sample results from subject RA are shown in Figure 4. Both bimodal summation conditions produce an increment of sensitivity relative to the unimodal values, indicating that the perceptual performance was improved by the additional information from the second sensory channel. However, the increase in sensitivity increment was greater when the auditory and visual stimuli were in phase than when they were out of phase. Figure 5 shows the results for all subjects.

On average the in-phase condition yielded a relative sensitivity improvement of about 1.6 compared with 1.4 of the out-of-phase condition. Although this difference is
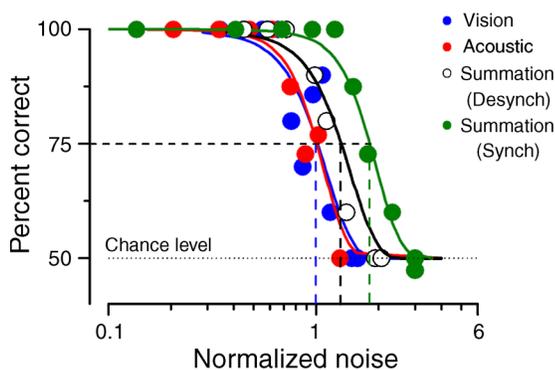


Figure 4. Normalized sensitivities (relative to the unimodal data) of subject RA for visual (blue symbols) and auditory (red symbols) conditions, and for the bimodal conditions with tap sounds in phase (green symbols) or out of phase (black open symbols). An increment in sensitivity was observed in both bimodal conditions (as shown by the rightward translation of the black and green psychometric functions), but it was actually found to be highest when stimuli of the two sensory modalities were synchronized.
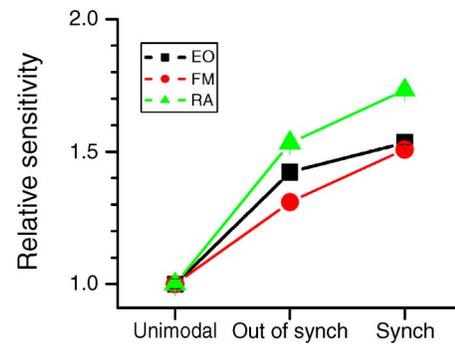


Figure 5. Normalized sensitivities (relative to the unimodal data) for all three subjects. On average the relative sensitivity increment for the out-of-phase condition was around 1.4 while it was 1.6 when auditory and visual signals were synchronized.

not great, it was statistically significant for all subjects to a bootstrap paired $t$-test (one-tail, 10000 repetitions, $p < 0.05$, $p < 0.01$, $p < 0.01$ for EO, FM, and RA, respectively).

Figure 6 plots the average thresholds as a two-dimensional representation, auditory against visual thresholds along with the theoretical threshold predictions derived by some summation models. The optimal statistical integration of auditory and visual signals provided by the Bayesian fusion model correctly accounts for results obtained when sounds were out of phase relative to visual streams. However, the further improvement achieved when auditory and visual information were synchronized could not be explained in statistical terms, thus suggesting that for temporally matched stimuli some kind of physiological summation actually occurs.
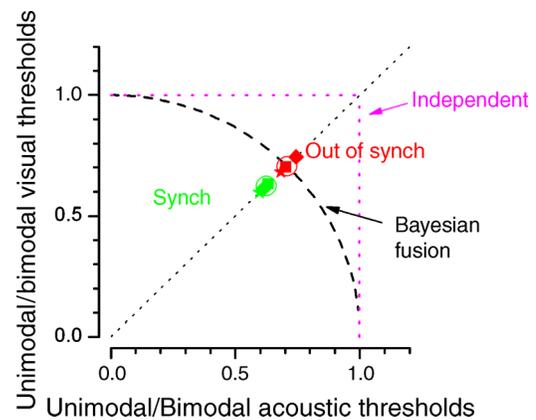


Figure 6. Results of the summation study, together with theoretical predictions. Single subject data (filled symbols) as well as averages (open symbols) are shown for both out-of-synchrony and synchronized experimental conditions. Out-of-synchrony auditory visual summation (red symbols) fits well with predictions obtained by a Bayesian integration (Equation 1). However, no simple statistical model can account for the results of the in-phase experimental condition (green symbols) suggesting that physio-logical mechanisms integrate the visual and the auditory signals.

