Multisensory integration, perception and ecological validity

Beatrice De Gelder[1,2] and Paul Bertelson[1,3]

[1]Cognitive and Affective Neurosciences Laboratory, Tilburg University, 5000 LE Tilburg, The Netherlands
[2]MGH-NMR Center, Harvard Medical School, Bldg 36, First Street, Charlestown, Boston, MA 02129, USA
[3]Laboratoire de Psychologie Experimentale, Universite Libre de Bruxelles, B-1050 Bruxelles, Belgium

Studies of multimodal integration have relied to a large extent on conflict situations, in which two sensory modalities receive incongruent data concerning one aspect of the source. Exposure to such situations produces immediate crossmodal biases as well as longer lasting aftereffects, revealing recalibrations of data-to-percept matches. In the natural environment, such phenomena might be adaptive, by reducing the perturbing effects of factors like noise or growth-induced changes in receptor organs, and by enriching the percept. However, experimental results generalize to real life only when they reflect automatic perceptual processes, and not response strategies adopted to satisfy the particular demands of laboratory tasks. Here, we focus on this issue and review ways of addressing it that have been developed recently.

Research on perceptual processing, whether behavioural or physiological, has generally considered one sense modality (sight, hearing, touch, smell, etc) at a time. Yet, most events in the natural environment generate stimulation to several modalities. An explosion simultaneously emits light, noise and heat, and experiencing all of these together make for a richer percept than each individually; a speaker produces facial movements in a predictable temporal relationship to corresponding speech sounds and experiencing both together can provide a more adequate percept. Such availability of partially redundant data about the same objects or events provides perceiving systems, whether biological or artificial, with opportunities for compensating for various modality-specific disturbances.

Research for more than 100 years on multisensory integration (MSI) has shown that animals and humans alike effectively take advantage of these opportunities [1]. In this article, we shall briefly summarize the main evidence from human behavioural work and discuss its relevance for the central issue of ecological function. We shall argue that to generalize to real life, experimental findings must reflect genuine perceptual processes rather than response strategies adopted to satisfy the specific demands of particular laboratory tasks.

Experimental conflict situations

One major instrument of MSI research has been the experimental conflict situation, in which two modalities receive incongruent data regarding one particular aspect of the environment. Contrary to a frequent misconception, such conflicts occur not only in the laboratory, but also in the natural environment, and artificial experimental conflicts can be seen as simulations of natural ones.

Natural conflicts result from two main types of factors. One is moment-to-moment variability, or noise, in either impinging stimulation or subsequent processing. This noise usually affects one modality specifically and thus produces transient crossmodal incongruence. Combining information across modalities can bring compensation for these disturbances. The other source of natural incongruence consists of longer lasting changes in bodily parameters affecting perceptual input. The classic example here is the effect of growth on inter-aural distance, and consequently on time and intensity differences at the two ears as the basis of sound localization. In situations of aural ambiguity, the seen location of the sound's source can be used to partially 'recalibrate' sound processing in the correct direction. Other long-lasting changes can result from sensory handicap, or from spontaneous drift in the processing rules.

In the laboratory, two main reactions to intermodal conflict have been demonstrated: 'immediate effects', observed in presence of the conflicting data, and 'after-effects', affecting responses to unimodal data following a period of exposure. They correspond to the two types of corrections – for transient noise and for longer-lasting deviations – that we considered for natural conflicts.

Spatio-temporal conflicts: immediate effects
A prime example of an immediate effect is the apparent attraction of a sound by a visual stimulus presented simultaneously in a separate location [2–5]. It is also called 'the ventriloquist effect', referring to the performing ventriloquist, who speaks without visible lip movements while agitating a puppet in synchrony with his speech. The successful ventriloquist creates in his audience the illusion that the words come from the puppet. The phenomenon has been studied in the laboratory using a 'selective unimodal localization task' [6,7]. Participants are instructed to point towards sound bursts whilst ignoring point flashes of light delivered synchronously at some

Corresponding author: Beatrice De Gelder (b.degelder@uvt.nl).

small distance. Pointing typically deviates towards the (to-be-ignored) flashes. The reverse effect, auditory bias of visual location, is much smaller, although sometimes above chance. Even the visual bias of auditory location is generally smaller than the objective distance between the stimuli, showing that the interaction does not reduce, as was sometimes believed, to a substitution of visual data for the auditory data, and implies real integration of information from the two modalities. In another task, 'discordance detection', subjects decide whether audio–visual stimulus pairs come from a single location or from separate ones [8–9]. A 'single' judgment indicates that the sum of the two biases has brought the apparent distance below some detection criterion.

Audio–visual biases depend crucially on synchroniz-ation of the bimodal inputs [10–12]. This dependence makes good functional sense. To play a useful role in the normal environment, crossmodal interactions must be applied mainly to multimodal inputs from a same source, which generally arrive simultaneously.

Spatial biases have also been demonstrated for other modality pairs: bias of proprioception by vision [13], touch by vision [14], and audition by touch [15–17]. Another recently demonstrated bias is the effect of a visual distracter on the apparent direction of motion of a target sound when the two stimuli are moving simultaneously in the opposite direction [18,19]. A new line of study currently focuses on audio–visual biases in the time dimension, or 'temporal ventriloquism' [20–22]. For example, the appa-rent temporal separation between asynchronous sound bursts and light flashes seemed shorter when these were presented in the same location rather than in separate ones [21]. It appeared also that, in this interaction, audition played the stronger role [20,22]. This finding brings new support for Welch's 'modality appropriateness hypothesis', which relates relative modality dominance to the accuracy with which the task's target dimension is measured in each modality [5]. An innovative development along a similar line has been the successful prediction of modality dominance by a model based on a maximum likelihood principle [23].

