Proceedings of 20th International Congress on Acoustics, ICA 2010

23–27 August 2010, Sydney, Australia

Influence of the ventriloquism effect on minimum audible angles assessed with wave field synthesis and intensity panning

Florian Völk, Maximilian Schmidhuber*, and Hugo Fastl

AG Technische Akustik, Lehrstuhl für Mensch-Maschine-Kommunikation, Technische Universität München, Germany

florian.voelk@mytum.de

PACS: 43.66.Qp, 43.60.Dh

ABSTRACT

The psychophysical phenomenon called ventriloquism effect describes a possibly occurring influence of a visual stimulus on the judgment of a sound source's perceived position. The term minimum audible angle usually denotes the just distinguishable horizontal angular deviation between two sound sources. It depends on the direction of the sound sources relative to the listener's head, the considered sound stimuli, and the presentation sequence. Within this paper, a possible influence of the ventriloquism effect on the minimum audible angle for sequentially presented broadband noise is assessed with wave field synthesis and intensity panning as reproduction methods in a reflective environment, using an adaptive two-alternative forced choice procedure. As baseline, the minimum audible angle without any intended visual stimuli is determined in complete darkness for both reproduction methods. These baseline data are compared to absolute localization trial results for validation purposes. As visual stimuli, one concurrent and one contradictory light spot are added to the sound stimuli. A possibly occurring ventriloquism effect can reduce or enlarge the minimum audible angle depending on whether the visual stimulus' position change is larger or smaller than the change in sound source position. It is shown that visual stimuli can influence minimum audible angles for both reproduction methods, at which considerable differences occur inter-individually and between the considered reproduction methods.

INTRODUCTION

In 1958, Mills defined the minimum audible angle (MAA) "as the smallest detectable difference between the azimuths of two identical sources of sound", in detail as

> the angle formed at the center of the head by lines projecting to two sources of sound whose positions are just noticeably different, when they are sounded in succession.

This angle is of particular interest in audio-visual presentation systems since it defines the maximum horizontal angular resolution required for the acoustical reproduction chain when considering the human ear as the communication receiver, as it is advisable obviously in any case dealing with human listeners (cf. Zwicker and Feldtkeller 1967, 1999, Fastl and Zwicker 2007). The assumption of the MAA representing the maximum resolution required in audio reproduction holds true for MAAs determined without any visual stimulation (e. g. in complete darkness), since additional visual information in a meaningful audio-visual scenario might lower the resolution required for the audio reproduction by shifting the auditory perceived direction towards the corresponding visual direction (e. g. Seeber 2002b). This latter effect is commonly referred to as ventriloquism effect (cf. Howard and Templeton 1966).

Seeber (2003) discussed among others the thesis that the amount of the shift of the auditory perceived direction due to the ventriloquism effect is dependent on the accuracy of physically present directional information. Seeber used real sources and individualized as well as non-individualized static binaural synthesis (cf. Völk 2009) to achieve different degrees of accuracy of directional information. As a consequence of the selected reproduction methods, he was not able to prove his thesis, especially due to the unrealistic auditory event position achievable with static non-individualized binaural synthesis. This lack of realism is primarily caused by the, compared to real sound sources, reduced perceived distance (cf. Völk et al. 2008).

There are two major aims of the paper on hand. One is to prove the existence of influences of visual stimuli on the minimum audible angle and to quantify them for different playback procedures. The other is to re-evaluate Seeber's hypothesis using wave field synthesis rendering and intensity panning. That way, different accuracy of directional acoustical information can be achieved while roughly preserving the auditory event position from the baseline scenario without visual stimulation (cf. Völk 2010).

In the first section, an overview of related work is given. Then, the setup and procedures used are introduced. On this basis, first the acquired minimum audible angle data and afterwards the influences of the ventriloquism effect are discussed. A summary concludes the given discussion.

PREVIOUS WORK

The authors are not aware of a study considering the influence of the ventriloquism effect on minimum audible angles. For that reason, related work on MAAs as well as on the ventriloquism effect is presented separately.

* The second author is now with Akustikbüro Schwartzenberger und Burkhart, Pöcking, Germany.

Minimum audible angle

Mills (1958) used pairs of tone pulses, each of them having a duration of 1 s with 70 ms rise and fall time and 1 s pause in between, to assess minimum audible angles in an anechoic chamber. As experimental paradigm, he chose a two-alternative forced choice (2-AFC) procedure. Every first pulse was presented from a reference position, while the second pulses occurred at different horizontal angular separations to the reference. The subjects had to indicate whether the second pulse originated from a point to the left or right of the first. Then, the minimum audible angle was taken as one half the difference between the 25% and 75% point of the psychometric function, the latter defined as line fit to the proportion of responses *right* over the change of the linearly scaled horizontal angle. The angular limits were selected symmetrical to the frontal direction. According to Hartmann and Rakerd (1989), this procedure is equivalent to directly taking the angle as minimum audible angle, where 75% of the responses are correct. Mills' results indicate that the minimum audible angle depends on the direction of sound incidence (frontal stimuli can be distinguished best) and on the stimulus' spectral content.

Hartmann and Rakerd (1989) showed that the procedure used by Mills (1958) is more likely an absolute than a relative localization task. For that reason, they suggest to use a method they call *two sources two intervals* (2S2I) in experiments aiming to measure the ability of listeners to discriminate between sound locations. Nevertheless, Hartmann and Rakerd assume the dependencies of the MAA on frequency and direction found by Mills to be correct since his procedure is suited for a comparison of different situations.

Perrott and Pacheco (1989) addressed a dependence of the minimum audible angle for broad band pink noise impulses (duration 10 ms, rise and decay time less than 0.1 ms) on the time between the stimulus onsets (inter-stimulus onset interval, ISOI). They used an adaptive 2-AFC procedure asking whether the second auditory event occurred to the left or right of the first, which can be regarded as variation of a 2S2I method (as defined by Hartmann and Rakerd 1989). The minimum audible angle was defined as the angular separation between two sources which could be resolved at a 70.7% correct response level. They found a decrease of the minimum audible angle from about 2.5° to about 1° when increasing the ISOI from 1 ms to 150 ms, and no further change for larger time differences.

