Adaptively Adjusting the Stereophonic Sweet Spot to the Listener's Position*

SEBASTIAN MERCHEL, *AES Student Member*, **AND STEPHAN GROTH**
(sebastian.merchel@tu-dresden.de) (stephan.groth@gmx.net)

Department of Communication Acoustics, Dresden University of Technology, 01062 Dresden, Germany

A stereophonic playback system designed to adjust the sweet spot to the listener's position is presented. The system includes an optical face tracker, which provides information about the listener's *x–y* position. The loudspeaker signals are manipulated in real time in order to move the sweet spot. The stereophonic perception with an adjusted sweet spot is theoretically investigated on the basis of several models of binaural hearing. The results indicate that an adjustment of signals corresponding to the center of the listener's head does improve localization over the entire listening area. Nevertheless, some localization error remains because of asymmetric signal paths for off-center listening positions, which can be estimated and compensated for.

0 INTRODUCTION

The spatial reproduction of sound in a conventional stereo system works only over the small area located on the symmetry axis between the loudspeakers—the so-called sweet spot. Beyond this area the spatial perception collapses and the stereo image moves toward the nearer loudspeaker since that signal arrives louder and sooner. Finally the stereo image is located completely in the nearer loudspeaker because of the precedence effect. Different studies have determined the area of stereophonic localization [1]–[3]. The optimal listening area depends on the maximum tolerable shift of the phantom source. Fig. 1(a) shows that the sweet spot is a stretched area rather than a spot. The constriction of the region in which correct audio localization is possible is one major disadvantage of stereophony.

In an attempt to reproduce correct auditive localization over a larger listening area, different playback methods have been developed (such as Ambisonics and WFS). In most cases new recording techniques are needed and the complexity of the reproduction system increases rapidly. However, stereophony is used widely and many stereophonic recordings are available. Releasing the stereophonic listener from its static hearing position would be a decisive advantage.

1 BROADENING THE SWEET SPOT

Present methods to broaden the sweet spot can be separated into two groups—those who try to adjust the radiation pattern of the loudspeakers and those who adjust the signals of the loudspeakers directly. Methods that remove localization information or which primarily increase localization blur will not be discussed here.

1.1 Adjustment of Loudspeaker Directivity

A number of studies exist that deal with broadening the area of stereophonic perception by adjusting the loudspeaker characteristics. In one of the early studies Bauer [3] explains that the level difference between two loudspeakers in a listening point is dependent on the room characteristics and the directivity of the loudspeakers. For monopoles there is only a small area close to the symmetry axis in which the level difference is less than 3 dB. Bauer proposes a system where the angle between the loudspeaker axes is approximately 120–130°. Frequencies above 250 Hz should be radiated using dipoles. In such a system the level difference between the loudspeakers remains almost constant over a wider area, as can be seen in Fig. 1(b).

An example of determining the optimal loudspeaker directivity using listening tests is given in Aarts [4]. Further publications dealing with loudspeaker directivity are [5]–[7].

Problematic with all methods mentioned is the frequency dependence of the loudspeaker radiation pattern, which cannot be adjusted arbitrarily. In addition all those studies do not address the real problem of the precedence effect, which is especially dominant for transient signals such as speech or music. Furthermore there are contradicting localization cues from interaural level and time differences, resulting in localization blur for excentric listening positions [3].

1.2 Adjustment of Loudspeaker Signals

Another approach is the direct adjustment of the loudspeaker signals. Beside the level differences, the

*Presented at the 126th Convention of the Audio Engineering Society, Munich, Germany, 2009 May 7–10 under the title "Analysis and Implementation of a Stereophonic Playback System for Adjusting the "Sweet Spot" to the Listener's Position"; revised 2010 May 5.

delay between the loudspeaker signals, caused by different distances from the listener's position is responsible for a localization shift. Several methods try to adjust the delay for a specific listening position so that the loudspeaker signals will reach the head of the listener at the same time. In many hi-fi systems the delay between loudspeakers can be adjusted manually. Others have automated this process using a measuring microphone. Aoki et al. [8] described an interesting system that uses groups of delayed directional loudspeakers. In all cases the systems are designed for static positions and do not respond to movements of the listener.

Another idea to estimate the location of the listener is stated in Kim et al. [9]. The position of the remote control is used to reproduce binaurally rendered surround signals via loudspeakers using a crosstalk system. However, the system is not designed for stereophonic reproduction and continuous adjustment of the sweet spot. Kyriakakis et al. [10] were the first to describe a system that uses head tracking and time delay between loudspeaker signals to move the sweet spot. Unfortunately there are no publications about emerging artifacts and the usefulness of time delay and amplitude adjustment in off-center listening.

2 ADAPTIVE SIGNAL ADJUSTMENT

The present work presents a playback system that manipulates the loudspeaker signals in real time, depending on the listener's position. The x–y position of the listener is tracked by a camera, and the delay is calculated in such a manner that the signals of both loudspeakers arrive at the center of the listener's head at the exact same time. In addition the amplitudes of the loudspeaker signals are adjusted to reduce the level difference at the listening position. As can be seen in Fig. 2 the signal paths from loudspeakers to ears become asymmetrical for off-center listening positions. For the situation in Fig. 2 the signal at the left ear originating from the right loudspeaker p_{RL} will be more attenuated because of stronger head shadowing than the signal at the right ear originating from the left loudspeaker, p_{LR}. The arrival time difference τ_R between the signal at the right ear originating from the right loudspeaker p_{RR} and the signal at the left ear originating from the right loudspeaker p_{RL} will be bigger than the arrival time difference τ_L between p_{LL} and p_{LR} ($\tau_L < \tau_R$). This asymmetry is important for correct off-center localization.

