

- that grammatical rules are processed by the procedural system *J. Cogn. Neurosci.* 9, 266–276
- 38 Manis, F. et al. (1997) Are speech perception deficits associated with developmental dyslexia? *J. Exp. Child Psychol.* 66, 211–235
- 39 Leonard, L.B. et al. (1987) Specific language impairment in children: a cross-linguistic study *Brain Lang.* 32, 233–252
- 40 Francis, W.N. and Kucera, H. (1982) *Frequency Analysis of English Usage: Lexicon and Grammar*, Houghton-Mifflin
- 41 Baddeley, A., Vallar, G. and Wilson, B. (1987) Sentence comprehension and phonological memory: some neuropsychological evidence, in *Attention and Performance Vol. 12: The Psychology of Reading* (Coltheart, M., ed.), pp. 509–529, Erlbaum
- 42 Just, M.A. and Carpenter, P.A. (1992) A capacity theory of comprehension: individual differences in working memory *Psych. Rev.* 99, 122–149
- 43 Waters, G. and Caplan, D. (1992) The capacity theory of sentence comprehension: critique of Just and Carpenter *Psych. Rev.* 103, 761–772
- 44 Montgomery, J.W. (1995) Examination of phonological working memory in specifically language-impaired children *Appl. Psycholinguist.* 16, 355–378
- 45 Christiansen, M.H. and MacDonald, M.C. Fractionated working memory: even in pebbles, it's still a soup stone *Behav. Brain Sci.* (in press)
- 46 Shankweiler, D. et al. (1995) Cognitive profiles of reading-disabled children: comparison of language skills in phonology, morphology and syntax *Psychol. Sci.* 6, 149–156

Crossmodal identification

Gemma A. Calvert, Michael J. Brammer and Susan D. Iversen

Everyday experience involves the continuous integration of information from multiple sensory inputs. Such crossmodal interactions are advantageous since the combined action of different sensory cues can provide information unavailable from their individual operation, reducing perceptual ambiguity and enhancing responsiveness. The behavioural consequences of such multimodal processes and their putative neural mechanisms have been investigated extensively with respect to orienting behaviour and, to a lesser extent, the crossmodal coordination of spatial attention. These operations are concerned mainly with the determination of stimulus location. However, information from different sensory streams can also be combined to assist stimulus identification. Psychophysical and physiological data indicate that these two crossmodal processes are subject to different temporal and spatial constraints both at the behavioural and neuronal level and involve the participation of distinct neural substrates. Here we review the evidence for such a dissociation and discuss recent neurophysiological, neuroanatomical and neuroimaging findings that shed light on the mechanisms underlying crossmodal identification, with specific reference to audio-visual speech perception.

Evolution has furnished humans with several different senses, each tuned to a distinct form of energy and providing a unique window through which to experience the environment. The possession of multiple sensory systems provides considerable behavioural flexibility since input from one modality can substitute for another under circumstances of specific sensory deprivation. In darkness, for example, auditory and tactile cues might supplant visual information. Such polysensory capability also permits the integration of different sensory streams. Combining sensory inputs is clearly advantageous since it supplies information about the environment that is unavailable from any single modality, influencing the perception of events in the surroundings and our subsequent responses.

The many behavioural consequences of multimodal integration have been investigated extensively with respect to

orienting and attentive behaviours, primarily concerned with the determination of stimulus location (for reviews, see Stein and Meredith¹; Driver and Spence², this issue). In addition to facilitating the detection of, and orientation to, stimuli in the environment^{3,4}, the integration of different sensory cues has also been shown to influence localization judgements. Specifically, when two or more sensory events are in close temporal proximity, albeit in slightly distinct spatial locations, they are generally perceived as emanating from a common source^{5–7}. Typically, the modality with the best spatial resolution (e.g. vision's superiority over audition) has the greatest influence on the location of the fused percept. Such crossmodal influences on localization are perhaps best typified by the ventriloquist's illusion. The ventriloquist speaks without moving his lips but it is his puppet that seems to be talking.

The assemblage of information across the senses also plays an important role in the perception of stimulus identity. For example, in noisy conditions, combining audible speech signals with the visible evidence of articulation (i.e. the accompanying lip and mouth movements) can improve comprehension to a degree equivalent to altering the acoustic signal-to-noise ratio by 15–20 decibels (Ref. 8). That such multisensory interactions are also capable of modifying the perceptual outcome is clearly illustrated by the McGurk effect⁹. When an audible syllable (e.g. 'ba') is dubbed onto videotape of a speaker mouthing a different syllable (e.g. 'ga') subjects typically report 'hearing' another syllable (usually 'da'). Frequently, the perceived syllable represents some form of combination of the two sensory inputs. Such intersensory effects on event identification have also been shown to extend to judgements of non-speech stimuli, specifically whilst watching and listening to the playing of stringed instruments¹⁰. Although less striking than the effect observed with conflicting audio-visual consonant-vowel (CV) combinations (typically employed during McGurk demonstrations), the intermodal influence was reportedly comparable to that elicited during the presentation of conflicting audio-visual vowels^{11,12}. Whilst further research is clearly necessary to determine whether the crossmodal mechanisms involved in the integration of speech and non-speech stimuli display important differences, such findings encourage the view that the critical distinction in cross-sensory operations resides between localization and identification judgements^{13,14}.

