To appear in : G. Calvert, C. Spence, & B. E. Stein (Eds.), Handbook of multisensory processes. MIT Press

Perceptual Effects of Cross-modal Stimulation: Ventriloquism and the Freezing Phenomenon

Jean Vroomen and Beatrice de Gelder

Tilburg University, the Netherlands

Running head: Perceptual effects of cross-modal stimulation

Correspondence address:

Jean Vroomen

Tilburg University, Dept. of Psychology

P.O. box 90153

5000 LE TILBURG

The Netherlands

Phone: +31-13-4662394

Fax: +31-13-4662370

mailto:j.vroomen@UvT.nl

Introduction

For readers of a book on multi-modal perception, it probably comes as no surprise to say that most events in real life consist of perceptual inputs in more than one modality and that sensory modalities may influence each other. For example, seeing a speaker not only provides auditory information about what is said, but also visual information about movements of the lips, face, and body, as well as visual cues about the origin of the sound. In handbooks on cognitive psychology, though, comparatively little attention is paid to this multimodal state of affairs and the different senses (e.g.,

seeing, hearing, smell, taste, touch) are treated as distinct and separate modules with little or no interaction. But in recent years, it has become increasingly clear that when the different senses receive correlated input about the same external object or event, information is often combined by our perceptual system to yield a multimodally determined percept.

An important issue is to characterize such multisensory interactions and their cross-modal effects. There are at least three different notions at stake here: One is that information is processed in a hierarchical and strictly feed-forward fashion. On this view, information from different sensory modalities converges into a multimodal representation in a feed-forward way. For example, in the fuzzy logic model of perception (Massaro, 1998) degrees of support for different alternatives from each modality (say audition and vision) are determined and then combined to give an overall degree of support. Information is propagated in a strictly feedforward fashion so that higher-order multimodal representations do not affect lower-order sensory-specific representations. There is thus no cross-talk between the sensory modalities such that, say, vision affects early processing stages of audition or vice versa. Cross-modal interactions in feed-forward models take place only at or beyond multimodal stages. An alternative possibility is that multimodal representations send feedback to primary sensory levels (e.g., Driver & Spence, 2000). On this view, higher-order multimodal levels can affect sensory levels. Vision might thus affect audition, but only via multimodal representations. Alternatively, it may also be the case that cross-modal interactions take place without multi-modal representations. For example, the senses may access each other directly from their sensory-specific systems (e.g., Ettlinger & Wilson, 1990). Vision may then affect audition without involvement of a multi-modal representation (see for example Falchier et al.; 2001 for recent neuroanatomical evidence showing that there are projections from primary auditory cortex to the ‘visual’ are V1).

The role of feedback in sensory processing has, of course, been debated for a long time (for example, the Interactive Activation Model of reading by Rumelhart & McClelland, 1982). However, as far as cross-modal effects are concerned, there is at

present no clear empirical evidence that allows distinguishing between feed-forward, feedback, and direct-access models. Feed-forward models predict that early sensory processing levels should be autonomous and unaffected by higher-order processing levels whereas feedback or direct-access models would, in principle, allow that vision affects auditory processes or vice versa. Although this theoretical distinction seems straightforward, empirical demonstration in favour of one or another alternative has proven to be difficult. One of the main problems is to find measures that are sufficiently unambiguous and that can be taken as 'pure indices' of an auditory or visual sensory process.

Among the minimal requirements to state that a cross-modal effect has perceptual consequences at early sensory stages, the phenomenon should at least be (1) robust, (2) not explainable as a strategic effect and (3) the effect should not occur at response-related processing stages. If one assumes stage-wise processing with sensation coming before attention, (e.g., the 'late-selection' view of attention), one might also want to argue that 4) cross-modal effects should be pre-attentive. If these minimal criteria are met, it becomes at least likely that cross-modal interactions occur at early perceptual processing stages and thus that models that allow access to primary processing levels (i.e. feedback or direct-access models) better describe the phenomenon. In our work on cross-modal perception, we investigated the extent to which such minimal criteria apply to some cases of audio-visual perception. One case concerns a situation where vision affects the localization of a sound (i.e. the "ventriloquism effect"), the other where an abrupt sound affects visual processing of a rapidly presented visual stimulus (the "freezing phenomenon"). In an accompanying chapter, we describe the case of cross-modal interactions in affect perception (de Gelder, Pourtois & Vroomen, this volume). Each of these phenomena we consider to be based on cross-modal interactions affecting early levels of perception.

Vision Affecting Sound Localization: The Ventriloquist Effect.

Presenting synchronous auditory and visual information in slightly separate locations creates the illusion that the location of the sound is shifted in the direction of the visual

stimulus. Although the effect is smaller, shifting of the visual percept in the direction of the sound has, at least in some studies, also been observed (Bertelson & Radeau, 1981). The auditory shift is usually measured by asking subjects to localize the sound by means of pointing or by fixating the eyes on the apparent location of the sound. When localization responses are compared to a control condition (e.g. a condition in which the sound is presented in isolation), one usually observes a shift of a few degrees in the direction of the visual stimulus. Reactions to such an audio-visual spatial conflict are designated by the term ventriloquism, because one of their most spectacular everyday examples is the illusion created by performing ventriloquists that the speech they produce without visible facial movements comes from a puppet they agitate in synchrony with the speech.

