

Vision affects how fast we hear sounds move

Joan López-Moliner

Grup d'Atenció, Acció i Percepció, GRNC, Parc Científic de Barcelona, Universitat de Barcelona, Catalonia, Spain



Salvador Soto-Faraco

Grup d'Atenció, Acció i Percepció, GRNC, ICREA & Parc Científic de Barcelona, Universitat de Barcelona, Catalonia, Spain



There is a growing body of knowledge about the behavioral and neural correlates of cross-modal interactions in the perception of motion direction, as well as about the computations that underlie unimodal visual speed processing. Yet, the multisensory contributions to the perception of motion speed remain largely uncharted. Here we show that visual motion information exerts a profound influence on the perception of auditory speed. Moreover, our results suggest that this influence is specifically caused by visual velocity rather than by earlier, more local, frequency-based components of visual motion. The way in which visual speed information affects how fast we hear a sound move can be well described by a weighted average model that takes into account the visual speed signal in the computation of auditory speed.

Keywords: audiovisual motion, speed discrimination, multisensory integration, weighted average

Citation: López-Moliner, J., & Soto-Faraco, S. (2007). Vision affects how fast we hear sounds move. *Journal of Vision*, 7(12):6, 1–7, <http://journalofvision.org/7/12/6/>, doi:10.1167/7.12.6.

Introduction

Estimating the velocity at which external objects move is essential for survival in most animal species, as it underlies vital abilities such as catching preys or avoiding collisions. A significant number of studies have been devoted to understand how the visual system processes speed information, revealing several important properties of these computations (e.g., Perrone & Thiele, 2001). However, current knowledge is mostly circumscribed to how the brain extracts velocity information from single sensory modalities, mostly vision, largely ignoring the potential contribution of other sensory channels carrying speed information about distal stimuli. Whether, and eventually how, the different sensory sources of velocity information influence each other during the perception of speed is therefore a question that has yet to be systematically addressed (but see Manabe & Riquimaroux, 2000).

In everyday life situations, the diversity of sensory signals originating from a single object are usually highly correlated (e.g., an approaching car produces an expanding image in the retina as well as a raising sound at the ear). Generally, these intersensory correlations are exploited by the brain to create robust representations of the environment, especially under impoverished input conditions (e.g., Ernst & Banks, 2002; Stein & Meredith, 1993). Motion is no exception, and several studies have already revealed strong mutual influences between vision and audition in the perception of motion direction. On the one hand, detection performance can improve when auditory and visual motion signals are available together.

For example, Wuerger, Hofbauer, and Meyer (2003) reported an improvement of motion detection for bimodal stimuli that could be predicted by a probability summation model and interpreted this result as the visual and the auditory signals being integrated at a decision stage. Alais and Burr (2004) also provided evidence for a decreased motion detection threshold when sound and vision were shown together that was consistent with both probabilistic summation and maximum likelihood. These two studies therefore support the general view that estimating a physical attribute (e.g., the speed of an audiovisual object) improves when more than one cue to that attribute is available. Although there is some debate about the processing level at which audiovisual interactions occur (see Soto-Faraco, Kingstone, & Spence, 2003), the view of an audiovisual integration mechanism operating after the stimuli have been processed in unimodal pathways has been favored by some authors (Alais & Burr, 2004; Burr & Alais, 2006) against earlier interactions such as those predicted by linear summation models.

Another successful approach to study multisensory influences on motion perception (and other attributes) has been intersensory conflict, addressing whether information incongruence in one sensory modality can exert an influence on the perception of motion in another modality. For example, vision can strongly affect the perception of auditory direction (e.g., Mays & Schirillo, 2005; Soto-Faraco, Spence, & Kingstone, 2004; for a review, see Soto-Faraco et al., 2003). Likewise, supra-threshold auditory motion can bias the perception of near-threshold visual motion direction in a way that is consistent with the direction of the auditory motion

(Meyer & Wuerger, 2001). Here, we precisely focus on this approach, namely, studying the effects of one modality (vision) on the other (audition), to address the nature and level of processing regarding multisensory integration of motion speed. It is worth noting that this approach deviates from the question of how estimating the speed of an audiovisual object improves by combining multisensory cues and focuses on gaining a better understanding of where and how auditory and visual information interact to compute speed.

