On auditory-visual interaction in real and virtual environments

Bernhard U. Seeber[+] *and Hugo Fastl*

AG Technische Akustik, MMK, TU München, Arcisstr. 21, 80333 München, Germany
[+] Now: Auditory Percept. Lab, UC Berkeley, Berkeley, CA 94720-1650, bernhard_seeber@gmx.de

Abstract

The ventriloquism-effect describes an attracting influence of visual objects on perceived auditory directions. In the ventriloquist example the spectators will perceive the sound as coming from the mouth of the puppet as it's lips make synchronous movements to the speech, although the ventriloquist speaks.

When simulating directions in auditory displays the question comes up how the auditory directional shift towards the visual object is affected by a reduction of auditory directional information, i.e. the omission of individualized auditory directional cues. Two opposite outcomings could occur: (1) the directional shift increases as the visual modality gains more weight against the less accurate and reliable, "weaker" auditory modality, or (2) the directional shift is reduced through a reduction of cognitive congruence between both objects.

Auditory-visual interaction was therefore investigated in three different listening environments: (1) real, anechoic space, (2) virtual acoustics using individual head-related transfer functions (HRTFs), or (3) selected non-individual HRTFs. The subjects task was to fixate visual objects while listening to auditory targets. Localization responses were collected as trial-by-trial aftereffects.

A new localization method utilizing a laser-pointer was developed which allows for a fast and accurate collection of localization responses. By using a trackball as an input device the interference of proprioceptive information could be decoupled from the auditory-visual interaction experiment.

The study showed auditory directional shifts of up to 7° towards visual objects in the real and in the individualized virtual environment. Directional shifts were statistically similar for both environments. Using selected non-individual HRTFs smaller shifts were observed. As experimental conditions were similar in all environments except for the directional presentation the results suggest that auditory-visual interaction is dependent on the presentation of auditory directional cues.

1. Introduction

The introduction of spatial auditory reproduction systems in everyday live, for example in home cinema systems or in user interfaces in a multimedia context, brings up questions about the interaction between the auditory and the visual modality. The ventriloquism-effect is a well known effect of auditory-visual directional interaction: although the ventriloquist speaks, the spectators will perceive the voice as coming from the mouth of the puppet. The ventriloquism-effect describes the perceptual fusion of the auditory and the visual object as well as the pure shift of the auditory direction towards the visual object. When auditory directions are reproduced in virtual environments the question comes up how the ventriloquism-effect is affected by a reduction of auditory directional information, i.e. the omission of individualized auditory directional cues. Two opposite tendencies could show up: (1) the directional shift increases as the visual modality gains more weight against the less accurate and reliable, "weaker" auditory modality, or (2) the directional shift is reduced through a reduction of cognitive congruence, i.e. compellingness of unity, between both objects.

To investigate this question a localization method was developed which uses a laser-pointer to allow for a fast and accurate collection of directional responses. By using a trackball as an input device the experiments can be laid out bimodally – the interference of proprioceptive information on the auditory-visual interaction experiment can thus be reduced [1, 2]. The interaction with the laser pointer spot can be minimized by an open-loop measurement and a random trial-by-trial variation of the initial position of the laser spot in the vicinity of the sound source position. Therefore the localization method is called: *Pro De Po* – *Pro*prioception *De*coupled *Po*inter. As experimental factors can be precisely controlled and methodical bias effects are reduced this laser-pointer method provides a new approach to localization and interaction studies.

The available auditory directional information was varied as follows: a study in real space provided natural cues, whereas using individual head-related transfer functions (HRTFs) and selected non-individual HRTFs the availability of natural directional cues was reduced. The usage of selected non-individual HRTFs nevertheless ensured for an individually optimized directional reproduction.

2. Methods

2.1. Apparatus and Localization Method

In the real environment, i.e. the anechoic chamber, the target sounds were played back through speakers mounted at the directions $-50°$ left to $+50°$ right in a $10°$-spacing at ear level and at a distance of 1.95 m. LEDs were placed concentric in front of the speakers to serve as fixation targets. The LEDs and speakers were covered by an acoustically transparent, but opaque curtain [1]. A laser-pointer method was used to gather the localization results. The laser-spot was projected with deflection mirrors on the curtain. Subjects adjusted the movable laser spot on the direction of the sound using a trackball. They confirmed their input by pressing one of the three buttons at the trackball, which coded the perceived sound position as "externalized in front", "inside the head", and "in the rear" [1].

