Ventriloquism Effect in Wave Field Synthesis: Aspects of Direction and Distance

Florian Völk, Ulrich Mühlbauer, Hugo Fastl
AG Technische Akustik, MMK, Technische Universität München, 80333 München, Germany
florian.voelk@mytum.de

Introduction

The psychophysical phenomenon usually referred to as ventriloquism effect describes a visually induced shift of hearing sensation positions. In this contribution, the stability of the hearing sensation positions corresponding to virtual sources generated by wave field synthesis with additional optic stimulation is addressed. Discussed are visually induced direction and distance shifts. It is shown that optic stimuli can influence the direction and distance of hearing sensations at which differences occur for varying optic stimulation.

Problem and Hypotheses

Humans integrate information received by different sensory modalities in perceiving their environment. In natural surroundings, plausible relations between the data collected visually, auditory, and by all other modalities result, combined with further information as for example experience or knowledge, in the impression of reality. The perceptual process continuously provides the currently most plausible representation of the surrounding environment. When humans aim at reporting only information collected by one modality (for example only what they hear), it is not clear whether this information actually has been collected auditory only or if other modalities were taken into account unconsciously. Putting this statement drastically, it can not be taken for granted that there is an auditory (only) perception – perception may always be influenced by multiple modalities.

Against this background, the visually induced shift of the typically less precise hearing sensation position (Alais and Burr 2004) commonly referred to as ventriloquism effect (Howard and Templeton 1966) appears plausible. A precondition for the ventriloquism effect to occur is that the temporal and spatial stimulus organization suggests a mutual relationship (Bertelson and Radeau 1981). Taking a closer look at possible visual influences on hearing sensation positions, the multimodal localization process shall be modeled by a *technical* measurement procedure with limited accuracy and precision. The overall precision may be taken as a measure of the plausibility of the overall result, while the unimodal variances indicate the plausibility of the unimodal percept. Therefore, a modality weighting inversely proportional to the unimodal variances appears meaningful in modeling the multimodal localization process (Bowen et al. 2011). Assuming further that normally distributed measurement results occur for a specific stimulus presented in a static scenario, many localization results of the "average subject" may show a Gaussian probability density function. The parameters of this Gaussian distribution (mean and standard deviation, representing accuracy and precision) depend on the measurement conditions (e. g. listening environment and stimuli) and the measurement system properties (the localization process). A static scenario may for example be a single sound source at a specific position in a completely darkened room.

For designing a localization experiment aiming at collecting a subject group's (internal) localization results, a procedure must be found that minimizes differences between the actual percepts and the reports thereof. A commonly applied procedure is having the subjects point at the perceived location, ideally using a proprioception decoupled pointer, for example a visual marker controlled by a trackball device (Seeber 2002, pointer method). While this method fails at collecting distance results, it has proven accurate regarding directional localization (Völk et al. 2010). Using a controllably moving (real or virtual) sound source, it is also possible to instruct the subjects to position the sound source so that the perceived location coincides with a visual target (adjustment method). Wave field synthesis (WFS) as a physically motivated virtual acoustics technology aims at creating virtual so-called primary point sources at predefined positions (Berkhout 1988). An interactive WFS system that allows controlling the source position by a trackball device is used here.

The *first hypothesis* discussed on that basis is that comparable average localization precisions and accuracies result from the pointer and adjustment methods. If both experiments are conducted in complete darkness (apart from the constantly lit marker or target), no visual contributions to the localization process are expected due to the missing temporal synchronization of the optic and acoustic stimuli. The *second hypothesis* and main topic of this study is that temporally synchronized and spatially close optic stimulation causes sound stimuli located outside the azimuth and distance range determined by the adjustment experiment to be localized at a visual target.

Methods and Stimuli

As acoustic stimuli, frontally located (0° azimuth) primary WFS point sources were implemented at different distances as described by Völk (2010b) based on the signal processing given by Völk and Fastl (2012), with the WFS optimization positions at the array midpoint. The custom made WFS rendering algorithm allowed to relocate the primary sources during the adjustment procedure. As the reference localization method, the pointing method according to Seeber (2002) was employed. The subjects had to indicate the hearing sensation azimuth with a red light point created by a computer controlled laser pointer on an acoustically permeable screen between the subjects and the primary source. The light point could be moved horizontally using a trackball device. Each localization procedure was started randomly at pointer positions randomly selected from the intervals $[-30°, -10°]$ and $[10°, 30°]$, with the aim of forcing the subjects to move the pointer in every case. Errors of both methods due to the uncertainties of the calibration and the step motors positioning the laser pointer were smaller than $\pm 0.1°$.