Identity biases

MSI and perceptual biases (which are only part of the MSI picture) do not occur exclusively along spatio-temporal dimensions, and can equally operate at the level of recognition of stimulus identity. A well-known case is the 'McGurk effect' [24,25]. When a spoken syllable like /ba/ is dubbed onto the visual presentation of a face articulating an incompatible syllable (e.g. /ga/, pronounced, unlike /ba/, without lip closure) subjects often report hearing a compatible syllable (e.g. /da/). The effect requires a minimum of crossmodal synchronization, in the same way as ventriloquism [26–27].

MSI between non-linguistic inputs has been relatively less studied. For example, categorization of musical sounds as 'bow' or 'pluck' along a bow-to-pluck continuum was biased by the sight of hands either bowing or plucking a cello [28]. This result could reflect either genuine perceptual interactions or post-perceptual decisions based on familiarity with musical instruments sounds,

or both. Other recently reported examples are the effect of an added sound on the interpretation of an ambiguous sequence of visual motion [29], and a surprising visual illusion induced by the presentation of incongruent numbers of auditory beeps [30].

Finally, a new series of investigations has documented mutual biases between visible and audible expressions of emotion, providing evidence for an 'emotional McGurk effect' [31–33]. For instance, faces from a continuum of expressions (e.g. sad to happy) were presented simul-taneously with a semantically neutral speech utterance pronounced in an emotional tone corresponding to one or other end-point of the face continuum (see Figure 1), or conversely, utterances from an affective continuum, together with faces from the end point [33]. The two pos-sible biases, of facial expression judgment by voice tone, and of voice tone judgment by facial expression, were obtained, despite instructions to focus on the target modality.

Aftereffects

Aftereffects are behavioural modifications induced by exposure to conflicting inputs. They have been studied mainly for cases of spatial conflict, ventriloquism and prismatically displaced vision of the subject's own body parts. This last situation was actually the focus of the first systematic movement of research on multimodality in the
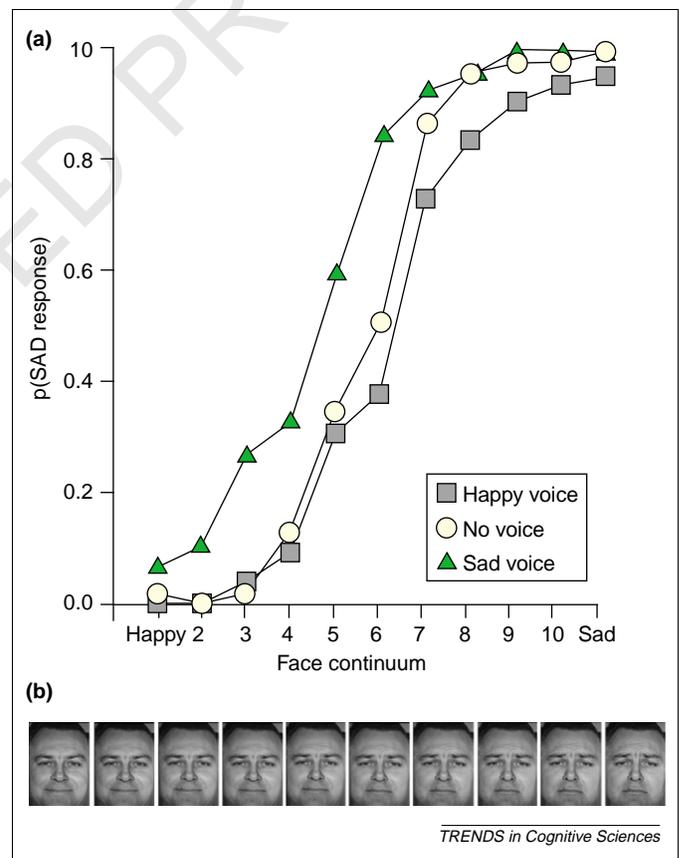
Figure 1. Emotional voices influence the categorization of facial expressions. Faces from a happy-to-sad morphed continuum were presented either alone or in syn-chrony with a voice pronouncing a semantically neutral sentence in a happy or a sad tone, and the subjects were instructed to judge whether the face was sad or happy, ignoring the voice. Nevertheless, the proportion of 'sad' judgments increased with the sad voice and decreased with the happy voice. Redrawn from [33].

1960s (see [34,35] and review in [36]). Extension to the ventriloquism situation was stimulated by theoretical issues raised by this tradition in prismatic displacement [2,3]. 'Aftereffects of ventriloquism' are typically studied by asking subjects to point to unimodal sound or light targets before and after a period of exposure to spatially separate but simultaneous sounds and light flashes. Post-exposure responses are generally shifted in the direction occupied by the distracter during exposure [37–39].

The occurrence of aftereffects shows that exposure to intermodal discrepancy 'recalibrates' sensory input-to-percept relationships. Recalibration is probably the mechanism through which the operation of the different modalities are kept coordinated in the face of changes resulting from spontaneous drift, growth or sensory handicap [3,34,35].