2.2. Interaction Study

Wide-band noise pulses (125 Hz – 20 kHz, 5 pulses of 30 ms duration, 70 ms pause) served as target sounds in all experiments. The temporal sequence of a single trial is depicted in figure 1. In each trial a fixation-LED at a randomly chosen direction from $-40°$, $0°$, $+40°$ lit up in the completely darkened anechoic chamber. 1 s later the target sound was played at a randomly chosen position from $-50°$, $-30°$, $-10°$, $+10°$, $+30°$, $+50°$, or from the direction of fixation. The fixation LED went off 250 ms after the sound was played. Further 250 ms later the laser-pointer spot appeared at a random horizontal position within $\pm 20°$ of the previous direction of the sound. The subjects adjusted the laser spot to the perceived direction of the sound and confirmed their input by pressing a button at the trackball. 21 trials were taken for each combination of the 7 sound directions and 3 fixation positions in 3 sessions. 9 subjects, age 24 – 28 years, participated in the experiments.

2.3. Baseline Study

A baseline study without visual fixation was conducted with the subjects of the interaction study. The baseline study was similar to the interaction study, but visual targets (LED's) were not displayed. It consisted of 20 trials for each direction $-50°$, $-40°$, ..., $+50°$, which were taken in two sessions. The results of the baseline study are reported in [2]. The current study shows results of the interaction experiments, from which the median results of the respective baseline study were subtracted.

2.4. Modification of Auditory Directional Cues

To investigate the ventriloquism effect as a function of individual adaptation of auditory cues, sounds were presented in real and virtual space. Sounds in real space transmit the maximum amount of individual information for localization. Localization cues presented with individual HRTFs closely mimic natural conditions whereas a reduction of individual localization cues can be expected using non-individual HRTFs. By an individual selection of HRTFs from non-individual HRTFs the externalization of virtual sound sources was facilitated [3]. The studies using virtual acoustics were similar to the ones in real space apart from sounds being played through an electrostatic headphone at virtual positions. The baseline study and the interaction study were conducted for all three environments.

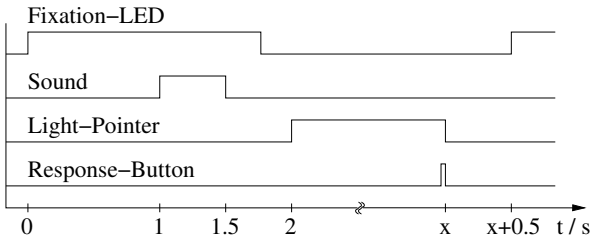

Figure 1: Schematic of a single trial's temporal sequence in the interaction experiments.

3. Results and Discussion

Localization results from the virtual environments with individual and selected non-individual HRTFs are given in figure 2. All results are regarded relative to the baseline study without visual targets. The results obtained in the real environment are not statistically different from the results in the virtual environment with individual HRTFs (at 5%, α-corrected Mann-Whitney-Wilcoxon U-test for 21 single tests). The localization results from the individualized virtual environment (fig. 2, left) show the attractive influence of the visual target on auditory directions. As an example, with individual HRTFs a sound source at $+10°$ is perceived shifted by $7°$ towards the visual target at $+40°(\square)$. The maximum absolute shift is observed for auditory-visual discrepancies of $30 - 50°$, i.e. a lateral fixation at $\pm 40°$ and sounds coming from -10 to $+10°$. The shifts are symmetrical for a fixation of targets at $+40°$ right or $-40°$ left. For a frontal fixation target auditory directions appear to be less shifted $(3°)$.

The localization results obtained with non-individual HRTFs are different (fig. 2, right): if, for example, the results for a $+40°$ Fixation $(\square)$ for sounds from $+40°$ are compared against the results for sounds from $-50°$ to $-10°$, no relative shift can be observed. This becomes more apparent if the results for fixation at $\pm 40°$ are combined, as in figure 3. Although the maximum average shift is obtained at an auditory-visual discrepancy of $30°$ in all environments, the effect is smaller with non-individual HRTFs and has already disappeared for discrepancies of $50°$. Whereas similar directional shifts are