The rationale of the present experiment relies on how visual motion can be decomposed. In particular, visual research has successfully addressed velocity processing by decomposing it into spatial frequency (SF) and temporal frequency (TF) (Watson & Ahumada, 1983), two frequency domains that can be conveniently separated using sinusoidal moving gratings. The velocity (v) of a grating is given by the ratio (TF/SF) between its TF (in Hz) and its SF (number of cycles per degree of visual angle). This spatiotemporal definition of stimulus space has been used to characterize the spectral receptive fields of neurons at various levels of the visual system in the monkey (e.g., Perrone & Thiele, 2001). For instance, many neurons in the middle temporal cortex (MT) encode velocity, as their preferred response stimuli lay on an elongated area oriented along an isovelocity line in the space defined by SF and TF. That is, these neurons are tuned to a given velocity and not to a particular value of TF or SF. Psychophysical evidence for a velocity-tuned mechanism has also been reported in humans (Reisbeck & Gegenfurtner, 1999). Unlike MT neurons, motion-sensitive neurons found in earlier stages of the visual system such as V1 do not show an invariant response across stimuli moving at the same velocity, but rather a TF response profile. This is regarded as evidence that V1 neurons are not tuned to speed but to local temporal frequencies.

Several plausible models have been put forward to explain how MT neurons integrate information from motion-related activity at lower levels of the visual system (such as V1) to compute speed (e.g., Perrone & Thiele, 2002; Priebe, Lisberger, & Movshon, 2006). However, the mechanisms underlying the combination of speed cues regarding a multisensory object remain largely unknown. Are speed signals conveyed by different sensory systems actually integrated and, if so, does this multisensory integration of speed depend on the local spatial structure of the moving stimuli? An affirmative answer could be regarded as evidence for very early audiovisual interactions, occurring before the velocity-tuned mechanisms in MT have combined signals motion detectors sensitive to the stimulus spatial structure (presumably separable mechanisms in V1). On the contrary, if the integration of audiovisual speed information depends on visual velocity independently of particular values of SF and TF, then one can conclude that these interactions take place later on, only after MT computations have been carried out.

Methods

Subjects

Four naive observers plus the two authors participated in this experiment ($N = 6$). All subjects reported normal hearing and normal or corrected-to-normal vision.

Stimuli and apparatus

The experiment was run in a sound-attenuated room with the walls padded with antireverberant foam. The auditory stimuli were generated online using a PRAAT script (<http://www.praat.org>) interfaced with a custom-written stimulus presentation program. The script created an amplitude-modulated (a major correlate of sound movement direction in the horizontal plane; Middlebrooks & Green, 1991) stereo file simulating a moving sound source. The trajectory of the sound was circumscribed between two loudspeaker cones located right at the sides of the computer monitor (subtending a separation of 40° azimuth). In the stereo sounds used to generate the acoustic motion, each channel contained a ramped 440-Hz tone of equivalent duration (variable, see [Methods](#)). The initial/final amplitude of the ramps determined the direction and the distance subtended by the sound in terms of azimuth. Because the loudspeakers were situated at 40° separation, when the initial–final amplitude of the ramped sounds was 0–100%, the distance traveled was effectively 40° . By varying the initial–final amplitudes of the ramps in a complementary way between channels/loudspeakers (i.e., 10–90%; 20–80%), we controlled the distance traveled (which was selected according to the duration of the sound, see below). Note that at any point in time, the total energy (summed amplitude at both channels) of the sound equalled 100%. Once the speed of the acoustic stimulus had been selected in a given trial (according to the QUEST procedure, see below), its duration was chosen pseudorandomly (between 0.6 and 5 s), and its spatial extent was then calculated appropriately as to range from 20° to 40° . This procedure was followed to prevent observers from focusing on spatial extent or duration rather than on speed in their responses (see, Carlile & Best, 2002).