A standard explanation of the ventriloquist effect is that in case auditory and visual stimuli occur in close temporal and spatial proximity, the perceptual system assumes that a single event occurred. The perceptual system then tries to reduce the conflict between the location of the visual and auditory data because there is an *a priori* constraint that an object or event can have only one location (e.g. Bedford, 1999). Shifting the auditory location in the direction of the visual event rather than the other way around would seem to be ecologically useful because spatial resolution in the visual modality is better than in the auditory one.

However, there are also other, more trivial explanations of the ventriloquist effect. One alternative is similar to Stroop-task interference: When two conflicting stimuli are presented together - like the word 'blue' written in red ink -, there is competition at the level of response selection rather than at a perceptual level per se. Stroop-like response competition may also be at stake in the ventriloquist situation. In that case, there would be no real attraction between sound and vision, but the ventriloquist illusion would be derived from the fact that subjects sometimes point to the visual stimulus instead of the sound by mistake. Strategic or cognitive factors may also play a role. For example, a subject may wonder why sounds and light are presented from different locations, and then adopt a post-perceptual response strategy that satisfies the experimenter's ideas of the task (Bertelson, 1999). Of course, these possibilities are not exclusive, and one

has to find ways to check or circumvent them.

In our research, we dealt with these and other aspects of the ventriloquist situation in the hope of showing that the apparent location of a sound is indeed shifted at a perceptual level of auditory space perception. More specifically, we asked whether a ventriloquist effect can be observed when (1) subjects are explicitly trained to ignore the visual distracter; (2) when cognitive strategies of the subject to respond in a particular way can be excluded; (3) when the visual distracter is not attended, either endogenously or exogenously; (4) when the visual distracter is not seen consciously; and (5) whether the ventriloquist effect as such is possibly a pre-attentive phenomenon.

1) A visual distracter cannot be ignored.

In a typical ventriloquist situation, subjects are asked to locate a sound while ignoring a visual distracter. Typically, subjects remain unaware of how well they perform during the experiment and how well they succeed in obeying instructions. In one of our experiments, though, we asked whether it is possible to train subjects explicitly to ignore the visual distracter (Vroomen et al., 1998). If despite training it is impossible to ignore the visual distracter, this speaks to the robustness of the effect. Subjects were trained to discriminate among sequences of tones that emanated either from a central location only or from alternating locations, in which case two speakers located next to a computer screen emitted the tones. With no visual input, this same/different location task was very easy, because the difference between the central and lateral locations was clearly noticeable. However, the task was much more difficult when, in synchrony with the tones, light flashes were alternated left and right on a computer screen. This condition created the strong impression that sounds from the central location now alternated between left and right, presumably because the light flashes attracted the apparent location of the sounds. We then tried to train subjects to discriminate centrally presented, but ventriloquized sounds from sounds that alternated physically between the left and right. Subjects were instructed to ignore the lights as much as possible (but without closing their eyes) and they received corrective feedback after each trial. The results were that the larger the separation between the lights, the more false alarms

occurred (responding ''alternating sound' on a centrally presented ventriloquized sound), presumably because the farther apart the lights, the farther apart was the perceived location of the sounds. Moreover, in spite of feedback provided on each trial, performance did not improve in the course of the experiment. Instructions and feedback could thus not overcome the effect of the visual distracter on sound localization, which indicates that the ventriloquist effect is indeed very robust.

2) A ventriloquist effect is obtained even when cognitive strategies can be excluded. When subjects are asked to point to the location of auditory stimuli while ignoring spatially discrepant visual distracters, subjects may be aware of the spatial discrepancy and adjust their response accordingly. The visual bias one obtains may then reflect postperceptual decisions rather than genuine perceptual effects. However, contamination by strategies can be prevented when ventriloquist effects are studied via a staircase procedure or when studied as an after-effect.

Bertelson and Aschersleben (1998) were the first to apply the staircase procedure in the ventriloquist situation. The advantage of a staircase procedure is that it is not transparent and that the effects are therefore more likely to reflect genuine perceptual processes. In the staircase procedure by Bertelson and Aschersleben, subjects had to judge the apparent origin of a stereophonically controlled sound as left or right of a median reference point. Unknown to the subjects, the location of the sound was changed as a function of their judgement, following the principal of the psychophysical staircase. After a 'left' judgement, the next sound on the same staircase was moved one step to the right, and vice versa. A staircase started with sounds coming from an extreme left or an extreme right position. At that stage, correct responses are generally given on each successive trial so that the target sounds move progressively towards the centre. Then, at some point, response reversals (i.e. responses different from the preceding one on the same staircase) begin to occur. From this point on, the subject is no longer certain regarding the location of the sound. The location at which these response reversals occur is the dependent variable. In the study by Bertelson and Aschersleben, sounds were delivered together with a visual distracter

(a light-emitting diode, LED) in a central location. When the LED was synchronized with the sound, response reversal occurred earlier than when the light was desynchronised with the sound. Apparently, the synchronized LED attracted the apparent location of the sound toward its central location so that response reversal occurred earlier on the staircase. Similar results have now been reported by Caclin et al. (in press) showing that a centrally located tactile stimulus attracts a peripheral sound towards the middle. Importantly, there is no way in which subjects can figure out a response strategy that might lead to this result, because once response reversal begin to occur, subjects do not know anymore whether a sound belongs to a left or right staircase. A conscious response strategy in this situation is thus extremely unlikely to account for the effect.

