Full text provided by www.sciencedirect.com

Seeing sounds: visual and auditory interactions in the brain

David A Bulkin and Jennifer M Groh

Objects and events can often be detected by more than one sensory system. Interactions between sensory systems can offer numerous benefits for the accuracy and completeness of the perception. Recent studies involving visual–auditory interactions have highlighted the perceptual advantages of combining information from these two modalities and have suggested that predominantly unimodal brain regions play a role in multisensory processing.

Addresses
Center for Cognitive Neuroscience, Department of Psychology and Neuroscience, Department of Neurobiology, Duke University, LSRC Room B203, Box 90999, Durham, NC 27708 USA

Corresponding author: Groh, Jennifer M (jmgroh@duke.edu)

Current Opinion in Neurobiology 2006, **16**:415–419

This review comes from a themed issue on
Sensory systems
Edited by Yang Dan and Richard D Mooney

Available online 11th July 2006

0959-4388/$ – see front matter

DOI 10.1016/j.conb.2006.06.008

Introduction

The focus of sensory research has expanded in recent years to include serious consideration of the nature and scope of interactions among sensory systems. In this review, we discuss recent findings concerning visual influences on auditory processing, at both the behavioral and the neurophysiological levels. These recent studies suggest that the perceptual benefits of integrating visual and auditory information are extensive, and that the neural substrate for these interactions involves recruitment of brain areas previously thought of as predominantly unimodal.

Visual–auditory interactions

The roles of the visual and auditory systems overlap. Both senses are concerned, at least in part, with stimuli at some distance from the body, which distinguishes them from the somatosensory system. Furthermore, both vision and hearing are capable of providing reasonably fine-grained spatial and temporal information, unlike the chemical senses. Thus, vision and audition are often charged with conveying information of a similar quality about the same (or at least related) objects and events.

Because of this overlap in function, interactions between the visual and the auditory systems can be beneficial in a number of ways. First, each sensory system can provide ‘missing pieces’. For example, the visual system, although generally excellent at providing spatial information, cannot detect the locations of occluded objects, objects located beyond the field of view, or objects in the dark. The auditory system is capable of filling in these lacunae. The complementary nature of visual and auditory spatial information provides obvious survival benefits, be it for humans engaged in the task of crossing a busy street or animals seeking prey or avoiding predators.

Second, when the two senses both provide information about exactly the same object or event, combining the signals from each modality can enhance the accuracy of the resulting percept. This type of interaction is best illustrated using illusions in which the visual and auditory information is actually slightly discrepant but the percept is unified. For example, in the McGurk effect, a video of a person’s mouth saying “ga” is dubbed with a soundtrack of a voice saying “ba”, resulting in a percept intermediate between the two, such as “da” [1]. This illusion reflects a strategy employed by the brain of combining visual and auditory information to provide a single percept consistent with both sources of information, a strategy that, under normal circumstances, produces greater accuracy and/or speed of speech comprehension when the two sources of input are matched [2•,3]. There appears to be a critical period for developing the ability to form these associations: deaf children with cochlear implants are more likely to show fusion of visual and auditory syllables if they have received their implant prior to the age of 2.5 years [4].

Persistent discrepancies between visual and auditory information can lead to persistent changes in the way that information from one of the two senses is perceived, with concomitant changes in neural processing. This was first shown by Knudsen and Knudsen [5] in the spatial domain: juvenile barn owls reared with visual displacing prisms gradually shift their sound localization behavior to agree with the mismatched visual scene. Recent work from this group has suggested that the underlying mechanism enabling these shifts is a remapping of projections between the central nucleus of the inferior colliculus and the external nucleus of the inferior colliculus [6,7]. The coexistence of multiple circuits explains the ability of prism-reared owls to easily switch back and forth between prism and normal conditions as adults [8].

Recent work by Recanzone and co-workers [9] has demonstrated that both humans and monkeys experience a similar recalibration of the auditory scene on the basis of

Glossary

Coordinate transformation: The process of converting a signal in one reference frame to another.