The fact that the ventriloquist's illusion and the McGurk effect both arise during the combination of auditory and visual inputs does not imply that they share common neural mechanisms or sites of sensory integration. Indeed, psychophysical studies that have explicitly compared the conditions for multisensory integration by exploiting these two artificially induced conflict phenomena have indicated that they are actually subject to rather different cognitive constraints^{15–17}. For example, judgements about stimulus location are consistently modified as the spatial disparity between the two inputs is increased. By contrast, spatial separation of the auditory and visual stimuli by up to 38 deg has little impact on the McGurk effect^{16–18}. Furthermore, whilst desynchronization of the auditory and visual information has been shown to have deleterious effects on the ventriloquist's illusion^{6,19,20}, the McGurk effect can still be elicited even when the auditory stimulus lags the visual input for up to 180 ms (Refs 21,22). Recent findings further suggest that the effect of temporal displacement on crossmodal (audio-visual) localization, but not identification, might be dependent on the nature of the stimuli²³. These discrepancies in spatial and temporal parameters required for the two processes of auditory-visual integration could reflect different underlying rules for stimulus combination at the neural level. Indeed, such a distinction between the integration of information for stimulus location ('where') and stimulus identification ('what') might parallel that described for the visual processing of information relating to object position and identity, the so-called dorsal and ventral visual streams²⁴.

In contrast to the amount of published work on the neural mechanisms involved in the crossmodal integration

of spatial information (see Refs 1,25), comparatively little is known about the nature and location of the mechanisms underlying the cross-sensory integration of inputs for the purposes of object and pattern recognition. Furthermore, whilst behavioural attempts to delineate the multimodal processes involved in localization and identification have predominantly involved direct comparison of audio-visual combinations, research on the neural basis of crossmodal spatial interactions has been preoccupied with the integration of visual, tactile and proprioceptive information. Such is the current state of knowledge in the field of crossmodal processing that a complete comparison between all sensory combinations at every level of analysis is clearly untenable. In this review, we will consider neurophysiological, neuroanatomical and neuroimaging findings that might elucidate the mechanisms underlying crossmodal identification (making comparisons with data on crossmodal localization where appropriate), focusing in particular on the integration of auditory and visual speech. Nevertheless, many of the principles discussed might be relevant to the crossmodal integration of other stimulus features, such as intensity and duration, and various combinations of sensory inputs across the different modalities (for a comprehensive review of intersensory effects, see Ref. 26).

What determines integration?

For two or more sensory inputs to be perceived as relating to a single object or event, some point of commonality must be detected between them. Across all combinations of sensory modalities, temporal and spatial proximity are clearly major determinants for co-registration^{27,28}. However, the relative importance of these factors might differ depending on the purpose for which these inputs are combined and the specific sensory streams being integrated. Another factor that could determine integration is the information content of the different sensory inputs. For stimuli of low information content (e.g. a short light flash and a brief burst of sound) there might be little shared information bar their simultaneous onset. Consequently, if the two sensory inputs are even slightly displaced in time, binding is often precluded. For information-rich stimuli, especially those with complex temporal microstructure, simultaneous onset or spatial contiguity might be less critical for integration of the inputs to occur. Instead, time-varying similarities in the patterning of information might prove a more salient feature for binding. For example, Summerfield²⁹ has proposed that heard and seen speech is possibly bound by shared temporal frequency and amplitude features, even though the nature of the information received in each modality is different (e.g. visible changes in the oral area are correlated with the amplitude and formant frequencies detected in acoustic speech input). These properties of speech could explain why integration persists despite slight disparities in the initial onset or spatial location of the auditory and visual inputs (see above).

Putative neural mechanisms

Irrespective of the sensory features that determine binding during any particular crossmodal combination, it is likely that the different sensory streams must ultimately gain

Box 1. Enhancement effects of multimodal integration in brain and behaviour

Combining information from different sensory inputs can enhance detection and orientation behaviours. In a study examining the contribution of auditory and visual cues to speech perception, Risberg and Lubker^a found that the effect on comprehension of integrating these different sensory channels was supra-additive. Figure A shows the mean results from five subjects when they were either listening to unfamiliar sentences that had been low-pass filtered (thus retaining only the fundamental frequency of the female voice used – around 180 Hz), or lip-reading the same stimuli in the absence of sound. During concurrent presentation of these cues, it can be seen that the percentage of words correctly perceived was substantially enhanced.

