# Spatial and temporal factors in visual-auditory interaction

Heike Heuermann & Hans Colonius
Universität Oldenburg, D-26129 Oldenburg
email: hans.colonius@uni-oldenburg.de

**Abstract**

*In two experiments, saccadic eye movements were investigated in a bimodal focused attention task with visual targets and auditory distractors under various spatial and temporal conditions. Different spatial effects on saccadic reaction time could be observed depending on whether the auditory distractor preceded or followed the visual target. While vertical interstimulus distance affected reaction time only when the auditory signal was presented first, effects of horizontal displacement seemed not to depend on the temporal interstimulus relationship. In an auditory localization experiment, eye movements to the perceived sound source were measured. In contrast to visually guided saccades, trajectories of eye movements evoked by auditory stimuli were frequently curved. Further analysis showed that vertical eye movements often start with a delay and are corrected once or twice. Both results indicate that different mechanisms for azimuth and elevation cue analysis exist in the auditory system. These mechanisms are also involved in auditory-oculomotor and visual-auditory sensory integration.*

Latencies to target stimuli are usually significantly smaller if an additional (non-informative) accessory stimulus is presented in close temporal and spatial relationship with the target. Various psychophysical and physiological studies have suggested different explanations for this intersensory facilitation effect (IFE), for example, attentional or warning effects or multisensory information integration.

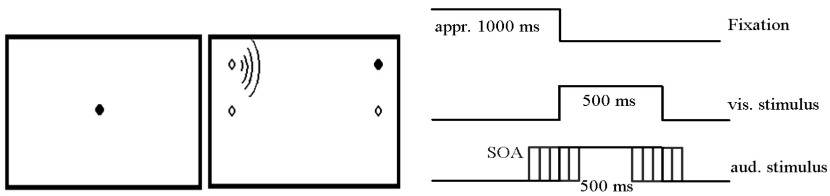
In humans, interaction between the visual and the auditory system is of special importance, with vision usually dominating perception while audition seems most important for the detection of warning signals. Temporal aspects of various visual-auditory interaction effects, particularly with respect to reaction times, have been investigated since the early 60ies. Several models reaching from simple statistical facilitation to attentional effects have been considered to explain the findings (for an early review see Nickerson 1973). More recently, quantitative analyses of the effect of spatial interstimulus relations on saccadic reaction time (SRT) have been performed (Frens, Van Opstal, and Van der Willigen 1995, Harrington and Peck 1998, Hughes, Nelson, and Aronchick 1998). A general observation is that the extent of intersensory facilitation increases with spatial proximity. Using pure tones or noise signals, Frens et al. (1995) showed that the perceived stimulus position has a significant influence on visual-auditory interaction. They suggested a linear relation between radial interstimulus distance and the amount of facilitation. Colonius and Arndt (2001) proposed a two-stage model describing both temporal and spatial aspects in visual-auditory interaction. In their study, saccadic reaction times toward visual targets decreased the more the auditory accessory preceded target presentation and the smaller the spatial distance between both stimuli was. This paper extends their study by introducing spatial stimulus configurations with different horizontal and vertical eccentricity. The goal of this study is to reveal those aspects of visual and auditory information integration that are involved in processing both azimuth and elevation cues.

A central role in visual-auditory interaction has been assigned to the Deep Layers of the Superior Colliculus (DLSC) (cf. Meredith and Stein 1996). The SC is a brain stem nucleus participating in integrative mechanisms in the visual and visuo-motor system and is of substantial importance for reflexive movements in response to a stimulus. Moreover, it has also been found to be a prominent stage in intersensory integration. Afferents from different modalities converge here, building spatial saliency maps which are in close register with each other. It should be noted here that the SC has so far been the only mammalian brain structure showing *topographically* organized auditory maps at all. It remains unclear how this map is constructed. Unlike the retinotopic maps of the visual and oculomotor system, an internal representation of the auditory environment is based upon the calculation of interaural intensity- and phase-differences and on the analysis of direction-specific spectral cues resulting in a craniocentric reference system.