As sound stimuli, broadband uniform exciting noise (UEN) pulses (Fastl and Zwicker 2007, section 6) were selected for containing equal intensity in all critical bands and thus providing the listener with all spectral localization cues at the same perceptual weight. For providing dynamic localization cues, 700 ms impulses with 20 ms Gaussian gating and 300 ms pause were used.

Procedure

The experiments described here were conducted in a slightly reverberant laboratory (50 ms average reverberation time, 6.8 m × 3.9 m × 3.3 m) at Lehrstuhl für Mensch-Maschine-Kommunikation of Technische Universität München. The WFS playback was realized using a circular loudspeaker array centered at the laboratory midpoint, consisting of 96 broadband loudspeaker boxes mounted at ear height (Bose Freespace 3 Satellite, loudspeaker spacing 8.5 cm, array radius 129.9 cm). Each loudspeaker's free-field response was equalized using an individually designed FIR-filter with the aim of a frequency independent absolute transfer function on the symmetry axis. The equalized loudspeakers were calibrated so that every single loudspeaker produced a level deviating less than ±0.1 dB from all others when reproducing broadband pink noise. The subjects were seated one after another in the darkened laboratory, aiming at applying (apart from the intended optic stimuli) only the optic stimulus darkness. No fixation was applied.

As a reference situation for verifying the adjustment procedure, the azimuth of a primary WFS point source was chosen. The experiment contained nine primary sources at different azimuth/distance combinations, but only the frontally located source at 2.1 m distance is discussed. First, the sources were localized three times each in random order using the pointer method. Then, real loudspeaker boxes (Bose Freespace 3 Satellite, 7.6 cm × 7.6 cm front plate) were positioned at the previous primary source positions. The subjects were asked to adjust the azimuth of primary WFS point sources synthesized at the fixed distances of the visual targets to the azimuth of the visual targets. The primary sources appeared at initial azimuth angles randomly selected from the intervals $[-30°, -10°]$ and $[10°, 30°]$ around the visual targets. Before the sound started, the optic stimulus was slightly illuminated, just to be visible against the dark background without illuminating other elements of the scene. The overall procedure required two times 10 minutes on average.

The main experiment was conducted in two steps of 12 minutes average duration each, beginning with the adjustment experiment for three frontally located targets at the distances 0.8 m, 2.1 m, and 3 m. The subjects had to adjust the azimuth as in the previous experiment and in addition the distance, starting randomly at distances deviating at least 10% from the target distances. Two different targets, light points (red LEDs) and loudspeaker boxes (Bose Freespace 3 Satellite) were used. In the second step, primary WFS point sources positioned at different azimuths and distances around the adjustment results were presented with simultaneous optic stimulation by the light points or by illuminating the loudspeakers. The subjects had to indicate in a yes/no procedure whether the sound was heard at the visual target.

Results and Discussion

Seven experienced normal hearing subjects between 23 and 32 years (average 28) participated in the study. The results of the localization experiments are shown graphically as the histograms of all values (step sizes 2° and 34% of the distance). Supplementary, the quartiles of the individual arithmetic means are tabulated. The yes/no procedure directly results in the relative frequency of the answer *yes*, with the step size of the chosen primary source positions. In addition, all relative frequencies are fit in a least-squares sense to the Gaussian distribution

$$w(x) = A\,\mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}. \tag{1}$$

The parameters A, σ, and μ are tabulated for every case together with the quartiles of the localization experiments.

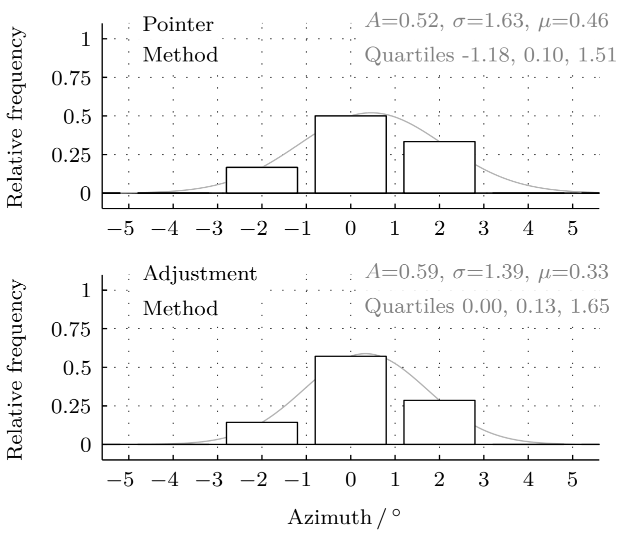

Figure 1: Histograms (bars) and Gaussian least-squares fits (gray contours) of the localization result azimuths for a wave field synthesis point source at 0° azimuth and 2.1 m distance.

