Auditory-visual Spatial Interactions

Vroomen, J.; Bertelson, P.; de Gelder, B.

Published in:
Out of Mind

Publication date:
2001

In B. De Gelder, E. de Haan, C. Heywood, (Eds.), Out of Mind (pp. 140-150). Oxford: Oxford University Press, (2001).

Auditory-visual spatial interactions: Automatic versus intentional components

*Jean Vroomen, *Paul Bertelson, and Beatrice de Gelder*
Tilburg University, The Netherlands
*also Universite Libre de Bruxelles

For a long time, we have known that in the initial stages of visual processing, information is broken down and analyzed in separate and specialized parts of the brain (e.g. color, movement, orientation). However, our impression of the world is not a bundle of separate features, but we see integrated color-form objects in their appropriate place. The question, then, is how the separately coded information is integrated so that a coherent representation of the world is achieved. In some way, it seems necessary that the system keeps track of which data belong to the same object so that one can see a red bal on his left, instead of something reddish, something round, and something on the left. This is known as the feature-integration problem (Treisman & Gelade, 1980). The same problem, though, arises not only in vision, but also in other sense modalities like hearing or touch, and even more important, also across modalities. For example, when I clap my hands, the perceptual system has to know that the motor-, proprioceptive-, auditory- and visual effects of the clapping all belong to the same event of clapping hands. This problem of information integration across the senses has been variously labeled the problem of pairing (Radeau & Bertelson, 1977; Radeau; 1994), binding (Pourtois, 2000), object identity (Bedford, 1999), or the unity assumption (Welch & Warren, 1980).

One of the crucial issues is on what criteria cross-modal pairing is based. For instance, what in the previous example, are the cues for treating the auditory and visual information as belonging to the same event? In what follows we will describe our research that is relevant to this question. Most of it is concerned with a very specific case of cross-modal interaction, the so-called ventriloquist illusion. Our hope, though, is that some of the same principles of cross-modal integration may apply to other examples of cross-modal pairing.

The ventriloquist illusion is created by presenting synchronous auditory and visual information in somewhat separate locations. The most important effect is a cross-modal bias, which occurs when a subject is instructed to localize the sound while ignoring the spatially discordant input in the visual modality. Usually, the perceived location of the sound is shifted in the direction of the visual stimulus. Reactions to such 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 approximate synchrony with the speech.

As a first approximation, the fact that vision biases audition may indicate that the perceptual system has treated the auditory and visual information as a single event. The most popular explanation is that the perceptual system reduces the conflict between the location of the visual and auditory data because there is an a priori constraint that an object can have only one location (e.g. Bedford, 1999). Shifting the auditory location in the direction of the visual event would seem to be ecologically useful because spatial resolution in the visual modality is better than the auditory one. Alternatively, though, the ventriloquist effect may also be accounted for by far less interesting possibilities such as a subject who sometimes points, by mistake, to the visual stimulus instead of the sound, or a subject who adopts a post-perceptual response strategy to satisfy the demands of the particular laboratory task. Of course, these possibilities are not exclusive, and one has to find ways to check against them.

Next, one can ask what the factors are that influence the ventriloquist effect, or, more in general, that contribute to the pairing decision. A commonly made distinction is between structural and cognitive factors. Structural factors are inherent properties of the inputs. Most prominent are similarity in temporal pattern (e.g., common onset time) and spatial location (Radeau, 1994). Data arriving at the same time and from the same location are, within limits, likely to originate from a single event. Cognitive factors are those that provide semantic knowledge about the situation or instructions that might inform or deceive the perceiver about the origins of the inputs. As an example, Welch (1999) lists >familiar association= as one of the factors that contribute to the pairing decision.

Do 'cognitive factors' influence cross-modal pairing?