The effect can be illustrated with an impulse phantom source positioned at the center between the loudspeakers. The two loudspeakers emit identical impulses, which are adjusted to reach the center of the head at the exact same time with the same amplitude. Fig. 3 shows the resulting signals at the left and right ears for

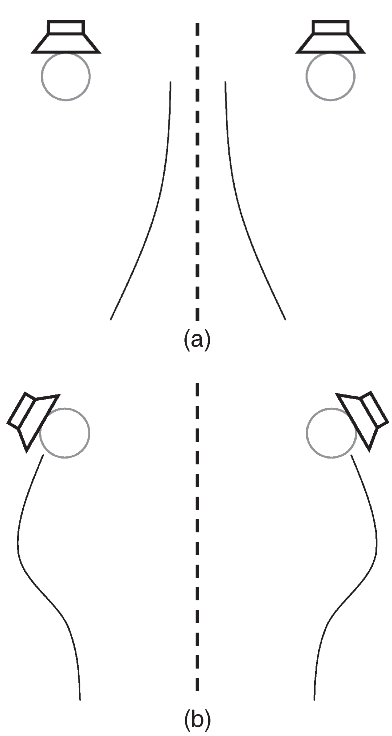

Fig. 1. Stereophonic listening area. Level difference between loudspeaker signals inside marked area is less than 3 dB ($f > 250$ Hz, dipole characteristic). After Bauer [3]. (a) Stereo setup with parallel aligned loudspeakers. (b) Stereo setup with rotated loudspeaker axes.

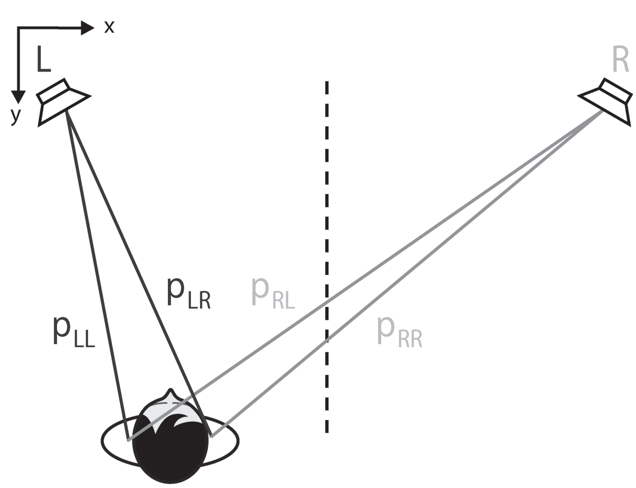

Fig. 2. Geometric layout of excentric listening position.

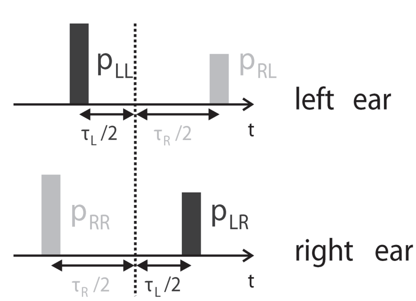

Fig. 3. Ear signals for transient stimuli. Loudspeaker signals are adjusted to reach the center of the head at exactly the same time ($t = 0$).

the off-center listening position, as depicted in Fig. 2. The dashed line represents the time when both signals in theory reach the center of the head. It can be seen that the sum signal at the left ear is delayed to the sum signal at the right ear. In addition the left ear signal might be louder than the signal at the right ear because of head shadowing. This leads to a phantom source shift to the right. From the listener's point of view this is toward the center, where ideally the phantom source should be localized. This thought experiment helps to understand the usefulness of adaptive signal adjustment. The following section analyzes the effect from a quantitative point of view.

3 ANALYSIS OF AUDIO LOCALIZATION WITH ADAPTIVE SIGNAL ADJUSTMENT

Different binaural models are used to study the utility of the system. As a first approximation an analytical approach by Lipshitz [11] is used. This approach analyzes the superimposed signals at the listener's ears if several sources emit low-frequency sine waves. The emerged phase difference is converted into an interaural time difference (ITD) and finally into a corresponding azimuth angle. In a more advanced approach a binaural model after Braasch [12], [13] is used. This model simulates the outer ear, the inner ear, and a decision device. The pathway to the ear is simulated by measured head-related transfer functions (HRTFs) for several angles. The decision device performs a cross-correlation analysis in several frequency bands and finally provides the ITD with maximum likelihood, which can be converted into an azimuth angle as well. This model can also handle broadband input signals such as bandpass noise.