Figure 1 shows the results of the verification experiment. The upper panel represents the data acquired with the visual pointer method, the lower panel the results of the sound source adjustment method. One factorial analysis of variance indicates no significant main effect of the method [F(1,6)= 0.7; p= 0.4342] and confirms the visible similarity of the results. The accordance is also

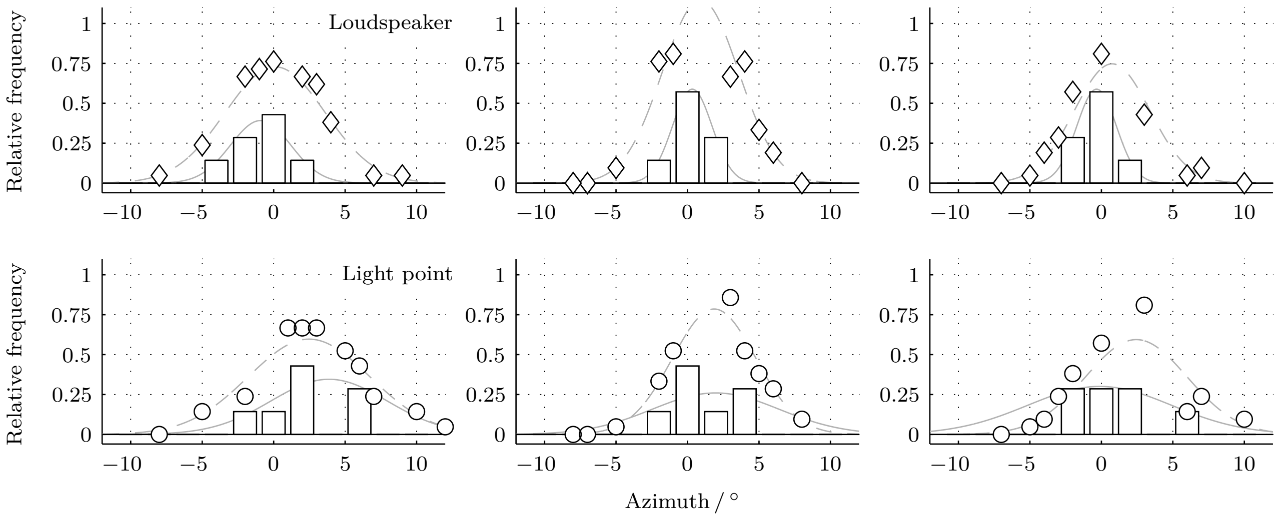

Figure 2: Histograms (bars) of the azimuths of wave field synthesis point sources adjusted to the inserted visual targets at the distances 0.8 m, 2.1 m, and 3 m and 0° azimuth. The symbols indicate the results of a yes/no procedure asking whether the sound was localized at the visual target. Gray contours represent Gaussian least-squares fits (parameters given by table 1).

reflected in almost identical parameters of the fitted Gaussian functions and in the quartiles (cf. inserts in figure 1). Consequently, the *first hypothesis* is confirmed, the visual pointer and the sound source adjustment methods, conducted in complete darkness, result in comparable localization precision and accuracy regarding the azimuth. Therefore, the sound source adjustment method is considered valid regarding azimuth and distance for the main experiment, while being verified only for the azimuth.

The results of the main experiment are shown separately for azimuth (figure 2) and distance (figure 3). Both figures contain two rows, the upper representing the optic stimulus loudspeaker (LS), the lower the optic stimulus light point (LP). Each row is composed of three panels depicting data acquired at three different distances of the optic and acoustic stimuli (0.8 m, 2.1 m, and 3 m from left to right). In all panels, the vertical bars show the histograms of the results from the adjustment procedure at the respective position. The solid gray contours indicate the corresponding Gaussian fits, represented also by the parameters and quartiles given by tables 1 and 2 (adj). The results of the yes/no procedure are indicated for the loudspeakers as optic stimuli by diamonds and for the light points by circles. The dashed gray contours represent the corresponding Gaussian fits, described by the parameters also summarized in tables 1 and 2 (y/n).