Sinusoidal gratings were shown on a CRT monitor (Phillips 22 in. Brilliance 202P4, 1154×864 pixels) at 100 Hz in synchrony with the vertical frame rate in audiovisual and cross-modal conditions. Monitor output was linearized for each gun independently. The gratings subtended a horizontal angle of 21.3° and a vertical angle of 16.75° . Luminance-based gratings were set to a mean luminance of 8.5 cd/m^2 and 100% Michelson contrast. Isoluminant gratings were color-modulated along a red-green axis. The maximum stimulation along the L–M axis ranged between $(x, y, Y = 0.63, 0.34, 8.5)$ and $(x, y, Y = 0.31, 0.59, 8.5)$. Gratings could move rightward or leftward

always in agreement with the contingent sound. Three visual speeds were used in the audiovisual condition (15, 30, and 45 deg/s), each one resulting from two different combinations of TF and SF (15: 7.5/0.5, 3.75/0.25; 30: 9.90/0.33, 6.0/0.2; 45: 11.25/0.25, 7.50/0.167). The same spatial frequencies were used in the gratings of the cross-modal condition with a varying TF according to an adaptive staircase (see below). The perceived speed of a grating is known to depend on variations of SF of more than 2 octaves (Chen, Bedell, & Frishman, 1998, but see Smith & Edgar, 1991, for larger effects) and also on prior information that weights low speeds more than higher ones (Stocker & Simoncelli, 2006). The use of both low spatial frequencies and a small range (less than 2 octaves) would prevent visual velocity encoding from deteriorating at high speeds and eliciting high differences in the perceived visual speed. Furthermore, our speeds are far from the range where prior information is used in the perception of visual speed (Stocker & Simoncelli, 2006).

Procedure

Audiovisual conditions

Subjects sat 50 cm from the screen/sound center, with their head resting on a chin rest. They were shown an auditory standard stimulus moving at 30 deg/s, and after a blank/silent 2-s interval, a test auditory stimulus synchronized with a moving grating. Subjects were then asked to judge in which interval the sound moved faster. The speed of the test sound was controlled by a Quest staircase procedure (Watson & Pelli, 1983). Six different staircases were interleaved within a single session, with each staircase pairing the test sound with one of the six different moving gratings. Luminance-based and isoluminant gratings were shown in different sessions. The direction of motion (left vs. right; always the same for the sounds and the gratings) was chosen at random at every trial to prevent any adaptation effect. In the unimodal condition (auditory only), we followed the same procedure as described in the audiovisual conditions except that the test sound was presented in the absence of any visual grating. As described above, stimulus duration or spatial extent varied pseudorandomly as to prevent subjects from using either source of information as a cue to velocity. As a consequence, the faster auditory stimulus took actually longer to travel the corresponding distance in 32% of the trials. After conducting a binomial test, we found no evidence that subjects were relying on duration rather than speed in the audiovisual and the auditory conditions.

Cross-modal conditions

Subjects were shown the same auditory standard stimulus as in the audiovisual conditions, and after the 2-s blank interval, a sinusoidal grating moving in the same direction as the standard sound was silently shown for a randomly

chosen interval in the range [0.5–1.3] s. Direction of motion was randomized as above. Six different staircases (one per SF) were randomly interleaved within the session. Subjects had to judge whether the auditory or the visual stimulus moved faster. The speed of the grating within each staircase was controlled by the Quest procedure by adjusting the TF to be combined with the fixed SF.