# Discussion

The primary aim of these experiments was to investigate if auditory and visual signals integrate in the perception of a form of audiovisual "biological motion": tap dancing. The first experiment showed that a non-informative audio sequence can facilitate the recognition of visual tap dance sequences, provided that the taps are in synchrony with the visual motion. There was no facilitation when the visual and auditory sequences were out of synchrony. Possibly, the simplest mechanism to account for this facilitation could be reduction of temporal uncertainty. Although they were uninformative in their own right, the synchronized auditory signals could have served as temporal references, thereby narrowing the uncertainty window about the timing of the visual signals (and helping to ignore the noise signals). As the tap auditory sequences used in this experiment were the original sequences (not substituted with a template), they contained many distinct features with high salience (for example, shuffles, pullbacks, and toe punches) that could provide further information. This process, maybe mediated by some attentional mechanisms, was so effective that on some trials subjects reported the dancing dots to be segregated from the noise, a sort of "pop-out" phenomenon. This illusion is similar to that described by Van der Burg, Olivers, Bronkhorst, and Theeuwes (2008), where temporally related (but spatially uncorrelated) auditory information is provided to subjects performing a visual search task. This confirms that synchrony between auditory and visual events cannot only improve visual performance (Dalton & Spence, 2007; Vroomen & de Gelder, 2000) but also affect visual perception phenomenologically. As the reduction of temporal uncertainty did not occur when sounds were out of synchrony relative to the visual patterns it is not surprising that in the out-of-synchrony condition no facilitation was observed and performance was identical to those obtained when any sounds were provided at all.

The second experiment provided stronger evidence of meaningful integration of auditory and visual information. When stimuli were presented with both visual and auditory signals, both providing information about which interval contained the dance sequence, there was an improvement in performance, both when the two senses were in phase and when they were out of phase. This is to be expected. Even when out of phase, both visual and auditory signals provide information, so thresholds should decrease if the perception system is capable of taking advantage of the separate information. As the visual and auditory thresholds were matched for their unimodal thresholds, the expected increase from the maximum likelihood model (Equation 1) is root two. Indeed, the average increase was 1.4:

$$\frac{1}{\sigma_{AV}^2} = \frac{1}{\sigma_A^2} + \frac{1}{\sigma_V^2}. \qquad (1)$$

While the $\sqrt{2}$ sensitivity improvement in the out-of-phase experimental condition is what would be expected by an optimum statistical integration, as has been shown in many other conditions (Alais & Burr, 2004b; Ernst & Banks, 2002), the further improvement found in the in-phase condition cannot be accounted for by statistical combination. This additional improvement in sensitivity is evidence for a more functional combination of visual and auditory information, in order to detect efficiently audio-visual events.

Synchrony of auditory and visual signals appears to play a key role in producing cross-modal facilitative effects as well as audiovisual integration. However, it raises a potential concern that subjects may have taken advantage of the synchrony and used the overall audio-visual synchrony (higher for the in-phase presentations) in performing the task. However, it is unlikely that subjects were able to use this cue, as there were typically about 20 times more noise dots than signal dots where visuo-auditory correlations occurred frequently (especially as perceptual synchrony occurs within a 50-ms time inter-val). Certainly, no subject ever reported being aware of temporal correlations.

To what extent are the effects reported in this paper specific to biological motion? At this stage we cannot say. Certainly some studies have failed to show integration of visual and auditory motion (Alais & Burr, 2001; Wuerger et al., 2003), but others have shown direction-specific motion (Meyer et al., 2005). At this stage we cannot determine how important biological motion is for the integration. In the past we have shown that perceived temporal synchrony of audiovisual stimuli does not depend on them being "natural", with artificial stimuli producing similar results to natural stimuli (Arrighi, Alais, & Burr, 2006).

