Title
Multisensory enhancement of localization with synergetic visual-auditory cues

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 26(26)

ISSN
1069-7977

Authors
Godfroy, Martin
Roumes, Corinne

Publication Date
2004

Peer reviewed

Multisensory enhancement of localization with synergetic visual-auditory cues

Martine Godfroy (mgodfroy@imassa.fr),
Cognitive Science Department, Institut de Médecine Aérospatiale du Service de Santé des Armées,
B.P.73, 91223 Brétigny sur Orge Cedex, France

Corinne Roumes (croumes@imassa.fr)
Cognitive Science Department, Institut de Médecine Aérospatiale du Service de Santé des Armées,
B.P.73, 91223 Brétigny sur Orge Cedex, France

Abstract

Enhanced behavioral performance mediated by multisensory stimuli has been shown using a variety of measures, including response times, orientation behavior and even simple stimulus detection. In the particular case of the study of saccadic response to unimodal or bimodal stimuli, Corneil et al. (2002) were able to show that the bimodal visual-auditory saccades benefited from the accuracy of visual saccades at saccadic response time (SRTs) typical of auditory saccades. However, there has been little evidence of multisensory mediated improvement in stimulus localization. Recently, Hairston et al. (2003) shows improvement in visual-auditory localization performance (variability) for induced myopia while no benefit was reported for normal vision. Using a similar experimental design, taking into account two space dimensions, azimuth vs. elevation, we examined the ability of human subjects to localize visual, auditory and combined visual-auditory targets for stimuli considered optimal for the given task. The results showed significant improvement in bimodal localization when compared with the more accurate modality, visual, as measured with multi criterion data (precision, dispersion and orientation of the response patterns). Furthermore, the 2D analysis of combined visual-auditory target localization performance, for azimuth and elevation response components, underlines the role of the auditory system in the determination of the response characteristics. The data suggested that visual-auditory localization performance benefited from the "*best of the two worlds*" (Corneil et al., 2002), in that it was improved only in the horizontal plane, and restricted to the response criterion where audition is more reliable than vision.

Introduction

The literature dealing with intersensory perception first dealt with the phenomena of sensory illusions, the most well known being the ventriloquism effect (Howard and Templeton, 1966) and the McGurk effect (McGurk and McDonald, 1976). Both these "on-line" effects result from discrepancies, either spatial and/or temporal between the two unimodal components of the stimulation. The much more ecological situation, in which visual and auditory signals are synergetic, i.e. in terms of spatial and temporal congruence, has been rarely investigated systematically in a localization task. Furthermore, to our knowledge, taking into consideration the two dimensions (azimuth and elevation) of the observer's perceptive field for a multimodal localization task was never explored. In addition, the simultaneous presentation of spatially congruent visual and auditory cues was mostly studied considering detection of a target (Frasinetti et al., 2002), orientation toward a target (Stein et al., 1988, 1989) or reduction in response latencies (Hugues et al., 1994; Frens et al., 1995, Colonius & Arndt, 2001) rather than purely localization capability. When shown, increase in precision of the localization was restricted to the analysis of an angular value, expressing the stimulus-response discrepancy in polar coordinates. The purpose of this experiment was to evaluate multisensory integration in a two-dimensional localization task and qualify the nature of a cross modal benefit that could be obtained when the spatial information in the two modalities was convergent. We suggested a separate analysis of the localization performance for the azimuth and elevation components of the response, as a function of target double pole coordinate system in which the origin coincides with the center of the head. This procedure should reveal the contribution of the auditory modality into the bimodal localization performance, given the initial differences in coding the position of an auditory target in azimuth (Interaural Time and Level differences) and in elevation (monaural spectral shape cues). Indeed, as a consequence of this specific coding, auditory resolution differs in the horizontal and the vertical dimension while the visual resolution, associated to a retinotopic coding, is isotropic in space. The investigation of criterion we assumed to be relevant for the task was performed. Centering, precision, dispersion and orientation of the responses were successively examined to determine a potential benefit and the modal contribution of a bimodal visual-auditory target presentation.