Results

Separable versus nonseparable spatiotemporal mechanisms

In a first experiment, participants compared the speed of an auditory test stimulus accompanied by a task-irrelevant luminance-modulated visual grating (AV stimulus), with a reference sound moving at 30 deg/s (A stimulus). The task-irrelevant gratings were chosen from a set of six particular TF/SF combinations in the frequency space representing three different velocities ($v = \text{TF}/\text{SF}$: 15, 30, and 45 deg/s). The proportion of AV-faster responses conformed well to cumulative Gaussians for each of the six AV conditions separately. Figure 1A shows the points of subjective equality (PSEs; the mean of the curve, representing the speed of the test AV stimulus that is perceived as moving as fast as the reference A stimulus) as a function of the grating velocity and split by TF. The perceived sound speed was systemically overestimated when the irrelevant grating moved at 45 deg/s (33% overestimation) and at 30 deg/s (12%), but it was underestimated (–16%) when combined with the 15 deg/s gratings. Crucially, the two different PSE estimates from each speed (due to different TF) were not significantly different from each other (the 95% CI clearly overlapped). The same outcome was held when we split the data by SF (not shown in Figure 1A). That is, the pattern of perceived sound velocity of the AV stimulus was clearly modulated by the speed of the visual stimulus, irrespective of the particular values of TF and SF used. This result lends support to the idea that no audiovisual integration of motion occurs prior to visual speed integration mechanisms in MT and renders the view of a later integration more likely (e.g., Burr & Alais, 2006).

Psychophysical responses to luminance-based gratings are known to be optimal to tap at velocity-tuned mechanisms, whereas isoluminant gratings tap more efficiently at the separable spatiotemporal mechanisms (Reisbeck & Gegenfurtner, 1999). Therefore, one could argue that luminance-modulated gratings are not the optimal stimuli to test local effects and that the possibility of revealing audiovisual interactions prior to the stage where visual speed information is integrated remains open. Therefore, in a second experiment, we tested the influence of visual motion on acoustic speed perception using isoluminant

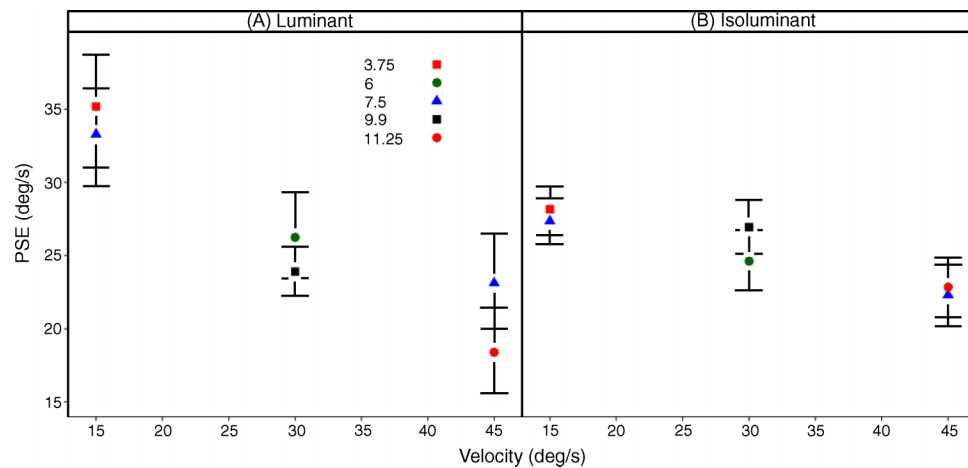


Figure 1. Point of subjective equality (PSE) obtained in the audiovisual condition for each grating velocity (15, 30, and 45 deg/s) and split by temporal frequency (TF) (see legend). Panels A and B show the data corresponding to the luminant and isoluminant gratings, respectively. Vertical bars for each PSE data point denote the 95% confidence intervals, obtained by conducting parametric bootstrap with 1000 simulations (Efron & Tibshirani, 1993).

gratings, a situation where the detection of local frequency characteristics is strengthened at the expense of the analysis of the visual speed profile, which is weakened. Figure 1B shows the pattern of perceived auditory velocities when sounds are combined with isoluminant gratings. Not surprisingly, isoluminant visual motion exerted a weaker effect on auditory speed perception, as reflected by the smaller differences of the PSEs across visual velocities. Yet, crucially, the basic pattern arising from the luminance-based gratings remained: The speed of auditory stimuli was perceived as equivalent when the accompanying gratings moved at the same speed, regardless of the particular values of TF and SF. In sum, all evidence converged to support the idea that the combination of auditory and visual speed occurs only after velocity information is computed from separable TF and SF detection mechanisms.