Interaural time- (or phase-) and intensity difference analysis can be assigned to the EE-ans EI-cells of the Superior Olivary Complex (SO), sending their efferents to the Inferior Colliculus which in turn projects to the SC. Hence, binaural information processing already takes place in subcortical areas, which means that its processing can be assumed to be more "hardwired" and faster. Unfortunately, the details of how auditory elevation judgment is performed and which neural mechanisms exactly are involved are not known yet. It seems obvious that the direction-dependent spectral modifications of the signal caused by the listener's pinna folds (Head Related Transfer Functions, HRTF), represent the substantial cue for localization in the elevation domain. Hence, we deal with a spectral pattern recognition problem of which physiological data indicate that it seems to be performed by a different neural pathway involving thalamic and cortical areas. This idea is supported by a behavioral study of Hofman, Van Riswick, and Van Opstal (1998) showing that participants were able to learn to adequately use a new set of HRTFs (corresponding to a pair of new ears) without losing the capability of correctly localizing with their "genuine ears". Hofman compared this effect with learning a new languages. In another psychophysical study, Frens and Van Opstal (1995) found that auditory saccades are often curved, in contrast to visually evoked eye-movements. Auditorily guided trajectories frequently show a strong horizontal trend at first which is supplied by an "elevation correction movement" after a period of about 30 msec. This, too, indicates certain temporal constraints in elevation determination (in contrast to azimuth estimation) and suggests separate mechanisms in the analysis of binaural and monaural location cues. If this holds true, temporal and spatial parameters in visual-auditory interaction should be seen as independent factors, but it can be expected that the amount of specific spatial interaction depend on the SOA actually chosen.

## Experimental Setup

Five paid volunteers (one female, four male, aged from 22 to 43 years) took part in the experiment. They were seated in a dark, sound proof chamber. Visual stimuli were white dots presented on a 37" monitor. Sound stimuli were white noise signals presented via headphones, using a virtual auditory environment. Stimulus duration was 500 msec, possible stimulus positions were ± 25 deg horizontal and 0 or 20 deg vertical eccentricity for both visual and auditory signals.




(a) (b)

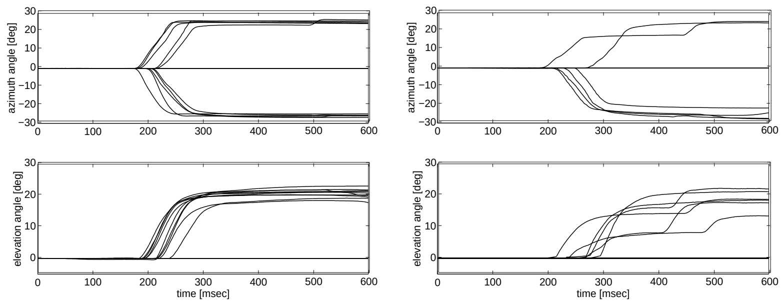
Figure 1: ***a:*** *Illustration of one bimodal trial. Visual and auditory stimuli could be presented from either of the four positions indicated by open circles.* ***b:*** *Chronological order of presentation.*

Participants were instructed to gaze at a central fixation point which disappeared as soon as a target from any of the four positions was presented (step condition). They should then perform a saccade toward the visual target as fast and as accurate as possible, while any auditory accessory signals had to be ignored (focused attention paradigm). In 80% of the cases, the visual stimulus was accompanied by an auditory accessory. In these bimodal trials, stimuli were presented at stimulus onset asynchronies (SOAs) from -60 msec to 40 msec (in 20 msec steps), with negative SOA values assigning sound preceding visual targets. Spatial configuration was also varied across all possible combinations. For each visual target position, this leads to 1 unimodal case and 4 bimodal combinations, the latter presented at 6 distinct SOAs each. Hence, with four distinct target positions we get 100 different stimulus conditions. Saccadic reaction times were measured 20 times for each condition with all conditions presented in randomized order across several days. Additionally, each subject participated in a (unimodal) auditory localization experiment in which eye movements to the perceived origin of an *auditory* target had to be performed. Stimulus conditions were the same as described above. The goal of this experiment was, apart from an examination of localization performance, a qualitative analysis of auditory eye movements compared to visually guided saccades. Stimuli were presented 80 times per position in randomized order across several days. Eye movements were recorded during the trials with an infrared eye movement registering system and stored in a computer. Saccadic responses were measured off-line afterwards using a velocity criterion. All calculations needed for the analysis of our data were performed with MATLAB.

## Results

### Eye Movements: Position Time Traces

A qualitative analysis of both types of saccades (Figure 2) reveals considerable differences. In auditory saccades, elevation movement often starts somewhat later and is corrected once or twice. Note that corrective eye movements are almost exclusively performed with respect to elevation. By contrast, visually guided saccades do not show this pattern, neither under unimodal, nor under bimodal visually guided conditions. This effect has already been described (e.g. Zambarbieri, Schmid, Magenes, and Prablanc 1982) and is often explained by a "multiple look strategy" (Hofman and Van Opstal 1998) due to successive vertical cue analysis. This study demonstrates that these eye movement patterns can be replicated well in a virtual auditory environment.