Materials and methods

Participants

Ten adults, aged 22 to 50 years, took part in the experiment. They all had a minimum of 20/20 visual acuity (if need be, corrected). Their audiometric capacities were also normal, with age related variations. All were naïve regarding the setup configuration (number and positions of the auditory sources).

Experimental setup

The participant sat in darkness in the center of an acoustically transparent semi-cylindrical vertical screen, 120 cm in radius and 145 cm high, with the head maintained by a chin-rest, as shown in Figure 1. A Liquid Crystal Display Philips Hopper SV10 video-projector was hung above and behind the observer, 245 cm from the screen, providing a 80° horizontal x 60° vertical green light field of view of 1.5 $cd.m^{-2}$ average luminance (Fig. 1). The color green (coordinates of the 1931 CIE system x = 0.267; y = 0.640) was used for the background and for the visual stimulus, and made it possible to obtain a maximum signal to noise contrast and maximum background homogeneity, given the characteristics of the optic device. A PC (Pentium III 300 MHz) equipped with a 128 SoundBlaster sound card and a Matrox G400 (32MB) video card generated the stimuli. It was connected to the video-projector on the one hand, and to the loudspeakers via an audio switch and its Velleman K8000 control module, on the other hand. Thirty five 10-cm-diameter loudspeakers (Fostex FE103 Sigma) were laid out behind the screen in a 7 x 5 matrix, with a 10° step. The speaker positions were defined in a two-dimensional polar coordinate system with the origin at the straight-ahead fixation position. Eccentricity in the perceptive field was referred in relation to this coordinate system. The speakers were positioned at azimuths 0°, ±10°, ±20°, ±30° and elevations 0°, ±10°, ±20° (Figure 2).

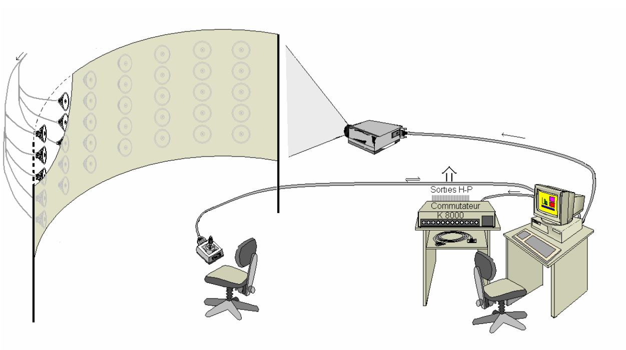


Figure 1: The experimental setup.

Visual stimuli consisted of a spot of light (100ms, 20 $cd.m^{-2}$), subtending a 1° of visual angle and auditory stimuli consisted of a pink noise burst (broadband noise, constant intensity per octave), 100ms duration (20ms fade-in and fade-out), at 49dB as measured at the subject's ear or hearing position, against a 38dB background noise (precision integrating sound level meter Brüel and Kjaer Model 2230). The device allows the precise superimposition of the visual and auditory stimulation for a combined presentation to the target, where the spot of light is exactly located at the center of the loudspeaker's cone. To perform localization judgments, participants used a track-ball, allowing for movements along all directions. Figure 2 describes the succession of the events in trial.

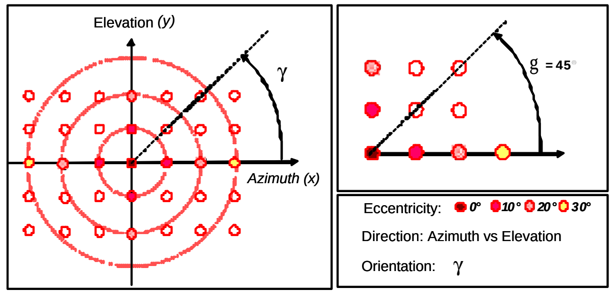


Figure 2: Definition of the independent variables used in the analyses and characterizing the target position. *Eccentricity* refers to the distance of the target from the center of the 2D perceptive field, *Direction* allow transforming target and response *Orientation* (?) in a two components position (azimuth and elevation).

1. At the beginning of each trial, a fixation cross was presented at the center of the screen, at (0°, 0°) coordinates, for 500 to 1500 ms for acquisition.
2. At the extinction of the cross, the visual, auditory or bimodal visual- auditory stimulus was presented randomly at one of the 35 positions during 100ms. The picture illustrates a - 20° to 0° visual stimuli.