3.1 Analytical Approach

To describe the pressure at the ears mathematically, the assumption of sinusoidal signals with low frequencies is useful. Thus head shadowing effects can be neglected. The signals at the loudspeakers can have an amplitude ratio of L/R and a time difference τ_l. It is assumed that delay and level differences due to different distances between loudspeakers and listener position are compensated for by using the method of signal adjustment as described above. The resulting signals from the left and right loudspeakers at the left ear are

$$p_{\mathrm{LL}} = L \exp\left(\mathrm{j}\omega \frac{\tau_l + \tau_L}{2}\right)$$

$$p_{\mathrm{RL}} = R \exp\left(-\mathrm{j}\omega \frac{\tau_l + \tau_R}{2}\right).$$

They can be superimposed onto the sum signal at the left ear,

$$\begin{aligned} p_L &= p_{\mathrm{LL}} + p_{\mathrm{RL}} \\ &= L\cos\left(\mathrm{j}\omega\frac{\tau_l+\tau_L}{2}\right) + \mathrm{j}L\sin\left(\mathrm{j}\omega\frac{\tau_l+\tau_L}{2}\right) \\ &\quad + R\cos\left(\mathrm{j}\omega\frac{\tau_l+\tau_R}{2}\right) - \mathrm{j}R\sin\left(\mathrm{j}\omega\frac{\tau_l+\tau_R}{2}\right). \end{aligned}$$

The resulting signals from the left and right loudspeakers at the right ear are

$$p_{\mathrm{LR}} = L \exp\left(\mathrm{j}\omega \frac{\tau_l - \tau_L}{2}\right)$$

$$p_{\mathrm{RR}} = R \exp\left(-\mathrm{j}\omega \frac{\tau_l - \tau_R}{2}\right).$$

They can be superimposed onto the sum signal at the right ear,

$$\begin{aligned} p_R &= p_{\mathrm{LR}} + p_{\mathrm{RR}} \\ &= L\cos\left(\mathrm{j}\omega\frac{\tau_l-\tau_L}{2}\right) + \mathrm{j}L\sin\left(\mathrm{j}\omega\frac{\tau_l-\tau_L}{2}\right) \\ &\quad + R\cos\left(\mathrm{j}\omega\frac{\tau_l-\tau_R}{2}\right) - \mathrm{j}R\sin\left(\mathrm{j}\omega\frac{\tau_l-\tau_R}{2}\right). \end{aligned}$$

The absolute value and the phase of the interaural transfer function $A(\omega) = p_L/p_R$ are related to ILD and ITD, which can be transferred into a localization angle.

3.1.1 Modeling a Center Source

If this model is used to simulate a center phantom source, the left and right loudspeaker get the same input ($L = R$, $\tau_l = 0$). Thus the absolute value of the interaural transfer function becomes

$$|A(\omega)| = 1$$

and the phase can be written as

$$\arg[A(\omega)] = \frac{\tau_L - \tau_R}{2}\omega.$$

It can be seen that, under the assumptions given, the ILD at both ears is independent of the listening position. The ITD depends on τ_L and τ_R and, hence, the listening position. Using the ITD a modeled localization angle Φ can be estimated. In Fig. 4 this modeled localization angle

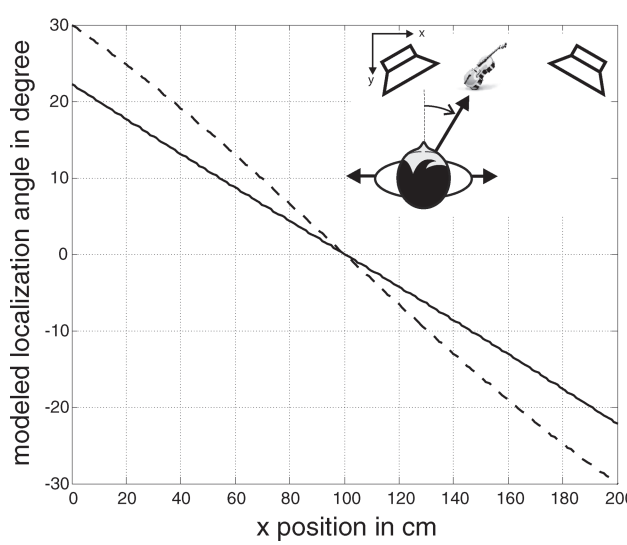

Fig. 4. Localization of center phantom source with signal adjustment versus x position of listener (y = 1.73 m, loudspeaker distance 2 m). —— perceived angle to phantom source (which should be at center), – – – actual angle to center. Perceived angle is modeled using the analytic approach.

Φ is compared with the actual target angle to the center in between the loudspeakers. A standard stereo setup with a loudspeaker distance of 2 m is used. The listening position is moved in the x direction with constant $y = 1.73$ m. The orientation of the listener is straight ahead, as was shown in Fig. 2. It can be seen that the angles match with a maximum deviation of 8° in the extreme positions in front of the left and right loudspeakers ($x = 0$ m and $x = 2$ m). This means that the perceived position of the center phantom source remains almost constant, with a slight shift in the direction of the listener's movement.

3.2 Binaural Model

To investigate more complex signals and effects such as head shadowing and reflections on head and torso, a binaural model after Braasch [12], [13] was adapted and implemented in MATLAB. The general model structure is shown in Fig. 5. It consists of two sources which generate the left and right channel of a stereophonic system. The outer and inner ear is modeled, and binaural localization cues (ITDs and ILDs) are estimated and analyzed. The corresponding localization angle was found with a remapping algorithm using measured HRTFs. A detailed description of the system can be found in [15]. The interaction of ITDs and ILDs in the localization process is still the object of research. Thus resulting localization angles will be examined separately. This paper describes the results for modeled localization angles using ITD cues. The results using ILD cues are similar. A discussion can be found in [14]–[16].