Such supra-additive effects on behaviour have also been shown to characterize the response of multimodal neurons to multisensory stimuli. Figure B illustrates the evoked responses of a single neuron in the cat superior colliculus to visual (V), auditory (A) and audio-visual (VA) stimulation. Visual or auditory stimulation alone produced only weak or unreliable responses but combined stimulation increased the number of impulses detected by 1207% (from Ref. b).

Together these studies illustrate that the response enhancement elicited by multimodal inputs at the behavioural level appears to have a demonstrable physiological correlate. Such

brain-behaviour relationships are known to characterize the temporal and spatial rules governing integration (see main text for further discussion).

References

- a Risberg, A. and Lubker, J. (1978) *Prosody and speech-reading STL-Quarterly Progress and Status Report 4*, 1–16
- b Stein, B.E. and Meredith, M.A. (1993) *The Merging of the Senses*, MIT Press

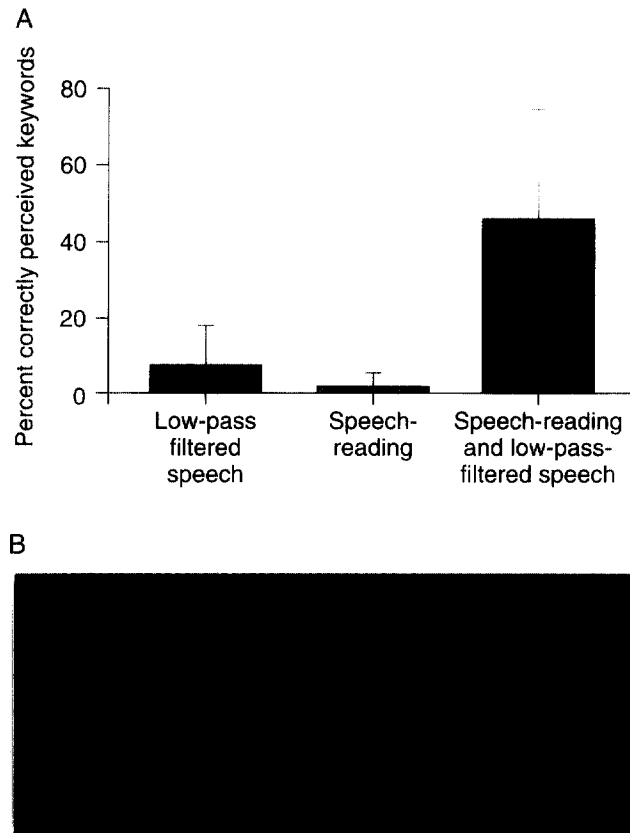


Fig. Comparison of multimodal effects at behavioural and neuronal levels. (A) The percentage of correctly perceived words when subjects listened to low-pass speech (left), lip-read the same stimuli without hearing any sound (centre), and with both auditory and visual presentation (right). (Modified from Ref. a.) **(B)** Evoked responses of a cat superior colliculus neuron to visual (V), auditory (A) and both visual and auditory (VA) stimuli. (Reproduced, with permission, from Ref. b.)

access to the same neurons (although see Massaro³⁰ for a discussion of integration by temporal synchrony). Indeed, electrophysiological studies have now identified many areas in the mammalian brain (see below) where the different sensory streams converge onto individual neurons responsive to stimulation in more than one modality, so-called 'multimodal' neurons. The rules governing multisensory integration have been investigated predominantly in deep layers of the superior colliculus of the cat¹, guinea pig³¹ and primate^{32,33} where multimodal cells are numerous. Although predominantly involved in the control of attention, localization and orientation to sensory stimuli, it is likely that some general principles of multimodal integration observed in the superior colliculus^{34,35} might extend to the crossmodal integration of identity information. For example, the supra-additive enhancement of every measurable characteristic of cellular responsiveness in the superior colliculus to multimodal inputs (response reliability, number of impulses evoked, peak impulse frequency and duration of the discharge train)

closely resembles the gain elicited at the behavioural level when visible speech cues are combined with the synchronous auditory speech information (see Box 1).

Furthermore, despite some latitude, similar temporal and spatial constraints to those operating during crossmodal localization have also been shown to apply during audio-visual speech integration (see above). These constraints on binding during localization appear to be based on the narrow time window for effective crossmodal integration at the neuronal level³⁶ and the close spatial correspondence between the different unimodal receptive fields of multisensory neurons in the superior colliculus³⁷. More recently it has been shown that similar principles of multisensory integration characterize binding in the cerebral cortex³⁸. There are, however, some interesting differences. Recording from the anterior ectosylvian fissure of the cat, Stein and Wallace found that the sensory receptive fields of multimodal neurons in this region were significantly larger than those in the superior colliculus³⁸. This, they argued, could reflect a greater emphasis on

non-spatial aspects of crossmodality integration in the cerebral cortex (as illustrated by the relative insensitivity of the McGurk effect to spatial separation of the auditory and visual inputs) than is required by the orientation and localization functions mediated (at least in part) by the superior colliculus.