Perrott and Saberi (1990) found a minimum audible angle of about 0.97° for 400 Hz click trains (50 ms duration, 500 ms inter-stimulus onset interval). They employed an adaptive 2-AFC procedure (question: second event to the left or right of the first) and defined the MAA at the 79.6% correct response level. Using the same procedure, Perrott (1993) found a MAA for noise, high-pass filtered at 1 kHz, of about of 1.2° (200 ms stimulus duration, 10 ms rise-/decay-time, 200 ms inter-stimulus interval).

To sum up this section: All studies discussed here find minimum audible angles of about 1° for broadband stimuli in frontal direction, when the ISOI exceeds about 150 ms. For shorter ISOI, reduced bandwidth, or lateral directions of sound incidence, the MAA increases.

Relation between the minimum audible angle and absolute localization

Recanzone et al. (1998) compared the relative and absolute localization ability with Gaussian noise as well as 1 kHz and 4 kHz tones (200 ms duration, 5 ms linear rise-/fall-time, 600 ms ISOI). Within their experiments, the relative localization threshold was a poor indicator of the range of estimates of the same stimulus in absolute space. The width of psychometric functions for relative localization was well correlated with the width of the distribution of estimates in the absolute localization task.

Ventriloquism effect

Seeber (2003) gives a comprehensive overview of the ventriloquism effect. This overview is adapted to the subject of the paper on hand. The demonstrative fundamental influence of the ventriloquism effect is the following: It moves the perceived direction of the ventriloquists voice in the region of the puppets mouth and leads to a fusion of the visual direction of the puppet mouth with the auditory event's direction (which makes the directional discrepancy not perceivable, Radeau 1974). Nevertheless, a directional shift might occur independent of the fusion (Bertelson and Radeau 1981). One possible explanation for the occurrence of the ventriloquism effect is that the perceptive system is expecting one auditory/visual object (*unity assumption*, Welch and Warren 1980, Vatakis and Spence 2007). Another explanation assumes that cross-referencing between modalities is used to reduce variability compared to single-modality processes (Vroomen et al. 2001).

To study the ventriloquism effect, an auditory object is played back from one direction simultaneously to a visual object from a different direction. These objects can be of natural origin (a video of a talker and his voice) or synthetically created (e. g. a light point and an acoustical noise signal). If the stimuli are too long (more than several seconds), the sensory system might recalibrate (Bertelson and Radeau 1981, Welch and Warren 1980) and the effects to be studied would disappear. There are influences between the auditory and visual modalities in both directions, but the visual modality is assumed to show the larger influence (visual dominance, for a more thorough discussion see Seeber 2003). The synchronization between auditory and visual stimuli influences their fusion in perception (leading auditory stimuli disturb fusion more than leading visual stimuli, see also Seeber 2003).

Hypotheses of the present study

Basis of this study was the hypothesis that the ventriloquism effect can influence the minimum audible angle. Since the additional visual stimulus is assumed to shift the auditory event position, theoretically an enlargement or reduction of the MAA should be possible. In the work discussed here, a light point was presented with each sound. Either the first or the second light point was located at one sound event position, the respective other light point at a position different from the respective other sound event position. This way, the occurrence of shifts in the auditory event position is intended to be verified.

A further aim of the study was to check for Seeber's hypothesis (cf. also Seeber 2002b) that the amount of the shift of the auditory perceived direction due to the ventriloquism effect is dependent on the accuracy of physically present directional information with different reproduction methods. This procedure appears to be promising since wave field synthesis as physically motivated reproduction method is able to generate the desired sound field at least approximately and therefore to deliver correct localization cues (cf. Völk 2010). Intensity panning on the contrary, as psychoacoustically motivated reproduction method (cf. Blauert 1997), is assumed to deliver less correct localization cues while, if arranged appropriately, causing the same auditory event position as wave field synthesis (without visual stimulation). As a consequence, more influence of additional visual stimulation on minimum audible angles is expected using intensity panning than using wave field synthesis. The two major differences to the study described in Seeber (2003) are the presentation methods used and the parameter attempted to be influenced by the ventriloquism effect. While Seeber assessed a shift in the absolute location, within this study, the ability to discriminate two close sound stimuli is of interest.

SETUP AND PROCEDURE

Within this section, the experimental setups for the different experiments are presented. At first, the psychoacoustical methods and the employed playback systems are described in detail. Then, the auditory and visual stimuli are introduced. Finally, the experimental room's acoustical characteristics are given.

Methods and definitions

Following Perrott and Marlborough (1989), a two-alternative forced choice (2-AFC, cf. Hellbrück and Ellermeier 2004) 2-down/1-up method (cf. Levitt 1971) is used as psychometric method for assessment of the minimum audible angle. For the step size adaption, PEST (Parameter Estimation by Sequential Testing, cf. Gelfand 1990) is employed. The listeners are asked to indicate whether the second auditory event occurred to the left or to the right of the first by pressing one of two buttons, while the presentation sequence (left/right or right/left) is chosen randomly. This procedure is repeated until both, the deviation between the last two minimum and the deviation between the last two maximum values remains smaller than 0.8°.

Since this procedure is converging towards the 70.7% point of the psychometric function, the MAA is defined here as the threshold in azimuth where about 71% of all judgments of the relative horizontal position of the sound sources are correct. This definition is comparable to earlier measurements defining the MAA in the region around the 75% point of the psychometric function. The audio-visual stimulus sequences were randomly interleaved, to avoid learning effects. Such learning effects might occur when the visual and auditory stimulus positions would change either always to the same direction or always contradictory.

