Multisensory integration of visual and auditory motion

Bingni Wen

CNS 120 Proposal
May 2005

Abstract

External signals from different sensory modalities are processed initially in sensory-specific brain areas, but integration between multiple modalities may be useful since these different signals often arise from common events in the external world. The existence of single cells that respond to stimuli in more than one sensory modality is well documented in sub-cortical and cortical areas, and there is recent evidence even typically sensory-specific brain areas may be affected by stimuli from other modalities. Perceptual interactions between visual and auditory systems have also been demonstrated. Here, I propose experiments to explore audiovisual interactions regarding the percept of motion. Using a random dot display at varying motion coherence values and a moving auditory noise source, I hope to find evidence for multisensory integration in an enhancement of bimodal motion detection threshold relative to the unimodal thresholds. Further, I propose to solicit subjective awareness of motion in both sensory modalities, which will shed light on the neuronal correlate of a merged audiovisual motion percept.

Introduction

In everyday life, we often experience events in the external environment that occur in more than one sensory modality: a moving car activates vision, audition, as well as tactile sensation from wind created as it speeds past. A common theme throughout much of neuroscience has been one of functional specialization – that distinct sensory functions are localized to distinct areas of the brain; such localization may be identified using functional magnetic resonance imaging (fMRI) in humans. Despite this modularity in our sensory processing, we nevertheless perceive the world as a coherent whole. Complementary information across multiple modalities is integrated, likely giving us a more complete internal representation of reality.

Cross-modal integration is often examined by making the information conveyed by two modalities in conflict with each other. Using such methods, vision and audition have been shown to interact powerfully to determine perception. In the McGurk effect, seeing the movement of lips changes how a simultaneous speech-sound is heard [1]. Shams *et al.* [2, 3] described a visual illusion induced by sound, where a single visual flash is incorrectly perceived as multiple flashes when accompanied by multiple auditory beeps. In addition, an auditory stimulus may enhance the perception of a visual stimulus. Sudden sound has been shown to improve detection of a subsequent masked visual flash at the same spatial location [4].

MSC 878, bingni@its.caltech.edu

An aspect of both vision and audition is their ability to determine direction and speed of motion. Returning to the example of the moving car, the visual and auditory information associated with this car is presumably merged to produce a unified percept – that of the car's movement within its environment. There are two features to be addressed in the study of cross-modal integration. First, how is information synthesized across multiple senses on the neural level? Second, does synthesis of complementary information (regarding the same stimulus) enhance performance, and what can this tell us about neuronal correlates of merging a multisensory percept? Each question is discussed briefly below; the proposed experiments will contribute to an understanding of the second question in the realms of visual and auditory perceptions of motion.

The neural basis of multisensory interactions has been investigated using electrophysiology in animals and fMRI in humans. Early studies on animals focused on converging feedforward projections from sensory-specific areas to multisensory areas, where neurons respond to stimuli from more than one modality. For example, Wallace *et al.* [5] have identified single neurons in the cat superior colliculus that respond more vigorously to simultaneous visual and auditory stimulus than the sum of each alone, a non-linear response known as "superadditivity" that is the hallmark of multisensory integration. Such neurons have also been discovered in cortical regions [6]. More recently, fMRI evidence from human brains has revealed another part of the story. Seemingly unimodal areas can also be affected by cross-modal interactions, suggesting the possible presence of back-projections from multisensory areas to unimodal brain regions [7].

In an imaging study of the human cortex during visual and auditory motion processing, Lewis *et al.* [8] identified areas whose activity was enhanced by cross-modal stimulation relative to unimodal tasks. Interactions were observed to involve both enhancing and suppressive effects in a network of both unimodal and multisensory cortical areas. The inferior and superior colliculus also contain neurons that respond selectively to motion signals [5].

Although the specific interaction of the two systems is unclear, it seems that they do interact. This brings us to the second question posed above, that of the perceptual effects of audiovisual interactions in motion detection. Meyer & Wuerger [9] demonstrated that salient auditory motion induces a bias in the perceived direction of visual motion consistent with the direction of auditory motion (audiovisual motion capture); further, Alais & Burr [10] observed bimodal improvement in an audiovisual motion/no motion discrimination task. Neural physiological evidence suggests the two sensory systems may interact and mutually enhance motion signals. Even so, it is unclear whether all observed audiovisual bimodal enhancements in performance reflect integration at the sensory level or the decision level. In this proposal, I hope to leverage from established experimental parameters and constraints found in previous works [9, 10, 11] to investigate further the interactions between the visual and auditory systems in motion detection, as well as how these interactions give rise to the conscious percept of motion, especially when both signals are at threshold. I plan to use a forced choice (left/right) motion detection paradigm in combination with a solicited judgment of whether the motion is seen, heard, both, or neither.

Specific Aims

1. Within the forced choice (left/right) motion detection paradigm, visual and auditory stimulus are presented simultaneously or alone. Is the motion detection sensitivity enhanced for correlated audiovisual stimulus? Does a sensitivity enhancement arise for longer stimulus durations that allow for saccadic eye movements?
2. If there is an enhancement, does it occur with or without increasing the subject's awareness of motion in either sensory modality?

