Auditory-Visual Speech Processing
(AVSP'98)
Terrigal - Sydney, Australia
December 4-6, 1998

A Visual Influence in the Discrimination of Auditory Location

Jean Vroomen, Paul Bertelson, and Beatrice de Gelder

Tilburg University, Dept of Psychology
Tilburg, The Netherlands

ABSTRACT

The compellingness of the interaction between vision and auditory localization (the ventriloquist effect) was investigated using a discrimination task. A tone sequence was presented either from the same location or from two locations that alternated along the horizontal plane. In synchrony with the tones, lights were presented either at the same or at alternating locations. Subjects had to decide whether the tones alternated or not, thereby ignoring the lights. The main result was that presenting non-alternating tones together with alternating lights increased the number of 'alternating'-judgements, even though subjects were told to ignore the lights, and even though they received corrective feed-back after each trial.

1. INTRODUCTION

In normal environments, our auditory and visual systems are dealing with complex mixtures of sounds and visual inputs derived from multiple sources that are simultaneously present. There have to be processes that parse this mixture so that features are assigned to the correct object. For auditory input, Bregman [1] has investigated how this might be accomplished via the streaming phenomenon. For the visual domain, the same has been done ever since Gestalt psychology. However, knowledge about the way in which the system assigns cross-modal auditory and visual information to objects has received much less attention.

A prominent example where audio-visual scene analysis has to play a role is what has been called the ventriloquist effect. The basic phenomenon is that the apparent origin of a sound is attracted towards a simultaneously presented displaced light [e.g., 2, 3, 4]. However, data on ventriloquism are only relevant for understanding audio-visual interactions to the extent that one can distinguish automatic perceptual processes from more controlled post-perceptual ones. This is of course a general problem in perceptual studies. Measures of cross-modal biases of vision on the localization of sound are, like any other perceptual response, susceptible to cognitive strategies. For example, in the unimodal pointing situation, subjects are asked to point towards a sound and ignore the light. One possibility is that subjects adjust their response strategy so that they respond to the light, and not the sound. In a similar way, [5] argued that the ventriloquist effect results essentially from a response bias. They applied signal detection analysis to the effect of timing on the performance of a same/different localization task, and found differences in beta only, but not in d-prime.

However, there are reasons to question whether d-prime and beta are pure measures of perceptual and response processes, respectively. In principle, changes in beta may reflect changes in a bias at a post-perceptual stage, but also at a perceptual stage in which the threshold for certain feature detectors is changed [e.g. 6]. Changes in beta are thus neutral as to whether the relevant effect originates at a perceptual or at a post-perceptual stage.

In the present study we present a new task in which subject are asked to focus on the veridical location of the auditory stimulus while ignoring the visual attractor. On each trial, subjects heard a sequence of alternating high and low tones. The tones either came from the center, or switched constantly between two loudspeakers, one to the left, and one to the right of the subject. Subjects were asked to decide, without any time constraint, whether the tones came from the same location or whether the tones alternated between two locations. In the experimental conditions, light flashes appeared on a screen in synchrony with the tones. The critical condition is were the sounds came form the center, but the lights were apart. A genuine ventriloquist effect should manifest itself in an increase in the number of ‘alternating’ responses when the distance between the lights increases, because the light flashes should capture the apparent location of the sounds. Thus, the farther the lights apart, the farther apart their perceived location. For trials with alternating sounds, there was no specific prediction for the effect of the lights. However, such trials had to be included to make the task a real discrimination one.

This task differs from the ones in previous studies by two aspects. First, on each trial, sounds were heard repeatedly until a response was given. The stimulus in the present experiment was thus not a usual short sound train of bursts whose location had to be remembered. This aspect eliminates the danger that the obtained ventriloquist effect originated in the *memory* for the location of the sound, rather than in perception per se. Secondly, subjects received corrective feedback after each trial. They were thus told what the veridical locations of the sounds was after a response was given. If the ventriloquist effect originates in a voluntary response strategy, one would expect subjects to adjust their response criterion so that a fast improvement can be observed. One the other hand, if we are dealing with a genuine perceptual phenomenon, one expects learning to be much slower or even absent.

2. METHOD

Participants. Eleven students from Tilburg university received course credits for their participation. None reported any hearing or seeing problem.

Experimental situation. Participants were tested individually in a sound shielded and dimly lit room. They sat at a distance of 70 cm in front of a 15 inch PC screen with on each side a loudspeaker. The centers of the loudspeakers were 45 cm apart (35 degrees). On each trial, a subject heard a continuous sequence of a low (1000 Hz) and a high tone (1259 Hz). The sequence lasted for 4 s, or until a response was given. The duration of the tones was 100 ms and the ISI between the tones was 60 ms. The tones were faded-in and faded-out to avoid clicks. Within a particular sequence, the tones either alternated between the left and right loudspeakers or they occurred constantly and simultaneously from both loudspeakers which gave the impression that they came from a central location. In the alternating case, the high tone came from the left and the low tone from the right on half of the trials, and the opposite arrangement was used in the other half of the trials. The alternating and non-alternating tones were presented at the same loudness level (62 dBa when measured at ear position).