For absolute localization judgments, the ProDePo light pointer method ac. to Seeber (2001) is employed in the current study (**Pro**prioception **De**coupled **Po**inter method, cf. also Seeber 2002a, 2003). Within this study, the subjects indicate the auditory event azimuth with a light spot, created by a computer controlled laser pointer on a small stripe of paper mounted directly on the loudspeakers. The light point could be moved horizontally using a trackball input device. Positioning errors of this procedure, due to calibration errors or uncertainties in the step motors positioning the laser pointer, are smaller than ±0.1°.

Sound stimuli

The sound stimuli used for all experiments presented in this paper are uniform exciting noise impulses (UEN, cf. Fastl and Zwicker 2007). UEN is selected as stimulus since it contains equal intensity in each critical band and is thus assumed to provide all spectral localization cues to the listener with the same perceptual weight.

For assessment of the minimum audible angle, two uniform exciting noise impulses positioned symmetrically to the left and right of the median plane (for the definition cf. Blauert 1997) are successively presented to the listeners. Following Blauert and Braasch (2007), a pulse duration above 200 ms allows for orientational head movements by providing so called dynamic localization cues. The first MAA and the ventriloquism effect experiments presented within this study are conducted using 700 ms impulse duration, 20 ms Gaussian gating, and 300 ms pause between the two stimuli. The sound stimuli for the verification experiment of the MAA discussed at the end of this paper are presented with 700 ms pulse duration, 20 ms Gaussian gating, and 1.3 s pause duration. The absolute localization measurements are conducted with a pulse sequence (700 ms pulse duration with 20 ms Gaussian gating and 300 ms pause duration) to add temporal localization cues besides the random temporal structure of the noise. The pulse sequence is presented for each stimulus as long as the localization process is not finished.

To sum up the selection of the sound signals discussed in this section: stimuli are used which are expected to provide the listeners with as many acoustical localization cues as possible. Therefore, the results acquired using these stimuli can be regarded as best case localization results for the situation considered in the respective case.

Intensity panning

The psychoacoustical phenomenon called summing localization (cf. Warncke 1941) is used within this study to generate auditory events – so called phantom sound sources – between two physically present loudspeakers by intensity panning (cf. Blauert 1997). In intensity panning, two loudspeakers are driven with coherent signals, possibly having different intensities. Under these circumstances and for certain geometrical arrangements of loudspeakers and listener, one auditory event arises somewhere in between the loudspeakers (cf. e. g. Theile and Plenge 1976). The horizontal angle of this auditory event can be controlled by the intensity difference between the loudspeakers' input signals (usually given as level difference) and depends further on the loudspeaker arrangement and the reproduction room's acoustical characteristics (cf. e. g. Cabot 1977).

There exist a number of studies (for example Aoki et al. 1990, Theile and Plenge 1976) on the auditory event positions for the so called standard stereo setup (two loudspeakers at ±30°, cf. ITU-R BS. 775-1), but the authors are not aware of a study discussing the auditory event positions for phantom sources in between closely spaced loudspeakers. Additionally, in cases where the reproduction room's influence can not be neglected, the auditory event positions may deviate considerably from comparable studies using identical loudspeaker and listener arrangements (cf. e. g. Queen 1979, Bech 1998). For these reasons, the azimuths of the auditory events in the situation discussed within this paper (two loudspeakers at 1.3 m distance, symmetrically arranged in front of the listener and separated by 3.75°) are measured for distinct level differences between the loudspeaker input signals. The aim of this procedure is to be able to repeatably generate sound sources spaced closer than the physical loudspeakers for the assessment of the MAA. The weighting of the loudspeaker input signals was carried out so that the desired level difference occurred while the sound intensity at the listening position remained unchanged. Figure 1 shows the inter-individual medians and inter-quartile ranges from the individual azimuth estimates of 13 normal hearing subjects (age 24 to 29 years). For each subject, the intra-individual medians over three repetitions per stimulus are regarded as individual result (random temporal presentation sequence, mean trial duration 17.2 minutes). The gray areas in the background indicate the physical extent of the loudspeaker boxes. Since the absolute position of the auditory events is of less interest than their relative positioning, the individual results were intra-individually normalized to the mean of the judgments corresponding to the real loudspeakers.

The results show a clear transition region where the auditory event judgments lie on average in between the physical loudspeakers. This transition region is expected to occur symmetrically around 0 dB level difference (cf. Blauert 1997), but is

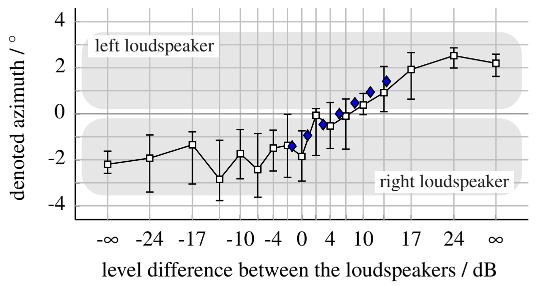

Figure 1: Denoted auditory event azimuths for broad band uniform exciting noise presented in a reflective listening environment with intensity panning (3.75° base angle, 1.3 m distance). Medians (squares) and inter-quartile ranges of intra-individual medians (intra-individually normalized to the denoted auditory event positions corresponding to the loudspeakers).

centered around a level difference of about 6 dB here. The reason for this shift could not be clarified definitely, but calibration errors or phase inversion in one of the loudspeaker input signals can be excluded. Possible reasons may be room influences or differences in the loudspeakers' transfer characteristics, despite individual calibration and equalization with the aim of a frequency independent absolute transfer function. Nevertheless, since the MAA assessment requires only relative positions of closely spaced auditory events, the results were used to compute seven level differences corresponding to neighboring auditory event position by linear fitting, which in turn were used for the MAA measurements with intensity panning. The level differences used and the assumed relative auditory event positions are shown as colored diamonds in figure 1.