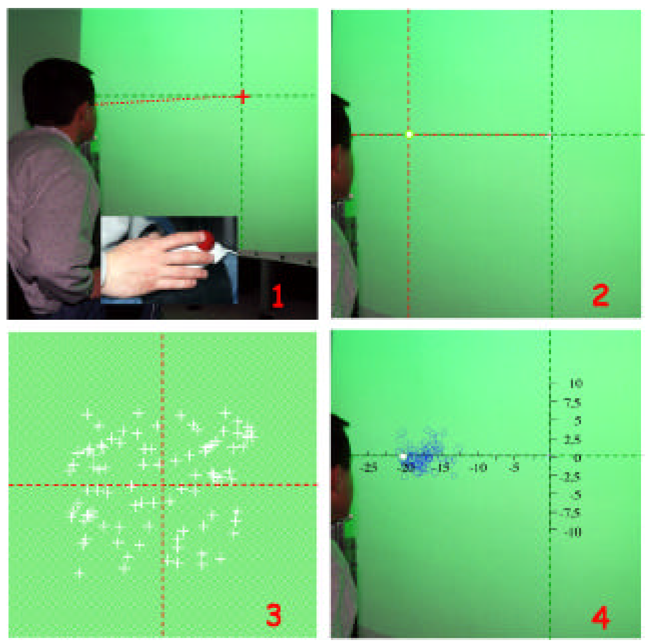


Figure 3: The experimental paradigm. 1. Presentation of a fixation cross at the center of the screen. 2. A visual stimulus at (-20°, 0° coordinates). 3. All the possible cursor position for the -20°, 0° target position. 4. Each dot stem from an individual localization response.

3. After the target disappears, a response cursor, associated to the further manipulation of the track-ball, appears randomly inside a 20° imaginary circle whose center is the position of the target with a minimum of 2.5° distance from it in both axes (azimuth and elevation).

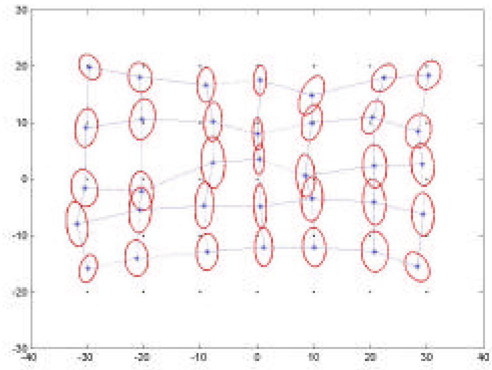
Subjects were instructed to localize the target as accurately as possible while pointing this cursor towards the perceived location of the target, the temporal constraint being secondary. The picture 4 illustrates the response distribution of the 10 subjects and 10 repetitions for the given location of the visual stimulation. The experiment consisted of 6 experimental sessions with 10 repetitions of each stimulus combination (3 stimulus conditions [Visual, auditory, bimodal] at 35 locations [7 azimuth values, 5 elevation values] presented in pseudo-random order) for a total of 175 trials per session, with a 1.5s inter-trial interval.
Prior to testing, 20 practice trials were performed to make the participant familiar with the task and the manipulation of the track-ball. The session lasted about 30 min. and a minimum 24-hour delay was observed between two sessions.