Conscious strategies are also unlikely to play a role when the effect of presenting auditory and visual stimuli at separate locations is measured as an after-effect. The after-effect is a shift in the apparent location of unimodally presented acoustic stimuli consequent on exposure to synchronous, but spatially disparate auditory-visual stimulus pairs. Initial studies used prisms to shift the relative locations of visual and auditory stimuli (Canon, 1971; Radeau & Bertelson, 1974). Participants localized acoustic targets before and after a period of adaptation. During the adaptation phase, there was a mismatch between the spatial locations of acoustic and visual stimuli. Typically, between pre- and post-test a shift of about 1°-4° was found in the direction of the visual attracter. Presumably, the spatial conflict between auditory and visual data during the adaptation phase was resolved by recalibration of the perceptual system, and this alteration lasted long enough to be detected as after-effects. Importantly, after-effects are measured by comparing uni-modal pointing to a sound before and after an adaptation phase. Stroop-like response competition between the auditory target and visual distracter during the test situation thus play no role because the test sound is presented without a visual distracter. Moreover, after-effects are usually obtained when the spatial discrepancy between auditory and visual stimuli is so small that subjects do not even notice the separation (Radeau & Bertelson, 1974; Vroomen et al., in prep.; Woods & Recanzone, in press). After-effects can therefore be interpreted as true

perceptual recalibration effects (e.g., Radeau, 1994).

3) Attention towards the visual distracter is not needed to obtain a ventriloquist effect

A relevant question is whether attention plays a role in the ventriloquist effect. One could argue, as Treisman and Gelade (1980) have done in feature-integration theory for the visual modality, that focussed attention might be the 'glue' that combines features across modalities. Could it be, then, that when a sound and visual distracter are attended, an integrated cross-modal event is perceived, but when unattended, two separate events are perceived that do not interact? If so, one might predict that a visual distracter would have a stronger effect on the apparent location of a sound when it is focused upon.

We considered the possibility that ventriloquism indeed requires or is modulated by this kind of focused attention (Bertelson et al., 2000b). The subjects' task was to localize trains of tones while monitoring visual events on a computer screen. On experimental trials, a bright square appeared on the left or on the right of the screen in exact synchrony with the tones. No square appeared on control trials. The attentional manipulation consisted of having subjects monitor either the centre of the display, in which case the attracter square was in the visual periphery, or the lateral square itself for occasional occurrences of a catch stimulus (a very small diamond that could only be detected when in fovea). The attentional hypothesis predicts that the attraction of the apparent location of the sound by the square would be stronger with attention focused on the attracter square than with attention focused on the centre. In fact, though, equal degrees of attraction were obtained in the two attention conditions. Focused attention did thus not modulate the ventriloquist effect.

However, the effect of attention might have been small and overruled by the bottom-up information from the laterally presented visual square. What would happen when the bottom up information would be more ambiguous? Would an effect of attention then appear? In a second experiment, we used bilateral squares that were flashed in synchrony with the sound so as to provide competing visual attracters. When the two squares were of equal size, auditory localization was unaffected by which side

participants monitored for visual targets, but when one square was larger than the other, auditory localization was reliably attracted towards the bigger square, again regardless of where visual monitoring was required. This led to the conclusion that the ventriloquist effect largely reflects automatic sensory interactions with little or no role for attention.

In discussing how attention might influence ventriloquism, though, one must distinguish several senses in which the term attention is used. One may attend to one sensory modality rather than another, regardless of location (Spence & Driver, 1997a), or one may attend to one particular location rather than another, regardless of modality. Furthermore, in the literature on spatial attention, two different means of the allocation of attention are generally distinguished. First, there is an endogenous process by which attention can be moved voluntarily. Second, there is an automatic or exogenous mechanism by which attention is reoriented automatically to stimuli in the environment with some special features. The study by Bertelson et al. (2000) manipulated endogenous attention by asking a subject to focus either on one or the other location. Yet, it may have been the case that the visual distracter received a certain amount of exogenous attention independent of where the subject was focusing. For that reason one might ask whether capture of exogenous attention by the visual distracter is essential to affect the perceived location of a sound.

To investigate this possibility, we tried to create a situation in which exogenous attention was captured in one direction whereas the apparent location of a sound was ventriloquized in the other direction. (Vroomen et al., 2001a). Our choice was influenced by earlier data showing that attention can be captured by a visual item differing substantially by one or several attributes (like colour, form, orientation, shape) from a set of identical items among which it is displayed (e.g., Treisman & Gelade, 1980). The unique item has been called the singleton, and its influence on attention is referred to as the singleton effect. If ventriloquism is mediated by exogenous attention, one predicts that presenting a sound in synchrony with a display that contains a singleton should shift the apparent location of the sound toward the singleton. Consequently, finding a singleton that would not shift the location of a sound in its direction would provide

evidence that exogenous attention can be dissociated from ventriloquism.