3.2.1 Modeling of a Center Source

As described above for a center source both loudspeakers are driven with identical signals ($L = R$, $\tau_l = 0$). Three different noise stimuli are used with bandwidth limitation, as shown in Table 1.

The modeled localization angle using ITDs is compared in Fig. 6 to the actual target angle to the center source. Again the standard setup is used as described. Fig. 6(a) shows the modeled localization angle without signal adjustment. The perceived position of the phantom source is shifted rapidly in the direction of movement until the precedence effect takes effect. Moving further to the left or right, the source is localized in the nearer loudspeaker. The absolute value of the localization angle decreases until the loudspeaker is straight ahead ($\Phi = 0°$ at $x = 0$ m and $x = 2$ m). The localization error (the difference between perceived angle to phantom source and actual angle to center) remains constant because the absolute value of the reference angle to the center increases. In Fig. 6(b) the perception with signal adjustment is modeled. The results from the analytical model are confirmed. The localization of a center source remains stable if the listener is moving in the x direction (e.g., from left to right), although the

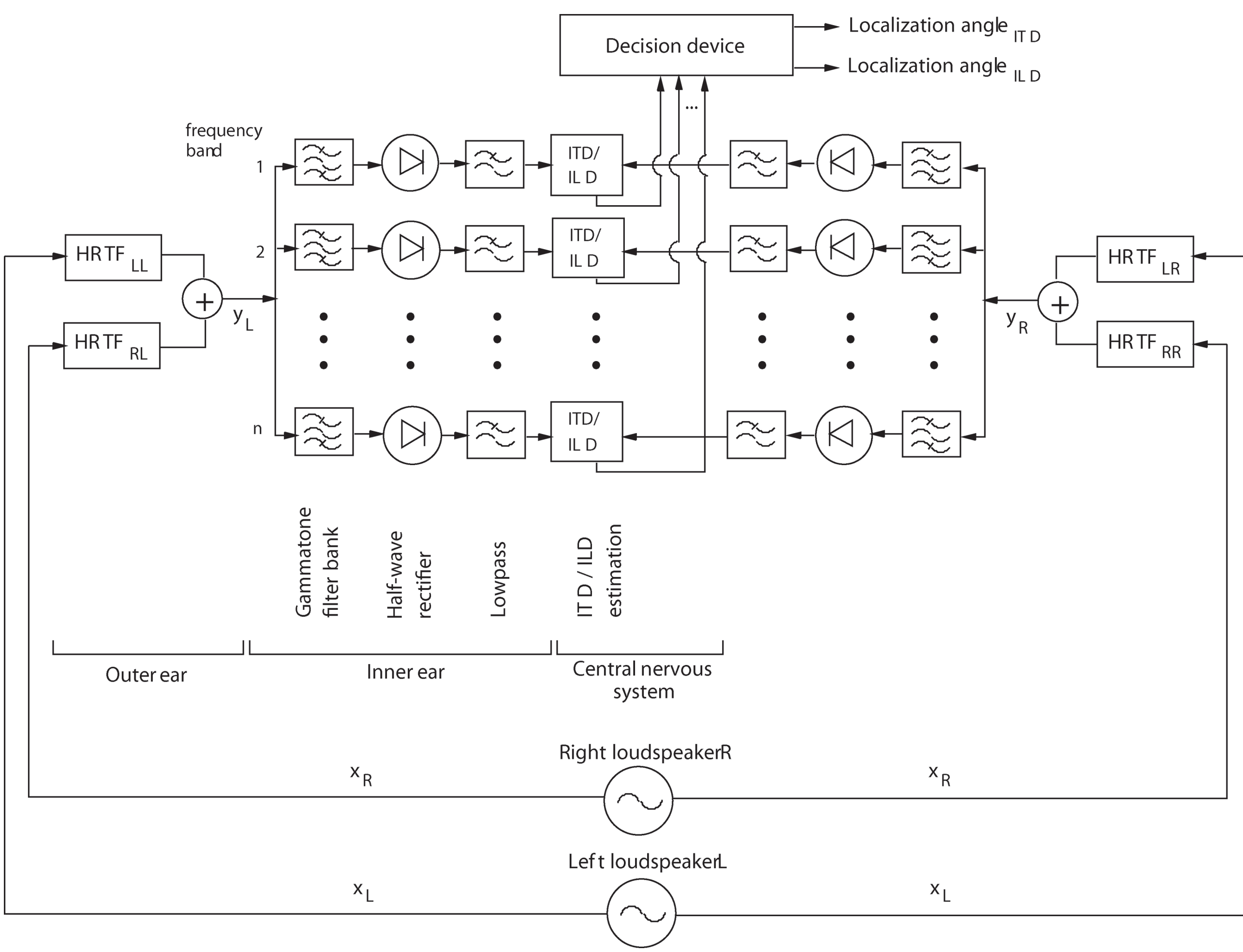

Fig. 5. Adapted binaural model for phantom source localization in a stereophonic setup. After Braasch [12], [13].

same image shift in the direction of movement can be seen. Again a small error of approximately 10° is found for extreme listening positions.