Eye-centered reference frame: A set of axes for defining the location of an object, with the eyes as the reference point. Visual spatial information arrives in this manner as the position of an image on the retina depends on its position with respect to the orientation of the eyes in space.

Head-centered reference frame: A set of axes for defining the location of an object, with the head as the reference point. Interaural timing and level differences provide information concerning sound source location in a head-centered reference frame.

Interaural level difference: The difference in amplitude or level of a sound wave across the two ears. The magnitude of an interaural level difference depends on the location of a sound source with respect to the head and ears.

Interaural time difference: The difference in the time of arrival of a sound wave across the two ears. The magnitude of an interaural time difference depends on the location of a sound source with respect to the head and ears.

Reference frame: A set of axes used to define a spatial location.

mismatched visual information. When a visual stimulus is presented at a location that is slightly shifted from an accompanying sound, this induces a shift in the perceived sound location that persists even when the sound is subsequently presented by itself [9,10•]. This phenomenon is known as the ventriloquism aftereffect, and is similar to the prism adaptation phenomenon observed in barn owls, except that the two phenomena have thus far been studied at quite different time scales; the ventriloquism aftereffect is apparent within a given behavioral testing session and prism adaptation lasts for weeks after removal of the spectacles. It is possible that the two phenomena will appear essentially the same when studied in the same fashion, or there might be genuine differences in the underlying neural mechanisms and behavioral properties.

Whereas the visual system excels at spatial acuity, the auditory system gives more precise temporal information and appears to dominate perception of when events occur. When subjects make judgments about the timing of temporally mismatched visual–auditory events, the auditory percept 'wins' [11–13]. It seems that the brain weights sources of sensory information according to their assumed reliability when producing a unified percept [14,15].

Neural computations for facilitating visual–auditory interactions

The type of visual–auditory interactions described in the preceding section requires a series of computations by the brain. Arguably the most important problem the brain must solve is to identify specifically which sights and sounds correspond to one another. Indeed, the ventriloquism and McGurk effects are examples in which this process goes slightly awry: the visual stimulus that 'captures' the sound is not actually the source of the sound, the soundtrack of the video is not actually the true sound of the person speaking. There are, in principle, three possible ways to 'bind' visual and auditory signals: first, spatially — signals from approximately the same location in space are presumed to arise from the same event; second, temporally — signals with the same temporal profile are presumed to arise from the same event; or third, through experience — signals that have been associated with each other in the past are presumed to arise from the same event on this occasion. Little work has been done on the temporal or experiential possibilities, but considerable effort has been devoted to exploring the spatial computations that are necessary for visual–auditory integration.

The crucial hurdle to integrating visual and auditory signals in the spatial domain is that the two sources initially employ different reference frames (see glossary). Visual spatial information is derived from the site of activation on the retina. The site of retinal activation depends on the location of the visual object with respect to the direction of gaze. By contrast, the locations of sounds must be inferred by comparing sound arrival time and level differences (see glossary) across the two ears. These cues specify the location of the sound with respect to the head and ears. In animals with mobile eyes, such as humans and monkeys (but not barn owls), there is no fixed correspondence between a location on the retina and a given set of interaural cue values. Rather, the relationship between visual space and auditory space is constantly changing as the eyes move around.

A recent behavioral study provided evidence that the brain somehow solves this reference frame discrepancy. Metzger *et al.* [16] compared the accuracy of eye movements to visual stimuli with those to auditory stimuli in monkeys making saccades from different initial eye positions. The results showed that monkeys do compensate for where the eyes are starting from when making saccades to sounds, just as they do for visual stimuli. This study suggests that by the time an oculomotor command is generated, visual and auditory information are encoded in a common coordinate system.

A series of physiological studies suggests that this process of computing a coordinate transformation (see glossary) on auditory signals has begun by a comparatively early point in the auditory pathway: many neurons in the primate inferior colliculus (IC) are sensitive to the position of the eyes in the orbits [17–19]. This sensitivity is also present in auditory cortex [20,21]. At present, it is unknown where the eye position signals might first arise in the auditory pathway, or in what direction they are propagated (i.e. through ascending versus descending connections).