Proposed Methods

Bimodal motion integration will be analyzed by measuring unimodal (visual/auditory) and bimodal motion detection thresholds. In addition to judgment of motion direction, subjects will be asked to report their awareness of motion in each sensory modality.

Stimuli

The auditory stimuli will be created based on what was described by Alais & Burr [10]. White noise will be low-pass filtered and then played at constant moderate loudness over two speakers placed on either side of a computer monitor. A sense of auditory motion will be created by playing the sound signal in one speaker at a time delay to the other speaker, so the listener gets an apparent sense of continuous auditory movement through interaural time difference. The exact time delay used will be calculated to match the apparent speed of visual motion displayed on the monitor. The motion signal may be diluted with a masking white noise played with no interaural time difference. This scheme of auditory motion generation has the advantage that the coherence of the signal may be easily manipulated by changing the proportion of total sound that carries the motion signal without varying total sound intensity, where 0% is noise only and 100% is all motion signal. The alternative, intensity-based method of generating auditory movement (where noise from one speaker waxes while the other wanes, used by Wuerger *et al.* [11]) has the limitation that the signal is bistable – the listener may perceive auditory movement or two independent speakers changing loudness.

The visual stimuli will be created following the dynamic random dot display used by Newsome & Paré [12]. The random dot display will appear on the monitor between the speakers such that the visual motion stimulus occupies 20° (± 10°) of the subject's visual angle. The stimulus will be composed of small dots, a variable percentage of which will move uniformly in the signal direction (left or right) while the rest act as visual motion masks and have random motion; 0% coherence corresponds to all random dot motion and 100% corresponds to all uniform signal motion.

The monitor and speakers will occupy the same plane, and the subject is placed such that he or she is directly facing this plane a fixed distance away. According to psychophysical studies [9, 10, 11] as well as single-neuron recordings [5], bimodal audiovisual enhancement only occurs when the two stimuli are spatially and temporally concordant. This is what the proposed stimuli are designed to achieve. When presented together in the same direction, the auditory and visual motion stimuli will appear to be co-localized in space and synchronized in time. The motion might feel similar to a long train running past the station platform.

Procedure

The study will solicit subjects with normal (or corrected-to-normal) vision and hearing. Each subject's motion detection threshold will first be measured for visual and auditory motion stimulus alone, which will be left or right at random. Threshold is defined, again following Newsome & Paré [12], as the coherence level where 82% accurate performance is achieved, as estimated from fitting motion detection performance with a psychometric function. To maintain data consistency with later bimodal measurements, both sensory modalities will be present during these unimodal measurements; for example, when measuring the visual motion threshold, white noise containing no motion signal will be played in the background. After each trial, the subject will be asked *1)* was the direction of motion left or right, and *2)* did you see/hear the motion or not? The second question essentially asks for the subject's confidence in the direction of motion response. At close to threshold coherences, the subject might perform at well above chance accuracy yet not be always aware of the direction of motion. This awareness response is assumed to be equivalent to the subject's consciousness of the motion percept in each sensory modality.

Next, the subject's threshold for bimodal motion detection will be measured, where visual and auditory motion stimuli will be presented at the same time, starting at signal coherence values around the subject's unimodal thresholds. Visual and auditory stimuli will be correlated (moving in the same direction) or anti-correlated at random. Again after each trial, the subject will be asked *1)* was the direction of motion left or right, and *2)* did you see the motion, did you hear the motion? The subject will be asked to judge the direction of motion of the combined audiovisual motion, not either one specifically, even though sometimes the stimuli are anticorrelated; this way, the subject must always attend to both motion stimuli.

The experimental procedure described above will be repeated separately for two stimulus durations, a short (500 ms) and a long (2000 ms). The short duration is to maintain comparability to related previous works, which have all used stimulus durations of under one second, presumably to minimize saccadic movement. The long duration will allow for saccadic eye movement and test whether a bimodal threshold enhancement may be observed for this altered parameter. All other conditions and parameters (apparatus position, apparent speed of motion) will remain identical for the two stimulus durations.

Interpretation of Results

Bimodal enhancement of performance

For each of the two stimulus durations, bimodal motion psychometric fits for correlated audiovisual stimuli may be compared to those for unimodal motion. It is expected that the bimodal curves will show shift to the left, corresponding to an improved lower detection threshold, which would be in agreement with previous results. When visual and auditory stimuli are anti-correlated, we may test if one sensory modality is consistently dominant over the other when both are near threshold. Such a perceptual dominance may confound results for the correlated case. The proposed experiment at the longer stimulus duration will test whether an improvement remains when the subject has time to make saccadic movements, which is a more ecologically relevant scenario.

Multisensory integration or combination of decision?