The improvement of the system with signal adjustment can be illustrated better using "quality" maps. They show the absolute value of the difference between target angle to phantom source and modeled localization angle. Fig. 7 shows the resulting error for a center phantom source with and without signal adjustment. Fig. 7(a) maps the sweet spot without signal adjustment, similar to Keibs [1] and Gaal [2]. It can be seen that modeling the precedence effect results in a considerably smaller listening area. Fig. 7(b) depicts the effectiveness of adaptive signal adjustment in an expanded listening area.

Even more intuitive is the visualization of the modeled localization angle using "vector" maps. In Fig. 8 the

Table 1. Stimuli used for binaural model.*

Stimulus	f_1	f_2
1	300 Hz	1300 Hz
2	300 Hz	2300 Hz
3†	300 Hz	4500 Hz

*Bandpass white noise with lower (f_1) and upper (f_2) frequency limits.
†Stimulus 3 covers the relevant frequency range for speech and music.

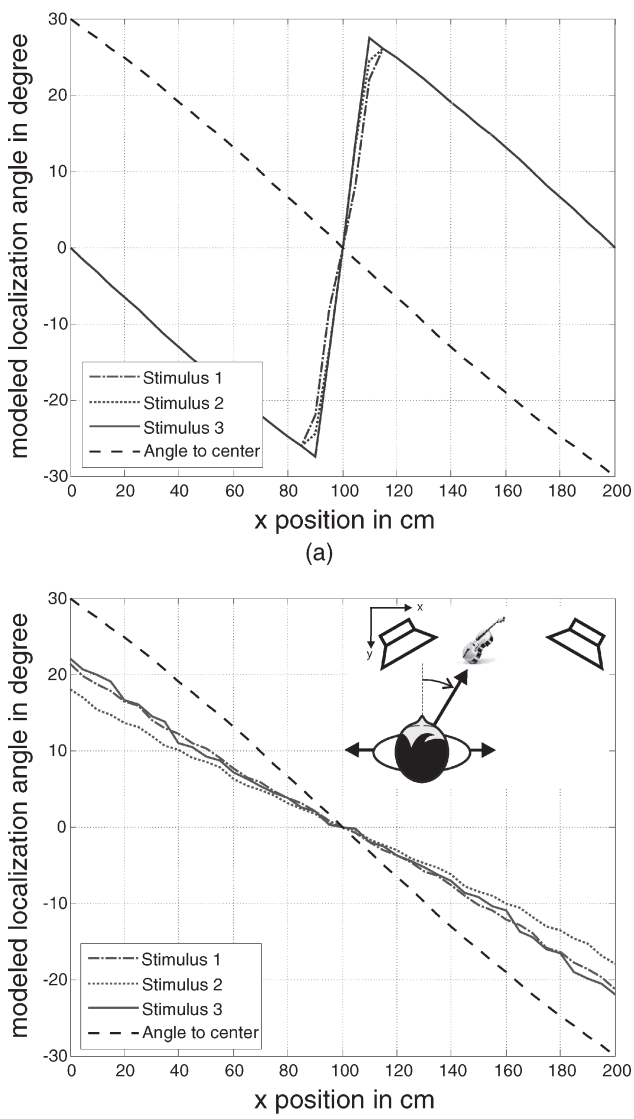

Fig. 6. Localization of center phantom source versus x position of listener ($y = 1.73$ m, loudspeaker distance 2 m). Perceived angle is estimated using the binaural model. Plot shows perceived angle to phantom source for three different stimuli and actual angle to center (dashed). (a) Without signal adjustment. (b) With signal adjustment.

modeled azimuth angle is plotted as a direction vector for a center source with and without signal adjustment. Fig. 8(a) shows the direction of the localized source depending on the listener position. If the signals are adjusted, the perceived source position remains almost stable, as can be seen in Fig. 8(b). It can be shown that this is also true for off-center phantom source positions using time or intensity stereo [16].

Both models lead to the same conclusion. They show that an adjusted sweet spot using interchannel time delay and amplitude adjustment improves the localization over the entire off-center listening area. The asymmetric signal paths for off-center listening positions are important to maintain correct stereophonic localization, although some localization error remains.

3.3 Remaining Error

It has been observed in listening tests that the localization of amplitude-panned virtual sources move to some degree as the head is turned [17]. These