What kind of coordinate transformation this neural circuit is performing is also unclear. The effect of eye position on auditory neurons in the IC and auditory cortex disrupts the head-centered-ness of the auditory spatial information, but does not produce eye-centered receptive fields. In fact, it is not certain that a head- to eye-centered coordinate transformation would necessarily be advantageous. It could be that a hybrid reference frame, such as that found in the IC and auditory cortex, is well-suited for preparing motor commands, especially if the motor command for generating a saccade occurs in a reference frame that is best described as mixed. Indeed, the hybrid nature of the representation of auditory information in the IC and auditory cortex appears to be similar to that found in the intraparietal cortex [22••,23••], a structure that has been specifically linked to the control of eye movements [24]. The auditory reference frame in all three areas — inferior colliculus, auditory cortex and intraparietal sulcus — might be similar to that of the primate superior colliculus (SC), which has auditory receptive fields that shift when the eyes move, but only about half the distance of the eye movement, therefore producing a hybrid representation [25,26].

Visual influences over auditory areas

Not only does the position of the eyes affect activity in auditory areas, but the stimulus on the retina does as well. Evidence for visual influence over auditory processing comes from a variety of sources, from anatomical and physiological to neuroimaging, and at a variety of levels of the two sensory pathways. Projections between the visual and the auditory cortices have recently been identified in primates [27•], and some auditory cortical neurons appear to have visual responses [28••]. Both event related potentials (ERP) and functional magnetic resonance imaging (fMRI) showed enhanced auditory cortical responses when tones were paired with an attended visual stimulus versus an unattended visual stimulus [29••]. In deaf individuals, auditory cortex shows fMRI activation to moving visual stimuli [30,31•]. There is also anatomical evidence for projections to the IC from the retina [32–34], visual cortex [35] and the superior colliculus [36,37]. However, in anesthetized cats visual responsiveness in the IC is rare (<10% of neurons), and occurs only among neurons that lack auditory responses ([38] see also [39]).

Top-down control mechanisms exert considerable influence over both visual and auditory response properties in auditory areas in barn owls. Disinhibition of the SC unmasks visual responses in the IC that are otherwise silent [40]. Microstimulation of the arcopallial gaze fields, analogous to the mammalian frontal eye fields, sharpens the tuning of auditory receptive fields in the SC [41].

Language and communication might be a particularly important instance of visual–auditory integration, requiring the recruitment of 'historically' auditory areas by visual stimuli. When monkeys match vocalizations with the appropriate facial expressions of conspecifics, local field potentials in both core and belt areas of auditory cortex show enhancement or suppression compared with responses to the auditory stimulus alone [42••]. In humans, the planum temporale, a region of the superior temporal gyrus, has been implicated in both auditory- and language- related processing. Planum temporale activation has been shown in hearing subjects during silent visual speech perception [43–45]. This visual activation might be specific to linguistic stimuli: hearing subjects passively viewing single elements of sign language (a non-linguistic stimulus to the non-signer) do not show activation in this area [46]. However, MacSweeney *et al.* [45] found responses in planum temporale in deaf and hearing signers as well as in hearing nonsigners while they were viewing both sign language gestures and Tic Tac (a racecourse betting code based on gestures but lacking phonological structure). In the MacSweeney study, participants were instructed to attempt to interpret the gestures. This suggests that planum temporale activation to visual stimuli is found in normal subjects only when they interpret the visual stimuli in a communication context [45]. In short, planum temporale might be best viewed not as having visual responses to just any kind of visual stimulus but as an auditory area that incorporates potential language-related material in general [46,47].

Conclusions

Clearly, we are just beginning to untangle the threads of interaction between the visual and the auditory systems. Much work needs to be done in both the perceptual and the neural domains. At the perceptual level, additional studies establishing the circumstances of cross modal binding and plasticity are needed, as are novel paradigms to investigate the nature of the information transfer between these systems. How does the visual system help the auditory system de-code cluttered auditory scenes, for example? At the neural level, further investigations into the extent and hierarchical organization of visual–auditory interactions are needed. Such studies will form the basis for a new understanding of how the visual and auditory systems function in concert.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

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

2. • Navarra J, Soto-Faraco S: **Hearing lips in a second language: visual articulatory information enables the perception of second language sounds**. *Psychol Res* 2005:1-9; DOI:10.1007/s00426-005-0031-5.