Many studies of ventriloquism have used fairly realistic situations, simulating real-life bimodal events such as a voice speaking and the concurrent sight of a face (e.g., Bertelson at. al., 1994; Radeau and Bertelson, 1977, Exps 2-3; Warren et al. 1981), the sight of whistling kettles (Jackson, 1952) or of beating drums (Radeau and Bertelson, 1977, Exp 1). For the time being, though, it is not clear whether such form of realism is really essential. For example, Radeau and Bertelson (1977) combined a voice with a realistic (the sight of the speaker) and a simplified visual input (light flashes in synchrony with the amplitude peaks of the speech) and found that exposure to these two situations produced comparable effects, suggesting that realism plays little if any role in ventriloquism. Admittedly though, faces and lights can hardly be compared directly and more evidence was needed. We (Bertelson, Vroomen, Wiegeraad, de Gelder, 1994) obtained more direct evidence in an experiment in which subjects heard, on each trial, an ambiguous >ama/ana= fragment from an array of seven hidden loudspeakers. At the same time, they saw on a centrally located screen a face, either upright or upside down, articulating >ama=, >ana=, or remaining still. Subjects had two tasks: they pointed to the apparent origin of the sound and reported what had been said. The orientation of the face had, as expected, an effect on >what= was perceived (i.e. the McGurk-effect, McGurk & MacDonald, 1976), but crucially, not on >where= the sound came from. Thus, even though the familiar orientation of a face is, without doubt, upright, there was no effect whatsoever of face inversion on the amount of ventriloquism. This suggests that >familiar association= does not play a role in ventriloquism, or at least that it is an ambiguous concept that needs refinement.

A somewhat related issue is the extent to which beliefs and instructions play a role in the cross-modal bias situation. One possibility is to rule out the impact of strategies by making the situation as less transparent as possible, as for instance in a psychophysical staircase procedure to be described later. If in this situation one still observes a cross-modal bias, it shows that intersensory bias occurs even when conscious strategies are unlikely to be at stake. Another way to explore this issue is to examine whether instructions can change the amount of cross-modal bias. If there is no effect of instruction, it speaks to the robustness of the effect. In one of our experiments, we trained subjects to discriminate a sequence of tones that emanated either from a central location or from alternating locations in which case the tones were emitted by two speakers located next to a computer screen (Vroomen, Bertelson, de Gelder, 1998). With the eyes closed, this was a very easy task. However, the task was much more difficult when light flashes were presented that alternated in synchrony with the tones between the left and the right on a 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 trained subjects to discriminate the two possible sound sequences and explicitly asked them to ignore the lights. In addition, corrective feedback (correct or incorrect response) was provided after each trial. The results were that the larger the separation between the lights, the more false alarms (responding >alternating sound= on a central sound) occurred, presumably because the farther apart were the lights, the farther apart was also the perceived location of the sounds. Moreover, in spite of the feedback provided on each trial, performance did not improve during the experiment. Instructions and feedback did thus not overcome the ventriloquist effect. Taken together, this led us to conclude that the effects of instructions on the ventriloquist effect are, at best, marginal.

Does attention influence cross-modal pairing?

An almost untouched issue is whether attention plays a role in the cross-modal pairing process. One could argue, as Treisman and Gelade (1980) had done for feature-integration in the visual modality, that attention is the >glue= that combines features across modalities. Could it be, then, that when a visual attractor is attended (or, in the terms of Treisman and Gelade, is focused), it is a better attractor for the apparent location of a sound than when it is not focused?

We considered the possibility that ventriloquism indeed requires such an attentional linkage between the two modalities (Bertelson, Vroomen, de Gelder, Driver, in press). The 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 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 center of the display, or the lateral square for occasional occurrences of a catch stimulus. The attentional hypothesis predicted that the attraction of the apparent location of the sound by the square would be stronger with attention focused on that square than with attention focused on the center. In fact, equal degrees of attraction were obtained in the two attention conditions. Focused attention did thus not seem to play a role.

However, maybe the effect of attention was small and overwhelmed by the bottom-up information from the laterally presented visual square. What would happen when the bottom up information would be more ambiguous? Would the 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 attractors. 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.

However, - one of us being stubborn - in discussing how attention might influence ventriloquism, 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, 1997), or one may attend to one particular location rather than another, regardless of modality. Furthermore, in the literature on spatial attention, two different means for the allocation of attention are generally distinguished. First, there is a <u>endogenous</u> process by which attention can be moved voluntarily. Second, there is an automatic or <u>exogenous</u> mechanism by which attention is reoriented automatically to stimuli in the environment with some special features. The previous study manipulated endogenous attention by asking a subject to focus either on one or the other location. However, it may have been the case that the visual stimulus received a certain amount of <u>exogenous</u> attention independent of where the subject was focusing. For that reason one might ask whether capture of exogenous attention by the visual stimulus is essential to affect the perceived location of a sound.