An important question, first considered by Bedford [40], is that of how specific the recalibration is to the particular conditions of the exposure situation. Bedford found that aftereffects of exposure to visual–proprioceptive conflict in one location generalized to a wide range of testing locations. More recently, Recanzone [39] reported that visual spatial recalibration induced with tones at a particular frequency did not generalize to frequencies two octaves away. This result has interesting implications concerning the mechanisms of sound localization, but was not replicated in a more extensive recent study [41].

The processing-levels issue

Until recently, most data purporting to demonstrate MSI consisted of voluntary responses of participants, such as verbal judgments, setting stimulus values or pointing. With such data, one is led to ask whether the observed interactions originate in automatic, mandatory perceptual processes, or in later decisional ones. This question has often been asked at some stage in the development of many areas of perception research, although convincing answers have not always followed.

Human responses are relevant to the general issue of intermodal coordination only if they reflect basic perceptual processes, rather than specific strategies adopted to satisfy the demands of particular laboratory tasks. The case of the selective-response paradigm, through which immediate bias effects were often studied serves to illustrate this point. The fact that biases occur in spite of instructions to ignore the non-target stimuli has sometimes been taken as supporting evidence for their automatic nature. The argument effectively suggests that responses are to some degree mandatory, but it is not a strong one. In fact, once discordance is detected consciously, it is still up to the subject to decide what to do with the experimenter's instructions. Indeed, these instructions can lead subjects to wonder why the distracter items they are asked to ignore were presented at all.

To overcome these difficulties, several ways of avoiding the contamination of MSI effects by post-perceptual factors have been introduced recently.

Undetected discrepancies

One approach is to measure interactions with intermodal discrepancies that are so small that participants have no awareness of their existence. Such a situation can be achieved using a modified psychophysical staircase method that was developed by Bertelson and Aschersleben [12] and used to demonstrate automaticity of the visual bias of sound location (see Box 1). The crucial aspect of this method is that the possibility of bias is examined at moments when the occurrence of response reversals demonstrates that the subject has become uncertain concerning the direction of the crossmodal discordance. The method has now also been applied to examine the effects of reflexive attention to the visual distracter in ventriloquism [42], a new bias of tactile location by auditory distracters [15], and temporal ventriloquism [21].

Indirect methods

In an experiment with the traditional 'cocktail party' situation, Driver [43] had listeners repeat one of two word sequences delivered simultaneously via a single loudspeaker. The target sequence was also presented as a video of the talker's face on a screen either close to the loudspeaker or at a distance from it. Performance was significantly better with the talker's face on the screen separated from the loudspeaker than on the proximal screen. Presumably, in that condition the target items were attracted towards the synchronously moving face, and this separated them spatially from the (non-synchronous) non-target items, a condition well known for facilitating selective listening. In this elegant experiment, a visual bias of auditory location was demonstrated indirectly through its effect in a non-spatial task, and this could therefore not be explained by any voluntary adjustment by the subject.

Several other applications of the indirect approach have been reported recently [44–46]. One interesting example concerns the visual bias of auditory motion direction [19]. This study took advantage of a contingent auditory motion aftereffect [47], in which listening to sounds with a falling pitch moving in one direction, alternating with sounds with a rising pitch moving in the opposite direction, resulted in the impression that stationary sounds moved in one or the opposite direction, depending on whether their pitch was rising or falling. This provides an auditory analogue of the well-known McCullough effect in vision [48]. Furthermore, adding a visual object (moving simultaneously with the sound in either the congruent or an incongruent direction) significantly enhanced the auditory aftereffect [19]. This result carries convincing proof of the perceptual nature of the visual bias of auditory motion, which earlier studies failed to provide [18].

Evidence from brain pathology

Patients who have lost the capacity to form conscious representations of particular forms of sensory stimulation (e.g. visual input) as a result of brain damage can sometimes still process (visual) stimuli unknowingly. Finding that sensory inputs of which a patient is not aware can bias perception in other modalities can have important implications for understanding the functional locus of MSI.

We have examined the visual bias of auditory location in cases of visual unilateral neglect, a condition characterized by an inability to attend to stimuli presented to one

Box 1. A modified psychophysical staircase method

On each trial, a train of three tone bursts is delivered in an apparent location controlled through time difference stereophony (time interval between presentations on two separate loudspeakers) and the participant indicates by pressing one of two keys whether the tones came from left or right of the display's centre (Figure Ia). Sound locations are chosen according to two randomly mixed staircases, one starting far to the left and the other far to the right (Figure Ib). When the 'left' response is given, the following sound targets on the same staircase are moved one step to the right, and vice versa after a 'right' response. This procedure necessarily results in the two staircases progressively converging towards the centre. Initially, the correct response is provided repeatedly on each staircase, so that, except for occasional mistakes, the convergence is monotonous. Then, at some point, 'response reversals', that is, responses different from the preceding one on the same staircase, begin to occur. From this point

on, the participant is uncertain regarding the location of the target sound, so that no particular response strategy could be reliably applied.

To examine the visual bias of apparent sound location, light flashes were produced in a central location in synchrony with the sounds. In this condition, reversals started occurring with the tones farther away from centre than in a control condition with no flashes, showing that the flashes effectively attracted the sounds. This attraction was observed at a time when, given the occurrence of reversals, it could not possibly be of voluntary origin. The effect entirely disappeared when flashes and sounds were desynchronized.