Neuroanatomical sites of multimodal integration

Anatomical, neurophysiological and behavioural studies in non-human primates (see Ref. 39 for a review), together with lesion and functional neuroimaging studies in humans^{40,41}, have identified several brain regions putatively involved in multisensory integration. These nominal 'heteromodal' regions have been defined largely on the basis of several shared properties: they receive convergent inputs from more than one unimodal area; their constituent neurons respond to stimulation in more than one modality (or if sensory-specific, are closely interspersed with neurons responsive to a different modality); and lesions to these areas in humans can result in multimodal behavioural deficits.

How these regions actually participate in crossmodal processing is much less clear. Indeed, relevant ablations in monkeys have typically failed to produce deficits in the crossmodal transfer or matching of information (see Ref. 42 for a review). Mesulam^{43,44} has suggested that whilst heteromodal cortex could ensure the binding of modality-specific information into multimodal representations, it might not be the repository of this information. Instead, heteromodal 'nodes' could act as critical gateways for the encoding and retrieval of knowledge but storage of the sensory-specific attributes of a semantic representation (i.e. the sound of a cat; the texture of its fur) might reside in the unimodal sensory cortices in which they were initially perceived. In this way, the constituent and qualitatively distinct sensory attributes of a representation could be retained during retrieval. In sum, whilst heteromodal cortices might act as convenient points of integration between the senses, the perceptual consequences of these multimodal interactions could be additionally realized in the relevant unimodal cortices.

Although heteromodal regions have been determined with most confidence in monkeys (see Ref. 45 for review), homologous regions have also been identified in man^{44,46}. These are thought to include parts of prefrontal cortex, posterior parietal cortex [posterior Brodmann's area (BA) 7, BA 39, BA 40] and regions within lateral temporal cortex extending into the depths of the superior temporal sulcus (zones within BA 37, BA 21). Outside the neocortex, putatively heteromodal areas include the posterior insula, the claustrum, the superior colliculus, portions of the parahippocampal gyrus (parts of BA 35, 36), the amygdaloid complex, the rhinal cortex and the supragenulate and medial pulvinar of the thalamus.

The specific involvement of these areas in multimodal integration will inevitably depend on the precise combination of sensory inputs and the purpose for which they are integrated. One potential site of convergence for auditory and visual speech cues is the cortex in the superior temporal sulcus (STS). This area has been shown in monkeys to receive inputs from auditory association cortex (BA 42/22) and from the ventral and dorsal visual brain areas concerned with the processing of form and motion^{47,48}. Although

analogies between humans and non-human primates are clearly problematic in the context of speech, lesions close to this area in humans have long been linked with receptive aphasia⁴⁹. In contrast, different areas have been implicated in the integration of information relating to stimulus location, including primarily the superior colliculus and posterior parietal cortex⁵⁰.

Insights from functional neuroimaging

Modern neuroimaging techniques have now made it possible to investigate multimodal processes in humans. Functional magnetic resonance imaging (fMRI) permits the identification of brain areas showing task-related cerebrovascular responses⁵¹. Using this technique, we investigated the brain regions involved in silent lip-reading in normal hearing subjects by comparison with those activated during auditory speech perception in the same individuals⁵². By superimposing the activations detected in the two experiments (each contrasted against a resting baseline) we aimed to identify areas that might be used for integration during normal audio-visual speech perception. Clearly, direct comparison of bimodal and unimodal speech perception would have precluded the detection of areas involved in both processes. Silent lip-reading was first shown to activate, in addition to visual association cortex (BA 18, 19) and areas of putative heteromodal cortex (BA 37, 39, 40, 21/22), specific regions of lateral temporal cortex (BA 41, 42, 21/22) extending into the STS. These temporal cortical regions, which include primary auditory cortex, overlapped considerably with those activated during heard speech, providing a possible physiological mechanism by which seen speech influences the perception of heard speech.

When audio-visual speech was subsequently contrasted directly with silent lip-reading, activation was also differentially observed in the ventral claustrum (Talairach co-ordinates $x = 32$, $y = 0$, $z = 3$). This area receives and gives rise to multimodal cortical projections (including those from the auditory and visual systems) and was first highlighted by Ertlinger and Wilson⁴², who recognized its possible role in multimodal integration. Specifically, these authors proposed that this structure might operate, not as a repository for amodal representations, but as a relay station whereby the senses can access each other directly from their sensory-specific systems. The claustrum has also been recently implicated by Hadjikhani and Roland in visuo-tactile crossmodal matching⁵³. The precise role of this structure in crossmodal integration clearly warrants substantial investigation.

Using a different imaging technique, magnetoencephalography (MEG), which detects the weak magnetic signals associated with neural currents, Sams and colleagues⁵⁴ analysed cortical activity during auditory and audio-visual speech processing. Although this technique cannot match the spatial resolution of fMRI, it does provide substantially superior temporal resolution (of the order of milliseconds). Exploiting this advantage, these investigators were able to demonstrate that the characteristic response (M100 wave) of the auditory cortex to heard speech could be modified by the inclusion of visible speech information. This modulation was characterized by the appearance of a second wave in the primary auditory cortex 220 ms after the M100 wave.