Visual effects on perceived auditory speed

Figure 2A shows the psychophysical curves obtained for the luminance-based gratings as a function of the velocity of the auditory test stimulus individually for each of the AV conditions, plus the curve corresponding to the auditory unimodal control condition (A, black squares), which provides a baseline measure of unimodal auditory speed discrimination. In a separate control experiment (cross-modal condition), we measured how fast these visual speeds of different SF are perceived when directly compared (in the absence of sound) to the auditory standard (results shown in Figure 3). For all the SF tested, the visual speed is overestimated with respect to the auditory speed (curves shifted to the left of the standard 30 deg/s). We also reproduce the results of the auditory unimodal condition in Figure 3 (dashed grey line) for the sake of comparison. One can see that the task for the

cross-modal condition is performed with less variability (average deviation around 6 deg/s) than that in the auditory unimodal one (deviation of 8 deg/s).

Taken together the results of Figures 2A and 3, it can be seen that the performance in the audiovisual condition does not follow the pattern that would be predicted by optimally integrating visual and auditory information. This is a logical consequence of the type of task used in the main experiment, which was designed to study the influence of one modality on the other. Indeed, because our participant's task was to report the auditory velocity rather than the audiovisual pattern itself, the predictions derived from using a combined percept cannot be adequately tested here. Yet, the apparent increase of deviation in the audiovisual condition remains to be explained.

Response bias as probability summation

It could be argued that the effects of visual motion on sound speed reported here are due to a response bias toward the task-irrelevant visual signal. In particular, the response bias would consist of an influence of the contingent visual speed when judgments about the auditory speed are made under a large uncertainty, but without influencing the actual perception of auditory velocity. If such a decision bias were true, the probability of responding that the audiovisual stimulus is faster than the auditory (standard) one could be modeled as a special case of probability summation:

$$P(\text{AV faster} | v_a, v_v) = u_a \cdot F_a(v_a, \mu_a, \sigma_a) + (1 - u_a) \cdot F_v(v_v, \mu_v, \sigma_v), \quad (1)$$

where v_a and v_v are the auditory and the visual speeds, respectively, F denotes the cumulative distribution function