Pooling locations for successive *reversals*, and not for successive trials (as is more usual when using staircases) is essential for the crucial purpose of this modified method, which is to isolate data recorded in the uncertainty region as opposed to pre-uncertainty regions.
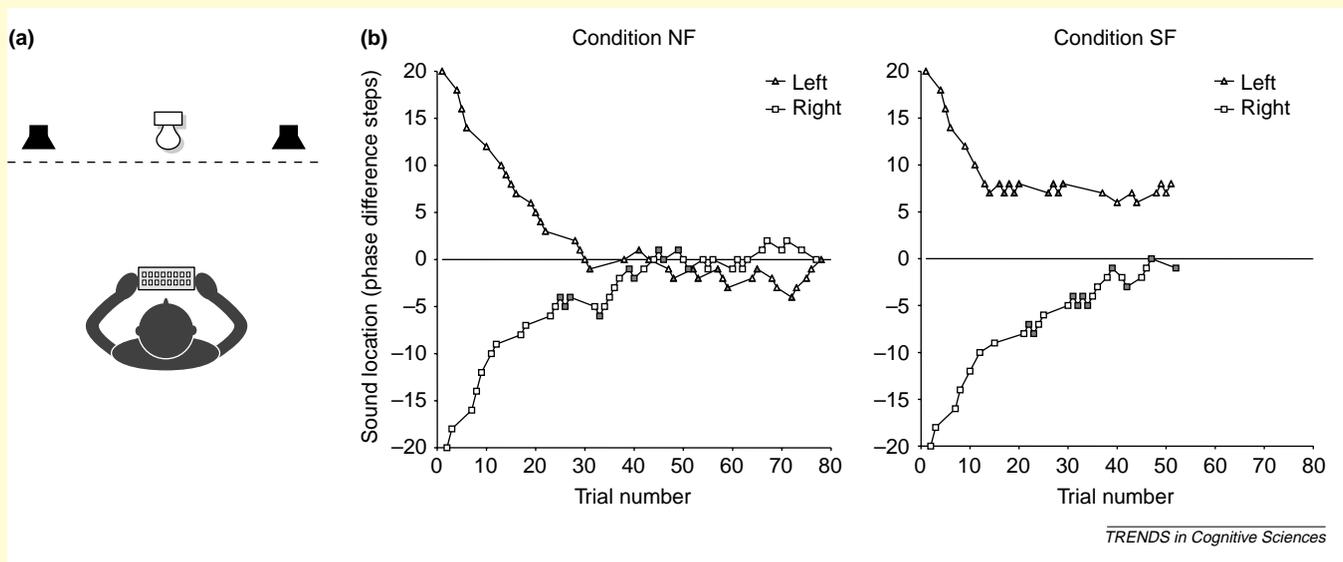
Figure I. Visual bias of perceived auditory location measured with psychophysical staircases. (a) Schema of experimental set-up. The subject indicates by pressing one of two keys whether the sound appeared to come from left or right of centre. A light flash can be synchronized or desynchronized with the sound. (b) Examples of explorations in control and experimental conditions. Abcissa: successive trials. Ordinate: sound location perceived by subject, in number of phase difference steps; up ? left; down ? right. After a 'left' response, the actual sound location is moved one step to the right for next trial, and vice versa after a 'right 'response. In a typical exploration in the control (no flash, NF) condition, each staircase moves monotonously towards the centre, until response reversals (marked by filled symbols) begin to occur. The exploration stops when 10 reversals have occurred on each staircase. In the synchronized flash (SF) condition, an exploration by same subject shows reversals beginning to occur when the staircases are further apart than in the no flash condition. Redrawn from [12].

side of the visual field, usually the left field [49]. Subjects were unable to detect bright light flashes delivered to the left of a fixation point. But when instructed to point to the location of auditory tones presented centrally in synchrony with such flashes, their responses showed small but significant attractions towards the neglected left flashes (Figure 2). By contrast, fully detected flashes to the right of fixation produced only erratic and non-significant effects, presumably the result of varying voluntary strategies.

Similar evidence about the effects of audiovisual integration from brain damage is provided by recent work on MSI of emotion (see Box 2). A patient with a face recognition deficit ('prosopagnosia'), and unable to identify facial expressions, was asked to name the emotional tone in which a spoken sentence was pronounced, a task she could do without difficulty. When a face displaying an incongruent expression was presented simultaneously with an emotional tone of voice (for example, an angry

face together with a fearful tone of voice), her judgment was biased in the direction of the unrecognized facial distracter [50].

Results with similar implications have been obtained in hemianopic patients with residual non-conscious vision. These patients have lost awareness of stimuli from the region of the visual field corresponding to the damaged region of cortex, but can nevertheless perform above chance on some indirect tasks based on such unrecognized items [51]. Recent findings obtained with two cortically blind patients indicated that a facial expression presented to the blind visual field influenced how the emotion expressed in a spoken sentence was processed [52]. Interestingly though, a fear-inspiring picture (e.g. a snake) influenced the auditory response only when presented in the intact field, suggesting that some multi-sensory interactions involve mechanisms related to conscious vision whereas others do not. This clearly shows that awareness of the stimuli does play a role in MSI.