Box 2. Neuroimaging studies of audio-visual speech perception

Modern neuroimaging techniques offer the potential to examine many aspects of the neural bases of multimodal interactions in human subjects. Although the initial integration of two or more sensory streams might depend on small clusters of multimodal neurons, the cognitive effects of such interactions could involve the participation of many other cortical and/or sub-cortical areas operating either in parallel or post-integration. Understanding the neural computations underlying the initial multisensory integrative process might ultimately depend on opportunistic electrophysiological inspection of multimodal neurons in pre-operative patients, together with comparative single-cell recording studies in other species. However, capturing the full range of activations elicited during multimodal processing, and the relative temporal involvement of different areas, will clearly benefit from the integration of neuroimaging techniques such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) (which offer whole head coverage and impressive spatial resolution), with methods such as electroencephalography (EEG), magnetoencephalography (MEG) and event-related potentials (ERPs), which provide millisecond-by-millisecond temporal resolution of neuro-electric activity.

The utility of combining data obtained from different imaging methodologies can be demonstrated by the experiments illustrated below, each of which examined different aspects of auditory and visual interactions during speech perception. Figure A displays the results of an fMRI study of silent lip-reading⁵. Five hearing subjects were scanned whilst viewing videotape of a lower half-face silently mouthing numbers and then, whilst listening to numbers being spoken, with the screen turned off. Median images were constructed⁶ showing areas of significant activation superimposed on a high-resolution structural MRI. Areas activated during silent lip-reading are shown in red and those activated whilst listening to speech shown in blue. Areas activated coincidentally during both experiments are shown in yellow. These regions of overlap fall in Brodmann areas 41 and 42 in primary auditory cortex and 22 and 21 in auditory association cortex.

Figure B shows the results from an MEG study by Sams and Imeda (pers. commun.). 122 neuromagnetic signals were measured during audio-visual stimulation and compared with the calculated summed individual response to auditory and visual stimuli. The yellow boxes indicate areas in the right temporoparietal cortex in which the responses to the combined stimuli differed significantly from the sum of the individual responses. Also apparent during audio-visual stimulation was a second wave of activity measured 220 ms after the M100 response elicited from auditory cortex in response to sound alone.

The results of both these experiments suggest that lip-reading contributes to speech perception by enhancing activity in the primary auditory regions, perhaps subsequent to integration of the two sensory streams in heteromodal regions proximal to the superior temporal sulcus. Activation of the primary auditory cortex by visible speech cues might proceed via back projections from heteromodal cortex.

The results from our experiments and those of Sams and his collaborators both suggest that the perception of visible speech can induce or modify activity in auditory cortex (see Box 2). The superior time resolution of MEG shows that processing of the visual component is somewhat delayed compared with that of pure acoustic speech. Such a delay is compatible with the possibility that after integration