Data analysis

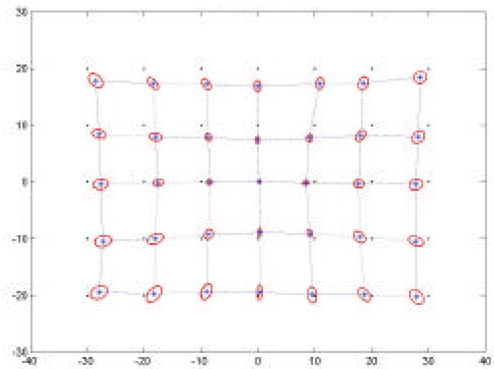
Localization errors were calculated as the difference, in degrees, between the localization judgment and the actual target location. Taking into consideration the azimuth and elevation components of the response, *centering* and *precision* of the responses were calculated from the raw data. *Centering* refers to the mean response, which the sign denotes a tendency to overshoot (positive values associated to errors eccentric to the target in reference to the reference coordinates) or undershoot (negative values associated to errors central to the target). *Precision* evaluate the amount of discrepancy (absolute value) from target to designation. The distribution of the response patterns were computed using a procedure of regression analysis for obtaining the regression slope that determines the major *orientation* of the response distribution. Estimation of the maximum and minimum variance of the distribution along the slope axis and the perpendicular one, respectively noted *b* and *a*, were used for *dispersion* analysis. By extension, in reference with Hofman et Van Opstal (Hofman et Van Opstal, 1998), a characterization of the response patterns under a geometrical approximation, i.e. ellipses, did allow a better comparison within and between modalities than the traditional methods using a two-dimensional discontinuous space analysis (Oldfield et Parker, 1984). In this way, the analysis of dispersion and orientation of the patterns would provide complementary data to those obtained with the use of the horizontal and vertical axis of the 2D coordinate system. To analyze the data, multiple 2-way within subjects ANOVAs were performed according to the specific hypothesis: Statistical comparisons were structured to examine the main effect of target modality (visual, auditory, combined visual-auditory) and target location (eccentricity range [0°, 10°, 20° and 30°] and direction [azimuth versus elevation]) as well as the possible interaction between the variables.

Results

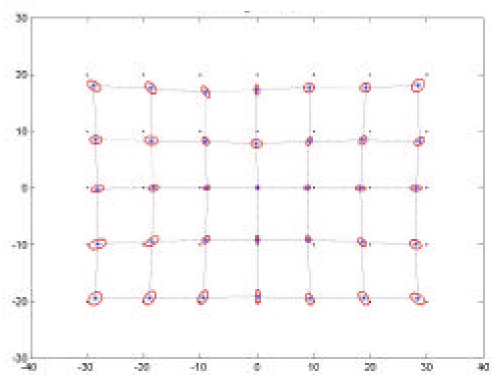
The results only consider here the comparison of response localization between modalities while a preliminary work



Auditory condition



Visual condition



Combined visual-auditory condition

Figure 4: Responses patterns as approximated by ellipses for the 3 conditions and the 35 target positions.

was performed on unimodal data to ensure the validity of the results. We shall now successively describe the data using the four variables mentioned in section Data Analysis.

Precision of the responses

A short look at the approximated data for each condition of presentation of the target for the 35 positions tested (Figure 4) underlines the specificity of the auditory system in terms of localization capability and the relative similar localization behavior between the visual and the bimodal conditions.

Table 1: Precision of localization between conditions

Modality	Azimuth		Elevation		A/E	$F_{1,3486}$	*P*
	Error	SD	Error	SD			
Auditory	2,96°	±2,6°	6,15°	±4,94°	A<E	1159	<0,0001
Visual	1,87°	±1,7°	1,86°	±1,63°	A=E	0,119	0,7304
Bimodal	1,53°	±1,5°	1,62°	±1,49°	A<E	7,7	0,0055

A more detailed analysis of mean of errors confirmed this first impression. A repeated measures ANOVA showed that the effect of modality condition is significant in azimuth ($F_{2,336}$=87.2; p<.0001) and in elevation ($F_{2,336}$=23.316 p<.0001). The much more interesting result concerned the significant improvement in bimodal localization compared to the visual one in azimuth, (Scheffe test, p=0.0302) but interestingly, not in elevation (Scheffe test, p=0.8355). When looking at the within-modality variations between error in azimuth and error in elevation, expressed by the Azimuth/Elevation precision relationship (A/E in table 1), it appears that the gain obtained in the bimodal condition follows the difference in precision of the auditory condition (with statistically significant values). This result is an argument for audition playing a structuring role in intersensory processing for a spatial task.

