CROSSMODAL SPACE AND CROSSMODAL ATTENTION

Edited by

CHARLES SPENCE

University of Oxford

and

JON DRIVER

University College London

OXFORD
UNIVERSITY PRESS

CHAPTER 7

THE PSYCHOLOGY OF MULTIMODAL PERCEPTION

PAUL BERTELSON AND BÉATRICE DE GELDER

Introduction

Sensory modalities are classically distinguished on the basis of the type of physical stimulation that they are most sensitive to—light for vision, sound for hearing, skin pressure for touch, molecules in air for smell, etc. Research on perception has generally considered each sensory modality in isolation. But most things that happen in the normal environment produce stimulation to several modalities simultaneously. The phenomena considered in the present chapter have their origin in the fact that some multimodal inputs yield related information about the same event or object. For example, an explosion produces approximately synchronized light, noise, heat, and pressure on a bystander's skin; somebody talking emits visible facial movements in predictable temporal relation to corresponding speech sounds. Such cases in which inputs to different sensory modalities bring more or less simultaneous information about the same external event can be called *valid co-occurrences*. Unavoidably, some *invalid co-occurrences*, in which inputs of independent origins come accidentally to coincide, must also occur.

The existence of valid co-occurrences of information in different sensory modalities creates for perceiving systems, whether natural or artificial, opportunities for improved performance. Perception can fail for two types of immediate reasons—irregularities in the incoming stimulation or in the subsequent processing. Whatever the case, generally only one sensory modality will be affected, so that taking into account the evidence collected by other sensory modalities can reduce the dysfunction. On a larger timescale, more durable discrepancies between the responses of different sensory modalities to inputs from the same events can emerge as a result of spontaneous drifts or of the growth of the body. The standard example of this is the increasing distance between the ears, which alters the correspondence between sound direction and the binaural disparities used for auditory localization (Held 1965). Recalibrating the now discordant processes against each other can to some extent move processing back toward veracity. Both short- and longer-term corrections will, of course, be more effective if the susceptibility of the respective sensory modalities to error is taken into account in their computation.

Research with both humans and animals has identified many cases of crossmodal interaction in which the interpretation of data in one sensory modality are influenced

by the data that are available in another sensory modality, thus showing that biological organisms effectively take advantage of the potential for intermodal redundancy. Work in this area has proceeded at both the behavioral and the physiological levels. Current developments with new brain-imaging techniques, which are reviewed elsewhere in this volume (see Chapters 9 and 10, this volume), hold the promise of important new insights. This chapter will focus on behavioral evidence from humans.

Behavioral studies of crossmodal perception: a historical survey

Cases of crossmodal influences were mentioned way back at the beginning of psychological science. Already in 1839, Brewster reported that observers who saw indented objects, such as engraved seals, through an optical device that inverted apparent concavity experienced the same inversion when they explored these objects simultaneously by touch. At around the same time, the German physiologist Johannes Müller, in the 1838 presentation of his famous 'law of the specific energies of nerves', cited the so-called 'ventriloquist illusion' as a possible exception. Müller's law stated that the particular phenomenal quality triggered by sensory stimulation depended on the nerve by which the corresponding impulses reached the brain. The ventriloquist illusion consists of the fact that the performing ventriloquist, by agitating a dummy in synchrony with the speech she or he produces without visible articulation, makes the audience experience that speech as coming from the dummy's mouth, instead of from the ventriloquist's own mouth. As Müller rightly realized, the existence of this effect implies the integration of information from two sensory modalities in forming an impression of spatial origin, indicating a limitation to these modalities' degree of autonomy.

Although scattered studies that were based, as in Brewster's (1839) observations, on experimental alteration (also called *re-arrangement*) of sensory inputs continued to be reported (for early reviews see Harris 1965; Howard and Templeton 1966; Welch 1978) and are occasionally quoted as precedents for more contemporary developments, they could not be said to constitute a coherent research effort. This is true even of the most widely quoted of these studies, those involving total inversion of the visual field (e.g. Stratton 1897), which, in spite of their innovative quality, fell short of achieving a coherent theoretical picture.

The first really systematic movement of behavioral research on multisensory perception can be traced back to the late 1950s and focused mainly on optical displacements of the visual field induced by the wearing of prisms. The work was modelled after an experiment reported by Helmholtz (1866) in which participants who pointed to a visual target, seen through a laterally displacing prism, initially landed to one side of the target but rapidly corrected their error, if they could see their hand through the prism. When, shortly afterward, the participants were asked to point to the visual target without the prism, they now misreached in the direction opposite to the earlier prismatic displacement. The occurrence of such *after-effects* was taken to show that the adaptation observed during prism exposure implied some (non-conscious) *recalibration* processes.

These recalibrations could, in principle, affect the registration of any of the functional articulations in the chain between retinal position and the felt location of the finger. It should be clear at this stage that the critical effect of prismatic displacement is the generation of a discrepancy between the proprioceptively felt location of the hand and its seen location, so that prism adaptation was effectively a case of multisensory perception, rather than of sensorimotor learning. Much research has been focused on locating the effects either in the felt position of the exposed limb, in visual localization, or in both (e.g. Harris 1965).

However, the question that received perhaps the greatest deal of attention was the role of *active movement*, as stressed by Richard Held and his MIT group on the basis of their data from prism adaptation in humans and also from the development of visuomotor coordination in animals (e.g. see Held 1965). With humans, the main finding was that adaptation occurred when the participant observed through the prisms her or his own hand in active movement, but not when the hand was moved by the experimenter or else remained simply immobile. It was proposed that the main condition for the occurrence of recalibration was exposure to rearranged *reafferent* stimulation, that is, consequent upon self produced movement. Following a familiar scenario, Held's seminal hypothesis triggered a number of falsification attempts that in turn led to alternative explanations of the effects of active movement and finally dispensed with reafference as a necessary condition for adaptation.

Among the arguments raised against the reafference hypothesis was the fact that adaptation could also result from exposure to spatial incongruence between purely afferent, or exteroceptive, stimuli. For body sensations, examples included the displaced sight of an immobile limb (Craske and Templeton 1968) or of an approaching tactile stimulator (Howard *et al.* 1965). Finally, it was shown that exposure to simultaneous noise bursts and prismatically displaced light flashes could result in recalibration of both auditory and visual location (e.g. Canon 1970; Radeau and Bertelson 1969, 1974, 1976).

In their 1974 experiment, Radeau and Bertelson had participants point to the location of either sound bursts or light flashes, before and after a period of exposure to simultaneous sounds and flashes presented with a constant 15° spatial separation. In comparison with pre-exposure pointing, post-exposure pointing responses were displaced toward the relative location of the irrelevant distractor during exposure. For instance, exposure to repeated presentations of synchronous sound-flash pairs with the sound on the left and the flash on the right shifted sound localization to the right and flash localization to the left. The important point about this kind of experiment is that the participant's only task during exposure was to monitor the inputs for occasional reductions in their intensity, with no localization required whatsoever, which ruled out any role of response processes during exposure in the generation of the observed effects.

These findings were therefore consistent with an alternative view of the basis of recalibration, which had already been proposed by Hans Wallach (1968). Working within the visual modality, Wallach had discovered several cases in which exposure to experimentally created conflicts, mainly between different cues to visual depth, produced

recalibration of one or both of the cues involved. Starting from this, he developed a general view of perceptual adaptation as based on 'informational discrepancy', which applied equally to conflicts regarding the location of one's own body parts as well as to that of external objects (see also Epstein 1975 for similar views).

Situations involving only exteroceptive sensory modalities presented many advantages for experimental analysis over those involving movement and proprioception. For instance, discordances could be created simply by separating the sources in external space, thus avoiding the intricacies of optical rearrangement (this advantage was not realized immediately, and several early studies with exteroceptive inputs still relied on prisms, e.g. Radeau and Bertelson 1969, 1974, 1976; Pick *et al.* 1969). But, more importantly, the experimenter had a degree of control over stimulus parameters, such as intensity or timing, which had not been available with reafferent proprioceptive inputs—proprioception cannot easily be switched on and off. Probably for these reasons, audiovisual spatial interaction (e.g. ventriloquism) has been studied in more detail than any other case of bimodal interaction. The role of input synchronization has for instance been demonstrated at the level of both the immediate manifestations of ventriloquism (Thomas 1941; Choe *et al.* 1975; Radeau and Bertelson 1987; Bertelson *et al.* 1997) and of its after effects (Radeau and Bertelson 1977), but only rarely in other situations (but see Held *et al.* 1966, for a prism adaptation study). Much of the current brain imaging work on crossmodal interactions has also used audiovisual inputs as well (e.g. Calvert 2001; Dolan *et al.* 2001).

Besides the extension to research on conflicts between exteroceptive modalities, another important development, which took place at about the same time, was the systematic study of immediate, or *on-line*, effects of multimodal stimulation. In the prism adaptation literature, participants' experience during exposure to the conflicting inputs had rarely been studied. Two main on-line reactions have been studied. One is the impression of *spatial fusion* of otherwise discordant inputs. This effect can be studied by having the participant judge the origins of the inputs as either the same or different. The method was used to show, for example, the fusion of heard speech with the seen face of the talker (Witkin *et al.* 1952), or to explore the effect on ventriloquism of the degree of synchronization of visual and auditory stimuli (Choe *et al.* 1975). A variant used by Jack and Thurlow (1973) was to have participants press a key whenever they experienced fusion.

The other on-line reaction is *immediate crossmodal bias*. This can be observed through a *selective localization* task in which the participant indicates, by pointing or some kind of spatial discrimination response, the apparent source of a stimulus in a target modality, while trying to ignore discordant distractors in the other modality (Klemm 1909; Thomas 1941). The now standard version of the task was introduced by Hay *et al.* (1965) to measure the visual bias of proprioception. In their experiment, participants were instructed to point with the unseen finger of the right hand to the *felt* location of a finger of the left hand, whose seen location was prismatically displaced. In comparison with a control condition in which the target finger was hidden from view,

proprioceptive localization was strongly biased toward the apparent visual location. In two other conditions, the instructions given to the participant were to point to the *seen* location of the same target finger, or of an independent external target, and little difference was observed. As discussed in the later section on 'Modality dominance', the quasi-total dominance of vision over proprioception that was suggested by these results has not been confirmed by subsequent research.

Such immediate bias measurements have now been applied to most documented cases of multimodal perception and have become the standard paradigm in the field. A problem with early studies of crossmodal biases, however, was that the direction of spatial discordance was kept constant throughout entire blocks of trials, just as during the exposure period of adaptation studies. Hence, the obtained measures probably reflected some progressive recalibration effect in addition to any immediate bias itself. This contamination can be avoided by presenting different directions of discrepancy in random alternation, a procedure introduced by Bertelson and Radeau (1981), which has been applied in the majority of subsequent studies (e.g. Bertelson *et al.* 2000*a*; Radeau 1992; Radeau and Bertelson 1987; Vroomen *et al.* 2001*a*).

The work considered so far has dealt exclusively with the perception of the spatial attributes of events, mainly their location. But crossmodal integration has also been demonstrated for the case of event identification. An especially active research movement has been concerned with the interaction between auditory and visual *speech recognition*. It has been known for some time that information useful for speech identification is available from the sight of the talker's facial movements as well as from the sound of his or her voice. Profoundly deaf people can be taught to use that information to some limited extent. The ability has come to be called *lip-reading* or *speech-reading*. Speech-reading is also used by normal hearing people in face-to-face conversation, where it can improve speech understanding, especially under conditions of poor auditory intelligibility, as in noisy environments (e.g. Sumby and Pollack 1954) or with difficult language (Reisberg *et al.* 1987). In view of this practical importance, the effects of crossmodal *congruence* (see papers in Dodd and Campbell 1987) have received more attention for the case of audiovisual speech than for other crossmodal interactions, where most of the research has been carried out with incongruent inputs. Nevertheless, understanding of the underlying mechanisms of audiovisual speech integration has also gained much from studying *incongruent* pairings, leading to cross modal identity conflict.

