Tilburg University

Exploring the relation between McGurk interference and ventriloquism

Bertelson, P.; Vroomen, J.H.M.; Wiegeraad, G.; de Gelder, B.L.M.F.

Published in:
Proceedings of the Third International Congress on Spoken Language Processing, Yokohama, Japan, September 18-22, 1994

Publication date:
1994

EXPLORING THE RELATION BETWEEN MCGURK INTERFERENCE AND VENTRILOQUISM

[1]Paul Bertelson, [2]Jean Vroomen, [2]Geert Wiegeraad, and [1,2]Béatrice de Gelder

[1]Université Libre de Bruxelles, Belgium
[2]Tilburg University, The Netherlands

ABSTRACT

An experiment is reported in which, on each trial, an ambiguous fragment of auditory speech was delivered on one of an array of hidden loudspeakers, and a face, either upright or upside-down, was shown on a centrally located screen, and articulated one of two different utterances or remained still. Subjects pointed to the apparent origin of the speech sounds and reported what had been said. Identification responses were strongly influenced by the nature of the seen movements (McGurk effect [1]) and localisation of laterally heard items was shifted toward the centrally located video monitor on trials with a moving face compared to those with a still face (ventriloquism [2,3]). Degree of McGurk interference was practically independent of spatial separation between auditory and visual source. Face inversion had no effect on ventriloquism, but it reduced the McGurk effect. This experimental dissociation suggests that the two phenomena originate in different components of the cognitive architecture.

I. INTRODUCTION

Two main lines of research have dealt with perceptual integration of visual and auditory data. One concerned the perception of location. When visual and auditory events occur in close temporal contiguity but in somewhat different locations, mutual attraction between data in the two modalities, manifest in single modality selective localisation (cross-modal bias) and impression of common origin, has typically been observed [2 to 4]. The attraction generally persists in the form of after-effects [5]. The term 'ventriloquism' has come to be used to designate these various phenomena, which are believed to form the basis of the illusion created by performing ventriloquists.

The other line dealt with speech recognition, in which information from the sight of the speaker's articulating movements has been shown to contribute, besides acoustic data, to identification performance. One striking finding was that the perception of a spoken utterance can in some cases be modified by simultaneous presentation of a face pronouncing a different utterance [1,6].

From the fact that integration of auditory and visual data occurs in both groups of phenomena, it does not follow that they originate in the same components of the cognitive architecture. The issue can be investigated by looking for experimental dissociations.

A situation was developed in which both identity conflict - conducive to McGurk interference - and spatial conflict - conducive to spatial bias - can be manipulated with the same material. It involved delivery of an auditory utterance on one of several loudspeakers laid out in an array in front of the subject and presentation, on a centrally located video screen, of a face articulating a different utterance. On each trial, the subject performed two tasks. He indicated the identity of the presented utterance and the location of the auditory source. Control trials, on which the face remained still, allowed the measurement of conflict-free identification and localisation performance.

In the present experiment, the main independent variable was the orientation of the face, which was shown either upright or upside-down. Earlier results suggested that, beyond temporal patterning, the identity of the visual input is of little importance for the occurrence of ventriloquism. For instance, the apparent location of a voice was affected by exposure to simultaneous but spatially discordant visual data to the same extent, whether the visual data were the face of the talking speaker or light flashes synchronised with the intensity peaks of the speech [7]. These considerations led us to expect no effect of face orientation on ventriloquism. It seemed likely, on the other hand, that McGurk interference would be affected by face inversion.

II. METHOD

Apparatus: The main component of the situation was a rectangular enclosure designed to house seven loudspeakers and an array of 108 response buttons. The buttons were laid out along an horizontal arc, so that the subject could reach them easily with the right hand through the open proximal side of the enclosure. The loudspeakers were positioned at regular intervals immediately behind the keys. They spanned a total angle from the subject's head of about 75 deg. A 14 x 11 cm video screen rested in central position on top of the enclosure.

Subjects: Twenty-four right-handed university students were tested individually for one session lasting approximately 60 min.

Material: A single auditory utterance was presented on each trial. It was an ambiguous utterance, intermediate between /ana/ and /ama/. The auditory utterance was always accompanied by the appearance on the screen of the face of a male speaker, who on experimental trials articulated either ANA or AMA, and on control trials remained still. On catch trials, which were interspersed among the other trials, a white patch appeared on the mouth. The successive appearances of the face were recorded on a video tape, one appearance every 8 sec. The auditory utterance had been dubbed on the same tape. The tape for the present experiment contained 14 practice trials, followed, in random order, by 210 experimental trials - 70 with visual ANA, 70 with visual AMA - and 70 control trials (with the face still). In half of the experimental trials the face orientation was up, in the other half the face was upside down.