	A	σ	μ	Quartiles
0.8 m				
LS adj	0.39	2.11	-0.91	-1.53,0.03,1.04
LP adj	0.35	4.07	3.88	1.34,3.17,5.30
LS y/n	0.73	3.50	0.10	
LP y/n	0.60	4.25	2.48	
2.1 m				
LS adj	0.59	1.39	0.33	-0.38,0.17,1.31
LP adj	0.26	4.44	1.97	-0.20,0.30,3.13
LS y/n	1.15	2.68	0.91	
LP y/n	0.79	2.92	1.86	
3 m				
LS adj	0.59	1.39	-0.33	-0.77,0.52,1.16
LP adj	0.30	5.04	-0.12	-1.07,1.09,2.40
LS y/n	0.75	2.65	0.75	
LP y/n	0.59	3.56	2.44	

Table 1: Parameters of Gaussian least-squares fits to the azimuth results shown in figure 2 for the optic stimuli loudspeaker (LS) and light point (LP) at different sound source distances. Given are results from the sound source adjustment (adj) and from a yes/no procedure (y/n) targeting ventriloquism.

Azimuth

The azimuth data indicate with both optic stimuli little dependence of the target distance, for the adjustment and for the yes/no procedure, that is regarding localization and ventriloquism. However, the average accuracy and precision depend on the optic stimulus: the loudspeaker azimuth is localized more precisely and more accurately than that of the light point. Ventriloquism is visible for both optic stimuli, meaning that sound sources outside the range covered by the adjustment results are perceived as being located at the position of the optic stimulus. This is reflected in the Gaussian fits (cf. table 1), confirming the *second hypothesis* regarding the azimuth. Thereby, the effect is somewhat more pronounced for the optic stimulus loudspeaker versus light point.

Distance

The distance adjustment results are globally as expected, taking into account the so-called distance compression (Zahorik et al. 2005, Völk 2010a): sources closer than about 1.5 m are rated too far, farther sources are judged too close. Therefore, the sources at 0.8 m distance (left column of figure 3) are expected to be adjusted closer than the optic stimulus, the other sources are expected to be adjusted farther than the respective optic stimuli, as confirmed by the results. Further, the data of table 2 indicate a with distance decaying precision (increasing σ and inter-quartile range). Regarding localization, no clear