Experiments in which participants were presented with short auditory utterances dubbed on to video recordings of different visual ones were first reported by McGurk and MacDonald in 1976. For some audiovisual pairs, the reports of heard speech utterances were clearly influenced by the visual inputs. The most spectacular case was that of the auditory disyllable /baba/, which, paired with visual /gaga/, was in the majority of cases heard as /dada/. The articulation of the consonant /b/ involves closing of the lips, while they stay open in the cases of both /g/ and /d/. A majority of studies concerning what is now called the *McGurk effect* have similarly been carried out with

audiovisual pairings of bilabial consonants (/p/, /b/) with non-bilabial dental (/t/, /d) or velar (/k/, /g/) ones. For pairs with an auditory bilabial and a visual non-bilabial, the typical illusory percept is of a single consonant, either the visual one itself (auditory /b/ plus visual /d/ giving the percept /d/) or a different one (as in the /b/ + /g/→/d/ example). Such percepts have been called *fusions*. This term 'fusion' is obviously appropriate in cases where the percept is different from both the auditory and the visual inputs, as with auditory /ba/ plus visual /ga/ giving the percept /da/. But when the same auditory /ba/ paired with visual /da/ is perceived as /da/, it is tempting to ask if all that happens is not a substitution of the seen phoneme to the heard one. The important point with this example is that speech-reading alone does not allow discrimination of the 'voicing' feature (/d/ versus /t/), so the fact that participants report /da/, and never /ta/, implies that the percept really includes an auditory contribution.

Another type of illusory report has been obtained when an auditory non-bilabial consonant (/da/, say) was combined with a visual bilabial (/ba/). It consisted of a juxtaposition of the two consonants (e.g. /bda/), generally called a *combination*. While fusions and combinations both imply some integration of auditory and visual data, they probably rely on different underlying mechanisms. Perhaps one should refer to 'the McGurk effects' instead of 'effect'.

Naive participants generally have little awareness of the discrepancy, and it has been shown that the effects survive instructions to report specifically the auditory impression (Massaro 1987), as well as explicit descriptions of the conflict situation (Manuel *et al.* 1983). These findings have been taken to imply that the integration of visual and auditory speech is to some extent mandatory, and not just the result of a deliberate strategy resorted to by the listener when confronted with deficient auditory information. Other manifestations of the effect's robustness are that it still occurs with discordant vowels (e.g. auditory /ba/ and visual /di/; Green 1996) or across gender differences (a male voice with a female face; Green *et al.* 1991).

An obvious question is whether similar identification interactions occur outside the linguistic domain. One would perhaps expect environmental events such as explosions, running animals, or passing vehicles to trigger crossmodal effects concerning identification, but the literature contains very few systematic analyses of such situations. Until recently, the main exception was a study in which the sight of hands either bowing or plucking a cello was combined with sounds from a continuum ranging from bow to pluck (Saldaña and Rosenblum 1993). The participants identified the sound as a bow or pluck, and their choices were slightly, but nevertheless significantly, biased in the direction of the seen action. This effect, which presumably depended on the participants' knowledge of the sounds produced by musical instruments, might, of course, reflect cognitive response strategies as well as genuine perceptual interactions.

Other relevant examples have been reported more recently. In a rare consideration of auditory–tactile interactions, Jousmäki and Hari (1998) had participants rub their hands together close to a microphone while they listened to the sound played back

through earphones. Dampening or accentuating the high-frequency components of these sounds resulted in a striking modulation of the experienced dryness/moisture of the palms. This effect, known as 'the parchment-skin illusion' appears as a beautiful case of crossmodal fusion (see also Guest *et al.* 2002). A slightly different case is the phenomenon studied by Sekuler *et al.* (1997) in a situation (first described by Metzger 1934) in which two identical objects are seen moving toward one another, coinciding, and then moving apart, all on the same linear path and at constant speed. When the visual display is presented by itself, participants typically reported either that the objects just crossed and continued in their original directions or, more rarely (on about 20% of trials), that they collided and bounced. Presenting a loud click at the time of coincidence increased the frequency of bouncing reports considerably (to about 60%). Subsequent research has shown that the same result is obtained when the click is replaced by a visual flash (Watanabe and Shimojo 2000), indicating that the effect does not have to be a crossmodal one. Another point worth making is that, in the bimodal version of the task, the resulting percept is a juxtaposition of data from the two sensory modalities, thus resembling the 'combinations' of the McGurk situation rather than the 'fusions'.

A question that has rarely been considered previously concerns whether exposure to identity conflicts can produce, besides immediate biases, after-effects indicative of crossmodal recalibrations. Some results from the speech adaptation literature (e.g. Roberts and Summerfield 1981) have been taken to show that exposure to incongruent pairs of audiovisual McGurk stimuli does not produce any after-effects (Rosenblum 1994). However, we have now shown that after-effects can be obtained by exposing participants to combinations of a visual speech token (/aba/ or /ada/) with an ambiguous auditory one (intermediate between /aba/ and /ada/; Bertelson *et al.* 2003). The preceding negative results were presumably due to selective speech adaptation caused by repeated presentations of non-ambiguous speech tokens, which was avoided by using ambiguous ones in Bertelson *et al.*'s study.

A new case of bimodality that was recently brought to the attention of the scientific community concerns the expression of emotions in seen faces and heard speech (de Gelder *et al.* 1995; Massaro and Egan 1996; de Gelder and Vroomen 2000). The paradigms used by de Gelder and her collaborators involved the simultaneous presentation of faces from a continuum of expressions (e.g. sad to happy) with a semantically neutral speech utterance pronounced in an affective tone corresponding to one or the other end-point of the face continuum (or, conversely, a continuum of voice tones with seen faces expressing extreme emotions). The two possible crossmodal biases—of facial expression judgements by voice tone, and of voice tone judgements by facial expression—were both obtained. These effects were not eliminated by instructions to focus on a particular sensory modality. Stronger arguments for the mandatoriness of the effect have recently been provided by studies with neurological patients (e.g. de Gelder *et al.* 2000, 2002; Dolan *et al.* 2001; see also the subsequent section on 'Evidence from brain-damaged patients').

In the following sections, some important questions raised by the phenomena that have just been surveyed will be discussed. Their choice, it will become evident, was strongly, if unavoidably, influenced by our own familiarity with particular lines of research as well as by our own strategic preferences.

Modality dominance

The modern interest in the interplay of the senses has its roots in speculations regarding the development of perception within eighteenth century philosophy, such as Berkeley's well-known thesis that visual perception of space is acquired on the basis of tactile experience. In reaction to such claims, some of the early experimental work on multisensory perception set out to demonstrate that, instead, vision dominates other senses in cases of disagreement (e.g. Rock and Harris 1967).

Some of the most influential arguments were derived from studies of crossmodal biases. The pioneering study by Hay *et al.* (1965), showing a near total dominance of vision over proprioception, as described above, is cited very frequently. However, subsequent work with the same tasks has produced a somewhat more complex picture, with a visual bias of proprioception constituting only part of the experimental discrepancy and a proprioceptive bias of vision often more substantial than in the original experiment (see review in Welch and Warren 1980, p. 641). A similar development has occurred in the case of ventriloquism. An early report of total visual dominance (Pick *et al.* 1969) was similarly followed by less dramatic observations of only partial dominance. For the 10–20° range of angular separations considered in the majority of studies, the visual bias of auditory location has, across the conditions of four separate studies (Bermant and Welch 1976; Bertelson and Radeau 1981; Radeau 1985, 1992), centered around 30% of the separation (range 15–42%). The auditory bias of visual location has more rarely been considered. It was smaller than the visual effect, but it reached significance in at least two studies (Bertelson and Radeau 1981, experiment 1; Radeau and Bertelson 1987; but not in Radeau 1985).

The initial results (e.g. Pick *et al.* 1969) suggesting total dominance of vision over the other sensory modalities allowed (though did not entail) a simplistic interpretation of crossmodal interaction by the *substitution* of the visual data for those with different origins. The fact that biases are generally only partial supports the alternative view of an *integration* process, in which the conflicting data receive different weights, but the less potent modalities can still exert some influence. This notion has implications for the interpretation of the spatial fusion experienced, for instance, by the spectators of performing ventriloquists (or manifested in 'same' judgements by participants in experiments with same/different origin tasks concerning incongruent audiovisual stimulus pairs). Such evidence is, in our experience, easily taken as showing that the sounds were perceived as coming from the location of the visual inputs. Actually, what the evidence implies is only that the combined effects of the two biases have brought the registered separation below the detection threshold.

Another core component of the notion of total modality dominance that subsequent empirical results have put into question is the idea that each of the sensory modalities can be characterized by a measurable general capacity to bias other sensory modalities and to resist their biases. This notion was, for instance, at the origin of a not very successful search for a hierarchy of sensory modalities governed by a principle of transitivity, according to which the size of the bias of modality C by modality A should be predictable once one knew the biases of B by A and of C by B (Pick *et al.* 1969; Warren *et al.* 1981). A major difficulty for such an enterprise is that the size of each particular bias depends on a number of parameters. For instance, the visual bias of auditory location is an inverse function of the loudness of the auditory target and a direct function of the brightness of the visual attractor (Radeau 1985). In another example, a displaced auditory distractor recalibrates visual location when the visual target, a point flash of light, is seen against a homogeneous background, but not when it is seen against a structured one (Radeau and Bertelson 1976). Such findings raise serious doubts about the applicability of any general measurement of dominance.

Finally, there is evidence that the degree of dominance depends on the dimension upon which the sensory modalities diverge. For instance, when vision and audition present conflicting information about timing, as with sounds and lights fluctuating at different rates, the auditory bias of the visual percept is much stronger than the visual bias of the auditory percept (Welch *et al.* 1986). Welch (1999*a*) has proposed that the biasing capacity of a particular sensory modality depends not so much on some intrinsic general accuracy as on its appropriateness to the particular task (the 'modality appropriateness hypothesis').

Other revealing evidence has come to light recently from a study in which participants were instructed to synchronize finger tapping with one or the other component of a series of bimodal stimulus pairs, each of which consisted of a light flash and a sound burst separated by one of several time intervals (Aschersleben and Bertelson 2003). In one experiment, the participants were instructed to synchronize their tapping with the flashes and to ignore the sounds. Nevertheless, in spite of these instructions, the taps were strongly attracted by the clicks, demonstrating a substantial auditory bias of visual apparent time of occurrence. The other experiment, in which the instructions were to synchronize with the sounds and to ignore the visual flashes, resulted in only a small (though significant) visual bias. Aschersleben and Bertelson's study therefore provides a convincing confirmation of the superiority of audition over vision for the case of temporal resolution.

A finding with similar implications was reported by Shams *et al.* (2000). Trains of one to four auditory beeps were presented simultaneously with trains of one to four visual flashes, and the participant's task was to report the number of flashes that they had seen. A single flash accompanied by two beeps was systematically reported as two flashes, just as were two flashes accompanied by a single beep or presented in silence.

Detecting event singleness: the pairing hypothesis

Discussions about multimodality often start, like the introduction to this chapter, with the consideration that simultaneous multimodal stimulation is the norm in our everyday environments, and that crossmodal interactions can be adaptive only when they apply primarily to data originating in the same distal events or objects (e.g. Radeau and Bertelson 1977; Stein and Meredith 1993; Bedford 1994, 1999; Radeau 1994*a*; Welch 1999*a*; Vroomen 1999; Driver and Spence 2000; Chapter 6, this volume). Consequently, the system must to some extent discriminate between valid single-origin co-occurring inputs versus invalid spurious ones, and ideally allow interactions to proceed in the valid cases only, or at least more often than in invalid cases.

In principle, one way of achieving this kind of discrimination is to examine intermodal congruence across the multimodal dimensions of the inputs. The important point (already mentioned in the 'Introduction') is that, when some accidental incongruence arises in real life within valid, same-origin pairs of stimuli, it generally concerns one particular dimension, but not others (e.g. stimulus location but not arrival time or rate of interruption). This limitation of incongruence to one particular dimension of the inputs in cases of valid multimodal co-occurrences could be used to distinguish these cases from spurious invalid ones.

The artificial conflict situations used in the laboratory to study crossmodal interaction similarly often involve bimodal stimulus pairs that are incongruent on one particular dimension while being congruent on several other dimensions. When a participant watches his hand through a prism, vision and proprioception signal incongruent locations, but congruent timing, direction, and speed for any movements. Similarly, in situations producing classical ventriloquism, locations are crossmodally incongruent but timing is congruent. We have already seen that a similar pattern exists in audiovisual situations leading to McGurk effects, in which the visual and the auditory components differ by one particular phonetic feature while sharing other ones, for example, the 'seen /da/–heard /ba/' pair differing in articulation place but sharing voicing and articulation manner.