Wave field synthesis rendering

For the wave field synthesis rendering, a circular loudspeaker array centered at the laboratory room's midpoint is used. The array consists of 96 broadband loudspeaker boxes (Bose Freespace 3 Satellit, loudspeaker spacing $\Delta x = 8.5\,\mathrm{cm}$, array radius $r = 129.9\,\mathrm{cm}$). Each loudspeaker's free-field response is 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 broad band pink noise.

Wave field synthesis is an audio reproduction method aiming to synthesize a sound field within a limited listening volume V (cf. Spors et al. 2008). The correct synthesis of a so called primary sound field within the listening volume is theoretically possible for example with an infinite number of monopole sources (secondary sources) on the listening volume's boundary ∂V based on the Kirchhoff-Helmholtz integral equation. Using loudspeaker boxes instead of monopole sources for realizations causes errors in the reproduced sound field, since loudspeaker boxes are neither point sources nor radiate spherical waves. Therefore, the reproduced wave field will not be correct at all spatial positions and for all frequencies. Nevertheless, wave field synthesis is capable to generate a sound field at a distinct listening position (reference point $\mathbf{x}_{\mathrm{ref}}$) within a limited frequency range.

$\hat{P}(\omega)$ denotes the Fourier-spectrum of the sound pressure time signal radiated from the primary source, $\mathbf{x}_0$ the position of the considered secondary source, Δx_0 the distance to the next secondary source, and $\mathbf{n}(\mathbf{x}_0)$ the normal vector on ∂V pointing from $\mathbf{x}_0$ into V (nomenclature adjusted from Spors et al. 2008). Using these conventions, it is possible to describe the loudspeaker input signals' Fourier-spectra (the so called driving functions) for synthesis of plane wave fronts (index pw, synthesis procedure adapted from Spors et al. 2008) as follows:

$$\begin{aligned} D_{2.5\mathrm{D,pw}}(\mathbf{x}_0,\omega) = a_{\mathrm{pw}}(\mathbf{x}_0)\sqrt{jk}\sqrt{2\pi|\mathbf{x}_{\mathrm{ref}}-\mathbf{x}_0|}\cdot \\ \cdot\, \mathbf{n}_{\mathrm{pw}}^{\mathrm{T}}\mathbf{n}(\mathbf{x}_0)\hat{P}(\omega)e^{-jk\mathbf{n}_{\mathrm{pw}}^{\mathrm{T}}\mathbf{x}_0}\Delta x_0 \end{aligned} \tag{1}$$

Here, $\mathbf{n}_{\mathrm{pw}}$ represents the normal vector of the wave front to be synthesized in propagation direction and the following equation applies:

$$a_{\mathrm{pw}}(\mathbf{x}_0) = \begin{cases} 1 & \text{for} \quad \mathbf{n}_{\mathrm{pw}}^{\mathrm{T}}\mathbf{n}(\mathbf{x}_0) > 0, \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Listening room environment

All experiments presented within this paper are conducted in a laboratory room (6.8 m × 3.9 m × 3.3 m) at Lehrstuhl für Mensch-Maschine-Kommunikation of Technische Universität München. Figure 2 shows its reverberation time averaged over measurements from each involved array loudspeaker at its position during the experiment to two microphones at the listening point (midpoint of the loudspeaker array) and 10 cm aside.

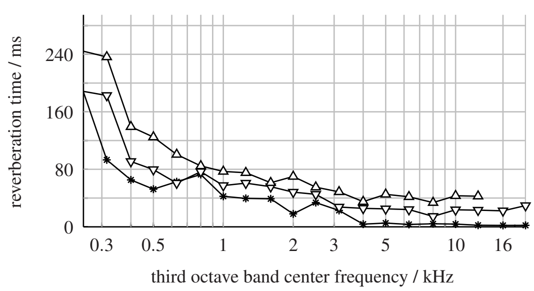

Figure 2: Reverberation time of the laboratory room where all listening experiments presented here took place. Early decay time (stars), T_{20} (downward pointing triangles), and T_{30} (upward pointing triangles).

The experiments are carried out in complete darkness. The laboratory is darkened after the subjects are instructed and positioned on a chair in such a way that the midpoint of their head is positioned at the reference point. The position is adjusted using a head tracking system, ensuring an accuracy in the plane defined by the loudspeaker array of ± 3 cm as well as an accuracy in height of ± 1 cm.

Visual stimuli

Within this study, 31 light emitting diodes (LEDs) are used as visual stimuli, producing red light spots in the dark surroundings. This same visual stimulus was also used by Seeber (2003). The LEDs are mounted directly in front of the loudspeaker boxes with an angular spacing of 0.47° with respect to the listening position. This spacing is selected based on the auditory event spacing corresponding to the phantom sound sources (cf. figure 1). That way, each phantom source can be represented by one LED and further LEDs are available to represent more lateral positions.

Audio-visual stimulus combination

For the assessment of the ventriloquism effect, different temporal and spatial combinations of visual and acoustical stimuli are possible. Figure 3 shows the temporal presentation sequences used here. The first audio-visual scenario consists of synchronized auditory and visual stimuli. The second second scenario

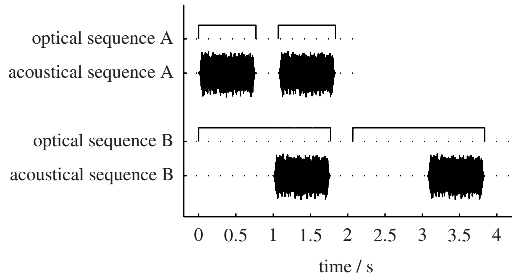

Figure 3: Temporal audio-visual stimulus sequences for assessment of the ventriloquism effect's influence on the minimum audible angle.

is set up to ensure that the visual object can be fixated before the sound stimulus is played back. According to Perrott (1993) as well as Shelton and Searle (1980), 1 s is a sufficient amount of time to allow for fixation even if some correcting saccades are required. Insertion of this additional time with exclusive visual stimulation causes the pause between the two sound stimuli to increase to 1.3 s.