Centering of the responses

One of the most well known characteristics of the auditory system is concerned with the differences in accuracy between azimuth and elevation, in relation to the differences in the initial information extraction process in the two directions of space (Oldfield & Parker, 1986; Hofman & Van Opstal, 1998). As a consequence, there is a strong response bias in the elevation responses, with a central compression of the auditory space related to a systematic undershoot of target eccentricity in this direction. No observable or statistical improvement in centering was obtained between the visual and the combined audio-visual conditions. On the other hand, the localization of an auditory target in azimuth is much less biased by eccentricity than for the visual and bimodal conditions, as illustrated in Figure 5 In this direction, the reduction of error is at the maximum when the direction of the visual and the auditory biases are in opposition of signs. When the sign of the bias is identical, no visible effect is observed. A statistical comparison between the visual and bimodal results fails to show any improvement, probably due to the arithmetic mean performed on data expressed in polar coordinates. Despite the lack of significance, the results did again suggest that the contribution of the auditory modality did enhance performance.

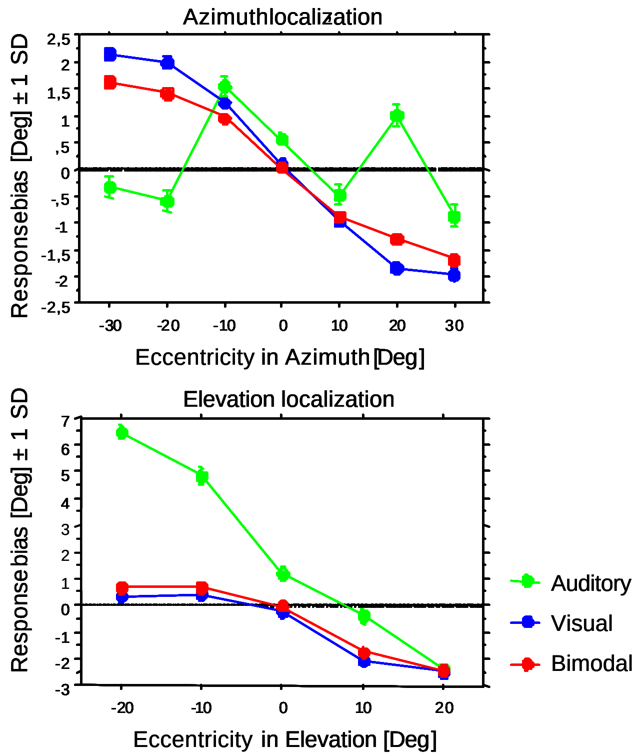


Figure 5: Centering of the responses for azimuth and elevation components of the localization responses. Improvement in performance is only visible in the horizontal plane.

Dispersion of the responses

The diverse responses are compared on the two characteristic axis of the responses patterns, *a* and *b* (Cf. Data analysis).

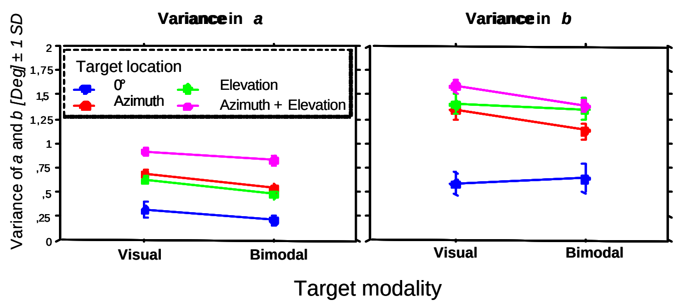


Figure 6: left: Decrease in variance in a between the visual and bimodal condition for all target locations. Right: Decrease in variance in b between the visual and bimodal condition for the targets that didn't belong to the median sagittal plane (0° and Elevation).