While it is always difficult to infer the neural substrates involved in audiovisual cross-modal integration from psychophysical data, it is nevertheless interesting to speculate. In theory, both sub-cortical as well as cortical areas could be involved. Most neurons of the superior colliculus (SC), an area of the mesencephalon, respond to stimuli of more than one sensory modality and the responses to signals of one modality are reciprocally affected by inputs of other modalities (Stein & Meredith, 1993; Wallace, Meredith, & Stein, 1993). Interestingly, these multimodal neurons are sensitive to temporal coincidence of incoming signals and show neural response enhancement only when visual and auditory stimuli occur at similar times (agreeing with our evidence for strong audiovisual integration only for synchronous signals). However, SC does not seem to be directly involved with the perception of biological motion but could be instrumental in the integration of auditory and visual signals prior to analysis of biological motion, although there is certainly no evidence for this at this stage.

At the cortical level two areas seem particularly interesting: pSTS and IPS. Both have been implicated in

the analysis of biological motion and show interesting multisensory properties (Bonda, Petrides, Ostry, & Evans, 1996; Brooks et al., 2007; Bruce, Desimone, & Gross, 1981; Grossman & Blake, 2002; Howard et al., 1996; Servos, Osu, Santi, & Kawato, 2002; Vaina, Solomon, Chowdhury, Sinha, & Belliveau, 2001). STS shows particularly interesting complex audiovisual interactions: for example it responds to a stimulus such as an object striking a surface, but neither the image nor the sound of the event alone will elicit a response; nor will the simultaneous presentation of simple stimuli like a flash and click (Bruce et al., 1981). These kinds of patterns—purposeful audiovisual stimuli—seem to have similar characteristics to the tap dance routines. However, the cortical structures most likely to be implicated in the recognition of tap dance patterns are those population of cross-modal neurons specifically tuned for perception (and/or production) of specific action patterns: audiovisual mirror neurons. As mentioned in the Introduction section, mirror neurons sensitive to visual as well as auditory patterns of a specific action have been localized in the rostral ventral premotor cortex of monkeys (Keysers et al., 2003; Kohler et al., 2002). These neurons seem to be implicated in the recognition and interpretation of actions by matching the auditory and visual patterns of external actions onto subject's internal motor coordinates. The response profiles of these neurons are so strictly tuned to specific action patterns that from their firing rate it is possible to discriminate among different actions with an error rate lower than 5% (Keysers et al., 2003). Together, all these evidences support the hypothesis that audiovisual mirror neurons play a central role in recognition of actions implicating visual and auditory patterns, such as tap dance. As evidence about the existence of cross-modal mirror neurons have been collected also in humans (Gazzola, Aziz-Zadeh, & Keysers, 2006) these cortical mechanisms could underlie the integration of auditory and visual information in the recognition of human actions, and thus play a key role in the perception of audiovisual motion profiles as tap dance routines. However, whatever the neural mechanisms involved, these results provide further demonstration of the flexibility and versatility of the human perceptual system in optimizing its performance.

## Acknowledgments

Commercial relationships: none.
Corresponding author: Roberto Arrighi.
Email: roberto.arrighi@gmail.com.
Address: Istituto Nazionale di Ottica Applicata (INOA), Firenze, Largo E. Fermi 6, Italy.

## References

Alais, D., & Burr, D. (2004a). No direction-specific bimodal facilitation for audiovisual motion detection. *19,* 185–194. [PubMed]

Alais, D., & Burr, D. (2004b). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14,* 257–262. [PubMed] [Article]

Alais, D., & Burr, D. C. (2001). No low-level interaction between visual and auditory motion. *Perception, 30,* 35 (Suppl.).