Interactions between sensory systems are often most evident when tested under challenging conditions. Here, the authors sought to test the value of visual information for subjects' comprehension of a second language. They tested Spanish and Catalan dominant bilingual speakers' comprehension of Catalan sounds. The authors report that although Catalan dominant bilinguals could easily differentiate two Catalan syllables using only auditory cues, Spanish dominant bilinguals were only able to do this when they could see the speaker's mouth.

3. Besle J, Fort A, Delpuech C, Giard MH: **Bimodal speech: early suppressive visual effects in human auditory cortex**. *Eur J Neurosci* 2004, **8**:2225-2234.

4. Schorr EA, Fox NA, van Wassenhove V, Knudsen EI: **Auditory-visual fusion in speech perception in children with cochlear implants**. *Proc Natl Acad Sci U S A* 2005, **102**:18748-18750.

5. Knudsen EI, Knudsen PF: **Vision guides the adjustment of auditory localization in young barn owls**. *Science* 1985, **230**:545-548.

6. Feldman DE, Knudsen EI: **An anatomical basis for visual calibration of the auditory space map in the barn owl's midbrain**. *J Neurosci* 1997, **17**:6820-6837.

7. DeBello WM, Feldman DE, Knudsen EI: **Adaptive axonal remodeling in the midbrain auditory space map**. *J Neurosci* 2001, **21**:3161-3174.

8. Knudsen EI, Knudsen PF: **Sensitive and critical periods for visual calibration of sound localization by barn owls**. *J Neurosci* 1990, **10**:222-232.

9. Recanzone GH: **Rapidly induced auditory plasticity: the ventriloquism aftereffect**. *Proc Natl Acad Sci USA* 1998, **95**:869-875.

10. • Woods TM, Recanzone GH: **Visually induced plasticity of auditory spatial perception in macaques**. *Curr Biol* 2004, **14**:1559-1564.
The authors show that spatial mismatches between visual and auditory stimuli produce persistent shifts in the perceived location of sounds in monkeys. The authors measured sound location judgments in two macaque monkeys after training them with combined visual and auditory targets that were mismatched in location. The monkeys shifted their sound location judgments in the direction of the visual stimulus. The effect persisted for at least 40 minutes. These results concur with literature regarding the ventriloquism aftereffect in humans.

11. Recanzone GH: **Auditory influences on visual temporal rate perception**. *J Neurophysiol* 2003, **89**:1078-1093.

12. Shams L, Kamitani Y, Shimojo S: **Illusions. What you see is what you hear**. *Nature* 2000, **408**:788.

13. Fendrich R, Corballis PM: **The temporal cross-capture of audition and vision**. *Percept Psychophys* 2001, **63**:719-725.

14. Shams L, Ma WJ, Beierholm U: **Sound-induced flash illusion as an optimal percept**. *Neuroreport* 2005, **16**:1923-1927.

15. Alais D, Burr D: **The ventriloquist effect results from near-optimal bimodal integration**. *Curr Biol* 2004, **14**:257-262.

16. Metzger RR, Mullette-Gillman OA, Underhill AM, Cohen YE, Groh JM: **Auditory saccades from different eye positions in the monkey: implications for coordinate transformations**. *J Neurophysiol* 2004, **92**:2622-2627.

17. Groh JM, Trause AS, Underhill AM, Clark KR, Inati S: **Eye position influences auditory responses in primate inferior colliculus**. *Neuron* 2001, **29**:509-518.

18. Zwiers MP, Versnel H, Van Opstal AJ: **Involvement of monkey inferior colliculus in spatial hearing**. *J Neurosci* 2004, **24**:4145-4156.

19. Porter KK, Metzger RR, Groh JM: **Representation of eye position in primate inferior colliculus**. *J Neurophysiol* 2006, **95**:1826-1842.