Procedure: The subject was instructed on each trial first to press the key immediately in front of the perceived source of the voice, and then to identify orally "what has been said". On catch trials, he had to give a specific detection response ("patch"). This was meant to insure that subjects kept looking at the screen.

III. RESULTS

Localisation: Spatial bias for each seen utterance (ANA/AMA) and each loudspeaker was measured by subtracting the mean response (measured in response buttons units, 1 being the left-most button and 108 the right-most one) on trials with the face immobile from the mean for the trials with the considered utterance. These differences are given in Fig. 1 separately for AMA and ANA trials, in the upright and in the upside-down condition. All four curves have positive values for loudspeakers 1 to 3, and negative values for loudspeakers 5 to 7. Thus, localisation of laterally presented auditory inputs is attracted toward the moving face. As overall measure of localisation bias, the mean bias for loudspeakers 5-7 was subtracted from the mean for loudspeakers 1-3. The mean values of that score for each visual input and each orientation are given in Table 1. By MANOVA, the effect of visual input is significant [$F(1,23) = 9.33$, $p = .006$]; spatial bias is stronger for the AMA than for the ANA input. The effect of orientation is non-significant [$F < 1$] as well as the visual input x orientation interaction [$F < 1$].

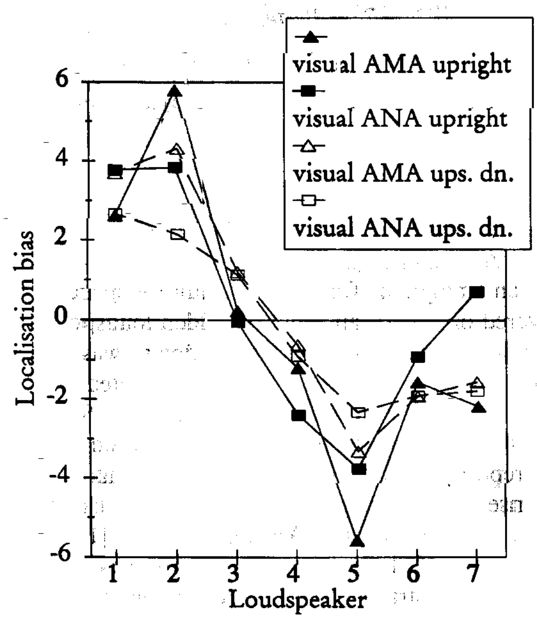

Figure 1. Localisation bias as a function of loudspeaker location, for each seen utterance and each orientation of the face. Bias = mean response button number for talking face minus button number for still face. Positive = to the right.

TABLE 1
Localisation bias in number of response buttons (SD in parentheses)

Visual input	Face orientation Up	Down
ANA	1.92 (3.03)	2.00 (2.15)
AMA	3.00 (2.97)	2.68 (2.30)

Identification: Subjects reported perceiving either ANA, AMA or AMNA. As appears in Table 2, with the speaker's face immobile, the auditory stimulus was identified as ANA on about 64% of the trials, and as AMNA on most of the remaining trials. AMA occurred very rarely. The proportion of ANA responses dropped to floor levels with the face saying AMA, and raised to high levels when the speaker's face said ANA (93% in the upright and 82% in the upside-down condition).

TABLE 2
Proportions of identification responses

Visual input	Orien-tation	Response		
		ANA	AMA	AMNA
Still	Up	.60	.08	.32
	Down	.66	.06	.28
ANA	Up	.93	.02	.06
	Down	.82	.03	.16
AMA	Up	.04	.15	.80
	Down	.10	.13	.77

Identification bias was measured by subtracting from the proportion of bilabial responses (AMNA + AMA) on each talking face condition the proportion obtained on the corresponding still-face condition. The means of these values appear in Table 3. By MANOVA, the effects of visual input [$F(1,23) = 41.0$, $p<.001$) and of orientation [$F(1,23) = 25.0$, $p<.001$) are highly significant and their interaction [$F(1,23) = 4.1$, $p=.055$) is marginally so. By *t*-test, the effect of orientation is non-significant for visual AMA [$t(23) < 1$] and highly significant [$t(23) = 3.74$, $p < .001$] for visual ANA. In the latter condition, bias is smaller with the face upside-down.