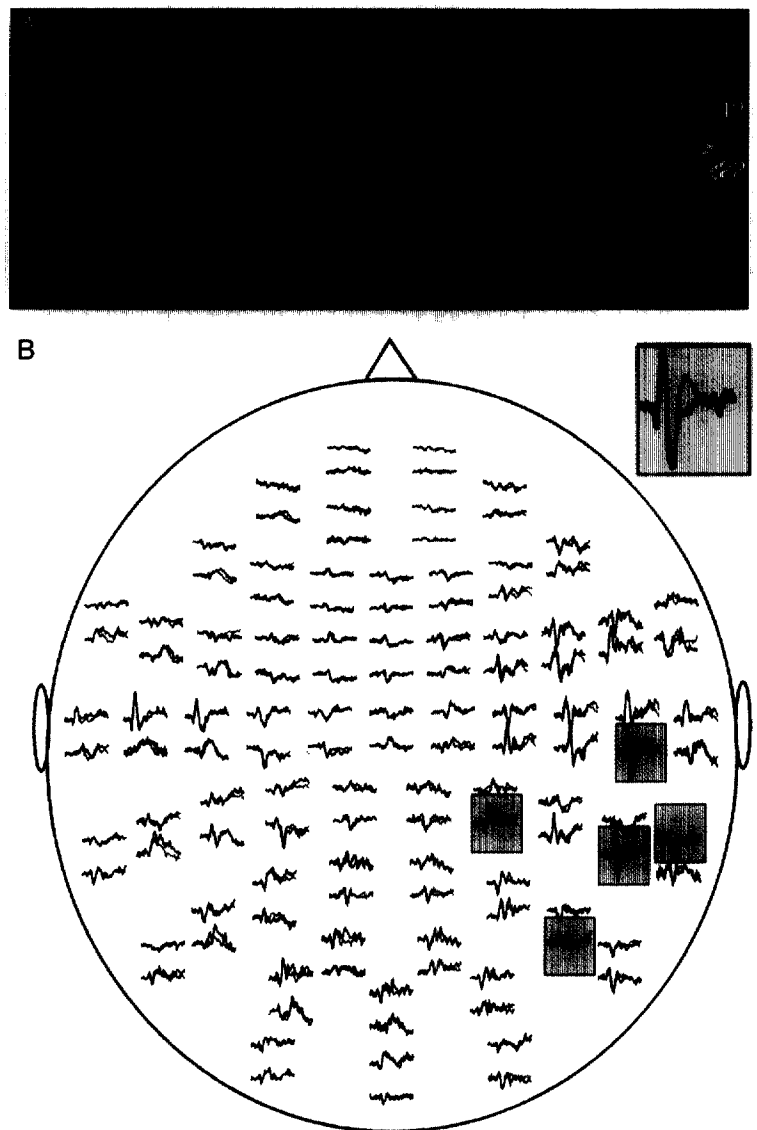


Fig. Audio-visual speech integration from fMRI and MEG. (A) Brain areas activated in an fMRI study whilst subjects viewed mouth movements without hearing any sound (shown in red), or whilst listening to speech (blue). Areas activated by both inputs are shown in yellow. (Modified from Ref. a.) **(B)** MEG signals measured during audio-visual presentation of speech. Signals surrounded by yellow boxes are those to both auditory and visual stimuli that differed significantly from responses to either modality alone. A second peak is apparent in the response in auditory cortex during audio-visual stimulation (shown in red; see inset) which is not present with auditory stimulation alone.

References

- a Calvert, G.A. et al. (1997) Activation of auditory cortex during silent lipreading *Science* 276, 593-596
- b Brammer, M.J. et al. (1997) Generic brain-activation mapping in functional magnetic resonance imaging: a non-parametric approach *Magn. Resonance Imag.* 15, 763-770

of the visible and auditory signals in association cortex (e.g. close to and including the superior temporal sulcus), information then feeds back to the auditory speech areas via back projections³⁵. This formulation gains feasibility in the light of psychological observations that the addition of visible speech is subjectively experienced as an improvement in hearing. Indeed, such a hypothesis is also consistent with

Outstanding questions

- Primary sensory cortex has now been shown to be activated by exogenous stimulation in its own modality, exogenous stimulation by another modality and (in the visual system at least) during mental imagery, yet all produce distinct perceptual experiences. What are the neural correlates of these differing experiences?
- Studies have shown that neonates are also capable of pairing different sensory inputs that occur in close spatial proximity, move together simultaneously, at the same rhythm and with the appropriate temporal microstructure. These findings prompt the question: to what extent is the capability to use sensory-specific information interchangeably and to combine multiple sensory inputs during perception innately determined or dependent upon environmental experience?
- How far do the principles underlying the neural integration of speech information generalize to other instances of crossmodal combination of non-speech stimuli?

Mesulam's suggestion^{43,44} that heteromodal cortex acts as a gateway through which one sensory modality can gain access to another.

Conclusions

As our understanding of the neural processes underlying modality-specific sensory perception increases, there is growing interest in the mechanisms by which information is combined between senses. In this review, we have focused on only one area of multimodal integration; that of audio-visual speech perception. However, it is clear from this discussion that our ability to probe and unravel its neural basis and those underlying the combination of information in other sensory modalities will necessitate a multifaceted approach integrating state-of-the-art techniques and traditional psychological methods. Combining the excellent temporal resolution of MEG or event related potentials (ERPs) with the superior spatial resolution of fMRI promises to be of great benefit in establishing the time course and route by which multisensory interactions occur in humans. However, a more complete understanding of the computations underlying these combinatory processes is likely to involve the integration of neuroimaging data in humans with electrophysiological studies in non-human primates, particularly for non-speech stimuli. We believe that such an approach holds great promise for revealing the details of the multisensory processes that characterize our interactions with the environment.

Acknowledgements

Oxford University Pump Priming and the Bethlem and Maudsley Trust supported the functional imaging research reported in this paper. G.C. acknowledges the support of the Medical Research Council, UK. We also thank the referees for their important and helpful comments on the manuscript.