The preceding considerations lead to the hypothesis that processing of multimodal inputs may involve an early assessment of the degree of concordance of the total input with a unitary source. At this stage, the incongruences registered on some dimensions would be weighted against the congruences on the remaining dimensions. Crossmodal interactions like biases or recalibrations might then be triggered when the total congruence exceeds some criterion. In earlier papers (Radeau and Bertelson 1977, 1987; Radeau 1994*a*; Bertelson 1998), this hypothetical operation was called *pairing*. (Welch 1999*a*; Welch and Warren 1980) has used the term *unity assumption* to denote the same function, albeit with more cognitive implications (see also, Chapter 6, this volume). As explained in an earlier paper (Bertelson 1999), pairing can be seen as an intermodal analog of perceptual grouping, or unit formation, resembling Gestalt phenomena described earlier for vision (Wertheimer 1923; Pomerantz 1981) and audition (Bregman 1990).

An important point is that pairing should be distinguished from the conscious impression of common origin that was called here *spatial fusion*, because some of its manifestations can occur in the absence of such fusion. Bertelson and Radeau demonstrated this by presenting participants with synchronous but spatially discordant sound-and-flash pairs, and having them on each trial both point to the location of the sounds and give same/different location judgements (Bertelson and Radeau 1981, experiment 2). Crossmodal biases could thus be measured separately for those trials on which the participants gave 'same location' and 'different locations' judgements. However, before valid comparisons could be carried out on the obtained values, these had to be corrected for a possible confounding factor, namely, an influence on the same–different location decision of the particular size reached by the bias on each trial. After application of such correction (for details of the procedure, see Bertelson and Radeau 1981, their Appendix), the visual bias of sound location was still found to be significant on both no fusion and fusion trials, although it was larger on the latter. This implies that pairing can occur at an unconscious level, and is sometimes accompanied by conscious fusion, sometimes not.

The failure to make the distinction between pairing and spatial fusion has often led to confusion in the literature. A single-factor view of conflict resolution, which equates pairing with fusion, cannot deal with the fact that both components of the conflict situation, incongruence on some dimension(s) and congruence on other ones, must be registered. The problem is avoided with the present notion that, whatever happens at the level of conscious perception, any incongruence is still registered at the preconscious level at which pairing occurs.

Processing levels: perception or post-perceptual judgement?

Most arguments for crossmodal integration were, until recently, based on the voluntary responses of participants, whether verbal descriptions, absolute judgements, settings of stimulus values, or pointing responses. As for all similarly approached phenomena, one could ask if the observed effects originated in automatic, mandatory perceptual processes or in post-perceptual judgemental processes instead. It should be clear that human responses are relevant to the general issue of intermodal coordination only if they reflect basic perceptual processes, rather than mere voluntary response strategies adopted to satisfy the demands of a particular laboratory task.

Tasks in which participants judge some aspect of the bimodal input on-line, are particularly susceptible to voluntary influences, to the extent that the situation is *transparent*, that is, that most information required for any explicit deliberations such as the inputs to each sensory modality—their location, relative timing, and even semantic context—is open to conscious inspection. In the selective responding paradigm (through which immediate crossmodal biases, whether in localization or identification, have generally been studied), the fact that biases occur in spite of instructions to report inputs to the target modality and ignore the others has often been taken as proving their

automatic nature. While the argument does effectively suggest at least some degree of mandatoriness, it is not a strong one. As a matter of fact, once any discordance is detected, it is still up to the participant to decide what to do with the experimenter's instructions. Furthermore, these instructions can draw the participant's attention to a discordance he or she might otherwise have ignored.

The after-effect paradigm does not present that danger to the same extent, because in its case the critical measures (the pre- and post-exposure responses) are obtained through a straightforward unimodal task, which leaves less room for deliberate strategies. That consideration was at the origin of the predominant use of adaptation paradigms in studies of prismatic displacement, as well as in our own early work on ventriloquism (Radeau and Bertelson 1969, 1974, 1976, 1977; Radeau 1973; see also Recanzone 1998). In the latter case, the hypothesis of a reduced susceptibility of after-effects to variable strategies received support from a study in which both the visual bias of sound localization and the auditory after-effect of exposure to sound-flash discordance were measured in parallel with the same inputs and the same participants (Bertelson and Radeau 1987; Radeau 1992). On-line biases showed systematically larger variability than after-effects. However, the issue of transparency can still be raised when participants become aware of the discordance during the exposure phase of the experiments. As Welch (1978, p. 20) noted in his discussion of the problem as it arises for prism adaptation, much depends on the reasons the participants may have to continue applying a voluntary correction during the subsequent post-test phase (for instance, on whether or not they believe that the visual displacing device is still in place).

The first direct attempt at elucidating the site of crossmodal interaction in the functional architecture was made by Choe *et al.* (1975), in their paper reporting the influence of temporal synchronization on the frequency with which sound/flash pairs of stimuli presented in the same or in different locations were judged as coming from the same place. Application of detection theory showed that synchronization affected the decision criterion 'β and not d''. The authors concluded from this that ventriloquism originates in a response bias, and not in shifts of the perceived separation of the inputs. In a comment on the paper, however, we showed that the particular sensory effect on synchronization that the detection analysis had discarded was from the start implausible, and that a more likely one was compatible with the results (Bertelson and Radeau 1976). We proposed that an effect of synchronization on the registered locations of the inputs was still a valid possibility, although contributions from post-perceptual factors like response biases or even deliberate judgements could not be excluded for results obtained in the habitual transparent situations.

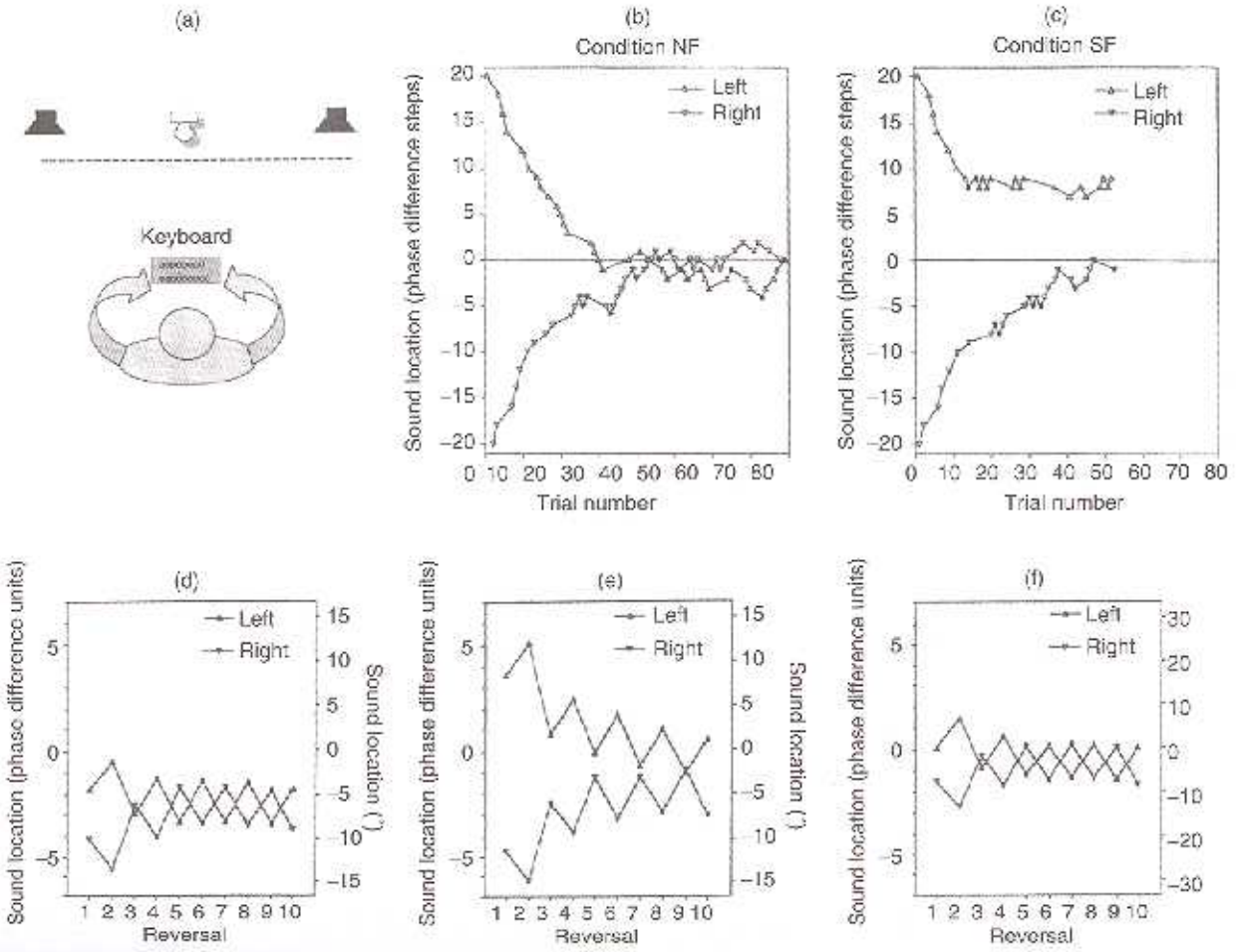
It follows from the preceding discussion that more convincing arguments for genuine perceptual crossmodal integration would be obtained in non-transparent situations, in which the participant has no awareness of the intermodal discrepancy. An approach that seems to satisfy this requirement has been developed in recent years and applied to the case of the visual bias of auditory location (Bertelson and Aschersleben 1998; and see Caclin *et al.* 2002, for a tactile bias of sound location). It is based on a new version

of the classical psychophysical staircase procedure. On each trial, a train of sound bursts is delivered from a stereophonically controlled apparent location, and the participant indicates by pressing one of two keys whether the sounds came from the left or right of the median plane (see Fig. 7.1(a)). Sound locations are chosen according to two randomly mixed staircases, one starting far to the left and the other far to the right (see Fig. 7.1(b)). 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 toward a median location. Initially, the correct response is provided repeatedly on each staircase so that, except for occasional mistakes, the progression toward the center 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 apparently uncertain as to whether the target sound is on the left or right, so that the non-transparency requirement should be met.

To examine any visual bias of auditory localization, light flashes were presented from a central location in synchrony with the sound bursts. If the sounds were attracted by these flashes, the uncertainty points should be reached at sound locations further away from the center with synchronized flashes than without them, or with less effective visual inputs. The experiments involved separate 'explorations', each starting with sounds in the extreme left and right positions, and stopping automatically when 10 reversals had been recorded on each staircase. The prediction concerning attraction by the light was tested on the basis of the mean sound locations (across explorations and participants) at which the 10 first reversals occurred on each staircase. These locations were significantly further apart on left and right staircases, respectively, with synchronized flashes (see Fig. 7.1(c)) than with no visual input (see Fig. 7.1(d)). Thus, attraction of the apparent location of the sounds by the light occurred at times when the participant was no longer aware of the sound's lateral deviation, and was thus presumably unable to apply any response strategy in systematic fashion. That implies that visual biasing of auditory location can occur automatically, presumably representing a true perceptual effect rather than just some post-perceptual correction or response bias.

We would argue that the critical feature of our method is the focusing of the analysis on data collected during the uncertainty phase of each exploration, as indexed by the occurrence of response reversals. The elimination of post-perceptual interpretations is conditional on that focusing. If data from different staircases are pooled instead on the basis of successive *trials* (as in traditional applications of psychophysical staircases to the measurement of detection thresholds, or in Soto-Faraco and colleagues' studies of the apparent visual capture of auditory motion; see Soto-Faraco and Kingstone, in press, for a review) instead of successive *reversals*, we would suggest that the argument for a non-transparent situation, and thus for automatic perceptual effects rather than post-perceptual strategies, may no longer hold.

In a second experiment with their staircase method, Bertelson and Aschersleben (1998) showed that automatic attraction depends on the timing of the presented


(a)
Keyboard
(b)
Condition NF
Left
Right
Sound location (phase difference steps)
Trial number
(c)
Condition SF
Left
Right
Sound location (phase difference steps)
Trial number
(d)
Left
Right
Sound location (phase difference units)
Sound location (°)
Reversal
(e)
Left
Right
Sound location (phase difference units)
Sound location (°)
Reversal
(f)
Left
Right
Sound location (phase difference units)
Reversal

bimodal pairs of stimuli: introducing long and variable sound-flash time intervals eliminated the effect (see Fig. 7.1(f)). This result implies that, contrary to the conclusion of Choe *et al.* (1975), timing effects can originate at a genuinely perceptual level.