TABLE 3
Identification bias (Proportion of bilabial responses for each seen utterance subtracted from still-face control condition), SD in parentheses

Visual input	Up	Down
ANA	-.31 (.19)	-.14 (.16)
AMA	.56 (.22)	.56 (.14)

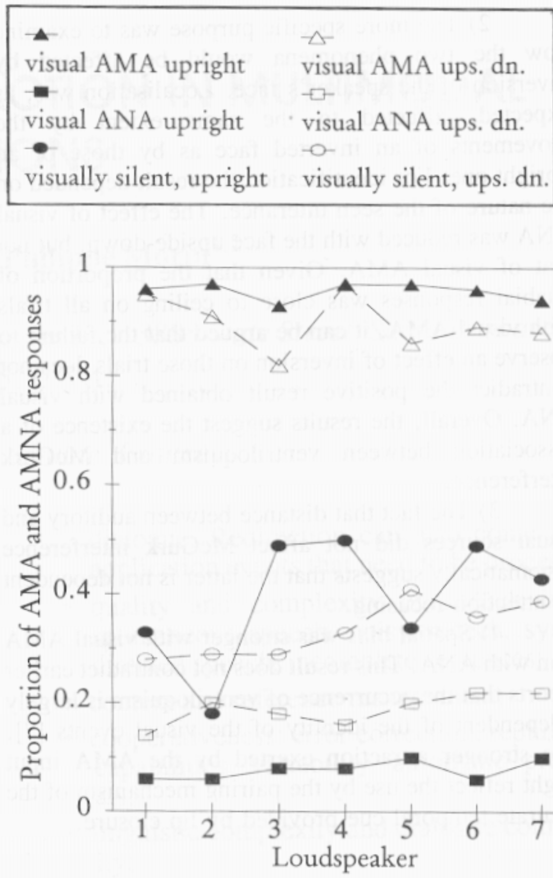

Figure 2. Proportion of bilabial identification responses (AMA + AMNA) as a function of loudspeaker location for each seen utterance and each orientation of the face.

Effect of location on identification: A further MANOVA, with location in addition to visual input and face orientation as within-subjects factor, revealed a significant overall effect of location [$F(6,138) = 2.80$, $p < .02$] on identification bias, and a significant visual input by location interaction [$F(6,138) = 4.44$, $p < .001$). In fig.2, which shows proportion of identification responses in the different types of trials as functions of loudspeaker location, no systematic pattern emerges. Whatever their origin, the complex effects of location are not a matter of distance between visual and auditory sources. Paradoxically, it is in the control still-face conditions that the strongest location effects are observed.

IV. DISCUSSION

1) The most general outcome of the study concerns the feasibility of the approach: ventriloquism and McGurk interference were observed together in the same situation.

2) The more specific purpose was to examine how the two phenomena would be affected by inversion of the speaker's face. Localisation was, as expected, affected to the same extent by the movements of an inverted face as by those of an upright one. For identification the result depended on the nature of the seen utterance. The effect of visual ANA was reduced with the face upside-down, but not that of visual AMA. Given that the proportion of bilabial responses was close to ceiling on all trials with visual AMA, it can be argued that the failure to observe an effect of inversion on those trials does not contradict the positive result obtained with visual ANA. Overall, the results suggest the existence of a dissociation between ventriloquism and McGurk interference.

3) The fact that distance between auditory and visual sources did not affect McGurk interference systematically suggests that the latter is not dependent on attention focusing.

4) Spatial bias was stronger with visual AMA than with ANA. This result does not contradict earlier reports that the occurrence of ventriloquism is largely independent of the identity of the visual events [7]. The stronger attraction exerted by the AMA input might reflect the use by the pairing mechanism of the accurate temporal cue provided by lip closure.

ACKNOWLEDGMENTS

The present work was partly supported by the Ministry of Scientific research of the French-speaking Community of Belgium (Concerted Research Action 91/96 - 148). The assistance of the Technical Department of Tilburg University is gratefully acknowledged.

REFERENCES

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

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

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

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

[5]: Radeau, M. & Bertelson, P. (1974). The aftereffects of ventriloquism. *The Quarterly Journal of Experimental Psychology, 26*, 63-71.

[6]: Massaro, D.M. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, N.J.: Lawrence Erlbaum Ass.

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