References

- 1 Stein, B.E. and Meredith, M.A. (1993) *The Merging of the Senses*, MIT Press
- 2 Driver, J. and Spence, C. (1998) Attention and the crossmodal construction of space *Trends Cognit. Sci.* 2, 254-262
- 3 Stein, B.E., Huneycutt, W.S. and Meredith, M.A. (1988) Neurons and behaviour: the same rules of multisensory integration apply *Brain Res.* 448, 355-358
- 4 Stein, B.E. et al. (1989) Behavioural indices of multisensory integration: orientation to visual cues is affected by auditory stimuli *J. Cogn. Neurosci.* 1, 12-24
- 5 Berman, R.I. and Welch, R.B. (1976) The effect of degree of visual-auditory stimulus separation and eye position upon spatial interaction of vision and audition *Percept. Motor Skills* 43, 487-493
- 6 Choe, C.S. et al. (1975) The 'ventriloquist effect': visual dominance or response bias? *Percept. Psychophys.* 18, 55-60
- 7 Bertelson, P. and Radeau, M. (1981) Crossmodal bias and perceptual fusion with auditory-visual spatial discordance *Percept. Psychophys.* 29, 578-584
- 8 Sumby, W.H. and Polack, I. (1954) Visual contribution to speech intelligibility in noise *J. Acoust. Soc. Am.* 26, 212-215
- 9 McGurk, H. and MacDonald, J.W. (1976) Hearing lips and seeing voices *Nature* 264, 746-748
- 10 Saldana, H.M. and Rosenblum, L.D. (1993) Visual influences on auditory pluck and bow judgements *Percept. Psychophys.* 54, 406-416
- 11 Massaro, D.W. and Cohen, M.M. (1993) Perceiving asynchronous bimodal speech in consonant-vowel and vowel syllables *Speech Commun.* 13(1-2), 127-134
- 12 Summerfield, A.Q. and McGrath, M. (1984) Detection and resolution of audio-visual incompatibility in the perception of vowels *Q. J. Exp. Psychol.* 36A, 51-74
- 13 Bertelson, P. (1994) The cognitive architecture behind auditory-visual interaction in scene analysis and speech identification *Cahiers de Psychologie Cognitive* 13, 69-75
- 14 Stein, B.E. et al. (1996) Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis *J. Cogn. Neurosci.* 8, 497-506
- 15 Fisher, B.D. (1992) Integration of visual and auditory information in the perception of speech events *Dissertation Abstracts International*, 52, 33248
- 16 Bertelson, P. et al. (1994) Exploring the relation between McGurk interference and ventriloquism *Proc. 1994 Int. Conf. on Spoken Lang. Process.* 2, 559-562
- 17 Jones, J.A. and Munhall, K.G. (1997) The effects of separating auditory and visual sources on audio-visual integration of speech *Canadian Acoustics* 2, 13-19
- 18 Fisher, B.D. and Pylyshyn, Z.W. (1994) The cognitive architecture of bimodal event perception: a commentary and addendum to Radeau, 1994 *Cahiers de Psychologie Cognitive* 13, 92-96
- 19 Jack, C.E. and Thurlow, W.R. (1973) Effects of degree of visual association and angle of displacement on the 'ventriloquism' effect *Percept. Motor Skills* 37, 967-979
- 20 Welch, R.B. and Warren, D.H. (1980) Immediate perceptual response to intersensory discrepancy *Psychol. Bull.* 88, 638-667
- 21 Massaro, D.W., Cohen, M.M. and Smeele, P.M. (1996) Perception of asynchronous and conflicting visual and auditory speech *J. Acoust. Soc. Am.* 100, 1777-1786
- 22 Munhall, K.G. et al. (1996) Temporal constraints on the McGurk effect *Percept. Psychophys.* 58, 351-362
- 23 Bertelson, P., Vroomen, J. and de Gelder, B. (1997) Auditory-visual interaction in voice localization and in bimodal speech recognition: the effect of desynchronisation, in *Proc. Workshop on Audio-visual Speech Processing, Cognitive and Computational Approaches* (Benoit, C. and Campbell, R., eds), Rhodes, 26-27 September, 1997, pp. 97-100
- 24 Ungerleider, L.G. and Mishkin, M. (1982) Two cortical visual systems, in *The Analysis of Visual Behavior* (Ingle, D.J., Mansfield, R.J.W. and Goodale, M.D., eds), pp. 549-586, MIT Press
- 25 Knudsen, E.I. and Brainard, M.S. (1995) Creating a unified representation of visual and auditory space in the brain *Annu. Rev. Neurosci.* 18, 19-43
- 26 Welch, R.B. and Warren, D.H. (1986) Intersensory interactions, in *Handbook of Perception and Human Performance (Vol. 1: Sensory Processes and Perception)* (Boof, K.R., Kaufman, L. and Thomas, J.P., eds), pp. 1-36, John Wiley & Sons
- 27 Ellis, W.D. (1938) *A Source Book of Gestalt Psychology*, Kegan Paul, Trench, Trubner
- 28 Radeau, M. (1994) Auditory-visual spatial interaction and modularity. *Cahiers de Psychologie Cognitive* 13, 3-51
- 29 Summerfield, Q. (1987) Some preliminaries to a comprehensive account of audio-visual speech perception, in *Hearing by Eye: The*