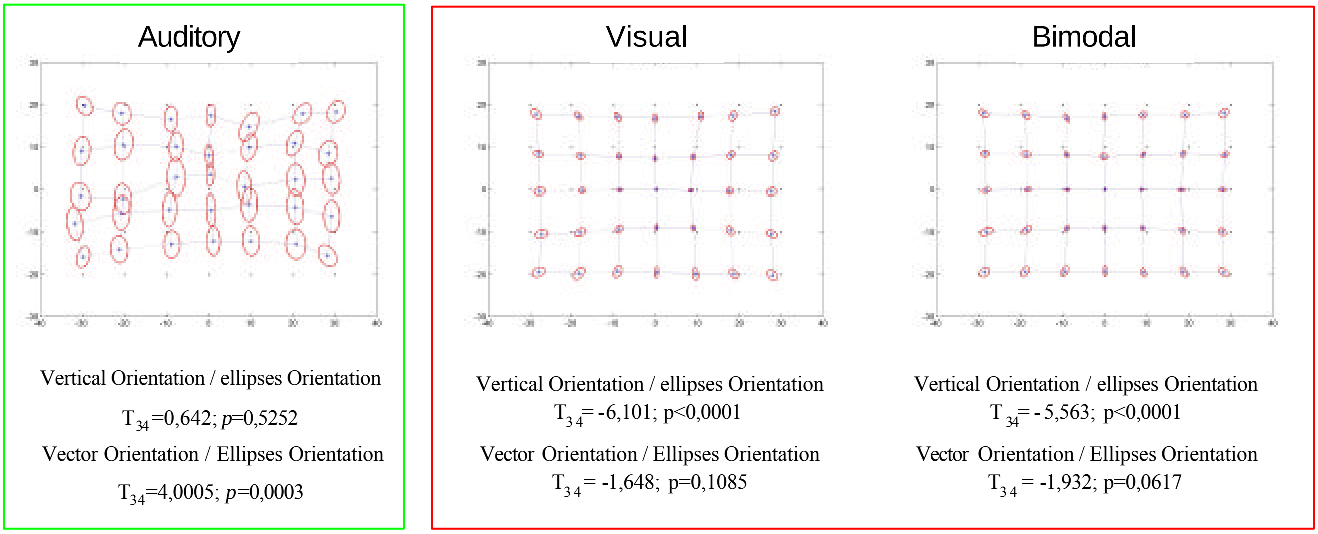


Figure 8: Orientation of the responses in relation with target location in the 2D perceptive field. In the auditory condition, the response patterns are vertically oriented (90°-180° axis) while visual and bimodal response patterns exhibit a vector distribution with the ellipses oriented centrifugally (toward the center of the perceptive field).

The minimum variance axis, *a*, diverges significantly according to the modality condition for target presentation (repeated measures ANOVA: ($F_{2,338}$=43.055 p<.0001), with the comparison of visual and bimodal conditions being also significant ($F_{1,169}$=23.356 p<.0001). Similar results are obtained for the *b* axis, with a slightly different behavior in relation to target location, expressed by the belonging or not to a specific plane (0°, Azimuth, Elevation, or combined eccentricity in Azimuth and Elevation). Indeed, only the targets that are not located on the median sagittal plane (0° and Elevation only) did benefit of a significant variance reduction (Figure 6).

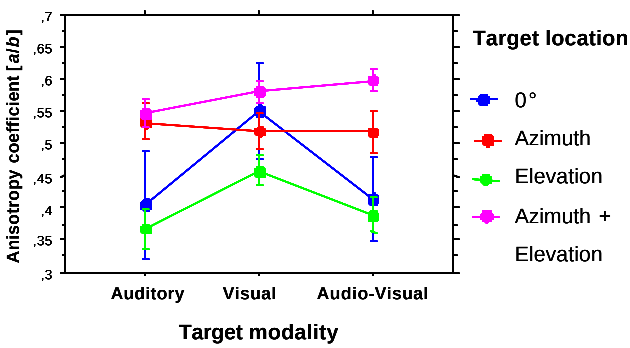


Figure 7: Anisotropy coefficient variations according to target modality and location in the 2D perceptive field. Note that the coefficient varies in the same way for auditory and bimodal conditions.

We also calculated an anisotropy coefficient, corresponding to the *a*/*b* ratio (a "1" value corresponding to an homogeneous distribution along the two axis), and looked at the variations of this coefficient according to the target location in space. It can be seen in Figure 7 that the value of the coefficient follows the same variations in the auditory and the combined visual-auditory condition. Once again, these data pointed out the role of the auditory modality into the multimodal spatial perception, not only in performance improvement, but also in representation structuring.