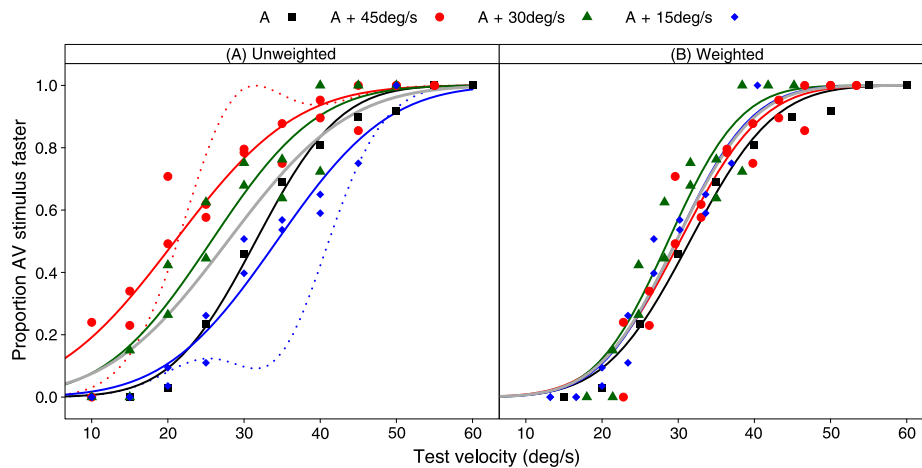


Figure 2. Psychometric curves. Proportion of faster responses for the test AV stimulus as a function of (A) the speed of the Auditory component alone (unweighted) and (B) the weighted combination ($w_a \times a + w_v \times v$) between the speed of the auditory a and the visual v components in Experiment 1. The weights (w) for which data points scatter less (according to the overall fitting squared error) are $w_a = 0.68$ and $w_v = 0.32$. A similar solution was found for the isoluminant gratings with weights $w_a = 0.89$ and $w_v = 0.11$. An animated version of the weighting process is available at <http://www.ub.edu/pbasic/visualperception/joan/demos.html>. Black squares represent the results of the unimodal condition (A alone) and the rest for the different AV conditions (see color key in the graph). Solid lines denote the best cumulative Gaussian fits to the psychophysical results. The solid gray curve is the best fit to the data points pooled over the AV conditions. The dotted lines represent the predicted pattern of a response bias model assuming a decision bias toward the visual speed when the auditory information has high uncertainty (see text for details; red: bias produced by 45 deg/s gratings; blue: bias produced by 15 deg/s).

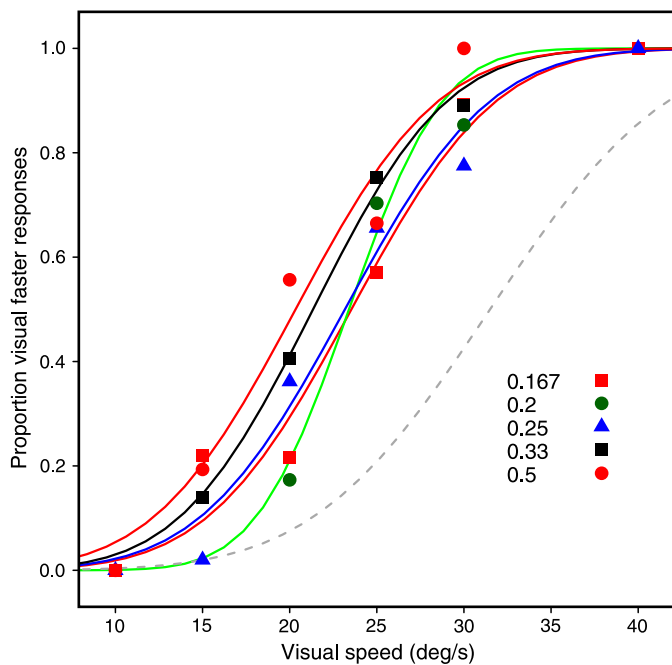


Figure 3. Proportion of responses in which a visual grating is perceived to move faster than a standard auditory stimulus (30 deg/s) in the cross-modal experiment. Results are split by spatial frequency (SF) of the gratings (which defined different staircases), represented by different symbols/colors (see legend). Solid lines are the best least mean square fits for each SF condition. The solid gray line corresponds to the model fit for the unimodal (auditory) control condition (reproduced from Figure 2 for ease of comparison).

for the auditory stimulus (F_a) and the visual one (F_v), and finally u_a is the weight given to the auditory stimulus, which would be modulated by the degree of uncertainty.

When the test auditory speed v_a is close to the standard stimulus (30 deg/s), the uncertainty is large and the auditory speed would be weakly weighted according to

$$u_a = 1 - e^{-(v_a - \mu_a)^2 / 2\sigma_a^2} \tag{2}$$

And, as a consequence, the response would then be mainly based on the visual speed. By using the means and the deviations fitted in the auditory control condition and in the cross-modal condition, we computed the result of Equation 1 for gratings moving at $v_v = 15$ and 45 deg/s and plotted the corresponding predictions (blue and red dotted lines) in Figure 2A.

The predictions obtained from the actual data in the AV conditions have a shallower slope (i.e., larger variability) than the curves obtained from the unimodal control condition and the “response bias” prediction. Therefore, this decision-level type of account can be ruled out. But, if subjects did not simply used the visual velocity and actually responded based on the auditory stimuli, then where did the increment in variability come from?