Box 2. The neural basis of audiovisual perception of emotion

Functional imaging studies in humans have started to identify some of the brain regions involved in MSI. At present generalizations are still difficult because the phenomena considered are quite diverse. Audio–visual situations studied include arbitrarily associated minimalist pairs, audio–visual speech [55,56] and audio–visual emotion [57]. The choice of control conditions and tasks is also rather diverse. Sometimes arbitrary audio–visual pairs have been used as control conditions for audio–visual speech [58], whereas other studies used meaningless grimaces [59]. Significant diversity also results from ways in which MSI is operationalized. One approach is to compare unimodal and bimodal situations. Inspired by animal electrophysiology [60], MSI effects in brain imaging studies have been quantified as response enhancement and applied to the MSI of speech [61]. By contrast, EEG and MEG studies tend to find suppressive effects [56]. Another approach is to compare bimodal congruent with bimodal incongruent scenarios, in which case crossmodal inhibition is observed for the incongruent case [52]. Depending on the situation studied, specific hypotheses can be formulated about the underlying mechanism. In an fMRI study, a

crucial component of the mechanism for crossmodal binding in the case of fearful face and voice pairs was found in the amygdala [57]. Subjects heard auditory fragments paired with either a congruent or an incongruent facial expression (happiness or fear) and subjects were asked to judge the emotion from the face. When fearful faces were accompanied by short sentence fragments spoken in a fearful tone of voice an increase in activation was observed in brain regions that included the amygdala and the fusiform gyrus (Figure I), suggesting an important role of the amygdala for binding of face and voice expressions.

An intriguing possibility is that presentation in one modality activates areas typically associated with stimulation in the other modality. Thus activation of auditory cortex has been observed for visual speech and visual cortex activation was observed for emotional voices [62]. An interesting topic for future research is whether this pattern of a 'crossed activation by implicit association' obtains specifically for naturalistic pairs where co-occurrence of the two components is the norm, as is the case for situations like audio–visual speech and emotion.
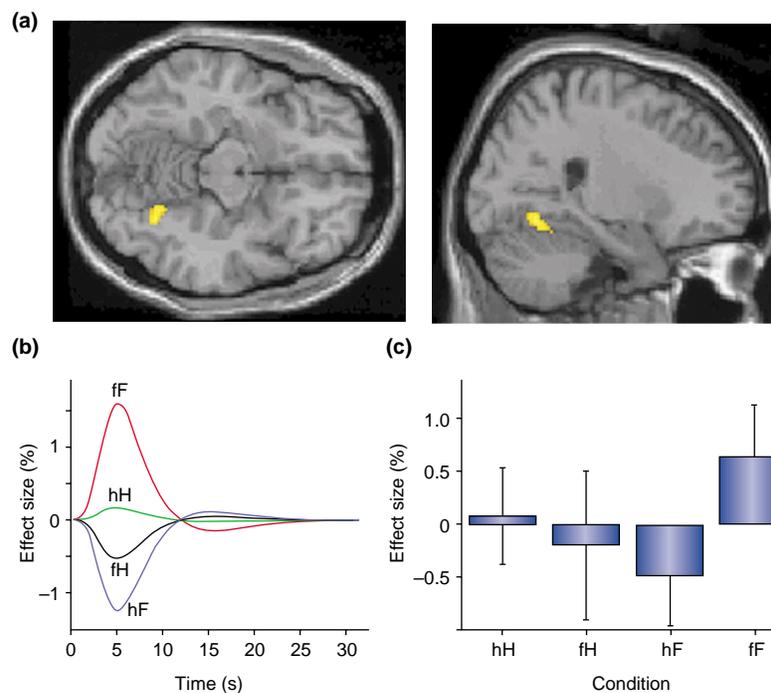
Figure I. Crossmodal binding of visual and auditory stimuli in the fusiform gyrus. Subjects listened to a sentence spoken in either a happy or fearful tone of voice, and at the same time were presented with either a happy or fearful face. (a) fMRI activation in the right fusiform gyrus (yellow), shown in transverse and sagittal views. The level of fusiform gyrus activation depended on the congruence or incongruence of the stimuli. The largest increase was observed when a fearful voice was presented with a fearful face (fF condition), compared with the other conditions: hH, happy voice with happy face; fH, fearful voice with happy face; hF, happy voice with fearful face. (b) Plots of the fitted haemodynamic responses in fusiform gyrus in the four experimantal conditions. (c) Estimates of the responses at the voxel of maximum significance. Modified from [57].

Future directions

Should theories of MSI proceed in a bottom-up fashion in the hope that understanding simple cue combinations yields the recipe for understanding complex naturalistic situations? The old debate on whether audio–visual MSI required realistic situations was closed with the consensus that situations using minimalist stimulus pairs using simple cues were all that was needed to elicit robust MSI [1]. This suggested that ecological validity did not hinge on the use of naturalistic stimuli. In line with animal studies,