Another (as yet unpublished) experiment has shown that the visual bias effect can survive *constant* visual leads or lags of 300 ms, suggesting that crossmodal pairs of inputs with constant time intervals (within some small range of durations) can, just as for simultaneous pairs, function as perceptual units giving rise to pairing.

A new question to which we have also applied our staircase method is whether ventriloquism can operate on the time dimension as well as on the spatial one (Bertelson and Aschersleben 2003). In classical spatial ventriloquism, the apparent spatial separation between auditory and visual inputs seems to be underestimated, provided these occur close enough in time. One could wonder if the reverse phenomenon, underestimation of the perceived temporal separation between the bimodal inputs on the condition that they occur at sufficiently close locations, also occurs. We used a temporal order judgement (TOJ) task, in which the participant decided on each trial which of two stimuli, a sound burst and a light flash, had occurred first. Successive stimulus onset asynchronies (SOAs) were chosen from two staircases, one starting with the sound a long SOA (250 ms) ahead and the other with the sound the same SOA after the flash. Through this procedure, we could thus determine the SOA values at which the participant started producing response reversals. In one condition, sounds and flashes were presented in the same frontal apparent location, and in two other conditions, the flashes were still presented from in front and the sounds from either far to the left or far to the right. The critical result was that response reversals in the TOJ task began occurring at significantly larger SOAs in the 'same location' condition than in the 'different locations' conditions, showing that sounds and lights were effectively perceived as occurring closer together in time when they were also close together in space, rather than farther apart.

Fig. 7.1 Visual bias of perceived auditory location measured with psychophysical staircases (after Bertelson and Aschersleben 1998). (a) Schematic illustration of the testing situation. (b), (c) Examples of explorations. Abscissas, successive trials; ordinates: sound location in number of phase difference steps; up = left; down = right. After a 'left' response, the sound location is moved one step to the right for the next trial, and vice versa after a 'right' response. (b) One typical exploration in the condition without a flash. Each staircase moves monotonously toward the center, until response reversals (marked in black) begin to occur. The exploration stops as soon as 10 reversals have been recorded on each staircase. (c) One exploration by the same participant in the condition with a central flash synchronized with each sound. Reversals began to occur when the staircases were still further apart than in (b). (d)–(f) Sound locations at 10 first reversals: means across explorations and participants. (d) Condition without flash. (e) Condition with synchronized flash. (f) Condition with de-synchronized flash. The larger separation in (e) than in (d) between reversal locations on the left and right staircases shows that flashes attract perceived sound location. The vanishing of the effect in (f) proves that the attraction depends on audiovisual synchronization.

The same influence of spatial separation on temporal order judgements has also been reported for visuotactile stimulus pairs by Spence *et al.* (2001). These authors, however, used the 'method of constant stimuli' procedure, which on our argument does not provide the protection from post-perceptual influences that our staircase procedure should ensure.

Evidence from indirect methods

Most of the evidence generally quoted for the existence of crossmodal interactions comes from situations in which the participants reported on aspects of the inputs that directly concerned the predictions of the putative crossmodal influence to be tested—the identity or location of some target. But useful evidence can also be provided indirectly by effects of the interaction on an apparently unrelated task. Effects of this type may also sidestep response-bias issues.

One convincing example was provided by an experiment with a traditional 'cocktail party' situation, in which Driver (1996) 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 shown on a screen either close to the loudspeaker or farther away. Performance was significantly better with the face on the distant screen than on the one close by. Presumably, the apparent location of the auditory target items was ventriloquized toward the screen, and this created an apparent spatial separation between the simultaneous auditory inputs, a condition that has long been known to facilitate selective listening (for the case of true rather than illusory spatial separation; Broadbent 1958). The non-target items, which were not synchronized with the movements of the face, were presumably not attracted, or at least not to the same extent. The identification gain from presenting matching lip movements on the distant screen rather than on a screen at the actual location of the two auditory speech-streams provides an indirect demonstration of a crossmodal effect that is apparently not amenable to any strategic explanation.

Another example of the indirect approach comes from a recent study of visual bias of auditory motion direction (Vroomen and de Gelder 2003). This took advantage of a newly discovered contingent auditory-motion after-effect, in which listening to sounds with a falling pitch moving in one direction, alternated with sounds with a raising 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 raising or falling (Dong *et al.* 1999). This is, in fact, an auditory analog of the well-known McCullough effect in vision (McCullough 1965). Vroomen and de Gelder added to the inducing situation a visual object moving simultaneously with the sound in either the same direction (congruent condition) or the opposite direction (incongruent condition). The auditory after effect was significantly enhanced by congruent visual movement, and not only reduced, but even reversed, by the incongruent visual movement. This result provides a convincing demonstration of the perceptual origin of the visual

bias of auditory motion identification. Earlier studies of that effect by Soto-Faraco and his collaborators failed to control effectively for possible post-perceptual influences (Soto-Faraco and Kingstone, in press; Soto-Faraco *et al.* 2002).

Evidence from brain-damaged patients

Patients who as a result of brain damage have lost the capacity to form conscious representations of particular forms of stimulation can sometimes nevertheless make effective use of such stimulation, as can be demonstrated through indirect methods. These cases of implicit processing offer opportunities for addressing the role of consciousness in certain aspects of perception. For the particular question of the locus of crossmodal interactions, finding that inputs of which a patient is not aware can nevertheless bias perception in other sensory modalities would evidently have important implications regarding automaticity.

In the case of ventriloquism, we have examined the visual bias of auditory location in patients with severe left unilateral neglect. These patients were totally unable to detect any of the bright flashes of light presented to their left visual hemifield, but their localization of target sounds (delivered frontally in synchrony with the flashes) was nevertheless shifted significantly in the direction of these undetected left flashes (Bertelson *et al.* 2000*b*). This finding that some visual bias of auditory localization can take place without awareness even of the *occurrence* of the visual distractor provides a demonstration of its automaticity that is still stronger than the demonstration from staircase procedures in normals, for which the participants were aware of the visual distractor's *presence* but only unsure of its location relative to the auditory target.

Reactions of neglect patients to another type of audiovisual conflict have been considered in a study by Soroker *et al.* (1995), but this concerned a question completely different from the one examined by Bertelson *et al.* (2000*b*). The patients, who recognized spoken syllables less accurately in their impaired left auditory space than in their right one, had this left-side inferiority reduced when a dummy loudspeaker was made continuously visible in their right visual hemifield. Thus, while our study focused on the remaining biasing capacities of visual stimuli presented in the impaired contralesional field (which, for that reason, went undetected), Soroker *et al.*'s study dealt with the effect of the presence in the ipsilesional field of a (presumably well identified) suggestive visual object on impaired auditory processing in the contralesional hemispace. Their results thus do not carry the same implications regarding the issue of automaticity for ventriloquism. The nature of the dummy loudspeaker effect will be considered in a later section.

Evidence from brain damage has been used extensively in recent work on crossmodal affect (i.e. emotion) integration (de Gelder *et al.*, in press). Part of that work involved so-called *prosopagnosic* patients, who are characterized by a more or less specific inability to identify faces. One such patient was asked to identify the affective tone in which a sentence was pronounced. Although she was totally unable to report at the conscious level the emotions carried by seen faces, presenting an emotional visual face at the same time as the

spoken sentence still biased her judgement of voice tone in the direction of the facial expression that she could not explicitly identify (de Gelder *et al.* 2000). This striking result strongly suggests that crossmodal affect integration can be a genuine perceptual effect that cannot be explained by voluntary strategies or any other post perceptual factor.

Results with similar implications have been obtained in studies of the *blindsight* syndrome. This syndrome is observed in *hemianopic* patients who, as a result of damage to particular brain areas including the primary visual cortex, have lost awareness of stimuli projected to the corresponding region of the visual field, but whose performance on some indirect tasks can nevertheless be enhanced in the presence of such stimuli (e.g. Weiskrantz 1997). De Gelder *et al.* (2002) have examined the effects on cortical responses to a fearful voice tone of the emotion displayed by a face presented in each of two blind-sight patients' blind fields. Using event-related brain potentials, they found increased auditory responses following presentation of a congruent (fearful) facial expression in the blind field, as well as in the intact field. By contrast, a fear-inspiring picture (e.g. a snake or a spider) enhanced the same auditory response only when presented in the intact field.

Processing levels: semantic influences

When considering the data that might be used by a system to decide which stimuli to pair, it is tempting to include knowledge of environmental regularities. Much work on crossmodality has involved familiar situations with well-known capacities to produce multimodal inputs; such as, for the case of prism adaptation, the felt and the seen location of some part of one's own body. Early work on ventriloquism similarly resorted to simulations of real-life audiovisual situations: speech and the moving face of the talker (Witkin *et al.* 1952), or the sound and the sight of sound-producing objects like steam whistles (Jackson 1953), a loudspeaker (Pick *et al.* 1969), or a door bell (Canon 1970). That such realism was instrumental in causing the observed interactions was often taken for granted (Welch and Warren 1980; see also Bedford 2001 and commentary by Bertelson *et al.* 2001).

With exteroceptive conflicts, like audiovisual ones, it is however easy to produce audiovisual discrepancies with stripped-down bimodal pairs reduced to sound bursts synchronized with point flashes of light. All the classical manifestations of ventriloquism, such as perceptual fusion (Choe *et al.* 1975; Bertelson and Radeau 1981), immediate bias (Thomas 1941; Bermant and Welch 1976; Bertelson and Radeau 1981; Warren *et al.* 1981; Radeau and Bertelson 1987; Radeau 1992; Bertelson and Aschersleben 1998; Bertelson *et al.* 2000*a*), and after-effects (Radeau and Bertelson 1969, 1974, 1976; Bermant and Welch 1976), have been obtained with such inputs. Documented effective factors are synchronization (Thomas 1941; Choe *et al.* 1975; Warren *et al.* 1981; Radeau and Bertelson 1987; Bertelson and Aschersleben 1998), spatial separation (Bermant and Welch 1976; Bertelson and Radeau 1981), loudness of the auditory target (Radeau 1985), and size of the visual attractor (Bertelson *et al.* 2000*a*; Vroomen *et al.* 2001*a*). These results establish the

most important point regarding the generation of crossmodal interactions: *semantic contributions from familiar bimodal contexts are not a necessary precondition for their occurrence.*

The next question is whether top–down effects from familiar contexts can nevertheless enhance the demonstrated bottom–up effects of sensory factors. Several classic studies have often been quoted as supporting a positive answer. One such study resorted to a picturesque experimental situation, in each trial of which a sound was produced in one of several locations by an unseen whistle, while participants were asked to match it with one of several visible steam-whistle kettles, from one of which steam was seen to emanate (Jackson 1953). Participants' choices were strongly biased toward the steaming kettle, which was chosen on nearly every trial with the sound at a 30° angular distance, and still on more than 60% of trials with a 60° angular separation. In another experiment, a ringing sound was matched with one of several bulbs, one of which was lit, and here the cued location was chosen over only a much smaller range of separation.

In two other studies (Jack and Thurlow 1973; Thurlow and Jack 1973), participants watched videotaped scenes with varying degrees of audiovisual appropriateness, and were told to press a key (or in other cases to start a stopwatch) whenever the sounds were experienced as originating in the represented action. For instance, sequences of tones were paired with either the sight of a finger pushing a button or that of a face counting. Fusion was reported for considerably longer durations for the finger (150 seconds in 5 minutes) than for the counting face (42 seconds).

On the other hand, results raising some doubts about the proposed role of familiarity were reported by Radeau and Bertelson (1977). In one experiment (see Fig. 7.2), participants were exposed to percussion sounds paired with the sight on a displaced screen of either the hands playing the instruments or of light flashes synchronized with the beats, and the resulting after-effects on unimodal localization of the sounds were of comparable magnitude in the two conditions. In a second experiment, auditory speech was paired with either the sight of the talker's moving face, or again with flashes synchronized with the amplitude peaks of the speech, and here also the after-effects were indistinguishable. Thus, in these two experiments, realism apparently had no effect, beyond that of temporal correlation between low-level components of the two inputs. In a third experiment, perceptual fusion was measured by the key-pressing method during exposure to the speech–face and speech–flashes situations, and more fusion was reported for the former than for the latter condition. This result is consistent with the already mentioned notion that on-line effects may be more susceptible than after-effects to contamination by post-perceptual factors.