To investigate this possibility, we first tried to create a situation in which exogenous attention was captured in one direction, whereas the apparent location of a sound was unaffected. 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 color, 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 <u>singleton</u>, and its influence on attention 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 which would not bias the sound location 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). In a typical study on ventriloquism, subjects are asked to point to the location of the auditory stimuli and to ignore the visual distractor. A difficulty with this procedure is that the subject is often aware of the spatial discrepancy between the auditory and visual stimuli, and can adjust his response accordingly. The visual bias one obtains may then reflect postperceptual decisions rather than genuine perceptual effects. A staircase procedure, though, is much less transparent than the pointing task, and the effects are more likely to reflect genuine perceptual processes. In our study, subjects had to judge the apparent origin of a stereophonically controlled sound as left or right of a median reference point (thus, a left/right decision was required rather than pointing). 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 center. 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 apparently uncertain regarding the location of the sound. The sound location at which these response reversals occur is the dependent variable.

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. 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 was 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 attractor size effect that we previously obtained (Bertelson et al., in press), occurred with the present visual display as well. This result thus suggested that attraction of a sound is 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 whose 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, presumably because attention was attracted away from the target letter and towards the singleton. Taken together, these two experiments demonstrate a dissociation between ventriloquism and exogenous attention: the apparent location of the sounds were shifted towards the two big squares, and not towards a singleton, but 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., in press) thus support the conclusion that ventriloquism is not affected by the direction of attention.

The same conclusion is further corroborated by our work on patients with unilateral visual neglect (Bertelson, Paviani, Ladavas, Vroomen, de Gelder, in press). The neglect syndrome is usually interpreted as an attentional deficit as reflected in a reduced capacity to report visual stimuli in the contra lateral visual hemifield. We addressed whether a stimulus that goes undetected because it is presented in the bad field, nevertheless shifts the apparent location of a synchronous sound in its direction. Patients with left visual neglect consistently failed to detect a stimulus presented in their left visual field, but nevertheless, their pointing responses to the sound where shifted in the direction of the visual stimulus. This is thus another demonstration that ventriloquism is not depending on attention.

Is the spatial scene on which attention operates cross-modally reorganized?

The previous studies led us to conclude that bimodal interactions take place at a stage of processing anterior to attentional selection. 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 distractor by a moving face. This result thus implies that focused attention operates on a representation of the external scene which has already been spatially reorganized by cross-modal interactions.

The question we addressed is whether a similar cross-modal reorganization of external space occurs when exogenous rather than focused attention is at stake. To do so, we used the cross-modal cueing task introduced by Spence and Driver (1997). 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 uninformative cues 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 a visual cue does not affect responses to auditory targets (Driver & Spence, 1998). This then opens an intriguing possibility: what happens with an auditory cue whose veridical location is in the center, but whose apparent location is attracted towards a simultaneous light in the periphery. Can such a 'ventriloquized' cue attract exogenous attention? Figure 1 presents a schematic outline of the situation. Subjects were either presented a visual cue, or a ventriloquized cue. The latter consisted of a tone presented from an invisible central speaker synchronized with the 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 appeared from one of the upper or from on of the lower speakers. Our results verified, as in Spence and Driver, that visual cues had no effect whatsoever on auditory target detection. More important, with ventriloquized cues, there was no cueing effect at 100 ms SOA, but the effect slowly appeared 300 and 500 ms SOA. A ventriloquized cue thus can, unlike a visual cue, direct auditory exogenous attention to a specific location thereby suggesting that ventriloquism can reorganize the space on which exogenous attention operates.

At which processing level do cross-modal interactions take place?

Cross-modal pairing in the ventriloquist situation seems to occur at a pre-attentive stage because 1) the effect is largely independent of where attention is focused (endogenous) or attracted (exogenous), and 2) because ventriloquism can reorganize the space on which endogenous and exogenous attention operates. An intriguing issue, then, is to describe more precisely the stage at which such cross-modal interactions occur. In animal studies, neurophysiological evidence has been found indicating that cross-modal interactions can take place at very early stages of sensory processing. Probably one of the best known sites of multi-modal convergence and integration is the superior colliculus, a midbrain structure known to play a fundamental role in attentive and orientation behavior (see Stein & Meredith, 1993 for review). In humans as well, neurophysiological evidence of very early cross-modal interactions has been found. For example, we (de Gelder, & Vroomen, 2000) had found previously that identification of the emotion displayed by a face is biassed in the direction of a simultaneously presented emotional tone of a voice, or vice versa. We (de Gelder, et al., 1999) employed this phenomenon in a typical mismatch negativity task (MMN) in which subject hear repeatedly the same sound (the standard), but sometimes a deviant which evokes a typical negative electric brain response. Our standard stimulus was an angry voice combined with a congruous angry face. In the deviant condition, the voice was kept the same, but the face changed to an incongruous sad face. This change in the visual stimulus evoked the MMN, which is usually evoked only by auditory deviations, thus showing an early interaction between vision and audition in the processing of affect.