We used a psychophysical staircase procedure as in Bertelson and Aschersleben (1998). The occurrence of visual bias was examined by presenting a display in synchrony with the sound. We tried to shift the apparent location of the sound in the opposite direction of the singleton by using a display that consisted of four horizontally aligned squares; two big squares on one side, and a big square and a small square (the singleton) at the other side (see Figure 1). The singleton was either in the far left or in the far right position. A visual bias dependent on the position of the singleton should manifest itself at the level of the locations at which reversals begin to occur on the staircases for the two visual displays. If, for instance, the apparent location of the sound were attracted toward the singleton, reversals would first occur at locations more to the left for the display with the singleton on the right than for the display with the singleton on the left

The results of this experiment were very straightforward. The apparent origin of the sound was not shifted towards the singleton, but actually in the opposite direction, i.e. towards the two big squares. Apparently, the two big squares on one side of the display were attracting the apparent origin of the sound more strongly than the small and big square at the other side. Thus, the attracter size effect that we previously obtained (Bertelson et al., 2000b) occurred with the present visual display as well. This result thus suggested that attraction of a sound was not mediated through exogenous attention capture. However, before that conclusion could be drawn, it was necessary to check that the visual display had the capacity to attract attention towards the singleton. We therefore ran a control experiment in which the principle was to measure the attention attraction capacity of the small square through its effect on the discrimination of targets presented elsewhere in the display. In the singleton condition, participants were shown the previously used display with the three big squares and the small one. A target letter X or O, calling for a choice reaction, was displayed in the most peripheral big square opposite the singleton. In the control condition, the display consisted of four equally sized big squares. Discrimination performance was worse in the singleton condition than in the control condition, thus showing that attention was attracted away

from the target letter and toward the singleton. Nevertheless, one might still argue that a singleton in the sound localization task did not capture attention because subjects were paying attention to audition, and not vision. In a third experiment, we therefore randomised sound localization trials with visual X/O discrimination trials so that subjects did not know in advance which task they had to perform. When subjects saw an X or an O, they pressed as fast as possible a corresponding key; otherwise, when no letter was detected, they decided whether the sound had come from the left or right of the central reference. With this mixed design, results were still exactly as before: Attention was attracted toward the singleton while the sound was shifted away from the singleton. Strategic differences between an auditory and visual task were thus unlikely to explain the result. Rather, we demonstrated a dissociation between ventriloquism and exogenous attention: The apparent location of the sound was shifted towards the two big squares (or the 'centre of gravity' of the visual display), while the singleton attracted exogenous attention. The findings from the studies concerning the role of exogenous attention together with those of the earlier one showing the independence of ventriloquism from the direction of endogenous attention (Bertelson et al., 2000b) thus support the conclusion that ventriloquism is not affected by the direction of attention.

4: The ventriloquist effect is still obtained when the visual distracter is not seen consciously.

The conclusion that attention is not needed to obtain a ventriloquist effect is further corroborated by our work on patients with unilateral visual neglect (Bertelson, et al., 2000a). The neglect syndrome is usually interpreted as an attentional deficit and reflected in a reduced capacity to report stimuli in the contra-lateral side (usually the left). Previously, it had been reported that ventriloquism could improve the attentional deficit. Soroker, et al. (1995) showed that inferior identification of syllables delivered through a loudspeaker on the left (auditory neglect) could be improved when the same stimuli on the left were administered in the presence of a fictitious loudspeaker on the right. The authors attributed this improvement to a 'ventriloquist' effect, even though their setting was very different from the usual ventriloquist situation. That is, their visual

stimulus was stationary, whereas typically the onset and offset of an auditory and visual stimulus are synchronized. The effect was therefore probably mediated by higher-order knowledge about the fact that sounds can be delivered through loudspeakers.

In our research, we used the more typical ventriloquist situation (a light and sound presented simultaneously) and asked whether a visual stimulus that remains undetected because it is presented in the neglected field, nevertheless shifts the apparent location of the sound towards its location. This may occur because although perceptual awareness is compromised in neglect, much perceptual processing can still proceed unconsciously for the affected side. Our patients with left visual neglect consistently failed to detect a stimulus presented in their left visual field, but nevertheless, their pointing to a sound was shifted in the direction of the visual stimulus. This is thus another demonstration that ventriloquism is not depending on attention or even awareness of the visual distracter.

5) The ventriloquist effect is a pre-attentive phenomenon

The previous studies led us to conclude that cross-modal interactions take place without the need of attention. This stage is presumably one concerned with the initial analysis of the spatial scene (Bertelson, 1994). The presumption receives additional support from the findings by Driver (1996) in which the visual bias of auditory location was measured in the classical "cocktail party" situation through its effect in facilitating the focusing of attention on one of two simultaneous spoken messages. Subjects found the shadowing task easier when the apparent location of the target sound was attracted away from the distracter by a moving face. This result thus implies that focused attention operates on a representation of the external scene that has already been spatially reorganized by cross-modal interactions.