A bimodal enhancement to motion detection threshold may indicate multisensory integration in processing of visual and auditory motion. However, it is important to keep in mind that an enhancement may be observed even if bimodal sensory integration does not occur, but rather the decisions from two independent pathways are combined at a higher level – the probability of detecting at least one stimulus near threshold coherence is greater when two are presented. To differentiate between the two possibilities, consider a correlated bimodal motion stimulus, where the visual and auditory motion signals are at coherences c_v and c_a such that the subject performs at p_v and p_a if motion signals were presented unimodally. Let the null hypothesis be that the motion signals are processed independently in the two sensory modalities. If we make the further conservative assumption that, at the decision level where the motion senses combine, only one stimulus needs to be correctly identified for a correct overall judgment, then the subject is expected to perform at an overall $p' = p_v + (1 - p_v)p_a$ accuracy. The null hypothesis may be rejected if p_{av}, the observed bimodal accuracy, is greater than p' ($p_{av} > p'$). Unfortunately, this simple model is unrealistic in that there is a $2p_v p_a$ chance the independent pathways disagree, and the decision level would not be able to tell which is the correct judgment. If motion detection is done randomly whenever the two pathways disagree, then the bimodal enhancement through statistic combination is lost and expected overall performance becomes $p'' = (p_v + p_a)/2$.

In short, proving multisensory integration at the processing level will be difficult, especially if bimodal threshold improvement is not dramatic. The temporal resolution of functional imaging techniques limits their applicability in resolving this issue. Different designs for direction of motion tasks may be necessary to differentiate between true multisensory integration and statistical combination of decisions, perhaps lengthening the motion trials so that an imaging study would be feasible.

An interesting result would be if the longer duration stimulus shows multisensory integration while the shorter duration stimulus does not. Since the superior colliculus has been observed to have bimodal motion integration cells and is also implicated in modulation of saccades, it is possible that sensory processing integration will gain a bimodal advantage only when saccades are possible.

Alterations to visual and auditory motion awareness

From the unimodal motion awareness data, we may compare a subject's performance accuracy in motion detection to the percentage of times the same subject actually feels confident about his or her motion judgment. How well does behavior correspond to conscious awareness? Are the trials where the subject judges direction of motion accurately the same trials when he or she reports being aware of the motion?

If the data does indeed suggest multisensory integration of motion, it would be interesting to see if the improvement to motion threshold is accompanied by an improvement to motion awareness in either or both sensory modalities. Such an enhanced awareness would be consistent with evidence from human fMRI that point to increased activity in sensory-specific regions with multi-modal stimulus. Should a bimodal stimulus elicits the same conscious sensation of motion in a specific modality as a more salient (coherent) unimodal stimulus, that would be a convincing demonstration of the malleability of a unimodal motion percept. Does multisensory integration occur to the same extent with or without conscious awareness? If no difference in motion awareness is found but behavior is enhanced, then the neuronal correlate of the combined motion percept likely resides beyond sensory-specific regions in higher cortical areas.

Future Directions

The proposed experiments will contribute to an understanding of the interactions between visual and auditory sensory systems, which are the primary channels through which we remotely sense our external environment. Furthering the evidence for back-projections from multisensory regions of the brain to influence sensory-specific regions would help explain the multisensory binding problem, where one attributes different sensory percepts processed in separate pathways correctly to the same object. An interesting question that also arises is that, assuming these cross-modal back-projecting interactions do occur, how does a normal brain keep the different senses apart and prevent perceptual phenomena similar to synaesthesia? Such cross-modal interactions may be central to our perception of the world.

References

[1] McGurk, H. & MacDonald, J. Hearing lips and seeing voices. *Nature* **264**: 746-748 (1976).

[2] Shams, L., Kamitani, Y. & Shimojo, S. What you see is what you hear. *Nature* **408**: 788 (2000).

[3] Shams, L., Kamitani, Y. & Shimojo, S. Visual illusion induced by sound. *Cognitive Brain Research* **14**: 147-152 (2002).

[4] McDonald, J.J., Teder-Sälejärvi, W.A. & Hillyard, .S.A. Involuntary orienting to sound improves visual perception. *Nature* **407**: 906-908 (2000).

[5] Wallace, M.T., Meredith, M.A. & Stein, B.E. Multisensory integration in the superior colliculus of the alert cat. *J. Neurophysiology* **80**: 1006-1010 (1998).

[6] Graziano, M.S. & Gross, C.G. The representation of extrapersonal space: a possible role for bimodal, visuo-tactile neurons. In *The Cognitive Neurosciences* (Gazzaniga, M.S., ed.), MIT press, pp 1021-1034.

[7] Macaluso, E. & Driver, J. Multisensory spatial interactions: a window onto functional integration in the human brain. *Trends in Neurosciences* **28**: 264-271 (2005).

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

[9] Meyer, G.F. & Wuerger, S.M. Cross-modal integration of auditory and visual motion signals. *NeuroReport* **12**: 2557-2560 (2001).

[10] Alais, D. & Burr, D. No direction-specific bimodal facilitation for audiovidsual motion detection. *Cognitive Brain Research* **19**: 185-194 (2004).

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

[12] Newsome, W.T. & Paré, E.B. A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *J of Neuroscience* **8**:2201-2211 (1988).