20. Werner-Reiss U, Kelly KA, Trause AS, Underhill AM, Groh JM: **Eye position affects activity in primary auditory cortex of primates**. *Curr Biol* 2003, **13**:554-562.

21. Fu KM, Shah AS, O'Connell MN, McGinnis T, Eckholdt H, Lakatos P, Smiley J, Schroeder CE: **Timing and laminar profile of eye-position effects on auditory responses in primate auditory cortex**. *J Neurophysiol* 2004, **92**:3522-3531.

22. •• Mullette-Gillman OA, Cohen YE, Groh JM: **Eye-centered, head-centered, and complex coding of visual and auditory targets in the intraparietal sulcus**. *J Neurophysiol* 2005, **94**:2331-2352.
The authors investigated the reference frame of visual and auditory neurons in the lateral and medial intraparietal areas. They found that space was encoded in a continuum of head- to eye- centered coordinates. Although some neurons had more eye-centered frames of reference, and some had more head-centered frames of reference, most neurons (of both modalities) were encoded in a manner that was neither eye- nor head- centered. The results were similar to those of Schlack *et al.* [23••] in the adjacent fundus of the intraparietal sulcus.

23. •• Schlack A, Sterbing-D'Angelo SJ, Hartung K, Hoffmann KP, Bremmer F: **Multisensory space representations in the macaque ventral intraparietal area**. *J Neurosci* 2005, **25**:4616-4625.
The authors showed that neurons in the macaque ventral intraparietal area respond to auditory stimuli. The responses were dependent on the location of the stimulus in space, and their receptive fields were fairly narrow. Visual and auditory receptive fields typically overlapped and employed a common reference frame. The results were similar to those of Mullette-Gillman *et al.* [22••] in the adjacent banks of the intraparietal sulcus.

24. Bisley JW, Goldberg ME: **The role of the parietal cortex in the neural processing of saccadic eye movements**. *Adv Neurol* 2003, **93**:141-157.

25. Jay MF, Sparks DL: **Auditory receptive fields in primate superior colliculus shift with changes in eye position**. *Nature* 1984, **309**:345-347.

26. Jay MF, Sparks DL: **Sensorimotor integration in the primate superior colliculus. II. coordinates of auditory signals**. *J Neurophysiol* 1987, **57**:35-55.

27. • Cappe C, Barone P: **Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey**. *Eur J Neurosci* 2005, **22**:2886-2902.
The authors provide evidence that anatomical pathways supporting multisensory integration exist at low levels of cortical processing. The authors made injections of retrograde tracers into early visual (V2 and MT), somatosensory (1/3b) and auditory (core) cortical areas in marmosets. Projections from visual areas to the core auditory cortex and somatosensory cortex, and from somatosensory to auditory cortex were found, supporting the view that inputs from different modalities are sent to cortical areas that are classically considered to be unimodal.

28. •• Brosch M, Selezneva E, Scheich H: **Nonauditory events of a behavioral procedure activate auditory cortex of highly trained monkeys**. *J Neurosci* 2005, **25**:6797-6806.
The authors reported visual responses in about 14% of multiunit clusters in auditory cortex of monkeys performing a behavioral task involving a visual cue light. Additional task-related signals were also identified.

29. •• Busse L, Roberts KC, Crist RE, Weissman DH, Woldorff MG: **The spread of attention across modalities and space in a multisensory object**. *Proc Natl Acad Sci USA* 2005, **102**:18751-18756.
The authors report that auditory cortical responses to sounds are influenced by visually cued attention, as revealed by a combination of ERP and fMRI in human subjects. The spatial locations of the visual and auditory stimuli were discrepant, and the results are interpreted in relation to the grouping of sensory signals to form supramodal, multisensory 'objects' and the ventriloquism effect. The authors propose that attention, initially cued by the visual modality, spreads ''across modalities and space to encompass the concomitant auditory signals''.

30. Finney EM, Fine I, Dobkins KR: **Visual stimuli activate auditory cortex in the deaf**. *Nat Neurosci* 2001, **4**:1171-1173.