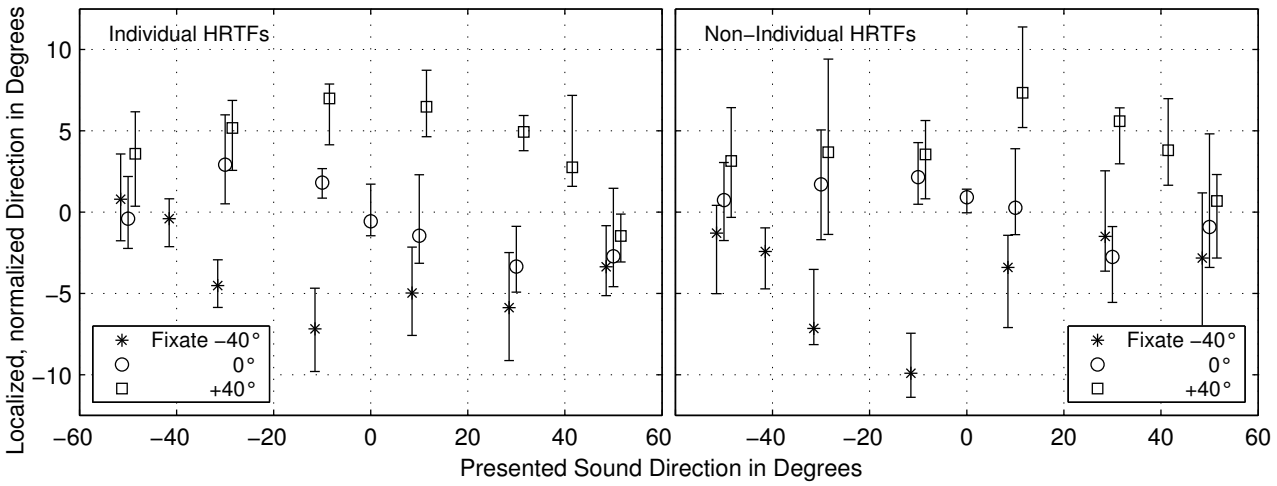

Figure 2: Localization of sounds in virtual space using individual (left) and selected non-individual HRTFs (right) with concurrent fixation of visual targets at $-40°(*)$, $0°(\circ)$, and $+40°(\square)$. Results are presented as deviation against the respective baseline study without visual fixation [2]. Medians of individual medians and quartiles are given.

observed in the real and individualized virtual environment, the shifts are in general smaller with non-individual HRTFs. The results with non-individual HRTFs differ significantly from the results in the real environment (at 0.01%, α-corrected U-test). Although visual bias effects might influence absolute directional shifts, the observed relative shifts between environments will only be slightly affected.

In table 1 a numerical summary of the results is given for: relative error, quartiles, bias, the ratio of the number of inside-the-head localizations as well as front-back-confusions for the experiments with and without visual fixation, cf. [2]. The average bias effects again show that directional shifts are smaller for a fixation of frontal directions, and are also smaller in the virtual environment with non-individual HRTFs compared to the other environments.

A further influence of visual fixation of frontal targets is apparent: The number of inside-the-head localizations and front-back-confusions is clearly reduced with visual fixation compared to the baseline study without visual targets (table 1). A 50% reduction of the number of front-back-confusions with non-individual HRTFs was observed. The number of inside-the-head localizations was also reduced by visual fixation: whereas in 6.2% of the trials obtained in darkness the sound was perceived as inside the head, this was reported in only 4.9% of the trials with visual fixation.

4. Discussion

Previous studies on ventriloquism were conducted in the real environment and showed the attractive influence of visual fixation targets on auditory directions. Similar effects to the current study were found by Weerts and Thurlow [4] for the fixation of a visible loudspeaker at $22°$ and sound being played at $0°$ using a hand-pointer method: bias effects were $9°$ for a closed-loop condition, whereas after-effects reached $2-4°$. Bohlander [5] reported shifts of $1.5-5.9°$ at $45°$ discrepancy for a positionable sound source in the median plane. Bertelson and Radeau [6] obtained an attractive bias of $4°/6.3°/8.2°$ for separations of $7°/15°/25°$ when using synchronized stimuli.

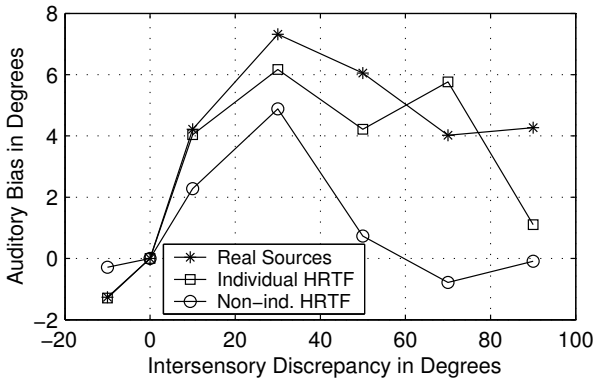

Figure 3: Average localization results for fixation at $\pm 40°$ for all three listening environments, evaluated against the baseline study, cf. fig. 2.