Given that such cross-modal electrophysiological effects arise early in time, it seems at least possible that intersensory interactions can occur at primitive levels of perceptual organization. Recently we have shown that the perceptual system indeed utilizes information from one sensory modality to organize the perceptual array in the other modality. We found an illusion that occurs when an abrupt sound is presented during a rapidly changing visual display (Vroomen & de Gelder, in press). Phenomenally, it looks as if the sound was pinning the visual stimulus for a short moment so that the visual display 'freezes'. In our experiments, we used as an estimate of the apparent visual persistence the performance on a speeded detection task. Subjects 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 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. Subjects in the experimental condition heard at the target display a high tone, but at the other distracting displays a low tone. In the control condition, participants heard only low tones. The results showed that the detectability of the target could be improved by the high tone synchronized with the target. Subjects were thus faster and more correct when a high tone was presented at target onset. In a second experiment we synchronized the high tone with the distractor display that preceded the target. In this case, performance got actually worse, presumably because the visibility of the distractor display was enhanced thereby increasing interference. This result excludes the possibility that a deviant tone simply acted as a warning signal giving subjects prior information about when to expect the target.

However, the most important result was that we could show that the perceptual organization of the tone sequence determined the cross-modal enhancement. Auditory stream segregation is one of the best-known examples of perceptual organization. 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 (Bregman, 1990). The principle which underlies this phenomenon is that abrupt sounds segregate more easily than less abrupt sounds. Our question was whether segregation in the auditory modality would cause segregation in the visual modality as well. Segregation of the high tone was prevented when it was part of the beginning of the tune AFrère Jacques@. In that case, visual target detection task was much more difficult than in a control condition in which the same tone was more abrupt. The abruptness of the tone thus increased the visibility of the target display thus showing that cross-modal interactions can occur at primitive levels of perceptual organization. Apparently, segregation in the auditory modality can cause segregation in the visual modality.

What is cross-modal perception good for?

In laboratory situations ventriloquism leads to *mis*-localization of the auditory source; so one may ask: why use visual information? The answer is, of course, an obvious one. In the real world, the auditory sources typically will correspond in location to matching visual events rather than being discrepant. Hence the cross-modal interactions that we have been studying should normally be adaptive, tending to favor veridical rather than illusory spatial perception. Cross-modal perception may therefore serve the ecologically useful role of attenuating the effects of variability, due to noise or drift, in the performance of modality-specific processes by cross-reference to the other modality.

It also makes sense that ventriloquism is not affected by where the subject is focusing attention. If auditory localization was affected by wherever a subject chose to attend to visually, then the apparent location of a fixed auditory source would change every time the person decided to shift his visual attention. A similar case can be made for exogenous attention. If auditory localization would change whenever a visual stimulus captured exogenous attention, it would presumably be impossible to keep an internal representation of space consistent with external reality. It thus seems better that cross-modal spatial interactions are driven by stimulus factors in a bottom-up fashion, rather than being susceptible to wherever attention is focused or captured.

More generally, it seems that cross-modal interactions are consistent with a perceptual mechanism that tries to make a coherent interpretation about auditory and visual data that are likely to originate from a single object or event. Multi-sensory stimulation that covaries in place and time is likely to originate from a single object. Perceptual evaluation in one modality may then have consequences in other modalities so that coherence is maintained. Ventriloquism and our freezing phenomenon are just a few demonstrations of this principle.

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. Müsseler (Eds.), Cognitive Contributions to the Perception of Spatial and Temporal Events (pp 347-362). Amsterdam: Elseviers.

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

Bertelson, P., Pavani, F., Ladavas, E., Vroomen, J., & de Gelder, B. (in press). Ventriloquism in patients with unilateral visual neglect. *Neuropsychologia.*

Bertelson, P. & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. Perception & Psychophysics, 29, 578-587.