Apart from the results presented as suggesting some *additional* effects of semantic congruence (possibly due to influences on strategic response-biases), there is in the literature one result that might be taken as demonstrating a pure effect of semantic factors. Pick *et al.* (1969) reported a visual bias of the apparent origin of sounds produced by simply seeing a small dummy loudspeaker in a prismatically displaced

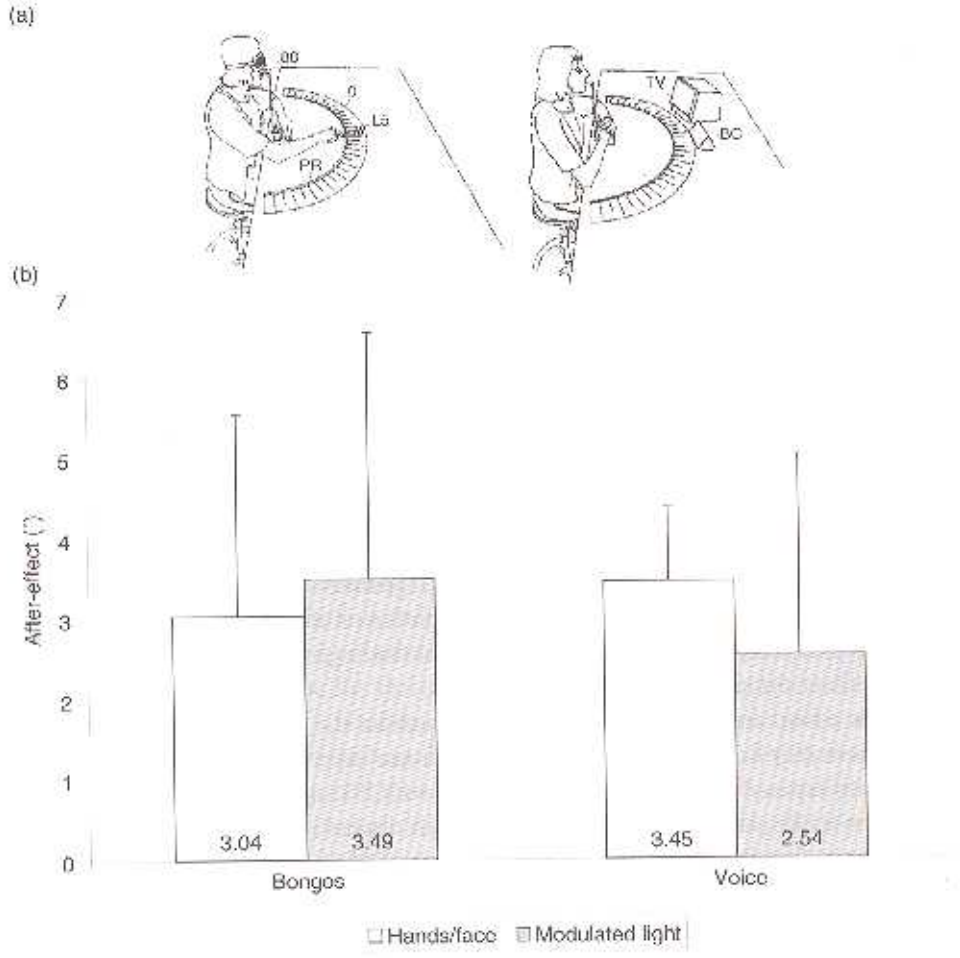


Fig. 7.2 After-effects on auditory localization of exposure to audiovisual spatial discordance, as a function of paired events (after Radeau and Bertelson 1977). (a) Experimental situation. Left, pre- and post-tests: The participant (blindfolded, head constrained by bite-board) points to the apparent location of auditory stimuli (experiment 1, bongo sounds; experiment 2, speech). Right, exposure phase: the participant (bite-board, no blindfold), monitor screen displaced left or right 20° from central auditory source. Visual events on screen for experiment 1 comprise: familiar conditions, hands playing the bongos, synchronized with sounds; non familiar conditions, light flashes synchronized with amplitude peaks of bongo sounds (i.e. beats). Visual events on screen for experiment 2 comprise: familiar conditions, face of speaker; non-familiar conditions, light flashes synchronized with amplitude peaks of speech (i.e. syllables). (b) After-effects in degrees (mean post tests minus mean pre-tests). White, familiar conditions (hands/face); gray, non-familiar conditions (flashes). Note that there was no significant effect of familiarity.

location. There were no transient visual changes whose correlation with the sounds would have provided a sensory explanation of the result. Thus, a possible implication (which the authors actually did not draw) would have been a pure effect of audiovisual symbolic congruence on auditory location. To check this possibility, we have run measures of the visual bias of target sound bursts by, on the one hand, the sight of a displaced dummy loudspeaker and, on the other, the usual synchronized flashes (Bertelson and Radeau 1987; Radeau 1992). The habitual significant bias was obtained with the flashes, but no bias whatever was found with the loudspeaker. After-effect measures run in the two conditions produced exactly the same pattern of results (i.e. a crossmodal effect only with synchronized flashes). Thus, the dummy loudspeaker effect can apparently be obtained only in combination with some particular conditions, probably instructional ones, which research has not yet elucidated. This presumably may also apply to the technique's apparent capacity to restore contralesional performance in neglect patients, as reported by Soroker *et al.* (1995).

The existing evidence concerning the effectiveness of context familiarity in producing or enhancing crossmodal interaction is thus replete with contradictions, with some studies producing positive results and others not. There are two possible reasons for this state of affairs. First, most studies on familiarity used transparent tasks, allowing variable strategic effects to influence the results. For example, the fact that Jackson's (1953) steam-emitting whistle was often chosen as the source of the sound does not necessarily mean that a displacement of the sound was experienced by participants in his studies. An alternative and more likely explanation is a post-perceptual combination of the location impression with the knowledge of the principle of steam kettle whistles. The results of Pick *et al.* (1969) or Soroker *et al.* (1995) with the dummy loudspeaker, and those of Thurlow and Jack (1973) and of Radeau and Bertelson (1977, Experiment 3) with a key-pressing task are open to a similar post-perceptual interpretation. As suggested by Welch (1999*a*, *b*), the effects of familiarity should be examined under non-transparent conditions such as those of the staircase procedure described earlier.

The other potential explanation is that many manipulations of context familiarity may also have inadvertently affected the critical sensory parameters of the inputs, such as the intensity, visual contrast, or saliency of stimuli in the combined modalities. For instance, in Thurlow and Jack's (1973) experiment with a pushing finger or a counting face synchronized with auditory tones, the observed superiority of the 'appropriate' finger tone condition may reflect low-level sensory factors as well as, or instead of, semantic congruence. The existence of the same problem for Radeau and Bertelson's (1977) comparisons between seen hands (or faces) and light flashes must also be admitted here. To assess the seriousness of the problem, the experiments (or similar ones) should be run with flash intensity varied over a wide range of values. Welch (1999*a*, p. 73) has drawn attention to another factor that also can be easily confounded with the degree of familiarity, namely, 'the number of amodal stimulus properties that are shared by the two sensory modalities'.

Studies with illusory limbs

In 1937, the French neurologist Tastevin reported that participants easily took a seen dummy finger to be one of their own fingers. In an ingenious study, Welch (1972, 1999a) took advantage of the phenomenon to explore the effect of knowledge on adaptation to visuoproprioceptive discordance. Participants had to move their hand to a visually specified location, at which their forefinger should become visible, but what they actually saw there was the experimenter's finger positioned at the typical 15° angular distance from their own finger. The participants in one group were told that they would see their own finger, and those in another group were warned that the sight of that finger could sometimes be displaced by an optical device. Pointing to visual targets with the hidden hand was measured before and after working in this simulated prismatic situation, and the obtained after-effects were larger for the participants receiving the deceptive instructions than for those who were more correctly informed. This result certainly demonstrates the influence of knowledge on observed response shifts, but it calls for two remarks. First, the smaller after-effects of the correctly informed participants were nevertheless significant, which shows that, here also, the belief in a unique source is not a necessary condition for recalibration to take place. Second, the superior after-effects of the deceived participants might well be due to some post-perceptual voluntary correction of responses in the post-tests. Thus, a simple tentative interpretation of the findings would be that the after effect observed in the informed group measures automatic recalibration, and the additional effect apparent in the deceived group reveals in addition the influence of belief on responses.

This kind of interpretation has recently received important support from a new application of the illusory limb approach by Pavani *et al.* (2000; see also Chapter 8, this volume). Participants rested both of their hands on the table, under a screen occluding them from sight, each hand holding between thumb and forefinger foam cubes in which were embedded tactile stimulators that could deliver vibrations to either finger or thumb. The participant's task was to respond to the tactile stimulation by a spatially compatible foot movement. Two stuffed rubber gloves, each holding a foam block identical to those held by the participant, were displayed on the table, immediately above the participant's own hands. Two small lights were positioned on the two foam blocks, each close to the thumb or the forefinger of the glove. In the bimodal condition, each vibrotactile stimulus was accompanied synchronously by a distracting flash of light in either the corresponding location on the gloves, or in any one of the three other locations. The principal finding was that the response to tactile targets was slowed down when these were located near the dummy rubber hands (see also Chapter 8, this volume). From our present point of view, the important point to note is that the participants were fully aware that the gloves, which had been put into position in their full view, were not their own hands. The effect thus occurred in spite of conscious knowledge to the contrary. On the other hand, it disappeared when, in another condition, the rubber gloves were placed in an orientation perpendicular to the participant's own

hands. What these remarkable results show is that a sufficiently well simulated body part seen in a location and an orientation compatible with proprioceptive information is mandatorily incorporated into the body schema.

Processing levels: the role of spatial attention

Another issue connected to levels of processing is the possible role of attention in determining crossmodal interactions (see also Chapter 8, this volume). Attention is often invoked cursorily to account for any otherwise unexplained difference in perceptual efficiency, and crossmodal interactions have proven no exception to this. There are, however, also more serious reasons to examine the possibility that attention might be involved in their generation. The extensive recent work on attentional phenomena and their crossmodal spatial interactions (reviewed in Chapters 8–11, this volume) has revealed many instances of interdependence. For instance, Driver and Spence (1994) and Spence and Driver (1996) have shown that it is difficult to orient visual and auditory endogenous attention simultaneously in different directions, thus suggesting the existence of linkages between spatial attention-orienting processes in the two modalities.

Some of our recent work has focused specifically on the possible role of the *orientation of spatial attention* in the causation of ventriloquism. It dealt with both endogenous (or deliberate) orientation and exogenous (or automatic) orientation. The effect of deliberate orienting was examined in the classical visual-bias-of-auditory-localization situation, by having the participant monitor for occasional visual catch events either at the location of the potential visual attractor or at another location (Bertelson *et al.* 2000*a*). The participant's task was to localize trains of tones while ignoring bright squares occurring, on a random half of the trials, at either the left or the right of the center of a screen, in exact temporal synchrony with the tones (see Fig. 7.3(a)). The attentional manipulation consisted of instructing the participant to monitor either the screen center or the attractor square for occasional changes of a small square into a diamond (see Fig. 7.3(b)). It focused therefore on the direction of overt (attention plus gaze) deliberate visual attention. The prediction from the attentional hypothesis was that orienting attention to the square should enhance attraction by that square above the level obtained with central orientation. In fact, strong shifts of sound localization toward the flashing square were obtained, but they were of comparable magnitude, regardless of the orientation of attention (see Fig. 7.3(c)). In a second experiment, two squares, one on either side, were presented on every trial and in synchrony with the sounds, and the same catch events to be monitored for were located on one of them. Attraction toward the attended side would have provided a pure demonstration that attention to a potential attractor increases its attraction capacity. But no such effect of the direction of attention was observed.