Arrighi, R., Alais, D., & Burr, D. (2006). Perceptual synchrony of audiovisual streams for natural and artificial motion sequences. *Journal of Vision, 6*(3):6, 260–268, http://journalofvision.org/6/3/6/, doi:10.1167/6.3.6. [PubMed] [Article]

Bonda, E., Petrides, M., Ostry, D., & Evans, A. (1996). Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *Journal of Neuroscience, 16,* 3737–3744. [PubMed] [Article]

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. [PubMed]

Bruce, C., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *Journal of Neurophysiology, 46,* 369–384. [PubMed]

Dalton, P., & Spence, C. (2007). Attentional capture in serial audiovisual search tasks. *Perception & Psychophysics, 69,* 422–438. [PubMed]

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415,* 429–433. [PubMed]

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends Cognitive Science, 8,* 162–169. [PubMed]

Frassinetti, F., Bolognini, N., & Làdavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research, 147,* 332–343. [PubMed] [Article]

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119,* 593–609. [PubMed] [Article]

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

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

Howard, R. J., Brammer, M., Wright, I., Woodruff, P. W., Bullmore, E. T., & Zeki, S. (1996). A direct demonstration of functional specialization within motion-related visual and auditory cortex of the human brain. *Current Biology, 6,* 1015–1019. [PubMed] [Article]

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14,* 201–211.

Keysers, C., Kohler, E., Umiltà, M. A., Nanetti, L., Fogassi, L., & Gallese, V. (2003). Audiovisual mirror neurons and action recognition. *Experimental Brain Research, 153,* 628–636. [PubMed] [Article]

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. [PubMed]

Lewis, J. W., Beauchamp, M. S., & DeYoe, E. A. (2000). A comparison of visual and auditory motion processing in human cerebral cortex. *Cerebral Cortex, 10,* 873–888. [PubMed] [Article]

Meredith, M. A., & Stein, B. E. (1986). Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Research, 365,* 350–354. [PubMed]

Meredith, M. A., & Stein, B. E. (1996). Spatial determinants of multisensory integration in cat superior colliculus neurons. *Journal of Neurophysiology, 75,* 1843–1857. [PubMed]

Meyer, G. F., Wuerger, S. M., Röhrbein, F., & Zetzsche, C. (2005). Low-level integration of auditory and visual motion signals requires spatial co-localisation. *Experimental Brain Research, 166,* 538–547. [PubMed] [Article]

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Brain Research: Cognitive Brain Research, 3,* 131–141. [PubMed]

Servos, P., Osu, R., Santi, A., & Kawato, M. (2002). The neural substrates of biological motion perception: An fMRI study. *Cerebral Cortex, 12,* 772–782. [PubMed] [Article]

Stein, B. E., Huneycutt, W. S., & Meredith, M. A. (1988). Neurons and behavior: The same rules of multisensory integration apply. *Brain Research, 448,* 355–358. [PubMed]

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

Stein, B. E., Meredith, M. A., & Wallace, M. T. (1993). The visually responsive neuron and beyond: Multisensory integration in cat and monkey. *Progress in Brain Research, 95,* 79–90. [PubMed]

Vaina, L. M., Solomon, J., Chowdhury, S., Sinha, P., & Belliveau, J. W. (2001). Functional neuroanatomy of biological motion perception in humans. *Proceedings of the National Academy of Sciences of the United States of America, 98,* 11656–11661. [PubMed] [Article]

Van der Burg, E., Olivers, C. N., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: Nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 1053–1065. [PubMed]

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

Wallace, M. T., Meredith, M. A., & Stein, B. E. (1992). Integration of multiple sensory modalities in cat cortex. *Experimental Brain Research, 91,* 484–488. [PubMed]

Wallace, M. T., Meredith, M. A., & Stein, B. E. (1993). Converging influences from visual, auditory, and somatosensory cortices onto output neurons of the superior colliculus. *Journal of Neurophysiology, 69,* 1797–1809. [PubMed]

Wallace, M. T., & Stein, B. E. (1997). Development of multisensory neurons and multisensory integration in cat superior colliculus. *Journal of Neuroscience, 17,* 2429–2444. [PubMed] [Article]

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics, 33,* 113–120. [PubMed]

Wuerger, S. M., Hofbauer, M., & Meyer, G. F. (2003). The integration of auditory and visual motion signals at threshold. *Perception & Psychophysics, 65,* 1188–1196. [PubMed]