human MSI research is now also beginning to explore naturalistic situations, including communicative signals exchanged by biological organisms whose brains evolved for dealing with such environments. The challenge for the near future will be to design perceptual experiments simulating MSI in naturalistic environments. Sophisticated modelling of MSI will go hand-in-hand with much needed progress in theory. We will end by briefly listing what we see as some important steps along that path.
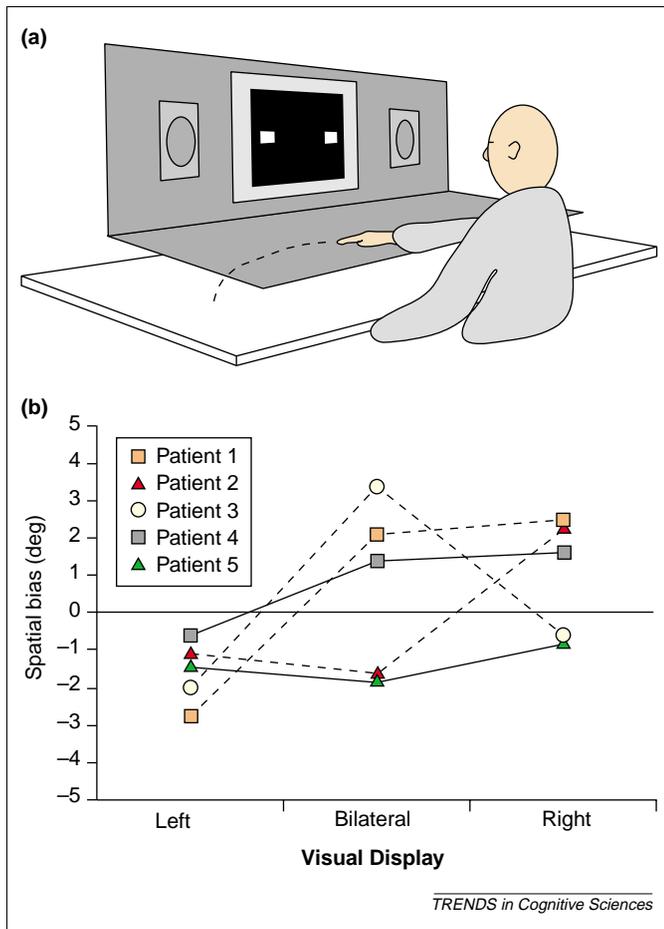
Figure 2. Visual bias of perceived sound location in patients with severe unilateral neglect. **(a)** Experimental set-up: the subject points with hidden hand to apparent location of a tone delivered centrally, in synchrony with presentation of a small digit (in centre of screen), either alone or together with a bright white square to left, right or both sides of digit. The subject also reports the digit and squares ('left', 'right', 'bilateral' or 'none'). The left square is practically never reported, probably because of extinction caused by attention to central digit. **(b)** Results of five neglect patients, showing visual bias of sound location (mean pointed location on trials with each visual display minus no-square trials). Positive signifies a bias to the right. There is a small but systematic attraction towards the undetected left square. Effects of the detected right square on both 'right' and 'bilateral' trials were erratic (and therefore non-significant), probably reflecting individual differences in post-perceptual strategies. Healthy subjects tested in the same situation had significant biases towards both left and right distracters, and none with bilateral ones. Redrawn from [49].

A framework for classifying multisensory signals

With the research field expanding, what is needed most urgently is a classification of different kinds of MSI situations. Traditionally, quantification and modelling of simple cue-redundancy-based MSI has attracted most of the attention. The adaptive function of MSI includes response enhancement, taking advantage of redundancy (still the most studied case), but also contributing to enrichment of the percept (e.g. the combination of colour and smell stimulate appetite more than each alone). Moreover, different cases of cue redundancy must be distinguished even for spatio−temporally paired cues, because effects of cue pairing can be additive, cumulative, inhibitory or lead to an emergent and novel percept. For example, the McGurk effect in fact consists of two quite different effects because some audio−visual speech combinations give rise to blends (with combining components still intact, visual /ba/ and auditory /da/ is perceived as

/abda/) and others to fusions (visual /ga/ and auditory /ba/ is perceived as /da/, an emergent percept), amounting to a qualitative difference between the two situations. More generally, cases of MSI-based enrichment of percepts or of emergent properties may require a different approach from cases of cue redundancy.

Another aspect of MSI that should be part of any classification concerns the kind of relationship that exists between the members of a pair. 'Arbitrary' pairs are those created for the purpose of an experiment (for example, a low-frequency tone paired with a square and a high-frequency tone with a rectangle). By contrast, naturalistic pairs consist of components that tend to co-occur in natural environments. It might be that ignoring these distinctions for the sake of an abstract theory and a generalized method is one way of making progress, but the alternative is equally worth considering. In the latter case, only detailed descriptions of each specific MSI situation will tell us how best to approach each particular case.

MSI and neural mechanisms

How does integration at the perceptual level relate to underlying neural mechanisms? A classification of types of MSI will be very useful for future hypotheses about underlying mechanisms and how best to model and measure these. At present MSI is measured variously at the cellular level, at the cell population level, and at the whole brain level. Neurophysiological mechanisms proposed for MSI include summation, suppression, inhibition and neural synchrony. Although all of these belong under the umbrella of integration, each means something specific only within the context of the neurophysiological method used. Will rules of translation be found that bridge all these levels and combine spatial with temporal information, the cellular level with the functional neuroanatomy? Perhaps, but the specific MSI situation (redundancy or enrichment, cue summation or inhibition, minimalist or naturalistic pairs) is likely to be an important factor in formulating specific predictions and applying the best methods. Thus, a classification of MSI scenarios will provide a roadmap for research by indicating what type of effects to expect and which method and baseline to choose.