We asked whether a similar cross-modal reorganization of external space occurs when exogenous rather than focused attention is at stake (Vroomen et al., 2001b). To do so, we used the orthogonal cross-modal cueing task introduced by Spence and Driver (1997b). In this task, subjects have to judge the elevation (up vs. down, regardless of whether it is on the left or right of fixation) of peripheral targets in either

audition, vision, or touch following an uninformative cue in either one of these modalities. In general, cueing effects (i.e. faster responses when the cue is on the same side as the target) have been found across all modalities, except that visual cues do not affect responses to auditory targets (Driver & Spence, 1998; but see McDonald et al., 2001; Spence, 2001). This then opens an intriguing possibility: What happens with an auditory cue whose veridical location is in the centre, but whose apparent location is ventriloquized towards a simultaneous light in the periphery. Can such a ventriloquized cue affect responses to auditory targets? The ventriloquized cue consisted of a tone presented from an invisible central speaker synchronized with a visual cue presented on the left or right. Depending on SOA (100, 300, 500 ms), a target sound (white noise bursts) was delivered with equal probabilities from one of the four target speakers. Subjects made a speeded decision about whether the target had been delivered through one of the upper or one of the lower speakers. Results showed that visual cues had no effect on auditory target detection (see also Spence & Driver, 1997b). More important, ventriloquized cues had no cueing effect at 100 ms SOA, but the facilitatory effect appeared at 300 and 500 ms SOA. This suggests that a ventriloquized cue directed auditory exogenous attention to the perceived rather than the physical auditory location, implying that the cross-modal interaction between vision and audition reorganized space on which auditory exogenous attention operates. Spence and Driver (2000) reported similar cueing-effects (with a somewhat different time-course) in their study of ventriloquized cues in the vertical dimension. They showed that a visual cue presented above or below fixation led to a vertical shift of auditory attention when it was paired with a tone presented at fixation. Attentional capture can thus be directed to the apparent location of a ventriloquized sound, suggesting that cross-modal integration precedes or at least co-occurs with reflexive shifts of covert attention.

To summarize the case of ventriloquism, the apparent location of a sound can be shifted in the direction of a visual stimulus that is synchronized with the sound. It is unlikely that this robust effect can be explained solely by voluntary response strategies, as it is obtained with a psychophysical staircase procedure and can be observed as an after-effect. The effect is even obtained when the visual distracter is not seen

consciously as in patients with hemi-neglect. Moreover, the ventriloquist effect does not require attention because it is not affected by whether a visual distracter is focussed upon or not, and the direction of ventriloquism can be dissociated from where visual attention is captured. In fact, the ventriloquist effect may be a pre-attentive phenomenon as auditory attention can be captured at the ventriloquized location of a sound. Taken together, this shows that the ventriloquist effect is perceptually 'real'.

Sound Affecting Vision: The 'Freezing Phenomenon'.

Recently, we described a case where sound affects vision (Vroomen & de Gelder, 2000). The basic phenomenon is that when subjects are shown a rapidly changing visual display, an abrupt sound may 'freeze' the display with which the sound is synchronized. Perceptually, it looks as if the display is brighter or shown for a longer time. We described this phenomenon against the background of 'scene analysis'.

Scene analysis refers to the notion that information arriving at the sense organs must be parsed into objects and events. In vision, scene analysis succeeds despite partial occlusion of one object by the other, the presence of shadows extending across object boundaries, and deformations of the retinal image produced by moving objects. Vision is not the only modality in which object segregation occurs. Auditory object segregation has also been demonstrated (Bregman, 1990). It occurs, for instance, when a sequence of alternating high- and low frequency tones is played at a certain rate. When the frequency difference between the tones is small, or when they are played at a slow rate, listeners are able to follow the entire sequence of tones. But at bigger frequency differences or higher rates, the sequence splits into two streams; one high and one low in pitch. While it is possible to shift attention between the two streams, it is difficult to report the order of the tones in the entire sequence. Auditory stream segregation appears to follow, like apparent motion in vision, Korte's third law (Korte, 1915). When the difference in frequency between the tones increases, stream segregation occurs at longer stimulus onset asynchronies.

Bregman (1990) described a number of Gestalt principles for auditory scene analysis in which he stressed the resemblance between audition and vision, since

principles of perceptual organization such as similarity (in volume, timbre, spatial location), good continuation, and common fate seem to play similar roles in the two modalities. Such a correspondence between principles of visual and auditory organization raises the question of whether the perceptual system utilizes information from one sensory modality to organize the perceptual array in the other modality. Or, in other words, is scene analysis itself a cross-modal phenomenon?

Previously, O'Leary and Rhodes (1984) showed that perceptual segmentation in one modality could influence the concomitant segmentation in another modality. They used a display of six dots, three high and three low. The dots were displayed one-by-one, alternating between the high and low positions and moving from left-to-right. At slow rates, a single dot appeared to move up and down, while at faster rates two dots were seen as moving horizontally, one above the other. A sequence that was perceived as two dots caused a concurrent auditory sequence to be perceived as two tones as well at a rate that would yield a single perceptual object when the accompanying visual sequence was perceived as a single object. The number of objects seen thus influenced the number of objects heard. They also found the opposite influence from audition to vision. Segmentation in one modality thus affected segmentation in the other modality. To us, though, it was not clear whether the cross-modal effect was truly perceptual, or whether it occurred because participants deliberately changed their interpretation about the sounds and dots. It is well known that there is a broad range of rates/tones at which listeners can hear, at will, one or two streams (van Noorden 1975). O'Leary and Rhodes presented ambiguous sequences, and this raises the possibility that a cross-modal influence was found because perceivers changed their interpretation about the sounds and dots, but the perception may have been the same. For example, participants under the impression of hearing two streams instead of one may infer that in vision there should also be two streams instead of one. Such a conscious strategy would explain the observations of the cross-modal influence without the need for a direct perceptual link between audition and vision.