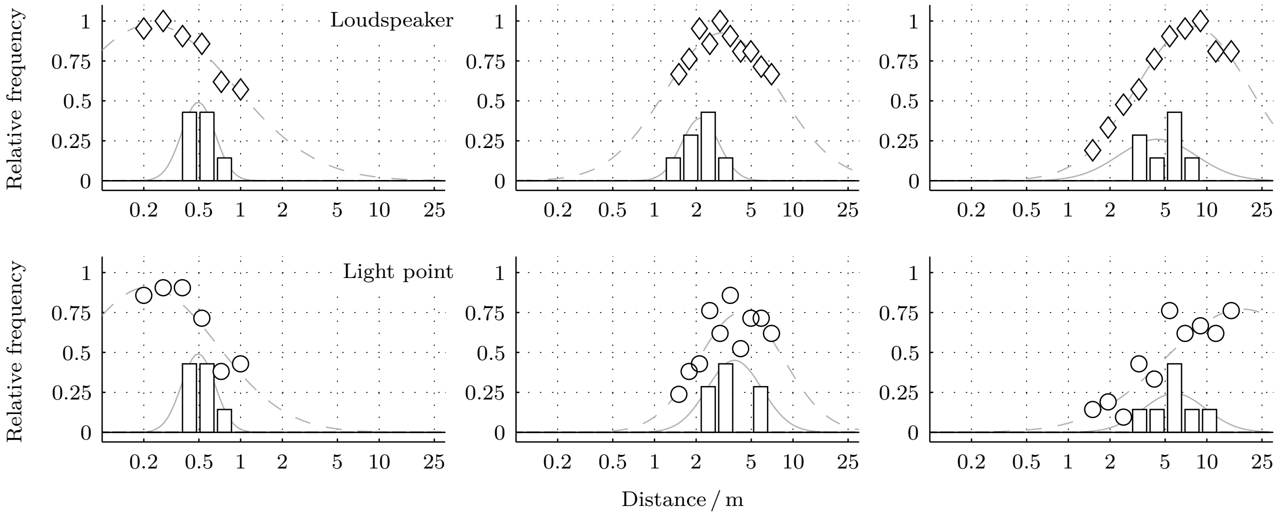

Figure 3: Histograms (bars) of the distances of wave field synthesis point sources adjusted to the inserted visual targets at the distances 0.8 m, 2.1 m, and 3 m and 0° azimuth. The symbols indicate the results of a yes/no procedure asking whether the sound was localized at the visual target. Gray contours represent Gaussian least-squares fits (parameters given by table 2).

	A	σ	μ	Quartiles
0.8 m				
LS adj	0.49	0.28	0.49	0.42,0.52,0.59
LP adj	0.49	0.28	0.49	0.46,0.55,0.64
LS y/n	0.98	1.37	0.23	
LP y/n	0.91	1.15	0.21	
2.1 m				
LS adj	0.39	0.31	2.14	2.06,2.31,2.73
LP adj	0.45	0.46	3.78	3.39,3.56,5.13
LS y/n	0.92	0.95	3.09	
LP y/n	0.75	0.71	4.22	
3 m				
LS adj	0.26	0.65	4.39	3.86,5.14,6.08
LP adj	0.24	0.52	5.85	4.69,6.04,7.81
LS y/n	0.97	0.94	7.94	
LP y/n	0.77	1.27	18.02	

Table 2: Parameters of Gaussian least-squares fits to the distance results shown in figure 3 for the optic stimuli loudspeaker (LS) and light point (LP) at different sound source distances. Given are results from the sound source adjustment (adj) and from a yes/no procedure (y/n) targeting ventriloquism.

effect of the optic stimulus is visible. Ventriloquism occurs for all conditions and both optic stimuli in that sources farther or closer than the ranges of the adjustment results are localized at the visual targets. The effect is reflected in the parameters of the Gaussian fits (table 2), confirming the *second hypothesis* also regarding distance, where a more pronounced effect occurs for the optic stimulus loudspeaker versus light point.

The authors are indebted to Deutsche Forschungsgemeinschaft (DFG) who supported parts of this work by grant FA 140/4.

References

Alais D., D. Burr: The Ventriloquist Effect Results from Near-Optimal Bimodal Integration. Current Biology **14**, 257–262 (2004)

Berkhout A. J.: A Holographic Approach to Acoustic Control. J. Audio Eng. Soc. **36**, 977–995 (1988)

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

Bowen A. L., R. Ramachandran, J. A. Muday, J. A. Schirillo: Visual signals bias auditory targets in azimuth and depth. Exp. Brain Res. **214**, 403–414 (2011)

Fastl H., E. Zwicker: *Psychoacoustics – Facts and Models.* 3rd Edition (Springer, Berlin, Heidelberg, 2007)

Howard I. P., W. B. Templeton: *Human Spatial Orientation.* (Wiley, New York, 1966)

Seeber B. U.: A New Method for Localization Studies. Acta Acustica united with Acustica **88**, 446–450 (2002)

Völk F.: Messtechnische Verifizierung eines datenbasierten binauralen Synthesesystems. In *Fortschritte der Akustik, DAGA 2010*, 1049–1050 (Dt. Gesell. für Akustik e. V., Berlin, 2010a)

Völk F.: Psychoakustische Experimente zur Distanz mittels Wellenfeldsynthese erzeugter Hörereignisse. In *Fortschritte der Akustik, DAGA 2010*, 1065–1066 (Dt. Gesell. für Akustik e. V., Berlin, 2010b)

Völk F., H. Fastl: Wave Field Synthesis with Primary Source Correction: Theory, Simulation Results, and Comparison to Earlier Approaches. In *133rd AES Convention* (2012) (Convention Paper 8717)

Völk F., M. Schmidhuber, H. Fastl: Influence of the ventriloquism effect on minimum audible angles assessed with wave field synthesis and intensity panning. In *20th Int. Congress on Acoustics (ICA)* (2010)

Zahorik P., D. S. Brungart, A. W. Bronkhorst: Auditory Distance Perception in Humans: A Summary of Past and Present Research. Acta Acustica united with Acustica **91**, 409–420 (2005)