MSI and consciousness

The notion of separate and qualitatively different routes for conscious and non-conscious processing is now actively investigated in visual perception (e.g. [51]). A similar distinction might be needed for other sensory systems. However, the relationship between MSI and consciousness has not yet received much attention. With different systems for conscious and non-conscious visual perception, MSI might well be different depending on whether or not the visual component of a stimulus pair is perceived consciously. As noted already, damage to primary visual cortex offers unique opportunities to investigate MSI in conditions of non-conscious awareness. In a recent fMRI study (following on from [52]) we investigated MSI for combinations of emotional voices paired with either face expressions or emotional scenes of which the subject was unaware. Preliminary results suggest that stimulus

awareness is an important factor for whether MSI is observed, and for which neural structures are involved.

MSI and qualia

Sensations have qualitative aspects, which contribute to making the representations they convey unique, as highlighted recently for example, in research on synaesthesia. An important difference between MSI in situations of redundancy versus situations of enrichment is that in the latter the quality of the percept is sometimes modified. A future challenge is to integrate these qualitative aspects, including their reward function, into more complete models of MSI.

References

1 Bertelson, P. and de Gelder, B. The psychology of multimodal perception. In *Crossmodal Space and Crossmodal Attention* (Spence, C. and Driver, J., eds), Oxford University Press (in press)

2 Radeau, M. (1994) Auditory–visual spatial interaction and modularity. *Curr. Psychol. Cogn.* 13, 3–51

3 Bertelson, P. (1998) Starting from the ventriloquist: the perception of multimodal events. In *Advances in Psychological Science* (Sabourin, M. *et al.*, eds), pp. 419–439, Erlbaum

4 Bertelson, P. (1999) Ventriloquism: A case of crossmodal perceptual grouping. In *Cognitive Contributions to the Perception of Spatial and Temporal Events* (Aschersleben, G. *et al.*, eds), pp. 347–362, Elsevier

5 Welch, R.B. and Warren, D.H. (1980) Immediate perceptual response to intersensory discrepancy. *Psychol. Bull.* 88, 638–667

6 Bermant, R.I. and Welch, R.B. (1976) Effect of degree of separation of visual-auditory stimulus and eye position upon spatial interaction of vision and audition. *Percept. Mot. Skills* 42, 487–493

7 Bertelson, P. and Radeau, M. (1981) Cross-modal bias and perceptual fusion with auditory–visual spatial discordance. *Percept. Psychophys.* 29, 578–584

8 Jack, C.E. and Thurlow, W.R. (1973) Effects of degree of visual association and angle of displacement on the 'ventriloquism' effect. *Percept. Mot. Skills* 37, 967–979

9 Choe, C.S. *et al.* (1975) The 'ventriloquist effect': visual dominance or response bias? *Percept. Psychophys.* 18, 55–60

10 Thomas, G.J. (1940) Experimental study of the influence of vision on sound localization. *J. Exp. Psychol.* 28, 163–177

11 Radeau, M. and Bertelson, P. (1987) Auditory–visual interaction and the timing of inputs: Thomas (1941) revisited. *Psychol. Res.* 49, 17–22

12 Bertelson, P. and Aschersleben, G. (1998) Automatic visual bias of perceived auditory location. *Psychon. B. Rev.* 5, 482–489

13 Hay, J.C. *et al.* (1965) Visual capture produced by prism spectacles. *Psychonom. Sci.* 2, 215–216

14 Pavani, F. *et al.* (2000) Visual capture of touch: Out-of-the-body experiences with rubber gloves. *Psychol. Sci.* 11, 353–359

15 Caclin, A. *et al.* (2002) Tactile 'capture' of audition. *Percept. Psychophys.* 64, 616–630

16 Jousmaki, V. and Hari, R. (1998) Parchment-skin illusion: sound-biased touch. *Curr. Biol.* 8, R190

17 Guest, S. *et al.* (2002) Audiotactile interactions in roughness perception. *Exp. Brain Res.* 146, 161–171

18 Soto-Faraco, S. *et al.* (2002) The ventriloquist in motion: illusory capture of dynamic information across sensory modalities. *Cogn. Brain Res.* 14, 139–146

19 Vroomen, J. and de Gelder, B. (2003) Visual motion influences the contingent auditory motion aftereffect. *Psychol. Sci.* 14, 357–361

20 Fendrich, R. and Corballis, P.M. (2001) The temporal cross-capture of audition and vision. *Percept. Psychophys.* 63, 719–725

21 Bertelson, P. and Aschersleben, G. Temporal ventriloquism: cross-modal interaction on the time dimension. 1. Evidence from temporal order judgements. *Int. J. Psychophysiol.* (in press)

22 Morein-Zamir, S. *et al.* (2003) Auditory capture of vision: examining temporal ventriloquism. *Cogn. Brain Res.* 17, 154–163

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

24 McGurk, H. and MacDonald, J. (1976) Hearing lips and seeing voices. *Nature* 264, 746–748

25 Massaro, D.W. (1999) Speechreading: illusion or window into pattern recognition. *Trends Cogn. Sci.* 3, 310–317

26 Bertelson, P. *et al.* (1997) Auditory–visual interaction in voice localization and in bimodal speech recognition: the effects of desynchronization. In *Proceedings of Rhodes Workshop on Audio-Visual Speech Processing: Cognitive and Computational Approaches* (Benoit, C. and Campbell, R., eds), pp. 97–100, European Cooperation in Scientific and Technical Research