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

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

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

de Gelder, B., Böcker, K. B. E., Tuomainen, J., Hensen, M., & Vroomen, J. (1999). The combined perception of emotion from face and voice: early interaction revealed by human electric brain responses. Neuroscience Letters, 260, 133-136.

de Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. Cognition and Emotion, 14.

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

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

Giard, M. H., & Peronnet, F. (in press). Auditory-visual integration during multi-modal object recognition in humans: a behavioural and electrophysiological study. Journal of Cognitive Neuroscience.

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

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

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.

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

Stein, B. E., & Meredith, M. A. (1993). The merging of the senses. Cambridge, MA: The MIT Press.

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

Vroomen, J. (1999). Ventriloquism and the nature of the unity assumption. In G. Aschersleben, T. Bachmann, and J. Müsseler (Eds.) Cognitive contributions to the perception of spatial and temporal events (pp. 388-394). North-Holland: Elsevier

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. (submitted). Visual bias of auditory location and the role of exogenous automatic attention.

Vroomen, J., & de Gelder, B. (2000). Cross-modal integration: a good fit is no criterion. Trends in Cognitive Sciences, 4, 37-38.

Vroomen, J., and de Gelder, B. (in press). Sound enhances visual perception: Cross-modal effects of auditory organization on vision. Journal of Experimental Psychology: Human Perception and Performance.

Warren, D. H., Welch, R. B., & McCarthy, T. J. (1981). The role of visual-auditory Acompellingness@ in the ventriloquism effect: Implications for transitivity among the spatial senses. Perception and Psychophysics, 9, 557-564.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. Psychological Bulletin, 88, 638-667.

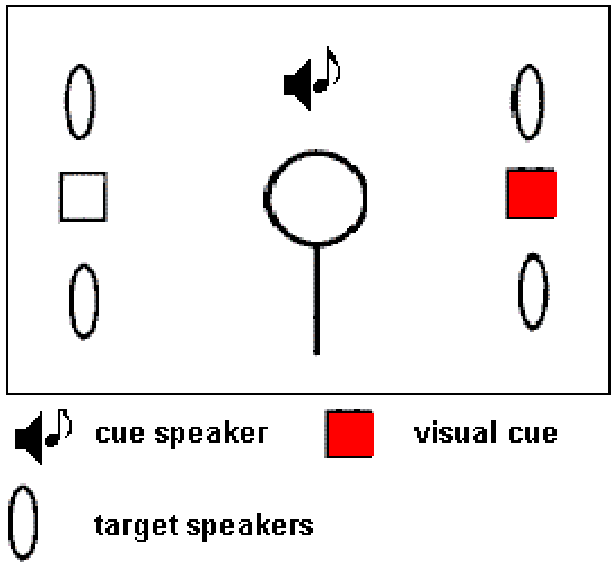

Figure 1. A schematic view of the position of the target loudspeakers, the cue lights and the (invisible) cue speaker as seen from behind the subject=s head. A tone (1000 Hz) from the cue speaker was synchronized with a cue light. Depending on SOA (100, 300, 500 ms), a target sound (white noise bursts) was emanated from one of the four target speakers.

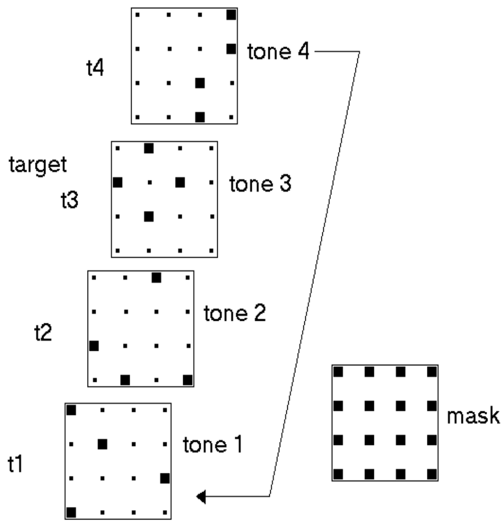

Figure 2. A simplified representation of a stimulus sequence. Big squares represent the dots shown at time t; small squares were actually not seen, but are only there to show the position of the dots within the matrix. The 4-dots displays were shown for 97 ms each. Not shown in the figure is that each 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) was presented at t3. The sequence of the four 4-dots displays was repeated without interruption until a response was given. Tones (97 ms in duration) were synchronized with the onset of the 4-dots displays.