Auditory speed percept results from combined audiovisual information

The psychophysical curves corresponding to the AV conditions for the luminance-based motion condition have

been plotted (see [Figure 2A](#)) using the velocity auditory component of the test stimuli alone, as this was the modality the participants were asked to respond to. This stimulus space, however, neglects the fact that change in the auditory percept of speed might have been produced due to a visual modulatory effect, the hypothesis we entertain here. Indeed, one could adopt a model where the representation of the auditory speed information that subjects were using to respond was actually affected by the speed of the visual gratings. Evidence for this would be obtained if we were able to account for all the data points with a single psychophysical curve when combining the auditory and the visual speed of the test stimulus. Let's assume a simple combination model where the effect of the visual speed v simply adds to the auditory speed a and that the magnitude of the effect is denoted by w_v , resulting in a weighted average $w_a \times a + w_v \times v$, where $w_a + w_v = 1$. Then, one should be able to find values for w_a and w_v such that when plotting the best fitting curves for the AV conditions as a function of the weighted sum, all conditions should collapse into a single curve with a mean and a deviation (i.e., slope) similar to the curve corresponding to the unimodal condition. Indeed, we found values of w_a and w_v that conform to such an outcome (see [Figure 2B](#); the same procedure was applied to the results of the isoluminant condition with equivalent results). The mean and the slope of the new psychophysical curve (solid gray curve) were statistically equivalent to the mean and the slope of the unimodal condition.

Discussion

The findings arising from this study reveal, for the first time, that visual motion affects how fast we hear sounds move. This multisensory phenomenon implies that a visual moving object can strongly modulate the heard speed of a moving sound, thereby influencing the auditory perception of velocity. This influence can be well described by combining auditory and visual speeds before they reach a decision level. To this extent, our results are consistent with low-level integration of auditory and visual motion signals when they coincide in space and time (Meyer, Wuerger, Rohrbein, & Zetzsche, 2005). At a first glance, our findings might seem then to be at odds with previous studies favoring audiovisual maximum likelihood or probability summation (e.g., Burr & Alais, 2006). However, this is not so if one considers the task at hand. By asking subjects to focus on one of the two modalities instead of responding to the audiovisual object, one cannot simply test whether they are combining information optimally. The optimal integration (Ernst & Banks, 2002; Ernst & Bühlhoff, 2004) model explains successfully how participants improve an estimate by optimally combining two sources of information (where

each source is weighted as a function of its individual reliability). Because the perceptual nature of the present cross-modal effect was measured as a visual modulation on auditory processing, the magnitude of the visual influence on the perceived velocity of sound should be limited by the sensory parameters that characterize unimodal auditory speed perception. As a consequence, if an observer is actually responding to the auditory component of the audiovisual condition, performance cannot, by definition, be more accurate than that in the auditory (control) one as overall acoustic sensitivity necessarily remains the same. The equivalent sensitivity of the weighted and unimodal curves confirms this prediction.

Yet, our results show evidence for the mandatory nature of the mechanisms combining the auditory and the visual information of speed because even if participants were attempting to respond to acoustic speed, their results were best explained when assuming combination of visual and acoustic information. The weighting average, which underlies the present results, might become a general principle that could not only explain how different speed components are integrated in vision (e.g., Priebe & Lisberger, 2004) but also could account for integration of speed signals across modalities. A further question that remains to be answered is whether this combination will turn out to be optimal when observers are asked to estimate velocity using both modalities at the same time.

Acknowledgments

This research was supported by Grant TINS2004-04363-C03-02 from the Ministerio de Educación y Ciencia (of the Spanish Government).

Commercial relationships: none.

Corresponding author: Joan López-Moliner.

Email: j.lopezmoliner@ub.edu.

Address: Departament de Psicologia Bàsica, Universitat de Barcelona, P. Vall d'Hebron 171, 08035 Barcelona, Catalonia, Spain.