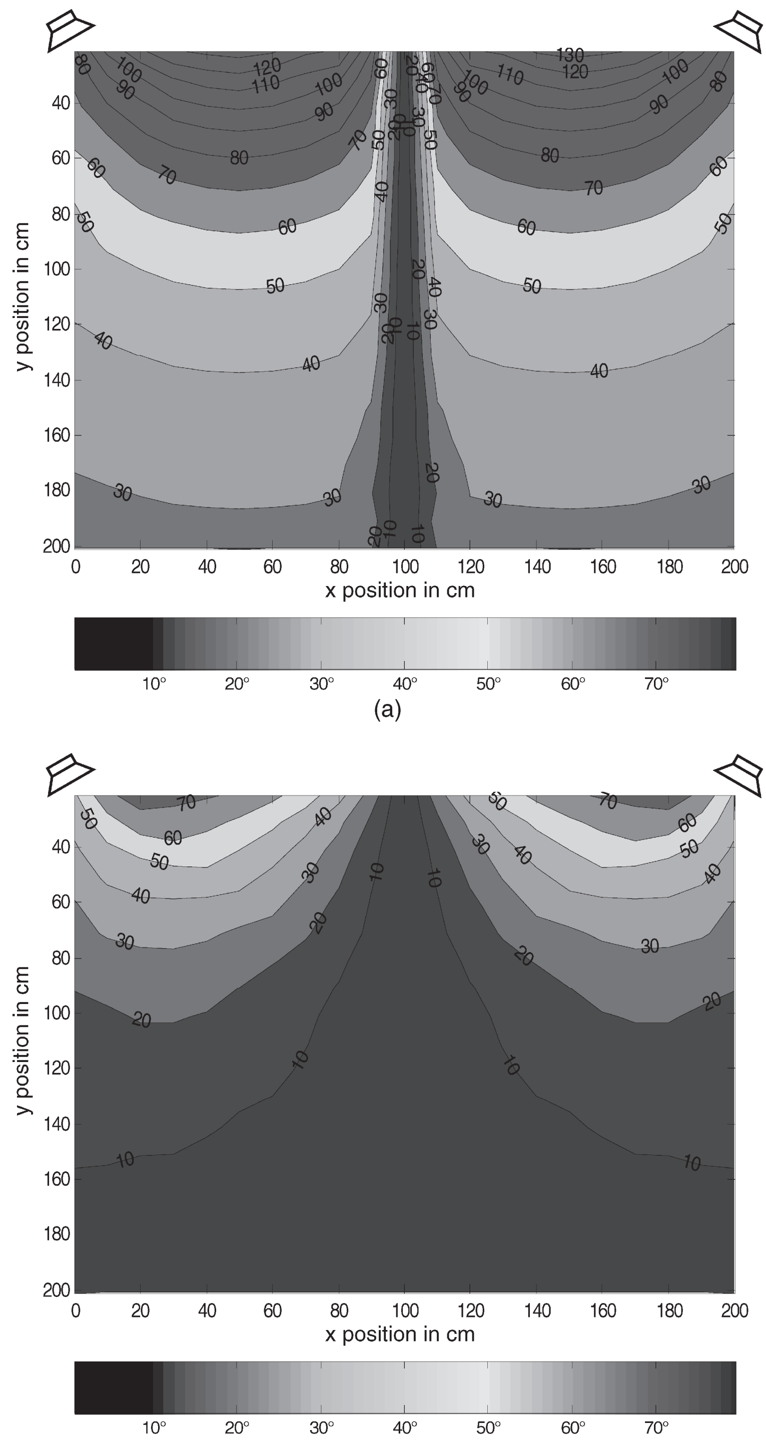

Fig. 7. "Quality" maps showing absolute value of difference between modeled localization angle using ITDs and target angle to center phantom source. (a) Narrow sweet spot without signal adjustment. (b) Large area with correct localization using signal adjustment.

observations are consistent with the results modeled here. Off-center listening with signal adjustment can be compared to head rotation in the standard sweet spot.

The small remaining error is dependent on the distance between the phantom source and the median plane of the listener. This is illustrated in Fig. 9. The further the phantom source is away from the viewing direction of the listener, the larger the error. This remaining error can be estimated and compensated using a binaural model or an averaged compensation function [18].

4 IMPLEMENTATION—SWEETSPOTTER

The system was implemented on a PC using C++. The resulting program SweetSpotter is able to replay stereophonic recordings while adaptively adjusting delay and amplitude between the loudspeakers (see Fig. 10). Thus the position of the sweet spot is updated at least every 100 ms. The spatial precision of the tracking system does not need to be highly accurate because the sweet spot is a small area rather than a point. To be suitable for daily use the tracking system must not use markers. Therefore a camera-based head tracker was used. Our current system runs in real time on a single laptop using an integrated camera. The practicability of the system was confirmed through informal listening tests.