We pursued this question in a study that led us to observe the freezing phenomenon. We first tried to determine whether the freezing of the display to which an

abrupt tone is synchronized is a perceptually genuine effect or not. Previously, Stein et al., (1996) had shown, with normal subjects, that a sound enhances the perceived visual intensity of a stimulus. The latter seemed to be a close analogue of the freezing phenomenon we wanted to create. However, Stein et al. used a somewhat indirect measure of visual intensity (a visual analogue scale in which participants judged the intensity of a light by rotating a dial), and they could not find an enhancement by a sound when the visual stimulus was presented subthreshold. It was therefore unclear whether their effect was truly perceptual rather than post-perceptual. In our experiments, we tried to avoid this difficulty by using a more direct estimate of visual persistence by measuring speeded performance on a detection task. Participants saw a four-by-four matrix of flickering dots that was created by rapidly presenting four different displays, each containing four dots in quasi-random positions (see Figure 2). Each display on its own was difficult to see, because it was shown only briefly and was immediately followed by a mask. One of the four displays contained a target to be detected. The target consisted of four dots that made up a diamond in the upper-left, upper-right, lower-left, or lower-right corner of the matrix. The task of the participants was to detect the position of the diamond as fast and as accurately as possible. We investigated whether the detectability of the target could be improved by an abrupt sound presented together with the target. The tones were delivered through a loudspeaker under the monitor. Participants in the experimental condition heard a high tone at the target display, and a low tone at the other four-dots displays (the distracters). In the control condition, participants heard only low tones. The idea was that the high tone in the sequence of tones segregated from the low tones, and that under these circumstances it would increase the detectability of the target display. The results indeed showed that the ease of detection of the target was improved when it was synchronized with the high tone. Subjects were faster and more accurate when a high tone was presented at target onset. Was it the case that the high tone simply acted as a warning signal that gave subjects information about when to expect the target? In a second experiment we controlled for this possibility and synchronized the high tone with a distracter display that immediately preceded the target. Subjects were informed about

the temporal relation between high tone and target display and thus knew that the target would be presented right after the high tone. Yet, despite the fact that subject were now also given a cue about when to expect the target, performance actually got worse. As reported by subjects, the reason is probably that the high tone contributed to higher visibility of the distracter display with which it was synchronized, thereby increasing interference.

However, the most important result was that we could show that the perceptual organization of the tone sequence determined the cross-modal enhancement. Our introspective observation was that visual detection was only improved when the high tone segregated from the tone sequence. In our next experiment we prevented segregation of the high tone by making it part of the beginning of the well-known tune 'Frère Jacques'. When subjects heard repetitively a Low-Middle-High-Low tone sequence while seeing the target on the third high tone, there was no enhancement of the visual display. Thus, the perceptual organization of the tone in the sequence increased the visibility of the target display rather than the high tone per se, showing that cross-modal interactions can occur at the level of scene analysis.

How to Qualify the Nature of Cross-modal Interactions?

We argued that ventriloquism and the freezing phenomenon are two examples of intersensory interactions with consequences at perceptual processing levels. They may therefore be likely candidates showing that cross-modal interactions can affect primary sensory levels. There is now also some preliminary neurophysiological evidence showing that brain areas that are usually considered to be 'unimodal' can be affected by input from different modalities. For example, with functional magnetic resonance imaging Calvert et al. (1997) found that lip-reading could affect primary auditory cortex. In a similar vein, Macaluso et al. (2000) showed that a tactile cue could enhance neural responses to a visual target in visual cortex. Giard and Peronnet (1999) also reported that tones synchronized with a visual stimulus affect event-related potentials (ERPs) in visual cortex. Pourtois et al. (2000) found early modulation of auditory ERPs when facial expressions of emotions were presented with auditory sentence fragments of which the

prosodic content was congruent or incongruent with the face. When the face-voice pairs were congruent, there was a bigger auditory N1 component at around 110 ms than when they were incongruent. All these findings are in line with the idea that there is feedback from multi-modal levels to unimodal levels of perception or with the notion that sensory modalities access each other directly.

Such cross-talk between primary sensory areas may also be related to the fact that the subjective experience of cross-modal interaction affects the target modality. In the McGurk-effect (McGurk & MacDonald, 1976), visual information provided by lip-reading changes the way a sound is heard. In the ventriloquist situation, when a sound is presented with a spatially conflicting light, the location of the sound is changed. With emotions: when a fearful voice is shown with a happy face, the voice sounds happier (de Gelder & Vroomen, 2000). There are other examples of such qualitative changes: For example, when a single flash of light is accompanied with multiple beeps, the light is seen as multiple flashes (Shams et al, 2000). These multimodally determined percepts thus have the unimodal qualia of the sensory input from the primary modality, and this may be due to the existence of back projections to the primary sensory areas.