References

- Alais, D., & Burr, D. (2004). No direction-specific bimodal facilitation for audiovisual motion detection. 185–194. [[PubMed](#)]
- Burr, D., & Alais, D. (2006). Combining visual and auditory information. *Progress in Brain Research*, 155, 243–258. [[PubMed](#)]
- Carlile, S., & Best, V. (2002). Discrimination of sound source velocity in human listeners. *Journal of the Acoustical Society of America*, 111, 1026–1035. [[PubMed](#)]

- Chen, Y., Bedell, H. E., & Frishman, L. J. (1998). The precision of velocity discrimination across spatial frequency. *Perception & Psychophysics*, *60*, 1329–1336. [[PubMed](#)]
- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.
- 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., & Bühlhoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*, 162–169. [[PubMed](#)]
- Manabe, K., & Riquimaroux, H. (2000). Sound controls velocity perception of visual apparent motion. *Journal of the Acoustical Society of Japan*, *21*, 171–174.
- Mays, A., & Schirillo, J. (2005). Lights can reverse illusory directional hearing. *Neuroscience Letters*, *26*, 336–338. [[PubMed](#)]
- Meyer, G. F., & Wuerger, S. M. (2001). Cross-modal integration of auditory and visual motion signals. *Neuroreport*, *12*, 2557–2560. [[PubMed](#)]
- Meyer, G. F., Wuerger, S. M., Rohrbein, F., & Zetsche, C. (2005). Low-level integration of auditory and visual motion signals requires spatial co-localisation. *Experimental Brain Research*, *166*, 538–547. [[PubMed](#)]
- Middlebrooks, J. C., & Green, D. M. (1991). Sound localisation by human listeners. *Annual Review of Psychology*, *42*, 135–159. [[PubMed](#)]
- Perrone, J. A., & Thiele, A. (2001). Speed skills: Measuring the visual speed analyzing properties of primate MT neurons. *Nature Neuroscience*, *4*, 526–532. [[PubMed](#)] [[Article](#)]
- Perrone, J. A., & Thiele, A. (2002). A model of speed tuning in MT neurons. *Vision Research*, *42*, 1035–1051. [[PubMed](#)]
- Priebe, N. J., & Lisberger, S. G. (2004). Estimating target speed from the population response in visual area MT. *Journal of Neuroscience*, *24*, 1907–1916. [[PubMed](#)] [[Article](#)]
- Priebe, N. J., Lisberger, S. G., & Movshon, J. A. (2006). Tuning for spatiotemporal frequency and speed in directionally selective neurons of macaque striate cortex. *Journal of Neuroscience*, *26*, 2941–2950. [[PubMed](#)] [[Article](#)]
- Smith, A. T., & Edgar, G. K. (1991). The separability of temporal frequency and velocity. *Vision Research*, *31*, 321–326. [[PubMed](#)]
- Soto-Faraco, S., Kingstone, A., & Spence, C. (2003). Multisensory contributions to the perception of motion. *Neuropsychologia*, *41*, 1847–1862. [[PubMed](#)]
- Soto-Faraco, S., Spence, C., & Kingstone, A. (2004). Cross-modal dynamic capture: Congruency effects in the perception of motion across sensory modalities. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 330–345. [[PubMed](#)]
- Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. MIT Press: Cambridge, MA.
- Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, *9*, 578–585. [[PubMed](#)]
- Reisbeck, T. E., & Gegenfurtner, K. R. (1999). Velocity tuned mechanisms in human motion processing. *Vision Research*, *39*, 3267–3285. [[PubMed](#)]
- Watson, A. B., & Ahumada, A. J. (1983). A look at motion in the frequency domain. In J. K. Tsotsos (Ed.), *Motion: Perception and representation* (pp. 1–10). New York: Association for Computing Machinery.
- Watson, A. B., & Pelli, D. G. (1983). Quest: A Bayesian adaptive psychophysical 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](#)] [[Article](#)]