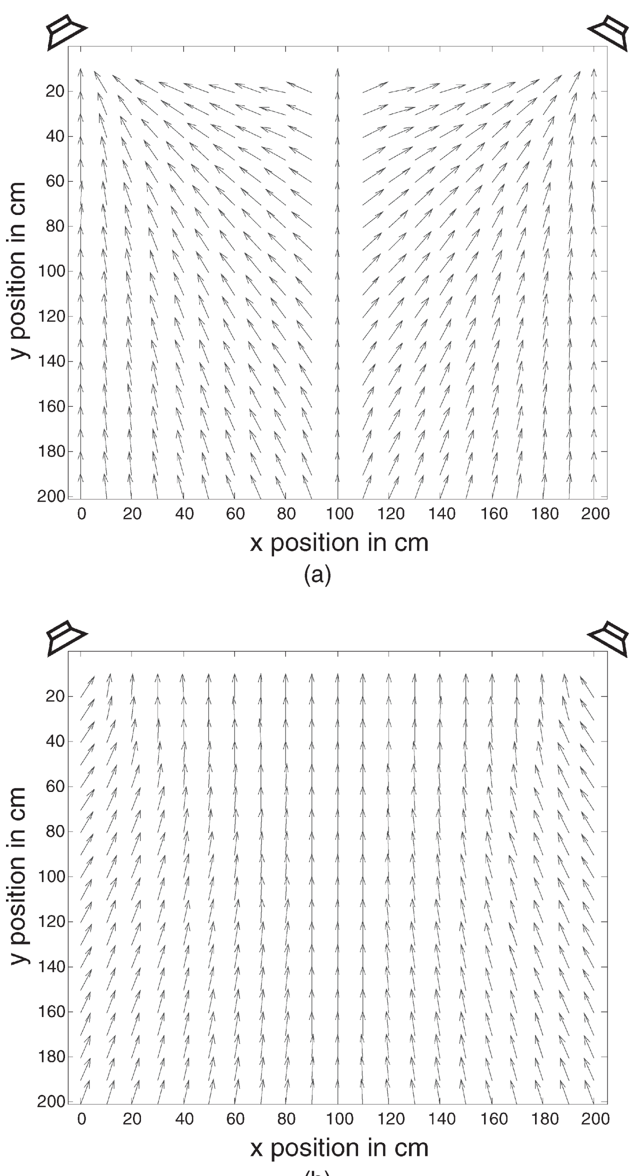

Fig. 8. "Vector" maps showing modeled localization angle of center phantom source using ITDs. (a) Without signal adjustment. Precedence effect distorts correct phantom source localization. (b) With signal adjustment. Localization is maintained over a wide listening area.

The area of application ranges from audio reproduction with desktop computers to teleconference systems or virtual realities. More and more television sets are replaced by multimedia computers with built-in cameras. Direct integration in loudspeakers and hi-fi systems is also possible. Because the system works only for a single person, it has to switch off automatically if multiple persons are entering the listening area. Alternatively, it is possible to adjust to the average position of the group.

5 CONCLUSION

The stereophonic perception in a system with adjusted sweet spot was theoretically investigated on the basis of an analytical model and an advanced binaural model. Both approaches indicate that an adaptive adjustment of the signals relative to the center of the listener's head improves the localization over the entire listening area. The remaining error can be estimated using the models described. A major advantage of adaptive signal adjustment is the compatibility with existing source material and equipment. It can be concluded that stereophonic multichannel reproduction remains an interesting topic and can be further improved using the methods described.

6 ACKNOWLEDGMENT

The authors wish to thank U. Jekosch, J. Blauert, and E. Altinsoy for supervision and informative discussions.

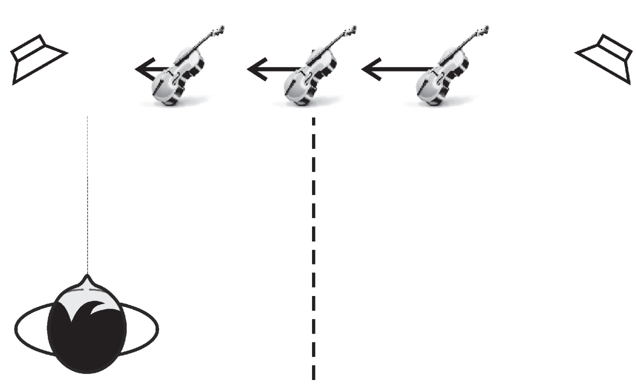

Fig. 9. Phantom source shift with signal adjustment for different phantom source positions. (This remaining error increases as the phantom source position deviates from the median plane of the listener.)

Thanks are also due to Tobias Pietzsch and Lars Beier for support with the implementation.

7 REFERENCES

[1] L. Keibs, "Perspektiven für eine raumbezogene Rundfunkübertragung," *Gravesaner Blätt.*, no. 22, pp. 2–40 (1961).

[2] D. Ga'al, "Calculation of the Stereophonic Localization Area," presented at the 53rd Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 24, p. 303 (1976 May), preprint B-4.

[3] B. Bauer, "Broadening the Area of Stereophonic Perception," *J. Audio Eng. Soc.*, vol. 8, pp. 91–94 (1960).

[4] R. M. Aarts, "Enlarging the Sweet Spot for Stereophony by Time/Intensity Trading," presented at the 94th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 41, p. 387 (1993 May), preprint 3473.

[5] J. M. Kates, "Optimum Loudspeaker Directional Patterns," *J. Audio Eng. Soc.*, vol. 28, pp. 787–794 (1980 Nov.).

[6] D. B. Keele, "A Loudspeaker Horn that Covers a Flat Rectangular Area from an Oblique Angle," presented at the 74th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 31, p. 964 (1983 Dec.), preprint 2052.

[7] M. F. Davis, "Loudspeaker Systems with Optimized Wide-Listening-Area Imaging," *J. Audio Eng. Soc.*, vol. 35, pp. 888–896 (1987 Nov.).

[8] S. Aoki, H. Miyata, and K. Sugiyama, "Stereo Reproduction with Good Localization over a Wide Listening Area," *J. Audio Eng. Soc.*, vol. 38, pp. 433–439 (1990 June).

[9] S. Kim, D. Kong, and S. Jang, "Adaptive Virtual Surround Sound Rendering System for an Arbitrary Listening Position," *J. Audio Eng. Soc.*, vol. 56, pp. 243–254 (2008 Apr.).

[10] C. Kyriakakis et al., "Signal Processing, Acoustics, and Psychoacoustics for High Quality Desktop Audio," *J. Vis. Commun. Image Represent.*, vol. 9, pp. 51–61 (1998).

[11] S. P. Lipshitz, "Stereo Microphone Techniques: Are the Purists Wrong?," *J. Audio Eng. Soc.*, vol. 34, pp. 716–744 (1986 Sept.).