The observation of an attractive bias for discrepancies beyond $50°$ is new to this study. Even for a discrepancy of $90°$ bias effects of about $4°$ were found in the real environment. The bias might me partly due to an effect of visual perception: After the fixation of lateral visual targets the normal position of the eyes is shifted slightly towards the preceding fixation side. Visual targets as the light pointer for the display of the auditory direction might thus be perceived as shifted towards the opposite side. The light spot will then be adjusted slightly

	Real sources	Individual HRTFs	Non-indiv. HRTFs
Rel. Error[a]	-2.4/0.1/4.9	-3.8/1.0/6.5	-5.3/0.2/4.7
Abs. Error[b]	4.4/3.4/4.9	4.8/2.8/6.5	6.4/4.3/5.7
Quartiles	2.4/2.0/2.1	2.5/2.2/1.8	2.4/2.4/3.0
Bias[c]	4.3/1.9/4.2	4.5/1.8/4.9	2.8/0.9/3.7
In-head[d]	-	-	0.86
Front/back[e]	-	0.31	0.50

[a] Relative error in degrees.
[b] Absolute error of responses against presented direction in degrees.
[c] Average absolute shift towards the fixated direction in localization results of the interaction study relative to results from the baseline study in degrees.
[d] Ratio of the number of inside-the-head localizations with and without visual fixation [2]. No inside-the-head localizations in real space and with individual HRTFs.
[e] Ratio of the number of front-back-confusions with and without visual fixation [2]. No data collection in real space.

Table 1: Localization results: error, quartiles, bias, and confusions. Error, quartile, and bias values are separately given for fixation at $-40°/0°/40°$.

towards the place of fixation. The effect increases with the duration of fixation, but reaches only $2°$ at $40°$ displacement after a fixation for 30 s [7].

Only few studies investigated so far the visual influence on the number of front-back-confusions. Jack and Thurlow [8] showed experimentally that a sound source from the rear is occasionally perceived as coming from the front. The effect was greatly reduced for lateral displacements. The current study provides numerical data for a reduction of the number of inside-the-head localizations as well as front-back-confusions through visual fixation. This reduction was observed although the "compellingness" [7] of the interaction situation was low: it can be assumed that the perceptual grouping of the LED fixation spot with the asynchronously presented wide-band-noise pulses is much lower than the auditory-visual grouping in many everyday situations, e.g. speech synchronous to lip movements. Since the use of a visual pointer method might also contribute to the reduction of confusions, the clear reduction seen through fixation suggests a strong visual effect. The fixation of frontal visual targets apparently supports the localization of auditory targets at a similar position, as described by the ventriloquism effect.

The reduction of bias in the non-individualized virtual environment is the most important finding of this study. This can be associated with the reduction of information in localization cues. Our data do not validate the first hypothesis, i.e. that visual dominance becomes greater with less auditory information presented. Instead, the results are consistent with the second hypothesis: The reduction of cognitive congruence between the auditory and the visual object reduces the interaction. It is known that the interaction decreases in general if the "compellingness" is reduced through unsynchronization, increased spatial discrepancy, or reduced contextual accordance [7]. Using non-individual HRTFs the width of the virtual auditory image usually increases (c.f. table 1, quartiles). This could reduce the congruence between the broader auditory and the focused visual object. An interaction study using narrow-band noises causing the same localization variance as observed with non-individual HRTFs showed the same visual bias as observed in the real environment [2]. Therefore, the increase in variance can not account for the decrease in interaction. Another cause for the decrease in interaction with non-individual HRTFs could be the closer distance of the auditory object compared to the other environments. This question is currently under investigation.

5. References

[1] B. Seeber, "A new method for localization studies," *Acta Acustica – Acustica*, 88(3):446–450, 2002.

[2] B. Seeber, "Untersuchung der auditiven Lokalisation mit einer Lichtzeigermethode (Investigation of auditory localization using a light-pointer method)," Dissertation, TU München, 2003, http://tumb1.biblio.tu-muenchen.de/publ/diss/ei/2003/seeber.html.

[3] B. Seeber and H. Fastl, "Subjective selection of non-individual head-related transfer functions," in *Proc. 9th Int. Conf. on Aud. Display*, E. Brazil and B. Shinn-Cunningham, Eds. Boston, USA: Boston Univ. Publications Prod. Dept., 2003, pp. 259–262.

[4] T. Weerts and W. Thurlow, "The effects of eye position and expectation on sound localisation," *Perception & Psychophysics*, 9(1A):35–39, 1971.

[5] R. Bohlander, "Eye position and visual attention influence perceived auditory direction," *Percep. Mot. Skills*, 59:483–510, 1984.

[6] P. Bertelson and M. Radeau, "Cross-modal bias and perceptual fusion with auditory-visual spatial discordance," *Perception & Psychophysics*, 29(6):578–584, 1981.

[7] R. Welch and D. Warren, "Intersensory interactions," in *Handbook of perception and human performance, I: Sensory processes and perception*, K. Boff and L. Kaufman, Eds. New York: Wiley, 1986, ch. 25.

[8] C. Jack and W. Thurlow, "Effects of degree of visual association and angle of displacement on the 'Ventriloquism' effect," *Percep. Mot. Skills*, 37:967–979, 1973.

This work was supported by the Deutsche Forschungsgemeinschaft GRK 267. We like to thank Dr. A.-M. Bonnel for comments on the manuscript.