In the other study reported by Vroomen *et al.* (2001*a*), the *automatic* orientation of attention was manipulated by resorting to the well known 'singleton phenomenon', by which visual attention is attracted toward one of several simultaneously presented visual

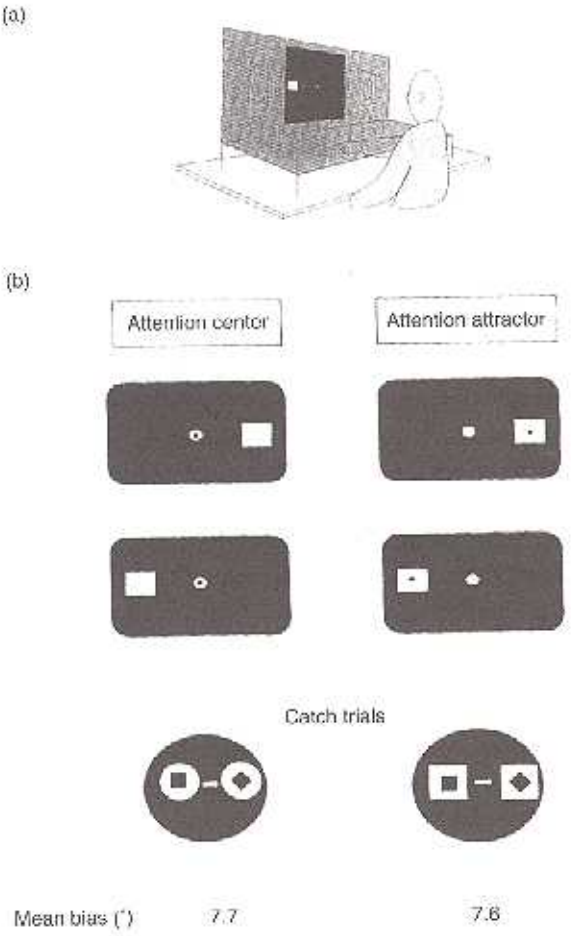


Fig. 7.3 Effect of deliberate overt orientation of visual attention on visual bias of apparent auditory location (after Bertelson *et al.* 2000*a*, experiment 1). (a) Experimental set-up. The participant points with hand out of sight to the apparent azimuthal location of sound trains (6 tones) synchronized with the presentation on the screen of a bright square to the left or right of the center. (b) Visual displays: central fixation white disk with (left) distractor square left or right of center, or without distractor (not represented on figure), and (right) to be monitored small black square on central disk or on distractor square. On catch trials (10%), square becomes diamond. Visual bias significant in both conditions, but there was *no effect of the direction of attention*.

items that differs from the rest by one particular feature (e.g. Treisman and Gelade 1980). Here, the critical idea was to use a singleton whose uniqueness consisted in being *smaller* than the other items in the display. The task was to judge as left or right a train of sounds accompanied by a horizontal row of four bright squares, three big plus a small one, the latter at either the left or the right extremity of the row (see Fig. 7.4). Using the

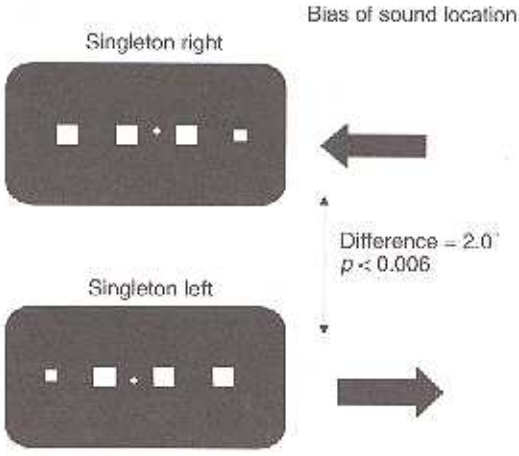


Fig. 7.4 Biasing effect of visual display with smaller singleton on apparent auditory location (after Vroomen *et al.* 2001*a*, experiment 1). Participants made left/right of center judgements concerning tones delivered in synchrony with one of the two displays (singleton left/right) presented at screen center. The stereophonically controlled location of the tones was varied in staircase mode (see Fig. 7.1). Locations at which response reversals occurred with each display showed that *apparent tone location was shifted away from the small singleton (i.e. toward the bigger squares at the other extremity of the display)*.

staircase procedure (Bertelson and Aschersleben 1998), it was shown that the sound was shifted *away* from the small singleton, that is, toward the big squares at the opposite end of the row. Thus, the small singleton did not attract the sound. On the other hand, it effectively captured visual attention, since discrimination of a letter calling for a two-choice reaction was slower when the letter was superimposed on the big square opposite the small singleton, than in the same position in a display with four identical big squares (see Fig. 7.5). Ventriloquism can thus be dissociated from automatic visual attention, and so cannot be mediated by attention. Taken together, the two studies converge on the conclusion that ventriloquism operates at a pre-attentive processing stage. In other words, crossmodal interaction reorganizes the auditory visual spatial scene on which selective attention later operates.

This notion of a pre-attentive site for crossmodal integration is also consistent with the already quoted finding by Driver (1996) that ventriloquism can facilitate attentional selection by moving the apparent location of an auditory target away from its competitors. It has received additional support recently from two studies run independently of each other that showed that sounds can cue visual attention to locations to which they were displaced through ventriloquism (Spence and Driver 2000; Vroomen *et al.* 2001*b*). These demonstrations took their starting point from earlier reports of an asymmetry in crossmodal attentional linkages, in which an auditory cue can capture visual attention, but the opposite effect, visual capture of auditory attention, does not arise, at least not in the specific paradigm used (e.g. Spence and Driver 1997; see also Chapter 11, this volume). Both studies used four loudspeakers, located at the corners of a virtual rectangle, to deliver target

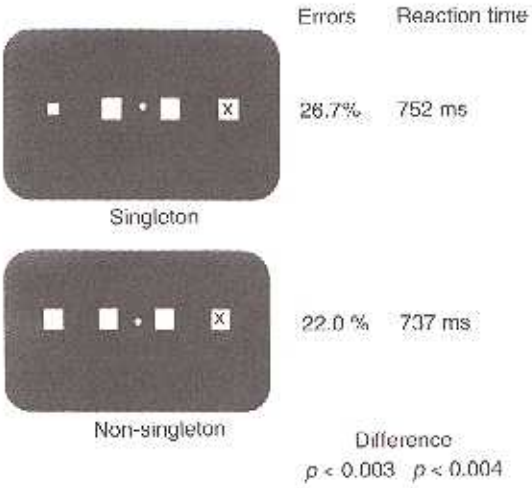


Fig. 7.5 Accelerated discrimination of targets on smaller singleton versus no-singleton display (after Vroomen *et al.* 2001*a*, experiment 2). Visual target (X or O) presented on corresponding extreme squares of displays with singleton versus no singleton (four identical squares). Both accuracy and speed of responding were lower for targets presented on big square opposite singleton on singleton displays, than in corresponding location on no-singleton display. This result shows that *the singleton draws selective attention away from the target.*

auditory stimuli that could appear unpredictably from any one of these four loudspeakers, plus one source in the center to deliver a priming sound. The participant's task was to classify the target as either left/right (Spence and Driver 2000) or high/low (Vroomen *et al.* 2001*b*). This target was preceded at a short interval by the central priming sound. In the critical ventriloquism condition, this priming sound was simultaneously accompanied by an attractor visual flash at a distance in the direction orthogonal to that of the task—thus high/low for Spence and Driver, and left/right in Vroomen *et al.*'s study. In both experiments, the main result was faster location discrimination performance when the location of the target in the (irrelevant) orthogonal direction matched that of the visual attractor. The Vroomen *et al.* (2001*b*) study involved a control condition with the same left or right visual distractors as in the main condition, but no central auditory prime, and the location of that distractor here had no effect whatsoever on performance. This check insured that the ventriloquized priming effect observed in the main condition was not caused directly by the visual attractor—a possibility that was raised by a recent report of visual priming of auditory attention, albeit in a somewhat different situation (Ward *et al.* 2000; again see Chapter 11, this volume). One may thus conclude that auditory attention can effectively be captured by the 'ventriloquized' location of an auditory prime, confirming the pre-attentional status of ventriloquism. Finally, to refer to an earlier point, the new results can also be seen as providing, once again, an indirect demonstration of the automaticity of auditory–visual bias.

The notion that crossmodal pairing can occur at a pre-attentive stage seems difficult to reconcile with Treisman's well-known hypothesis that binding together different

visual features requires attention (Treisman and Gelade 1980). Either crossmodal pairing is different from the kind of within-modality visual binding that Treisman has mostly been dealing with, or the kind of attention necessary for visual binding is different from those manipulated in the preceding experiments.

Modularity, impenetrability, and domains of crossmodal integration

The preceding discussions concerning the functional levels at which crossmodal interactions originate naturally take their place among more general issues concerning cognitive architecture and, in particular, those of modularity and autonomy of subparts of the processing system. These notions were already part of Zeitgeist before their strong formulation by Fodor (1983) and certainly provided the impetus for some of the work that has been described here. In her review papers, Radeau (1994*a*, *b*) discussed the applicability of current modularist proposals to crossmodal interactions, mainly ventriloquism and audiovisual speech recognition. She argued that each of the latter two phenomena display many of the attributes on Fodor's checklist of modularity criteria, such as automaticity, mandatoriness, and impenetrability, while also presenting specificities typical of separate modules.

The more recent developments that have been discussed here have brought further support for the autonomy of at least some cases of crossmodal interaction. There have been new demonstrations of the automaticity of some crossmodal biases (e.g. via psychophysical staircases, or the use of indirect methods) and new arguments against cognitive penetration (pre-attentive site of integration, effect of simulated hands). Of course, there are still areas of uncertainty. For instance, the evidence concerning semantic penetration of multimodal processing is still unclear.

The new developments have implications for the updated version of the modularist position presented by Pylyshyn (1999). This version is less ambitious than Fodor's, to the extent that perceptual autonomy—in particular, impenetrability—is claimed only for one subdomain, *early vision*. This would appear to exclude nonvisual influences at such an early stage. The possibility must be considered that Pylyshyn's thesis is too timid in its limitation to vision, and should be extended to initial processing in different sensory modalities and perhaps even to their interactions.

As our survey has shown, the bulk of behavioral work on multimodality published to date has dealt either with localization (whether of events in the external environment or of body parts) or with audiovisual speech recognition. Studies in the first group are generally related to problems of space perception, attention, sensorimotor coordination, and object recognition. Those in the second naturally refer mainly to the vast body of research on heard speech. As a result of these different affiliations, crosstalk between the two lines of research has largely been limited to occasional citations of well-known publications from the other side. One consequence is that different questions have generally been asked concerning the underlying mechanisms of the respective effects.

There have nevertheless been a few explorations of the relations between the McGurk and ventriloquism effects. One has used a situation that allowed the two phenomena to be examined simultaneously with the same materials (Bertelson *et al.* 1994). On each trial, an ambiguous synthetic speech segment (intermediate between /ama/ and /ana/) was delivered from one of a row of loudspeakers, and a face was displayed simultaneously on a centrally located screen, articulating the bilabial /ama/ or the non-bilabial /ana/ or staying still. The participants had two tasks—to point with a hidden hand to the location of the sound and to repeat 'what had been said'. Ventriloquism could thus be measured as any shift of pointing toward the location of the moving face, and the McGurk effect as any increase in the proportion of responses consistent with the visual distractor. The face was presented either upright or turned upside-down, and inversion had no effect on ventriloquism but reduced the McGurk effect. (The latter reduction was significant only in the condition with visual /ana/. But in a later experiment with more classical material—pairings of auditory/visual /ama/ and /ana/; see description in Bertelson (1998)—face inversion reduced the McGurk effect significantly for the two incongruent pairs of stimuli, while again leaving ventriloquism unaffected.)

Another suggestive result (common to the two experiments) was that, contrary to our expectation (Bertelson 1994), the McGurk effect was practically unaffected by the distance between voice and face, while ventriloquism was modulated by that factor. Using a different measure of ventriloquism (same–different origin judgements), Colin *et al.* (2001) have provided an even stronger demonstration of the contrasting effects of distance: they found a McGurk effect independent of spatial separation over a 160° range, while ventriloquism was practically non-existent beyond 40°. In that study, face inversion was also manipulated, and again affected the McGurk effect but not ventriloquism. Taken together, these results confirm what amounts to a double dissociation between phonetic and spatial audiovisual interactions. One possible implication of such a dissociation would be that the phonetic integration at the basis of the McGurk effect occurs at a later (or 'deeper') processing stage than ventriloquism, whose early determination is supported by much of the data discussed in the present chapter.

Clearly, comparisons between the two domains must be extended to other aspects. For instance, the possible roles of cognitive factors such as response strategies, attention, or familiarity have received more detailed consideration for ventriloquism (and to some extent visuoproprioceptive interaction) than for audiovisual speech.