27 Munhall, K.G. *et al.* (1996) Temporal constraints on the McGurk effect. *Percept. Psychophys.* 58, 351–362

28 Saldaña, H.M. and Rosenblum, L.D. (1993) Visual influences on auditory pluck and bow judgments. *Percept. Psychophys.* 54, 406–416

29 Sekuler, R. *et al.* (1997) Sound alters visual motion perception. *Nature* 385, 308

30 Shimojo, S. and Shams, L. (2001) Sensory modalities are not separate modalities: plasticity and interactions. *Curr. Opin. Neurobiol.* 11, 505–509

31 de Gelder, B. *et al.* (1995) Hearing smiles and seeing cries: the bimodal perception of emotions. *Bull. Psychonomic Soc.* 30, 196

32 Massaro, D.W. and Egan, P.B. (1996) Perceiving affect from the voice and the face. *Psychonomic Bull. Rev.* 3, 215–221

33 de Gelder, B. and Vroomen, J. (2000) The perception of emotion by ear and by eye. *Cogn. Emot.* 14, 289–311

34 Held, R. (1965) Plasticity in sensory-motor systems. *Sci. Am.* 213, 84–94

35 Harris, C.S. (1965) Perceptual adaptation to inverted, reversed, and displaced vision. *Psychol. Rev.* 72, 419–444

36 Welch, R.B. (1978) *Perceptual Modification: Adapting to Altered Sensory Environments*, Academic Press

37 Radeau, M. and Bertelson, P. (1974) The after-effects of ventriloquism. *Q. J. Exp. Psychol.* 26, 63–71

38 Radeau, M. and Bertelson, P. (1976) The effect of a textured visual field on modality dominance in a ventriloquism situation. *Percept. Psychophys.* 20, 227–235

39 Recanzone, G. (1998) Rapidly induced auditory plasticity: the ventriloquism aftereffect. *Proc. Natl. Acad. Sci. U. S. A.* 95, 869–875

40 Bedford, F. (1989) Constraints on learning new mappings between perceptual dimensions. *J. Exp. Psychol.: Human Percept. Perform.* 15, 232–248

41 Frissen, I. *et al.* (2003) The aftereffects of ventriloquism: are they sound-frequency specific? *Acta Psychol. (Amst.)* 113, 315–327

42 Vroomen, J. *et al.* (2001) The ventriloquist effect does not depend on the direction of automatic visual attention. *Percept. Psychophys.* 63, 651–659

43 Driver, J. (1996) Enhancement of selective listening by illusory mislocation of speech sounds due to lip reading. *Nature* 381, 66–68

44 Vroomen, J. and de Gelder, B. (2000) Sound enhances visual perception: cross-modal effects of auditory organization on vision. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 1583–1590

45 Spence, C. and Driver, J. (2000) Attracting attention to the illusory location of a sound: reflexive crossmodal orienting and ventriloquism. *Neuroreport* 11, 2057–2061

46 Vroomen, J. *et al.* (2001) Directing spatial attention towards the illusory location of a ventriloquized sound. *Acta Psychol. (Amst.)* 108, 21–33

47 Dong, C. *et al.* (1999) A contingent aftereffect in the auditory system. *Nat. Neurosci.* 2, 863–865

48 McCullough, C. (1965) Color adaptation of edge detectors in the human visual system. *Science* 149, 1113–1116

49 Bertelson, P. *et al.* (2000) Ventriloquism in patients with unilateral visual neglect. *Neuropsychologia* 38, 1634–1642

50 de Gelder, B. *et al.* (2000) Covert processing of faces in prosopagnosia is restricted to facial expressions: evidence from cross-modal bias. *Brain Cogn.* 44, 425–444

51 Weiskrantz, L. (1997) *Consciousness Lost and Found: A Neuropsychological Exploration*, Oxford University Press

52 de Gelder, B. *et al.* (2002) Fear recognition in the voice is modulated by unconsciously recognized facial expressions but not by unconsciously recognized affective pictures. *Proc. Natl. Acad. Sci. U. S. A.* 99, 4121–4126

55 Fuster, J.M. *et al.* (2000) Cross-modal and cross-temporal association in neurons of frontal cortex. *Nature* 405, 347–351

56 Calvert, G.A. *et al.* (1999) Response amplification in sensory-specific cortices during crossmodal binding. *Neuroreport* 10, 2619–2623

57 Dolan, R.J. *et al.* (2001) Crossmodal binding of fear in voice and face. *Proc. Natl. Acad. Sci. U. S. A.* 98, 10006–10010

58 Sams, M. and Hari, R. (1991) Magnetoencephalography in the study of human auditory information processing. *Ann. N. Y. Acad. Sci.* 620, 102–117

59 Calvert, G.A. *et al.* (1997) Activation of auditory cortex during silent lipreading. *Science* 276, 593–596

60 Stein, B.E. and Meredith, M.A. (1993) *The Merging of the Senses*, MIT Press

61 Calvert, G.A. *et al.* (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr. Biol.* 10, 649–657

62 de Gelder, B. *et al.* Multisensory perception of emotion, its time course and its neural basis. In *Handbook of Multisensory Perception* (Stein, B *et al.*, eds), MIT Press (in press)