Orientation of the responses

At this point, we shall remember that the orientation of the responses distributions are determined by the slope of the regression analysis computed for the 35 tested target positions and the 3 modalities. In each condition, the calculated orientation is compared to two models of sensory coding: an auditory coding using a Cartesian coordinates system on one hand, and a vector coding, which can reflect a saccadic component in the response, on the other hand. The data shown in Figure 8 allow mentioning that auditory response patterns are vertically oriented while visual and bimodal response patterns exhibit a vector distribution with the ellipses oriented centrifugally. These observations could reflect a possible different role of the saccadic system according to the target modality and the presence vs. absence of visual information in the perceptive field.

Discussion

This study investigated the localization performance to visual, auditory, and bimodal stimuli distributed throughout the 2D perceptive field. The result of the current study illustrates a significant multisensory enhancement of localization performance in precision and dispersion. Through a quantitative approach, the data allowed to parameterize the different dimensions which describe the perceptive field of an ideal observer and to attest to the relative contribution of each sensory modality into the bimodal perception. The results argue for an integrative process applying for synergetic presentations of visual and auditory stimuli, and cues considered as well suited for the given task. For all that, our result did not refute the very ecological principle of the "inverse effectiveness rule" (Stein & Meredith, 1993). They just underline the structuring role of the auditory system only when it is more reliable than the visual system, what can be shown only by the comparison in performance for the two directional components (azimuth and elevation) of the response. It is a strong argument to say that sensory integration in a localization (spatial) task rests on a tendency to optimization. Looking at the data obtained by Corneil et al. (2002), showing that bimodal visual-auditory saccades were at least as accurate as visual saccades, but also generated at saccadic response times (SRTs) shorter typical of auditory saccades, our result also go in the way of a very similar neural process applying. This tendency to optimize shall be considered as an economic and ecological process that drove the Central Nervous System (CNS) to use the sensory systems in relation with the specific contribution they can have. In the case of a localization task (spatial task), and given the reliability of each sensory system, we demonstrated an improvement in centering and a part correction of the variance attributed to audition, an increase in precision and possibly in structure of representation for vision.

Acknowledgments

We would like to acknowledge the assistance of A. Bichot, L. Pellieux, J. Plantier and P.M.B. Sandor in the technical phases of the work. Special thanks to G. Cooper for his daily encouragement. Financial support was provided by IMASSA.

References

Colonius, H. & Arndt, P. (2001). "A two-stage model for visual-auditory interaction in saccadic latencies." *Perception & Psychophysics, 63*: 126-147.

Corneil, B.D., van Wanrooij, M., Munoz, D.P., van Opstal, A.J. (2002). "Auditory-visual interactions subserving goal-directed saccades in a complex scene." *Journal of Neurophysiology, 88,* 438-454.

Frasinetti, F. et al. (2002). "Enhancement of visual perception by crossmodal visuo-auditory interactions." *Experimental Brain Research, 147,* 332-343.

Frens, M.A.; van Opstal, A.J. & van der Willigen, R.F. (1995). "Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements." *Perception & Psychophysics, 57(6),* 802-816.

Hairston, D.W. et al. (2003). "Multisensory enhancement of localization under conditions of induced myopia." *Experimental Brain Research , 152,* 404-408.

Hofman, P.M. & Van Opstal, A.J. (1998). "Spectro-temporal factors in two-dimensional human sound localization." *Journal of Acoustical Society of America 103,* 2634-2648.

Howard, I.P. & Templeton, W.B. (1966). *Human spatial orientation*. New York, NY: Wiley.

Hugues, H.C.; Reuter-Lorenz, P.A.; Nozawa, G. & Fendrich, R. (1994). "Visual-auditory interactions in sensorimotor processing: saccades versus manual responses." *Journal of Experimental Psychology: Human Perception and Performance 20(1),* 131-153.

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

Oldfield, S.R. & Parker, S.P. (1984). "Acuity of sound localisation: a topography of auditory space. I. Normal hearing conditions." *Perception, 13(5),* 581-600.

Stein, B.E.; Huneycutt, W.S. & Meredith, M.A. (1988). "Neurons and behaviour: the same rules of multisensory integration apply." *Brain Research, 448:* 355-358.

Stein, B.E.; Meredith, M.A.; Huneycutt, W.S. & McDade, L. (1989). "Behavioural indices of multisensory integration: orientation to visual cues is affected by auditory stimuli." *Journal of Cognitive Neuroscience 1(1),* 12-24.

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