Taking stock

1. One principal motive behind multimodality research is that most biologically significant situations in the environment provide inputs to several modalities simultaneously. Hence, examining the ways in which information from several modalities interacts is a necessary step for understanding their role in adaptive behavior.

2. Behavioral research with human participants has documented many cases in which judgements of data obtained in one sensory modality were influenced by data obtained simultaneously in other sensory modalities. To the classical cases of interactions in spatial localization, new cases concerning identification first of speech, then of spatiotemporal events, and, more recently, of emotional expression, have been added. While this diversification was certainly a positive development, as a guard against excessive dependence on specific conditions of particular cases, too little effort has been devoted so far at integrating results across domains. Different research questions have typically been asked about different cases.
3. Although the studies have generally claimed to deal with perceptual processes, few have addressed the possible role of post-perceptual factors such as response biases or even voluntary strategies in the generation of the observed effects. It is clear, of course, that results reflecting voluntary corrections performed to satisfy the particular demands of laboratory tasks instead of automatic sensory interactions cannot help us to understand the biological basis of perception (even if they can be of interest for other scientific pursuits, such as the social psychology of the experimenter–participant relation). We have argued here that post-perceptual influences are unavoidable with experimental tasks that are *transparent*, that is, that allow conscious access to critical parameters of the stimulation. The problem has often been recognized and discussed (e.g. Welch and Warren 1980), but ways of dealing with it have rarely been proposed.
4. Three approaches described in this chapter have allowed some success in separating perceptual from post-perceptual influences.

   (a) One was to use subthreshold values of the intermodal incongruence expected to trigger the interaction. This was the principle of the staircase method by which we have shown that ventriloquism occurs even when the participant is not aware of the audiovisual discrepancy (Bertelson and Aschersleben 1998).
   (b) In the *indirect* approach, the participant reports on a stimulus feature different from the one directly affected by the interaction, but the response still allows some valid inferences regarding the latter, as when selective listening to a target item was facilitated through ventriloquism (Driver 1996).
   (c) Finally, in the neuropsychological approach, crossmodal biases have sometimes been observed in patients who had lost all conscious perception of the biasing stimuli (e.g. de Gelder *et al.*, in press).

5. The problem of transparency arises in particular concerning the possible role of top–down knowledge, or 'semantics', in demonstrated interactions. It may be, together with the question of equating sensory factors across different experimental situations, one reason why the existing evidence on semantic influences still appears contradictory.

6. Another question related to processing levels is the role of attention. Recent experiments showing that the visual bias of auditory location appears to be independent of where visual attention is focused (on the visual attractor or somewhere else) imply that ventriloquism can occur at a pre-attentional processing stage.

7. Thus, recent work concerned with locating crossmodal interactions in the functional architecture has stressed the importance of data-driven, automatic, and impenetrable processes. However, in line with remarks in (2) above, before those conclusions can be given a more general extension, the necessary analytic approaches should be applied to a wider range of interactions.

8. As encouraged by the editors, this chapter has concentrated on the evidence from human behavioral studies, providing a historical perspective plus coverage of more recent advances. Chapters dealing with other methodological approaches should provide a wider perspective for the present views. Brain imaging studies, in particular, may provide powerful tests for proposals regarding the functional sites of multimodal interactions that have been proposed based on behavioral work (e.g. see Macaluso *et al.* 2000; also commentaries by de Gelder 2000, and McDonald *et al.* 2000; see also Chapters 9–11, this volume). Concerning the convergence of different approaches, it is important that the contributions of each of the involved subdisciplines should receive due attention. The essential contribution from behavioral work with humans lies in the importance of developing tasks designed according to state-of-the-art methodology, with full consideration of the various possible interpretations of ensuing results. So, our main message to all involved in multisensory work, for whatever purpose, is: don't underestimate the importance of a good task, nor the difficulty of getting one.

Acknowledgements

Work by the authors described in this chapter was partially supported by research grants from the Belgian National Fund for Collective Fundamental Research (Contracts 2.45.39.95 and 10759/2001 2870) and travel grants from the German Max-Planck Society. Thanks are due to the editors, Charles Spence and Jon Driver, and to Salvador Soto-Faraco, plus an anonymous reviewer, for constructive criticisms of an earlier draft. Correspondence to Paul Bertelson, CP 191, Université libre de Bruxelles, 50 Av. F.D. Roosevelt, B-1050 Bruxelles (Belgium) (e mail: plrtlsn@ulb.ac.be).

References

Aschersleben, G. and **Bertelson, P.** (2003). Temporal ventriloquism: crosssmodal interaction on the time dimension. (2) Evidence from sensory–motor synchronization. *International Journal of Psychophysiology* **50**, 157–63.

Bedford, F. (1994). A pair of paradoxes and the perceptual pairing processes. *Current Psychology of Cognition* **13**, 60–8.

Bedford, F. (2001). Towards a general law of numerical/object identity. *Current Psychology of Cognition* **20**, 113–76.

Bermant, R.I. and **Welch, R.B.** (1976). The effect of degree of visual–auditory stimulus separation and eye position upon the spatial interaction of vision and audition. *Perceptual and Motor Skills* **43**, 487–93.

Bertelson, P. (1994). The cognitive architecture behind auditory–visual interaction in scene analysis and speech identification [commentary on Radeau]. *Current Psychology of Cognition* **13**, 69–75.

Bertelson, P. (1998). Starting from the ventriloquist: the perception of multimodal events. In *Advances in psychological science*. Vol.2: *Biological and cognitive aspects* (ed. M. Sabourin, F.I.M. Craik, and M. Robert), pp. 419–39. Psychology Press, Hove, East Sussex.

Bertelson, P. (1999). Ventriloquism: a case of cross-modal perceptual grouping. In *Cognitive contributions to the perception of spatial and temporal events* (ed. G. Aschersleben, T. Bachmann, and J. Müsseler), pp. 347–62. Elsevier, Amsterdam.

Bertelson, P. and **Aschersleben, G.** (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin and Review* **5**, 482–9.

Bertelson, P. and **Aschersleben, G.** (2003). Temporal ventriloquism: crossmodal interaction on the time dimension. (1) Evidence from time order judgments. *International Journal of Psychophysiology* **50**, 147–55.

Bertelson, P. and **Radeau, M.** (1976). Ventriloquism, sensory interaction and response bias: remarks on the paper by Choe, Welch, Gilford and Juola. *Perception and Psychophysics* **19**, 531–5.

Bertelson, P. and **Radeau, M.** (1981). Cross modal bias and perceptual fusion with auditory–visual spatial discordance. *Perception and Psychophysics* **29**, 578–87.

Bertelson, P., Vroomen, J., Wiegeraad, G., and **de Gelder, B.** (1994). Exploring the relation between McGurk interference and ventriloquism. In *ICSLP 94* (Vol. 2, pp. 559–62). Acoustical Society of Japan, Yokohama.

Bertelson, P., Vroomen, J., and **de Gelder, B.** (1997). Auditory–visual interaction in voice localization and speech recognition: the effect of desynchronization. In *Proceedings of the Workshop on Audio visual Speech Processing: Cognitive and Computational Approaches*, Rhodes, Greece (ed. C. Benoit and R. Campbell), pp. 97–100.

Bertelson, P., Vroomen, J., de Gelder, B., and **Driver, J.** (2000*a*). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception and Psychophysics* **62**, 321–32.

Bertelson, P., Pavani, F., Ladavas, E., Vroomen, J., and **de Gelder, B.** (2000*b*). Ventriloquism in patients with unilateral visual neglect. *Neuropsychologia* **38**, 1634–42.

Bertelson, P., Vroomen, J., Aschersleben, G., and **de Gelder, B.** (2001). Object identity decisions: at what processing levels? Or: why the cantaloupe might work. *Current Psychology of Cognition* **20**, 177–82.

Bertelson, P., Vroomen, J., and **de Gelder, B.** (2003). Visual recalibration of auditory speech identification: a McGurk aftereffect. *Psychological Science* **14**, 592–7.

Bregman, A.S. (1990). *Auditory scene analysis: the perceptual organization of sound*. MIT Press, Cambridge, Massachusetts.

Brewster, D. (1839). *Letters on natural magic*. Harper, New York.

Broadbent, D.E. (1958). *Perception and communication*. Pergamon, London.

Caclin, A., Soto-Franco, S., Kingstone, A., and **Spence, C.** (2002). Tactile 'capture' of audition. *Perception and Psychophysics* **64**, 616–30.

Calvert, G.A. (2001). Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex* **11**, 1110–23.

Canon, L.K. (1970). Intermodality inconsistency of input and directed attention as determinants of the nature of adaptation. *Journal of Experimental Psychology* **84**, 141–7.

Choe, C.S., Welch, R.B., Gilford, R.M., and **Juola, J.F.** (1975). The 'ventriloquist effect': Visual dominance or response bias? *Perception and Psychophysics* **18**, 55–60.

Colin, C., Radeau, M., Deltenre, P., and **Morais, J.** (2001). Rules of intersensory integration in spatial scene analysis and speechreading. *Psychologica Belgica* **41**, 131–44.

Craske, B. and **Templeton, W.B.** (1968). Prolonged oscillation of the eyes induced by conflicting position input. *Journal of Experimental Psychology* **76**, 387–93.

de Gelder, B. (2000). More to seeing than meets the eye. *Science* **289**, 1148–9.

de Gelder, B. and **Vroomen, J.** (2000). Perceiving emotions by ear and by eye. *Cognition and Emotion* **14**, 289–311.

de Gelder, B., Vroomen, J., and **Teunisse, J.-P.** (1995). Hearing smiles and seeing cries: the bimodal perception of emotions. *Abstracts of the Psychonomic Society* **1**, 30.

de Gelder, B., Pourtois, G., Vroomen, J., and **Bachoud-Levi, A.-C.** (2000). Covert processing of faces in prosopagnosia is restricted to facial expressions: evidence from cross-modal bias. *Brain and Cognition* **44**, 425–44.

de Gelder, B., Pourtois, G., and **Weiskrantz, L.W.** (2002). Fear recognition in the voice is modulated by unconsciously recognized facial expressions but not by unconsciously recognized affective pictures. *Proceedings of the National Academy of Sciences* USA **99**, 4121–6.

de Gelder, B., Pourtois, G., and **Vroomen, J.** (in press). Multisensory perception of affect, its time course and its neural basis. In *Handbook of multisensory perception* (ed. G. Calvert, C. Spence, and B.E. Stein). MIT Press, Cambridge, Massachusetts.

Dodd, B. and **Campbell, R.** (1987). *Speech by ear and by eye: the psychology of lip-reading*. Erlbaum, Hillsdale, New Jersey.

Dolan, R., Morris, J., and **de Gelder, B.** (2001). Crossmodal binding of fear in voice and face. *Proceedings of the National Academy of Sciences USA* **98**, 10006–10.

Dong, C., Swindale, N.V., and **Cynader, M.S.** (1999). A contingent aftereffect in the auditory system. *Nature Neuroscience* **2**, 863–5.

Driver, J. (1996). Enhancement of listening by illusory mislocation of speech sounds due to lip reading. *Nature* **381**, 66–8.

Driver, J. and **Spence, C.J.** (1994). Spatial synergies between auditory and visual attention. In *Attention and performance 15: Conscious and nonconscious information processing* (ed. C. Umiltà and M. Moscovitch), pp. 311–31. MIT Press, Cambridge, Massachusetts.

Driver, J. and **Spence, C.** (2000). Multisensory perception: beyond modularity and convergence. *Current Biology* **10**, R731–R735.

Epstein, W. (1975). Recalibration by pairing: a process of perceptual learning. *Perception* **4**, 59–72.

Fodor, J. (1983). *The modularity of mind.* MIT Press, Cambridge, Massachusetts.

Green, K.P. (1996). The use of auditory and visual information in phonetic perception. In *Speechreading by humans and machines: models, systems and applications* (ed. D.G. Stork and M.E. Hennecke), pp. 55–78. Springer, Berlin.

Green, K.P., Kuhl, P.K., Meltzoff, A.M., and **Stevens, E.B.** (1991). Integrating speech information across talkers, gender, and sensory modality: female faces and male voices in the McGurk effect. *Perception and Psychophysics* **50**, 524–36.

Guest, S., Catmur, C., Lloyd, D., and **Spence, C.** (2002). Audiotactile interactions in roughness perception. *Experimental Brain Research* **146**, 161–71.