Besides the temporal organization of the audio-visual stimulus combinations, there exist different possibilities for the spatial distribution of visual and acoustical stimuli. Within this work, one of the light points is always presented at the position of one of the acoustical stimuli. The other visual stimulus is presented together with a sound occurring at a constant horizontal angular deviation of $\delta = 5.2°$ to the first sound. Therefore, the directional movement of the sound and light stimuli either occurs in the same direction or contradictory, since the light point is shifted by δ with regard to the corresponding sound stimulus either to the left or to the right. In table 1, the spatial presentation sequences studied within this paper are depicted. In addition, the related hypotheses are formulated, based on the thesis that the visually perceived stimulus position attracts the auditory perceived.

Table 1: Spatio-temporal audio-visual stimulus sequences for the assessment of the ventriloquism effect's influence on the minimum audible angle and related hypotheses.

1	no visual stimulation (complete darkness); *hypothesis: no ventriloquism effect occurs*
2	sound and light point positions first coincident, then different; the light point moves in sound direction, but further than the sound; *hypothesis: the ventriloquism effect causes a decrease of the MAA*
3	sound and light point positions first coincident, then different; the light point moves contradictory to the sound movement; *hypothesis: the ventriloquism effect causes an increase of the MAA*
4	sound and light point positions initially different, then coincident; the light point moves in sound direction, but further than the sound; *hypothesis: the ventriloquism effect causes a decrease of the MAA*
5	sound and light point positions initially different, then coincident; the light point moves contradictory to the sound movement; *hypothesis: the ventriloquism effect causes an increase of the MAA*

MINIMUM AUDIBLE ANGLES

Within this section, the minimum audible angles assessed without visual stimulation (in complete darkness) are presented. Initially, the adaptive method is verified. Afterwards, the MAAs are shown, compared to absolute localization judgments, and discussed eventually.

Verification of the procedure

Since implementing an adaptive listening experiment requires complex adaption rules, the algorithm used had to be checked before the actual experiment started. Therefore, the MAA resulting from the adaptive procedure is compared to the psychometric function acquired by a static two-alternative forced choice experiment for the same situation (wave field synthesis, frontal sound incidence, dark surroundings, reflective listening environment). For the static procedure, the results per stimulus condition are defined as inter-individual mean value of the intra-individual means over 20 repetitions. In both experiments, the same eight subjects participated (age between 22 and 30 years, mean duration about 9 minutes for the static and about 6 minutes for the adaptive procedure).

Figure 4 shows the outcome of both experiments. The results (inter-individual median and inter-quartile range over intra-individual medians) from the adaptive experiment are drawn at 70.7% of the psychometric function, assuming the 2-down/1-up procedure used converges to this point (cf. Hellbrück and Ellermeier 2004).

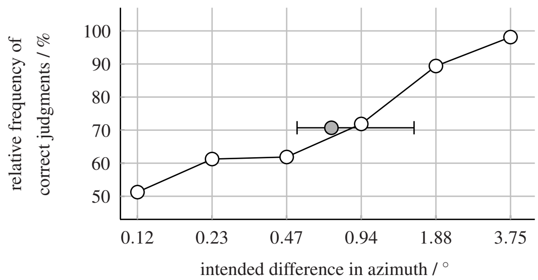

Figure 4: Minimum audible angles assessed with wave field synthesis in dark surroundings (without visual stimulation) with broad band uniform exciting noise. Inter-individual psychometric function as the result of a static 2-AFC procedure (open symbols) as well as median and inter-quartile range of intra-individual medians (filled), resulting from an adaptive 2-AFC procedure. Temporal presentation sequence A (cf. fig. 3).

It is clearly visible that the adaptive procedure converges to the 70.7% point of the psychometric function resulting from the experiment using the static method. For that reason, the implementation of the adaptive procedure is regarded as producing a valid approximation of the minimum audible angle as defined within this work (cf. section *Methods and definitions*), and therefore is used for all further experiments. In addition, the resulting minimum audible angle around 0.9° is in good agreement with MAAs reported for broad band stimuli and frontal sound incidence in earlier studies (cf. section *Minimum audible angle*).

Results and discussion

Figure 5 shows a comparison of the minimum audible angles acquired with intensity panning (IP) and wave field synthesis (WFS) for frontal sound incidence. The experiment with presentation sequence A (cf. fig. 3) was carried out by ten subjects aged 24 to 29 years. IP as presentation method caused a mean trial duration of 6.5 minutes, WFS about 7.3 minutes, both with three repetitions per stimulus. In an additional run, the experiment was repeated with WFS by four subjects aged between 25 and 29 years using presentation sequence B (cf. fig. 3). The mean duration resulted to about 9.3 minutes with three repetitions per stimulus. Filled symbols denote medians and inter-quartile ranges of intra-individual medians. Therefore, the filled symbols indicate the results to be expected for a typical subject, while the inter-quartile ranges give hints on how well the subject group considered agrees in judging the respective minimum audible angle. Open symbols denote medians and inter-quartile ranges of intra-individual inter-quartile ranges. That way, open symbols indicate the certainty of a typical subject's judgment. The respective inter-quartile ranges are a measure for the agreement in certainty of the subject group.

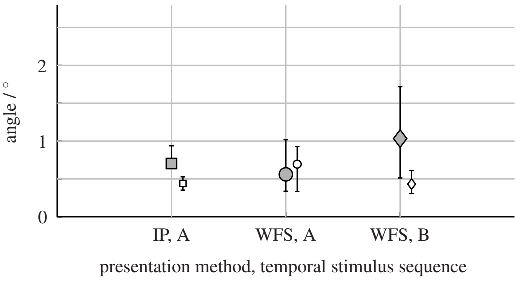

Figure 5: Minimum audible angles assessed with intensity panning (IP) and wave field synthesis (WFS) in dark surroundings with broad band uniform exciting noise. Shown are medians and inter-quartile ranges of intra-individual medians (filled symbols) and of intra-individual inter-quartile ranges (open symbols). Temporal presentation sequences according to figure 3.