In synchrony with the onset and the offset of the tones, two white squares (1.5 x 1.5 cm) were presented on the dark background of a PC screen. One square was synchronized with the high tone, the other with the low tone. The horizontal distance between the middle of the squares was 23 cm (18 degrees) in the ‘far’ condition, 2 cm (1.6 degrees) in the ‘close’ condition, and 0 cm in the ‘same’ condition (in which case the square flickered on the same location). On those trials on which both tones and squares alternated, the left tone was always synchronized with the left square, and the right tone with the right square. As a control condition, no lights were presented and the screen thus remained black.

The combination of the alternating/non-alternating sounds with the four possible displays (far, close, same, and control) resulted in eight conditions. Each condition consisted of 24 trials, so the whole session lasted 192 trials, divided into two blocks of 96 trials each The trials were presented in pseudo-random order, preceded by a practice session of 16 trials.

Subjects were asked to press one of two keys on a keyboard to indicate whether the sounds came from the same location or alternated between the loudspeakers. They were asked to ignore the lights, but they were not allowed to close the eyes. The instructions emphasized accuracy and subjects received corrective feedback (a green ‘correct’ or a red ‘wrong’) after each trial. Each trial was started by the subject pressing a special key.

Table 1.
Proportion of 'Alternating' Responses for Each Sound Sequence and Visual Display

Location of sounds	Distance between Visual Attractors			
	Far	Close	Same	No
Non-alternating	.31	.18	.12	.10
Alternating	.94	.89	.84	.83

3. RESULTS

Table 1 shows the mean proportions of 'Alternating' responses for the eight stimulus conditions. As explained in the introduction, the main prediction concerned the effect of visual inputs in trials with non-alternating sounds. These results were submitted to a one-way ANOVA. There was a main effect of the visual input, $F(3,30) = 12.14$, $p < .001$. As can be seen in the table, the smallest proportion of 'alternating' responses occurs in the no square condition. In the trials with squares, the number of 'alternating' responses increases with increasing separation between the squares. For the trials with alternating sounds, a similar ANOVA also showed a significant effect of the visual input, $F(3,30) = 4.45$, $p < .05$.

In order to determine whether subjects had learned to discriminate the non-alternating sound sequences during the experiment, we split the trials into the first and the second half of the experiment and ran another analysis with first/second half as an additional factor. In the 2 (half) x 4 (vision) ANOVA, neither the main effect of Half, nor the interaction with Vision was significant (all $\underline{p}$'s > .15). There were thus no practice effects during the experiment.

4. DISCUSSION

In the present experiment, subjects had to judge whether the locations of sounds alternated or not in the presence of light flashes. The main results was that on those trials in which the location of the sounds did not change, illusionary alternations were produced when the lights alternated. The bigger the distance between the lights, the more often the sounds were judged to alternate. The most likely explanation for this result is that the perceived location of the sound was attracted towards the simultaneously presented visual attractor.

It is worth noting that discrimination performance is far above chance and that subjects thus can distinguish alternating from non-alternating sound sequences, even in the presence of alternating light flashes. This implies that the visual bias is not complete. Instead of visual dominance, it is better to speak of a shift by the visual attractor. The shift is not complete, though, and there remains a perceptual difference between alternating and non-alternating tone sequences on which subjects can rely.

An interesting aspect of the data is that there was no improvement during testing. Despite the fact that subject received corrective feed-back on each trial, discrimination performance did not improve. This suggests that voluntary response strategies are not important for the cross-modal effects as observed in the present experiment. The results therefore concurs with other evidence that cross-modal bias effect are largely automatic [2, 7].

5. REFERENCES

1. Bregman, A. S. *Auditory scene analysis: The perceptual organization of soun*, London: MIT press, 1990.

2. Bertelson, P., and Radeau, M. "Cross-modal bias and perceptual fusion with auditory-visual spatial discordance," P & P, Vol. 19, 531-535, 1981.

3. Radeau, M., and Bertelson, P. "Auditory-visual interaction and the timing of inputs: Thomas (1941)

revisited," *Psych. Res., Vol. 49*, 17-22, 1987.

4. Radeau, M. "Cognitive impenetrability in auditory-visual interaction," In J. Alegria and D. Holender, (Ed.) *Analytic approaches to human cognition,* Amsterdam: North-Holland, pp. 41-55, 1992..

5. Choe, C. S., Welch, R. B., Gilford, R. M., and Juola, J. F. "The 'ventriloquist effect': Visual dominance or response bias?," *P & P, Vol. 18, 55-60,* 1975..

6. Bertelson, P., and Radeau, M. "Ventriloquism, sensory interaction, and response bias: Remarks on the paper by Choe, Welch, Gilford and Juola," *P & P, Vol. 19, 531-535*, 1976.

7. Bertelson, P., and Aschersleben, G. "Automatic visual bias of auditory location" *Psych. Bull. and Rev., in press.*