Harris, C.S. (1965). Perceptual adaptation to inverted, reversed, and displaced vision. *Psychological Review* **72**, 419–44.

Hay, J.C., Pick, H.L., and **Ikeda, K.** (1965). Visual capture produced by prism spectacles. *Psychonomic Science* **2**, 215–16.

Held, R. (1965). Plasticity in sensory-motor systems. *Scientific American* **213**, 84–94.

Held, R., Efstathiou, A., and **Greene, M.** (1966). Adaptation to displaced and delayed visual feedback from the hand. *Journal of Experimental Psychology* **72**, 887–91.

Helmholtz, H. von (1866). *Treatise on physiological optics*, Vol. 3. [English translation of 3rd German edition, 1962, Dover, New York.]

Howard, I.P. and **Templeton, W.B.** (1966). *Human spatial orientation.* Wiley, London.

Howard, I.P., Craske, B., and **Templeton, W.B.** (1965). Visuo-motor adaptation to discordant exafferent stimulation. *Journal of Experimental Psychology* **70**, 189–91.

Jack, C.E. and **Thurlow, W.R.** (1973). Effects of degree of visual association and angle of displacement on the 'ventriloquism' effect. *Perceptual and Motor Skills* **37**, 967–79.

Jackson, C.V. (1953). Visual factors in auditory localization. *Quarterly Journal of Experimental Psychology* **5**, 52–65.

Jousmäki, V. and **Hari, R.** (1998). Parchment-skin illusion: sound biased touch. *Current Biology* **8**, R190.

Klemm, O. (1909). Localisation von Sinneneindrücken bei disparaten Nebenreizen [Localization of sensory impressions with disparate distracters]. *Psychologische Studien (Wundt)* **5**, 73–161.

Macaluso, E., Frith, C.D., and **Driver, J.** (2000). Modulation of human visual cortex by crossmodal spatial attention. *Science* **289**, 1206–8.

Manuel, S.Y., Repp, B., Studdert-Kennedy, M., and **Liberman, A.** (1983). Exploring the 'McGurk effect' [abstract]. *Journal of the Acoustical Society of America* **74**, S66.

Massaro, D.W. (1987). *Speech perception by ear and eye: a paradigm for psychological inquiry*. Erlbaum, Hillsdale, New Jersey.

Massaro, D.W. and **Egan, P.B.** (1996). Perceiving affect from the voice and the face. *Psychonomic Bulletin and Review* **3**, 215–21.

McCullough, C. (1965). Color adaptation of edge detectors in the human visual system. *Science* **149**, 1113–16.

McDonald, J.J., Teder-Sälejärvi, W.A., and **Ward, L.M.** (2000). Multisensory integration and crossmodal attention effects in the human brain. *Science* **292**, 1791.

McGurk, H. and **MacDonald, J.** (1976). Hearing lips and seeing voices. *Nature* **264**, 746–8.

Metzger, W. (1934). Beobachtungen über phänomenale Identität [Observations regarding phenomenal identity]. *Psychologische Forschung* **19**, 1–60.

Müller, J. (1838). *Handbuch der Physiologie des Menschen*. II. [Cited by Boring, E.G. (1942). *Sensation and perception in the history of experimental psychology*. Appleton-Century-Crofts, New York.]

Pavani, F., Spence, C., and **Driver J.** (2000). Visual capture of touch: out-of-the-body experiences with rubber gloves. *Psychological Science* **5**, 353–9.

Pick, H.L., Warren D.H., and **Hay, J.C.** (1969). Sensory conflict in judgements of spatial direction. *Perception and Psychophysics* **6**, 203–5.

Pomerantz, J.R. (1981). Perceptual organization in information processing. In *Perceptual organization* (ed. M. Kubovy and J.R. Pomerantz), pp. 141–80. Erlbaum, Hillsdale, New Jersey.

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioural and Brain Sciences* **22**, 341–65.

Radeau, M. (1973). The locus of adaptation to auditory visual conflict. *Perception* **2**, 327–332.

Radeau, M. (1985). Signal intensity, task context, and auditory–visual interaction. *Perception* **14**, 571–7.

Radeau, M. (1992). Cognitive impenetrability in auditory–visual interaction. In *Analytic approaches to human cognition* (ed. J. Alegria, D. Holender, J. Morais, and M. Radeau), pp. 41–55. Elsevier, Amsterdam.

Radeau, M. (1994*a*). Auditory–visual interaction and modularity. *Current Psychology of Cognition* **13**, 3–51.

Radeau, M. (1994*b*). Ventriloquism against audio-visual speech: or where Japanese-speaking barn owls might help? *Current Psychology of Cognition* **13**, 124–40.

Radeau, M. and **Bertelson, P.** (1969). Adaptation à un déplacement prismatique sur la base de stimulations exafférentes en conflit. *Psychologica Belgica* **9**, 133–40.

Radeau, M. and **Bertelson, P.** (1974). The aftereffects of ventriloquism. *Quarterly Journal of Experimental Psychology* **26**, 63–71.

Radeau, M. and **Bertelson, P.** (1976). The effect of a textured visual field on modality dominance in a ventriloquism situation. *Perception and Psychophysics* **20**, 227–35.

Radeau, M. and **Bertelson, P.** (1977). Adaptation to auditory–visual discordance and ventriloquism in semi realistic situations. *Perception and Psychophysics* **22**, 137–46.

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

Reconzone, G.H. (1998). Rapidly induced auditory plasticity: the ventriloquism after effect. *Proceedings of the National Academy of Sciences* **95**, 869–75.

Reisberg, D., McLean, J., and **Goldfield, A.** (1987). Easy to hear but hard to understand: a lip-reading advantage with intact auditory stimuli. In *Hearing by eye: The psychology of lip-reading* (ed. B. Dodd and R. Campbell), pp. 97–114. Erlbaum, Hillsdale New Jersey.

Roberts, M. and **Summerfield, Q.** (1981). Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. *Perception and Psychophysics* **30**, 309–14.

Rock, I. and **Harris, C.S.** (1967). Vision and touch. *Scientific American* **216** (17 May), 96–104.

Rosenblum, L.D. (1994). How special is audiovisual speech integration? *Current Psychology of Cognition* **13**, 110–16.

Saldaña, H.M. and **Rosenblum, L.D.** (1993). Visual influences on auditory pluck and bow judgements. *Perception and Psychophysics* **54**, 406–16.

Sekuler, R., Sekuler, A.B., and **Lau, R.** (1997). Sound alters visual motion perception. *Nature* **385**, 308.

Shams, L., Kamitani, Y., and **Shimojo, S.** (2000). What you see is what you hear: sound induced visual flashing. *Nature* **408**, 788.

Soroker, N., Calamaro, N., and **Myslobodsky, M.S.** (1995). Ventriloquism effect reinstates responsiveness to auditory stimuli in the 'ignored' space in patients with hemispatial neglect. *Journal of Clinical and Experimental Neuropsychology* **17**, 243–55.

Soto-Faraco, S., Lyons, J., Gazzaniga, M.S., Spence, C. and **Kingstone, A.** (2002). The ventriloquist in motion: illusory capture of dynamic information across sensory modalities. *Cognitive Brain Research* **14**, 139–146.

Soto-Faraco, S. and **Kingstone, A.** (2004: in press). Multisensory integration of dynamic information. In *Handbook of Multisensory Perception* (ed. G. Calvert, C. Spence, and B.E. Stein). MIT Press, Cambridge, Massachusetts.

Spence, C. and **Driver, J.** (1996). Audiovisual links in endogenous covert spatial attention. *Journal of Experimental Psychology: Human Perception and Performance* **22**, 1005–30.

Spence, C. and **Driver, J.** (1997). Audiovisual links in exogenous covert spatial orienting. *Perception and Psychophysics* **59**, 1–22.

Spence, C. and **Driver, J.** (2000). Attracting attention to the illusory location of a sound: reflexive crossmodal orienting and ventriloquism. *NeuroReport* **11**, 2057–61.

Spence, C., Shore, D.I., and **Klein, R.M.** (2001). Multisensory prior entry. *Journal of Experimental Psychology: General* **130**, 799–832.

Stein, B.E. and **Meredith, M.A.** (1993). *The merging of the senses*. MIT Press, Cambridge, Massachusetts.

Stratton, G.M. (1897). Vision without inversion of the retinal image. *Psychological Review* **4**, 341–69, 463–81.

Sumby, W.H. and **Pollack, I.** (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America* **26**, 212–15.

Tastevin, J. (1937). En partant de l'expérience d'Aristote: Les déplacements artificiels des parties du corps ne sont pas suivis par le sentiment de ces parties ni pas les sensations qu'on peut y produire [Starting from Aristotle's illusion: The artificial displacements of parts of the body are not followed by feeling in these parts or by the sensations which can be produced there]. *L'Encephale* **1**, 57–84, 140–58.

Thomas, G.J. (1941). Experimental study of the influence of vision on sound localisation. *Journal of Experimental Psychology* **28**, 167–77.

Thurlow, W.R. and **Jack, C.E.** (1973). Certain determinants of the 'ventriloquism effect'. *Perceptual and Motor Skills* **36**, 1171–84.

Treisman, A. and **Gelade, G.** (1980). A feature-integration theory of attention. *Cognitive Psychology* **12**, 97–136.

Vroomen, J. (1999). Ventriloquism and the nature of the unity decision [commentary on Welch]. In *Cognitive contributions to the perception of spatial and temporal events* (ed. G. Aschersleben, T. Bachmann, and J. Müsseler), pp. 388–93. Elsevier, Amsterdam.

Vroomen, J. and **de Gelder, B.** (2003). Visual motion influences the contingent auditory motion aftereffect. *Psychological Science* **14**, 357–61.

Vroomen, J., Bertelson, P., and **de Gelder, B.** (2001*a*). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception and Psychophysics* **63**, 651–9.

Vroomen, J., Bertelson, P., and **de Gelder, B.** (2001*b*). Directing spatial attention towards the illusory location of a ventriloquized sound. *Acta Psychologica* **108**, 21–33.

Wallach, H. (1968). Informational discrepancy as a basis of perceptual adaptation. In *The neuropsychology of spatially oriented behaviour* (ed. S.J. Freeman), pp. 209–30. Dorsey, Homewood, Illinois.

Ward, L.M., MacDonald, J.J., and **Lin, D.** (2000). On asymmetries in crossmodal spatial attention orienting. *Perception and Psychophysics* **62**, 1258–64.

Warren, D.H., Welch, R.B., and **McCarthy, T.J.** (1981). The role of visual-auditory compellingness in the ventriloquism effect: implications for transitivity among the spatial senses. *Perception and Psychophysics* **9**, 557–64.

Watanabe, K. and **Shimojo, S.** (2000). Postcoincidence trajectory duration affects motion event perception. *Perception and Psychophysics* **63**, 16–28.

Weiskrantz, L. (1997). *Consciousness lost and found: a neuropsychological exploration.* Oxford University Press, Oxford.

Welch, R.B. (1972). The effect of experienced limb identity upon adaptation to simulated displacement of the visual field. *Perception and Psychophysics* **12**, 453–6.

Welch, R.B. (1978). *Perceptual modification: adaptation to altered sensory environments.* Academic Press, New York.

Welch, R.B. (1999*a*). Meaning, attention, and the unity assumption in the intersensory bias of spatial and temporal perceptions. In *Cognitive contributions to the perception of spatial and temporal events* (ed. G. Aschersleben, T. Bachmann, and J. Müsseler), pp. 371–87. Elsevier, Amsterdam.

Welch, R.B. (1999*b*). The advantages and limitations of the psychophysical staircases procedure in the study of intersensory bias [Commentary on Bertelson]. In *Cognitive contributions to the perception of spatial and temporal events* (ed. G. Aschersleben, T. Bachmann, and J. Müsseler), pp. 363–9. Elsevier, Amsterdam.

Welch, R.B. and **Warren, D.H.** (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin* **88**, 638–67.

Welch, R.B., DuttonHart, L.D., and **Warren, D.H.** (1986). Contributions of audition and vision to temporal rate perception. *Perception and Psychophysics* **39**, 294–300.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt, II [Investigations of Gestalt theory]. *Psychologische Forschung* **4**, 301–50.

Witkin, H.A., Wapner, S., and **Leventhal, T.** (1952). Sound localization with conflicting visual and auditory cues. *Journal of Experimental Psychology* **43**, 58–67.