[12] J. Braasch, "Modelling of Binaural Hearing," J. Blauert Ed., in *Communication Acoustics* (Springer, Berlin, 2005), pp. 75–108.

[13] J. Braasch, "Auditive Lokalisation und Detektion in Mehrschallquellen-Situationen," dissertation (VDI Verlag, Düsseldorf, Germany, 2002).

[14] S. Merchel and S. Groth, "Adaptive Adjustment of the "Sweet Spot" to the Listener's Position in a

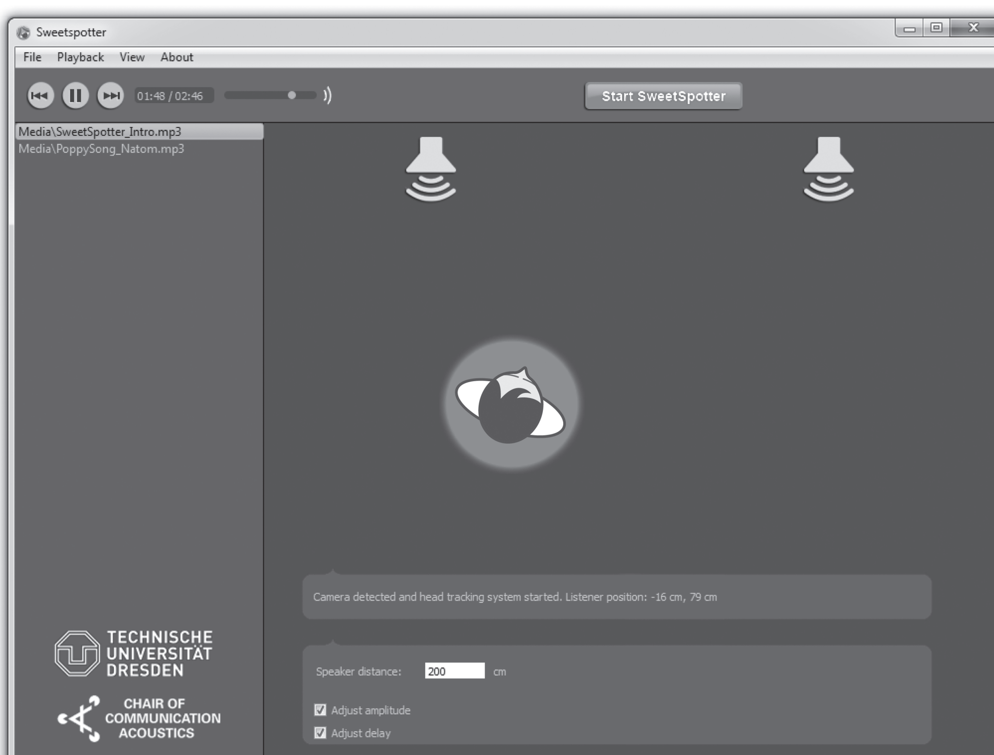

Fig. 10. SweetSpotter is a real-time C++ implementation running on Windows. Listener position is tracked using a camera and a face recognition algorithm. Delay and level are adjusted accordingly. Software is open source and can be downloaded at www.sweetspotter.de.

Stereophonic Play Back System," in *Proc. NAG/DAGA* (Rotterdam, The Netherlands, 2009).

[15] S. Groth, "Untersuchung eines Stereo-Systems mit Signalanpassung an die Hörposition," diploma thesis, Institute of Acoustics and Speech Communication, Technical University Dresden, Dresden, Germany (2008).

[16] S. Merchel and S. Groth, "Evaluation of a New Stereophonic Reproduction Method with Moving "Sweet Spot" Using a Binaural Model," in *Proc. ISAAR* (Helsingør, Denmark, 2009).

[17] V. Pulkki, "Compensating Displacement of Amplitude-Panned Virtual Sources," in *Proc. of AES 22nd Int. Conf* (Espoo, Finland, 2002, June 15–17).

[18] S. Merchel and S. Groth, "Automatische Anpassung des stereophonen Sweetspots bei Kopf-drehung," in *Proc. NAG/DAGA* (Berlin, Germany, 2010).

THE AUTHORS

S. Merchel

S. Groth

Sebastian Merchel was born in Grimma, Germany, in 1980. He received a Diploma degree in electrical engineering from the Dresden University of Technology, Dresden, Germany, in 2006. During his studies he visited Aalborg University in Denmark and participated in the "Master of Science Programme in Acoustics."

He is presently a research associate at the Institute of Acoustics and Speech Communication, Chair of Communication Acoustics, Dresden University of Technology. His research interests include audio–tactile perception, spatial audio, and digital signal processing. He is also interested in multimodal user interfaces and academic knowledge transfer.

He is a student member of the Audio Engineering Society, DEGA (German Acoustical Society), and VDE (Association for Electrical, Electronic and Information Technologies).

●

Stephan Groth was born in Mittweida, Germany, in 1983. He received a Diploma degree in mechatronics from the Dresden University of Technology, Dresden, Germany, in 2008. His thesis was related to spatial hearing in stereophonic playback systems. He is presently an acoustical engineer with Akustik Bureau Dresden and also a self-employed musician. He has specialized in room and building acoustics together with digital signal processing and digital audio. Moreover he is an expert in mediaeval and traditional music and instruments.