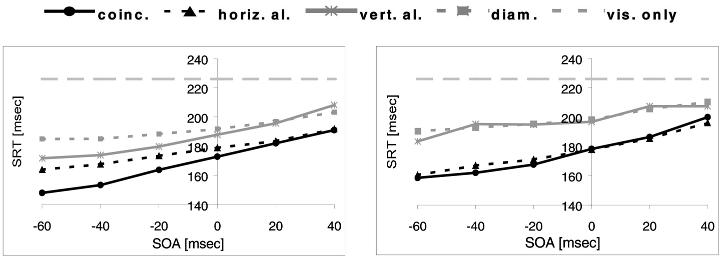



(a) Visual Saccades (b) Auditory Saccades

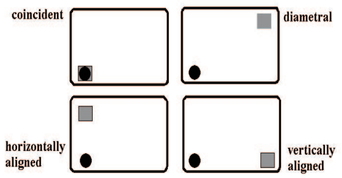
Figure 2: *Position time traces for visual (**a**) and auditory (**b**) saccades. Top panels: horizontal movement, bottom panels: vertical movement.*

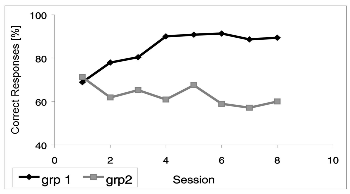
### Saccadic Reaction Times

As displayed in Figure 3, three main observations can be made with respect to bimodal saccadic reaction time. First, saccadic latencies are significantly shorter under bimodal stimulation than under unimodal visual stimulation for all participants. Second, the amount of this intersensory facilitation effect (IFE, defined as the difference between unimodal and bimodal latency) decreases monotonically with SOA. This holds for all spatial stimulus configurations used here. Third, IFE also decreases with increasing spatial distance, although this effect is somewhat more complex. Regarding spatial effects, two groups of participants can be discerned. In the first group (panel a), an effect of spatial distance can be found within both the horizontal and the vertical dimension: effects of vertical stimulus eccentricity are the more pronounced the more the auditory stimulus precedes target onset. With increasing SOA the SRTs in the spatial conditions "coincident" and "horizontally aligned" on the one hand, and "vertically aligned" and "diametral" on




(a) Bimodal SRT, group 1 (b) Bimodal SRT, group 2




(c) Stimulus configuration (d) Localization performance

Figure 3: ***a and b:*** *Saccadic latencies under different spatial (graphs) and temporal (x-axis) conditions for two groups of participants as described in text.* ***c:*** *Possible stimulus configurations are plotted for one target position. All four target positions were used.* ***d:*** *Mean localization accuracy (in % correct) of the participants in each group across sessions.*

the other, converge or even intersect. Two-way ANOVAs revealed highly significant main effects of both SOA and spatial stimulus configuration and an interaction. Subsequent Newman-Keuls tests ($\alpha = .05$) found an effect of vertical eccentricity only for negative SOAs, while for positive SOAs the conditions "coincident" and "horizontally aligned" form a homogeneous subgroup as do "vertically aligned" and "diametral". Hence, if the auditory stimulus follows the visual target onset, only azimuthal distance seems to play a role, while for negative SOAs both horizontal and vertical distance components are taken into account. In the other group of participants (panel b), only an effect of horizontal interstimulus distance was significant. This, however, is consistent with the finding that these participants are poor auditory localizers (panel d).

## References


Colonius, H., Arndt P. (2001). A two-stage model for visual-auditory interaction in saccadic latencies. *Perception & Psychophysics* **63** (1): 126-147

Frens, M.A., Van Opstal, A.J. (1995). A quantitative study of auditory-evoked saccadic eye movement in two dimensions. *Experimental Brain Research* **107**: 103-117

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

Harrington, L.K., Peck, C.K. (1998). Spatial disparity affects visual-auditory interactions in human sensorymotor processing. *Experimental Brain Research***122**: 247-252

Hofman, P.M., Van Opstal, A.J. (1998). Spectro-temporal factors in two-dimensional human sound localization. *Journal of the Acoustical Society of America* **103** (5): 2634-2648

Hofman, P.M., Van Riswick, J.G.A., Van Opstal, A.J. (1998). Relearning sound localization with new ears. *Nature Neuroscience* **1** (5): 417-421

Hughes, H.C., Nelson, M.D., Aronchick, D.M. (1998). Spatial characteristic of visual-auditory summation in human saccades. *Vision Research* **38**: 3955-3963

Nickerson, R.S. (1973). Intersensory Facilitation of Reaction Time: Energy Summation or Preparation Enhancement. *Psychological Review* **80** (6): 489-509

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

Zambarbieri, D., Schmid, R., Magenes, G., Prablanc, C. (1982). Saccadic Responses Evoked by Presentation of Visual and Auditory Targets. *Experimental Brain Research* **47**: 417-427