The measured MAAs are in good agreement with previous studies, which found MAAs for broadband stimuli of about 1° if the inter-stimulus onset interval exceeded 150 ms, as it is the case in this study (cf. section *Previous work*). When comparing the inter-quartile ranges of the intra-individual medians for IP and WFS (stimulus sequence A), a complete overlap is visible. This fact supports the validity of the procedure and reproduction methods for the assessment of the MAA. The somewhat lower inter-individual inter-quartile ranges for IP as presentation method have to be interpreted with respect to the limited number of discrete phantom sound source positions (cf. sec. *Intensity panning*), which might influence the inter-individual medians too.

The results for the different inter-stimulus onset intervals with WFS indicate a tendency for the MAA to increase for increasing inter-stimulus onset intervals. This increase is not a contradiction to Perrott and Pacheco's results, who found no change in minimum audible angles for inter-stimulus onset intervals larger than 150 ms (cf. section *Previous work*), since they considered only inter-stimulus onset intervals up to 200 ms.

Relation to absolute localization judgments

In section *Intensity panning*, absolute localization judgments for two loudspeakers and for phantom sound sources in between them are shown (cf. figure 1). Based on these results, a demonstrative attempt to deduce information about the auditory system's minimum directional resolution from the absolute localization results for real sound sources is presented. The intra-individual inter-quartile range is a measure for the uncertainty to be expected for one subject. Considering the absolute localization, it would be conceivable to regard the region of uncertainty around a mean auditory event position as the angular region within that sound source locations are not distinguishable. If this explanation holds true and the inter-quartile range is, as a first approximation, assumed to lie symmetrically around the median value, a sound source should become distinguishable from the source corresponding to the considered median if they are separated more than half the inter-quartile range (assuming in addition similar inter-quartile ranges for the results of both sources). This situation in turn equals the situation of two sound sources separated by the MAA. To verify this hypotheses, figure 6 shows again the MAAs measured for IP and WFS (medians and inter-quartile ranges, cf. fig. 5), and for comparison half the mean of the inter-individual median over the intra-individual inter-quartile ranges for the two real loudspeakers from the absolute localization experiment (figure 1), which is assumed to represent the MAA also.

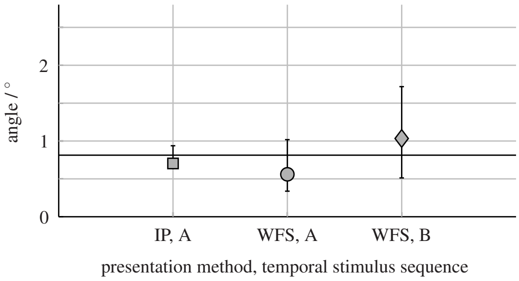

Figure 6: Minimum audible angles from figure 5. The black horizontal line represents half the median of the intra-individual inter-quartile ranges for absolute localization of real loudspeakers (mean over two loudspeakers, cf. figure 1).

While this comparison might not be representative, the results agree quite well. For that reason, at least for the study on hand, half the inter-quartile range of absolute ProDePo localization judgments is a good indicator for the MAA.

VENTRILOQUISM EFFECT

The object of the experiments on the ventriloquism effect was to cause changes in the minimum audible angle by additional visual stimulation. For that reason, the experimental procedure used for the MAA measurement was modified so that the results of both experiments are comparable. Therefore, not only the visual influence's direction but also its magnitude can be assessed. All subjects who took part in the experiments on the MAA for a distinct temporal presentation sequence carried out the corresponding ventriloquism effect experiment too. As with the MAA results, filled symbols indicate inter-individual medians and inter-quartile ranges of intra-individual medians, open symbols inter-individual medians and inter-quartile ranges of intra-individual inter-quartile ranges. The horizontal black line emphasizes the inter-individual median in situation 1 (without visual stimulation) for comparison purposes, since the hypotheses were formulated with respect to situation 1. Single stars indicate a statistically significant difference on a 5%, double stars on a 1% level.

Intensity panning, presentation sequence A

The mean duration for intensity panning resulted to 23.9 minutes. Figure 7 shows the experimental results for the five different spatio-temporal audio-visual stimulus combinations. A one factorial analysis of variance (ANOVA) indicates a significant main effect of the audio-visual stimulus condition $[F(4,36) = 3.75, p = 0.012]$. Post hoc comparison ac. to Bonferroni shows significant differences between situations 1 and 5 as well as 4 and 5.

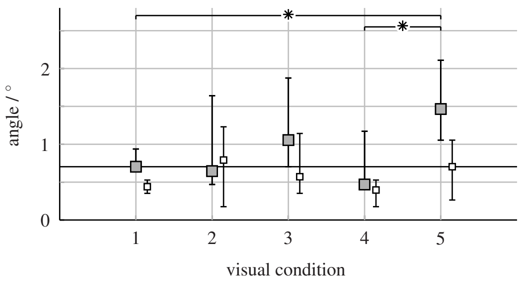

Figure 7: MAAs for different visual conditions (tab. 1), UEN and intensity panning. Medians and inter-quartile ranges of intra-individual medians (filled squares) and inter-quartile ranges (open squares). Temporal sequence A (fig. 3).

Wave field synthesis, presentation sequence A

With WFS, the mean duration was 25.1 minutes. The results for WFS playback are given in figure 8. One factorial ANOVA shows a highly significant main effect of the audio-visual stimulus condition $[F(4,36) = 5.84, p = 0.001]$. Here, significances occur between situations 2 and 5 as well as 3 and 4. A highly significant difference is visible between the conditions 4 and 5.