Yet, this does not mean to say that a percept resulting from cross-modal interactions is, in all its relevant aspects, equivalent to its unimodal counterpart. For example, is a ventriloquized sound in all its perceptual and neuro-physiological relevant dimensions the same as a sound played from the direction from where the ventriloquized sound was perceived? For McGurk-like stimulus combinations (i.e., hearing /ba/ and seeing /ga/), it has been shown that the auditory component can be dissociated from the perceived component, as the contrast effect in adaptation is driven by the auditory stimulus, and not by the perceived aspect of the audio-visual stimulus combination (Roberts & Summerfield, 1981). For other cross-modal phenomena such as ventriloquism, there are still a number of intriguing questions where it remains to be shown to which extent the illusory percept is indistinguishable from its veridical counterpart.

References.

Bedford, F. L. (1999). Keeping perception accurate. Trends in Cognitive Sciences, 2, 4-11.

Bertelson, P. (1994). The cognitive architecture behind auditory-visual interaction in scene analysis and speech identification. Current Psychology of Cognition, 13, 69-75.

Bertelson, P. (1999). Ventriloquism: A case of cross-modal perceptual grouping. In G. Aschersleben, T. Bachmann, & J. Müssler (Eds.). Cognitive contributions to the perception of spatial and temporal events. (pp. 347-362). Amsterdam: Elsevier Science.

Bertelson, P. & Aschersleben, G. (1998) Automatic visual bias of auditory location. Psychonomic Bulletin & Review, 5, 482-489.

Bertelson, P., Pavani, F., Ladavas, E., Vroomen, J., & de Gelder, B. (2000). Ventriloquism in patients with unilateral visual neglect. Neuropsychologia, 38, 1634-1642.

Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. Perception & Psychophysics, 29, 578-584.

Bertelson, P., Vroomen, J., de Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. Perception & Psychophysics, 62, 321-332.

Bregman, A. S. (1990). Auditory scene analysis. Cambridge, MA: The MIT Press.

Caclin, A., Soto-Faraco, S., Kingstone, A., & Spence, C. (in press). Tactile 'capture' of audition. Perception & Psychophysics.

Calvert, G. A., Campbell, R., Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of cross-modal binding in the human heteromodal cortex. Current Biology, 10, 649-657.

Driver, J. (1996). Enhancement of selective listening by illusory mislocalization of speech sounds due to lip-reading. Nature, 381, 66-68.

Driver, J., & Spence, C. (1998). Attention and the cross-modal construction of space. Trends in Cognitive Sciences, 2, 254-262.

Driver, J., & Spence, C. (2000). Multisensory perception: Beyond modularity and convergence. Current Biology, 10, 731-735.

Ettlinger, G., & Wilson, W. A. (1990). Cross-modal performance: Behavioural processes, phylogenetic considerations and neural mechanisms. Behavioural Brain Research, 40, 169-192.

Falchier, A., Renaud, L., Barone, P., Kennedy, H. (2001). Extensive projections from the primary auditory cortex and polysensory area STP to peripheral area V1 in the macaque. Society for Neuroscience Abstract, 27, 511-521.

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioural and electrophysiological study. Journal of Cognitive Neuroscience, 11, 473-490.

Korte, A. (1915). Kinematoscopische Untersuchungen. Zeitschrift für Psychologie der Sinnesorgane, 72, 193-296. [Kinematoscopic research. Journal of the Psychology of the Sense Organs].

Macaluso, E., Frith, C., & Driver, J. (2000). Modulation of human visual cortex by cross-modal spatial attention. Science, 289, 1206-1208.

Massaro, D. W. (1998). Perceiving talking faces: From speech perception to a behavioral principle. Cambridge: MA: MIT Press.

McDonald, J. J., Teder-Sälejärvi, W. A., Heraldez, D., & Hillyard, S. A. (2001). Electrophysiological evidence for the 'missing link' in cross-modal attention. Canadian Journal of Experimental Psychology, 55, 143-151.

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

Noorden, L. P. A. S. van (1975). Temporal coherence in the perception of tone sequences. Unpublished Doctoral Dissertation, Technische Hogeschool Eindhoven, The Netherlands.

O'Leary, A., & Rhodes, G. (1984). Cross-modal effects on visual and auditory object perception. Perception and Psychophysics, 35, 565-569.

Pourtois, G., de Gelder, B., Vroomen, J., Rossion, B., & Crommelink, M. (2000). The time-course of intermodal binding between seeing and hearing affective information. Neuroreport, 11, 1329-1333.

Radeau, M. (1994). Auditory-visual spatial interaction and modularity. Current Psychology of Cognition, 13, 3-51.

Radeau, M. & Bertelson, P. (1977). Adaptation to auditory-visual discordance and ventriloquism in semi-realistic situations. Perception and Psychophysics, 22, 137-146.

Roberts, M., & Summerfield, Q. (1981). Audiovisual presentation demonstrates that selective attention to speech is purely auditory. Perception & Psychophysics, 30, 309-314.

Rumelhart, D., & McClelland, J. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. Psychological Review, 89, 60-94.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. Nature, 408, 708.

Soroker, N., Calamaro, N., Myslobodsky, M. S. (1995). Ventriloquism reinstates responsiveness to auditory stimuli in the 'ignored' space in patients with hemispatial neglect. Journal of Clinical and Experimental Neuropsychology, 17, 243-255.

Spence, C. (2001). Cross-modal attentional capture: A controversy resolved? In C. Folk & B. Gibson (Eds.), Attention, Distraction and Action: Multiple perspectives on Attentional Capture (pp 231-262). Elsevier Science BV: Amsterdam.