31. • Fine I, Finney EM, Boynton GM, Dobkins KR: **Comparing the effects of auditory deprivation and sign language within the auditory and visual cortex**. *J Cogn Neurosci* 2005, **17**:1621-1637.
This study addresses plasticity in sensory processing in individuals with sensory deficits. The authors report fMRI activation in auditory cortex (especially the right hemisphere) of deaf subjects viewing moving dots. This effect did not occur in hearing (including both signers and non-signers) participants. Similar to the results of Busse *et al.* [29••], there was a greater visual response in auditory cortex to attended than to unattended stimuli.

32. Itaya SK, van Hoesen GW: **WGA-HRP as a transneuronal marker in the visual pathways of monkey and rat**. *Brain Res* 1982, **236**:199-204.

33. Paloff AM, Usunoff KG, Hinova-Palova DV, Ivanov DP: **Retinal innervation of the inferior colliculus in adult cats: electron microscopic observations**. *Neurosci Lett* 1985, **54**:339-344.

34. Yamauchi K, Yamadori T: **Retinal projection to the inferior colliculus in the rat**. *Acta Anat* 1982, **114**:355-360.

35. Cooper MH, Young PA: **Cortical projections to the inferior colliculus of the cat**. *Exp Neurol* 1976, **51**:488-502.

36. Doubell TP, Baron J, Skaliora I, King AJ: **Topographical projection from the superior colliculus to the nucleus of the brachium of the inferior colliculus in the ferret: convergence of visual and auditory information**. *Eur J Neurosci* 2000, **12**:4290-4308.

37. Harting JK: **Descending pathways from the superior colliculus: an autoradiographic analysis in the rhesus monkey (Macaca mulatta)**. *Journal of Comparative Neurology* 1977, **173**:583-612.

38. Mascetti GG, Strozzi L: **Visual cells in the inferior colliculus of the cat**. *Brain Research* 1988, **442**:387-390.

39. Tawil RN, Saade NE, Bitar M, Jabbur SJ: **Polysensory interactions on single neurons of cat inferior colliculus**. *Brain Res* 1983, **269**:149-152.

40. Gutfreund Y, Zheng W, Knudsen EI: **Gated visual input to the central auditory system**. *Science* 2002, **297**:1556-1559.

41. Winkowski DE, Knudsen EI: **Top-down gain control of the auditory space map by gaze control circuitry in the barn owl**. *Nature* 2006, **439**:336-339.

42. Ghazanfar AA, Maier JX, Hoffman KL, Logothetis NK:
•• **Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex**. *J Neurosci* 2005, **25**:5004-5012.

Rhesus monkeys communicate using vocalizations, and the mouth posture and movements associated with different calls are different, providing a visual cue for communication. These authors showed that the local field potentials evoked by auditory communication stimuli in core and lateral belt auditory cortex can be either enhanced or suppressed with the addition of a visual component. Generally, the visual stimulus alone did not evoke a response.

43. Calvert GA, Campbell R: **Reading speech from still and moving faces: the neural substrates of visible speech**. *J Cogn Neurosci* 2003, **15**:57-70.

44. Campbell R, MacSweeney M, Surguladze S, Calvert G, McGuire P, Suckling J, Brammer MJ, David AS: **Cortical substrates for the perception of face actions: an fMRI study of the specificity of activation for seen speech and for meaningless lower-face acts (gurning)**. *Brain Res Cogn Brain Res* 2001, **12**:233-243.

45. MacSweeney M, Campbell R, Woll B, Giampietro V, David AS, McGuire PK, Calvert GA, Brammer MJ: **Dissociating linguistic and nonlinguistic gestural communication in the brain**. *Neuroimage* 2004, **22**:1605-1618.

46. Petitto LA, Zatorre RJ, Gauna K, Nikelski EJ, Dostie D, Evans AC: **Speech-like cerebral activity in profoundly deaf people processing signed languages: implications for the neural basis of human language**. *Proc Natl Acad Sci USA* 2000, **97**:13961-13966.

47. Penhune VB, Cismaru R, Dorsaint-Pierre R, Petitto LA, Zatorre RJ: **The morphometry of auditory cortex in the congenitally deaf measured using MRI**. *Neuroimage* 2003, **20**:1215-1225.