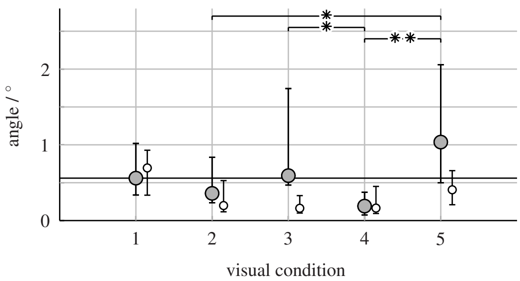

Figure 8: MAAs for different visual conditions (tab. 1), UEN and wave field synthesis. Medians and inter-quartile ranges of intra-individual medians (filled circles) and inter-quartile ranges (open circles). Temporal sequence A (fig. 3).

Wave field synthesis, presentation sequence B

The longer inter-stimulus onset interval in presentation sequence B causes the mean duration in the wave field synthesis situation to increase to 32.4 minutes. Figure 9 shows the corresponding results. ANOVA indicates a highly significant main effect of the stimulus condition $[F(4,12) = 5.63, p = 0.009]$. Situations 2 and 3 as well as 3 and 4 differ significantly.

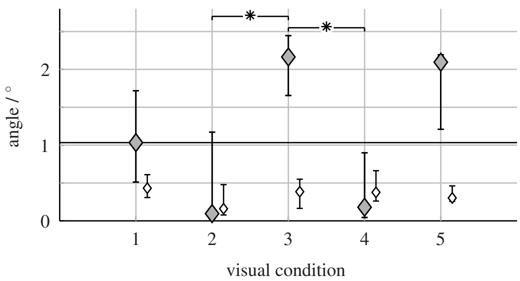

Figure 9: MAAs for different visual conditions (tab. 1), UEN and wave field synthesis. Medians and inter-quartile ranges of intra-individual medians (filled diamonds) and inter-quartile ranges (open diamonds). Temporal sequence B (fig. 3).

Discussion

The main hypotheses formulated in table 1 are: The ventriloquism effect causes a decrease of the minimum audible angle for light and sound shift in the same direction (situations 2 and 4) and an increase of the MAA for contradictory shifts (situations 3 and 5). Considering only the tendencies indicated by the medians of the intra-individual medians, these main hypotheses are confirmed for both presentation methods. All occurring significant effects support the hypotheses, even if not all differences are significant.

Two factorial ANOVA with the factors *visual condition* and *playback method* indicates a highly significant interaction between the factors $[F(4,36) = 11.00, p < 0.0001]$. For a more detailed inspection of the interactions, for temporal sequence A, the difference between visual condition 1 and each of the visual conditions 2 to 5 are shown for intensity panning (squares) and wave field synthesis (circles) in figure 10.

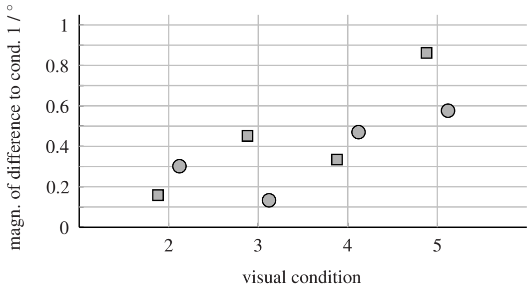

Figure 10: Absolute magnitude of the difference between the medians of the minimum audible angle without visual stimulation (condition 1) and the MAA with visual stimulation for different conditions. Temporal sequence A, IP (squares) and WFS (circles).

The magnitude of the change of the minimum audible angle due to the ventriloquism effect is somewhat larger for WFS than intensity panning in the visual conditions 2 and 4, that is when the light point moves in the same direction as the sound. In these cases, a decrease of the MAA is expected, since the ventriloquism effect would enlarge the distance between the auditory events. For IP, the ventriloquism effect's influence is clearly larger than for WFS in the visual conditions 3 and 5, which means the light and sound sources move to different directions. For these conditions, an increase of the MAA is expected. This antipodal behavior of the ventriloquism effect's magnitude for the two reproduction methods might indicate different strengths of the effect. Further, visual conditions 5 differs significantly from the situation without visual stimulation when using IP, whereas with WFS, only tendencies are visible. For the WFS presentation A, a group of six subjects shows no influence of the ventriloquism effect, while for IP all subjects show influence of the ventriloquism effect on the MAA. These facts can be regarded as additional indication for the auditory event positions in WFS to be more stable than those resulting from intensity panning. In sum, the highly significant interaction between the factors *playback method* and *visual conditions* can be regarded as confirmation of Seeber's hypothesis of the ventriloquism effect's influence to be dependent on the degree of accuracy of directional information.

A comparison of the results of WFS with presentation sequences A (fig. 8) and B (fig. 9) indicates a dependence of the influence of the ventriloquism effect on the stimulus sequence. The results for different stimulus sequences differ in magnitude, but still show the same tendencies.

SUMMARY

Significant influence of the ventriloquism effect on the minimum audible angle is found for intensity panning and wave field synthesis. Since the amount of variation is larger for intensity panning, the auditory event positions in wave field synthesis are assumed to be more stable. Further, the temporal presentation sequence can influence the amount of the ventriloquism effect.

ACKNOWLEDGMENTS

Parts of this work were supported by grant FA 140/4 of the Deutsche Forschungsgemeinschaft (DFG). Bose GmbH, Audio AG/RME, and Sound Service GmbH/Samson Tech. Corp. supported the setup of the wave field synthesis system.