Spence, C. & Driver, J. (1997a). On measuring selective attention to a specific sensory modality. Perception & Psychophysics, 59, 389-403.

Spence, C. & Driver, J. (1997b). Audiovisual links in exogenous covert spatial attention. Perception & Psychophysics, 59, 1-22.

Spence, C. & Driver, J. (2000). Attracting attention to the illusory location of a sound: Reflexive cross-modal orienting and ventriloquism. Neuroreport, 11, 2057-2061.

Stein, B. E., London, N., Wilkinson, L. K., & Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis. Journal of Cognitive Neuroscience, 8, 497-506.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. Cognitive Psychology, 5, 109-137.

Vroomen, J., Bertelson, P., & de Gelder, B. (1998). A visual influence in the discrimination of auditory location. Proceedings of the International Conference on Auditory-Visual Speech Processing (AVSP'98), (pp. 131-135), Terrigal-Sydney.

Vroomen, J., Bertelson, P., & de Gelder, B. (2001a). The ventriloquist effect does not depend on the direction of automatic visual attention. Perception & Psychophysics, 63, 651-659.

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

Vroomen, J., Bertelson, P., Frissen, I., & de Gelder, B. (in prep). A spatial gradient in the ventriloquism after-

effect.

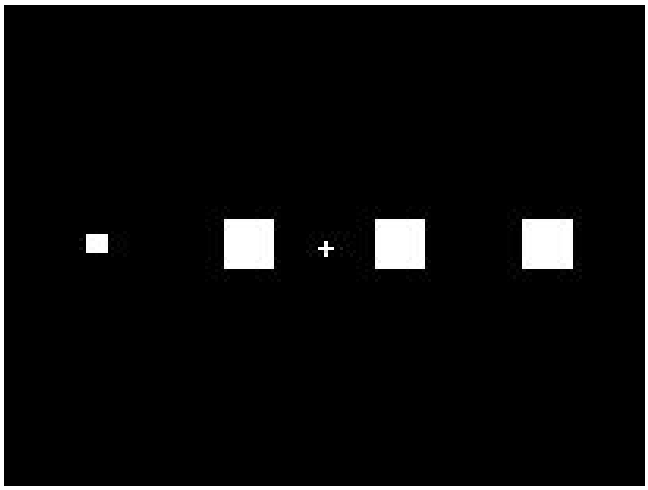
Vroomen, J., and de Gelder, B. (2000). Sound enhances visual perception: Cross-modal effects of auditory organisation on vision. Journal of Experimental Psychology: Human Perception and Performance, 26, 1583-1590.

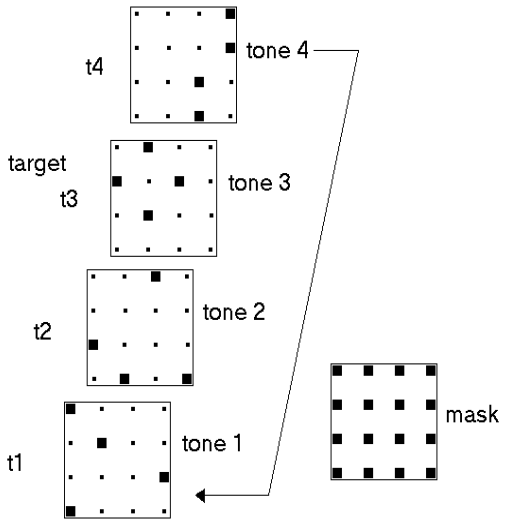
Woods, T. M., & Recanzone, G. H. Multimodal interactions evidenced by the ventriloquism effect in humans and monkeys. To appear in G. Calvert, C. Spence, & B. Stein, (Eds.), Handbook of multisensory processes. MIT Press.

Figure captions

Figure 1. An example of one of the displays used by Vroomen et al. (2001a). Subjects saw four squares with one of them (the singleton) smaller than the others. While the display was flashed on a computer screen, subjects heard a stereophonically controlled sound whose location had to be judged (left or right of the median fixation cross). The results showed that the apparent location of the sound was shifted in the direction of the two big squares, and not toward the singleton. On control trials, it was found that the singleton attracted visual attention. The direction in which a sound was ventriloquized was thus dissociated from where exogenous attention was captured.

Figure 2. A simplified representation of a stimulus sequence used in Vroomen & de Gelder (2000). Big squares represent the dots shown at time t; small squares were actually not presented to the viewers, but are only there to show the position of the dots within the 4x4 matrix. The four-dots displays were shown for 97 ms each. Not shown in the figure is that each display was immediately followed by a mask (the full matrix of 16 dots) for 97 ms, followed by a dark blank screen for 60 ms. The target display (in this example the diamond in the upper-left corner whose position had to be detected) was presented at t3. The sequence of the four-dots displays was repeated without interruption until a response was given. Tones (97 ms in duration) were synchronized with the onset of the four-dots displays. Results showed that when a tone was presented at t3 that segregated, target detection was enhanced presumably because the visibility of the target display was increased. When a segregating tone was presented at t2, target detection became worse because the visual distracter at t2 caused more interference. There was no enhancement when the tone at t3 did not segregate. The visibility of a display was thus increased when synchronized with an abrupt tone.




t4
tone 4
target
t3
tone 3
t2
tone 2
t1
tone 1
mask