Psychology of Lip-Reading (Dodd, B. and Campbell, R., eds), pp. 3-51, Erlbaum

- 30 Massaro, D.W. (1998) *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle*, MIT Press
- 31 King, A.J. and Palmer, A.R. (1985) Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus *Exp. Brain Res.* 60, 492-500
- 32 Frens, M.A. and van Opstal, A.J. (1997) Monkey superior colliculus activity during short-term saccadic adaptation *Brain Res. Bull.* 43, 473-483
- 33 Wallace, M.T., Wilkinson, L.K. and Stein, B.E. (1996) Representation and integration of multiple sensory inputs in primate superior colliculus *J. Neurophysiol.* 76, 1246-1266
- 34 Schneider, G.E. (1969) Two visual systems: brain mechanisms for localization and discrimination are dissociated by tectal and cortical lesions *Science* 163, 895-902
- 35 Wurtz, R.H. and Albano, J.E. (1980) Visual-motor function of the primate superior colliculus *Annu. Rev. Neurosci.* 3, 189-226
- 36 Meredith, M.A., Nemitz, J.W. and Stein, B.E. (1987) Determinants of multisensory integration in superior colliculus neurons: I. Temporal factors *J. Neurosci.* 10, 3215-3229
- 37 Meredith, M.A. and Stein, B.E. (1986) Spatial factors determine the activity of multisensory neurons in cat superior colliculus *Brain Res.* 365, 350-354
- 38 Stein, B.E. and Wallace, M.T. (1996) Comparisons of crossmodality integration in midbrain and cortex, in *Progress in Brain Research* (Norita, M., Bando, T. and Stein, B., eds), 112, 289-299
- 39 Pandya, D.N. and Seltzer, B. (1985) Association areas of the cerebral cortex *Trends Neurosci.* 5, 286-290
- 40 Banati, R.B. et al. (1997) The functional anatomy of sensory integration in man *Neuroimage* 5(4), S213
- 41 Chee, M.W. et al. (1997) Auditory and visual word processing studied with fMRI *Neuroimage* 5(4), S557
- 42 Ettlinger, G. and Wilson, W.A. (1990) Cross-modal performance: behavioural processes, phylogenetic considerations and neural mechanisms *Behav. Brain Res.* 40, 169-192
- 43 Mesulam, M-M. (1990) Large-scale neurocognitive networks and distributed processing for attention, language and memory *Ann. Neurol.* 28, 597-613
- 44 Mesulam, M-M. (1994) Neurocognitive networks and selectively distributed processing *Revue Neurol. (Paris)* 150, 564-569
- 45 Streicher, M. and Ettlinger, G. (1987) Crossmodal recognition of familiar and unfamiliar objects by the monkey: the effects of ablation of polysensory neocortex or of the amygdaloid complex *Behav. Brain Res.* 23, 95-107
- 46 Mesulam, M-M. et al. (1992) Differential cholinergic innervation within functional subdivisions of the human cerebral cortex: a choline acetyltransferase study *J. Comp. Neurol.* 318, 316-328
- 47 Seltzer, B. and Pandya, D.N. (1978) Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey *Brain Res.* 149, 1-24
- 48 Jones, E.G. and Powell, T.P.S. (1970) An anatomical study of converging sensory pathways within the cerebral cortex of the monkey *Brain* 93, 37-56
- 49 Wernicke, C. (1874) *Der Aphasische Symptomenkomplex*, M. Cohn and Weigert, Breslau, Poland
- 50 Graziano, M.S.A. and Gross, C.G. (1995) The representation of extrapersonal space: a possible role for bimodal, visual-tactile neurons, in *The Cognitive Neurosciences* (Gazzaniga, M., ed.), pp. 1021-1034, MIT Press
- 51 Ogawa, S. et al. (1990) Oxygenation-sensitive contrast in magnetic resonance image of a rodent brain at high magnetic fields *Magn. Resonance Med.* 14, 68-78
- 52 Calvert, G.A. et al. (1997) Activation of auditory cortex during silent lipreading *Science* 276, 593-596
- 53 Hadjikhani, N. and Roland, P.E. (1998) Cross-modal transfer of information between the tactile and the visual representations in the human brain: a positron emission tomographic study *J. Neurosci.* 18, 1072-1084
- 54 Sams, M. et al. (1991) Seeing speech: visual information from lip movements modified activity in the human cortex *Neurosci. Lett.* 127, 141-145
- 55 Zeki, S. and Shipp, S. (1988) The functional logic of cortical connections *Nature* 335, 311-317

Advanced Notice

Two Special Issues - *Trends in Cognitive Sciences* and *Trends in Neurosciences*

The increasing interest and understanding of the development, and motor and cognitive functions of the cerebellum is highlighted in a series of articles that have been commissioned for two special issues of *TICS* and *TINS*.

The special issues, commissioned with the assistance of the special guest editor Peter Strick, Syracuse, NY, USA, will present the latest information from leading scientists in the fields of anatomy, gene expression, development, conditioning, learning, neuroimaging, modelling, and cognitive function. The short review articles will provide a comprehensive introduction to the key issues in current cerebellar research for specialists and non-specialists alike.

Subjects areas covered in the special issues will include the following:

Development and developmental genetics of the cerebellum

Hereditary ataxias

Function of the inferior olive

Long-term depression

The cerebellum in motor learning and cognition

Conditioned reflexes and cerebellar learning

Neuroimaging of language, learning and memory in the cerebellum

Cerebellar dysfunction and cognition

Computational models of cerebellar function