REFERENCES

ITU-R BS.775-1, *Multichannel stereophonic sound system with and without accompanying picture, International Telecommunications Union, Geneva, Switzerland* (1992)

Aoki S., H. Miyata, K. Sugiyama: Stero Reproduction with Good Localization over a Wide Listening Area. J. Audio Eng. Soc. **38**, 433–439 (1990)

Bech S.: Calibration of Relative Level Differences of a Domestic Multichannel Sound Reproduction System. J. Audio Eng. Soc. **46**, 304–313 (1998)

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

Blauert J.: *Spatial Hearing. The Psychophysics of Human Sound Localization*. Revised ed. (The MIT Press, Cambridge, Massachusetts, London, England, 1997)

Blauert J., J. Braasch: *Chapter* Räumliches Hören (Spatial hearing). In *Handbuch der Audiotechnik (Handbook of audio technology)*, ed. by S. Weinzierl. (Springer, 2007)

Cabot R. C.: Sound Localization in 2 and 4 Channel Systems: A Comparison of Phantom Image Prediction Equations and Experimental Data. In *85th Conventino of the Audio Engineering Society* (1977) (Preprint 1295)

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

Gelfand S. A.: *Hearing. An introduction to psychological and physiological acoustics*. Second ed. (Marcel Dekker Inc., New York, 1990)

Hartmann W. M., B. Rakerd: On the minimum audible angle – A decision theory approach. J. Acoust. Soc. Am. **85**, 2031–2041 (1989)

Hellbrück J., W. Ellermeier: *Hören - Physiologie, Psychologie und Pathologie (Hearing - physiology, psychology and pathology)*. 2nd ed. (Hogrefe, 2004)

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

Levitt H.: Transformed Up-Down Methods in Psychoacoustics. J. Acoust. Soc. Am. **49**, 467–477 (1971)

Mills A. W.: On the minimum audible angle. J. Acoust. Soc. Am. **30**, 237–246 (1958)

Perrott D. R.: Auditory and Visual Localization: Two Modalities, One World. In *12th AES International Conference* (1993)

Perrott D. R., K. Marlborough: Minimum audible movement angle: Marking the end points of the path traveled by a moving sound source. J. Acoust. Soc. Am. **85**, 1773–1775 (1989)

Perrott D. R., S. Pacheco: Minimum audible angle thresholds for broadband noise as a function of the delay between the onset of the lead and lag signal. J. Acoust. Soc. Am. **85**, 2669–2672 (1989)

Perrott D. R., K. Saberi: Minimum audible angle thresholds for sources varying in both elevation and azimuth. J. Acoust. Soc. Am. **87**, 1728–1731 (1990)

Queen D.: The Effect of Loudspeaker Radiation Patterns on Stereo Imaging and Clarity. J. Audio Eng. Soc. **27**, 368–379 (1979)

Radeau M.: The after-effects of ventriloquism. Quarterly J. Experim. Psychol. **26**, 63–71 (1974)

Recanzone G. H., S. D. D. R. Makhamra, D. C. Guard: Comparison of relative and absolute sound localization ability in humans. J. Acoust. Soc. Am. **103**, 1085–1097 (1998)

Seeber B.: A New Method for Localization Studies. ACTA ACUSTICA UNITED WITH ACUSTICA **88**, 446–450 (2002a)

Seeber B. U.: Eine neue Messmethode für Lokalisationsuntersuchungen (A new measurement method for localization studies). In *Fortschritte der Akustik, DAGA '01*, 102–103 (Dt. Gesell. für Akustik e. V., Oldenburg, 2001)

Seeber B. U.: Zum Ventriloquismus-Effekt in realer und virtueller Hörumgebung (On the ventriloquism effect in real and virtual listening envriontments). In *Fortschritte der Akustik, DAGA '02* (Dt. Gesell. für Akustik e. V., Oldenburg, 2002b)

Seeber B. U.: *Untersuchung der auditiven Lokalisation mit einer Lichtzeigermethode (Investigation of auditory localization using a light pointer method)*, PhD thesis, Technische Universität München (2003)

Shelton B. R., C. L. Searle: The influence of vision on the absolute identification of sound-source position. Perception & Psychophysics **28**, 589–596 (1980)

Spors S., R. Rabenstein, J. Ahrens: The Theory of Wave Field Synthesis Revisited. In *124th AES Convention* (2008) (Convention Paper 7358)

Theile G., G. Plenge: Localization of Lateral Phantom-Sources. In *53rd AES Convention* (1976) (Preprint B-5)

Vatakis A., C. Spence: Crossmodal binding: Evaluating the "unity assumption" using audiovisual speech stimuli. Perception & Psychophysics **69**, 744–756 (2007)

Völk F.: Externalization in data-based binaural synthesis: effects of impulse response length. In *Proc. of Intern. Conf. on Acoustics NAG/DAGA 2009* (Dt. Gesell. für Akustik e. V., Berlin, 2009)

Völk F.: Psychoakustische Experimente zur Distanz mittels Wellenfeldsynthese erzeugter Hörereignisse (Psychoacoustical experiments on the distance of auditory events created by wave field synthesis). In *Fortschritte der Akustik, DAGA 2010* (Dt. Gesell. für Akustik e. V., Berlin, 2010)

Völk F., F. Heinemann, H. Fastl: Externalization in binaural synthesis: effects of recording environment and measurement procedure. In *Proceedings of Acoustics '08*, 6419–6424 (2008)

Vroomen J., P. Bertelson, B. de Gelder: The ventriloquist effect does not depend on the diection of automatic visual attention. Perception & Psychophysics **63**, 651–659 (2001)

Warncke H.: Die Grundlagen der raumbezüglichen stereophonischen Übertragung im Tonfilm (The basics of room related stereophonic transmission for the sound movie). Akust. Z. **6**, 174–188 (1941)

Welch R. B., D. H. Warren: Immediate perceptual response to intersensory discrepancy. Psychological Bulletin **88**, 638–667 (1980)

Zwicker E., R. Feldtkeller: *Das Ohr als Nachrichtenempfänger (The ear as a communication receiver)*. 2nd ed. (S. Hirzel Verlag Stuttgart, 1967)

Zwicker E., R. Feldtkeller: *The ear as a communication receiver* (English translation published for the Acoustical Society of America through the American Institute of Physics, 1999). – translated by Hannes Müsch, Søren Buus, and Mary Florentine, original German edition "Das Ohr als Nachrichtenempfänger." Second revised edition (1967), S. Hirzel Verlag Stuttgart