Audio Engineering Society

Convention Paper

Presented at the 126th Convention
2009 May 7–10 Munich, Germany

Analysis and Implementation of a Stereophonic Play Back System for Adjusting the "Sweet Spot" to the Listener's Position

Sebastian Merchel[1] and Stephan Groth[1]

[1] *Chair of Communication Acoustics, Dresden University of Technology, 01062 Dresden, Germany*

Correspondence should be addressed to Sebastian Merchel (`sebastian.merchel@tu-dresden.de`)

ABSTRACT

This paper focuses on a stereophonic play back system designed to adjust the "sweet spot" to the listener's position. The system includes an optical face tracker which provides information about the listener's x-y position. Accordingly, 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 the localization over the whole listening area. Although some localization error remains due to asymmetric signal paths for off-center listening positions, which can be estimated and compensated for.

1. INTRODUCTION

The spatial reproduction of sound in a conventional stereo system works only in a small area which is 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 to the nearer loudspeaker since the signal arrives both louder and sooner. Finally, the stereo image is completely located in the nearer loudspeaker due to the precedence effect. Different studies have determined the area of stereophonic localization [1, 2, 3]. The optimal listening area depends on the maximum tolerable shift of the phantom source. In Figure 1a it can be seen that the "sweet spot" is rather a stretched area than a spot. The constriction of the region in which correct audio

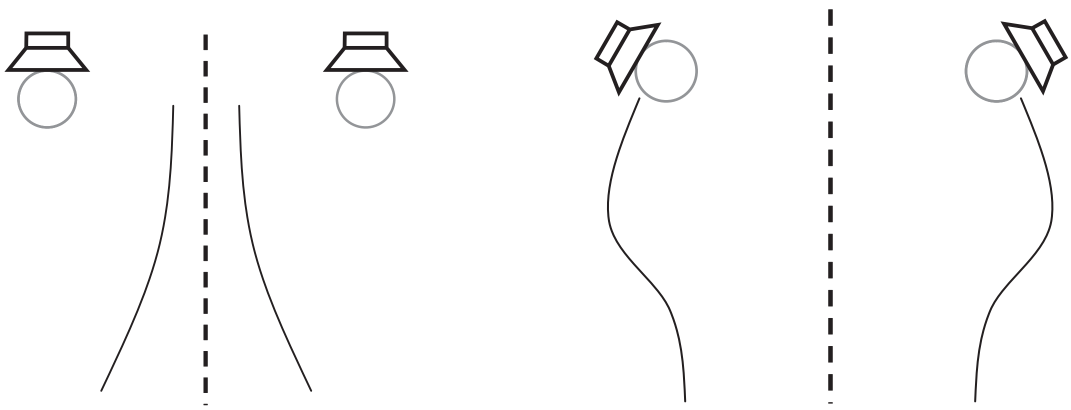


Fig. 1: Stereophonic listening area after Bauer [3]. The level difference between the loudspeaker signals inside the marked area is smaller than 3 dB ($f > 250$ Hz, dipole characteristic). **(a)** Stereo setup with parallel aligned loudspeakers. **(b)** Stereo setup with rotated loudspeaker axes.

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 (e.g. Ambisonics or WFS). In most cases new recording techniques are needed and the complexity of the reproduction system increases rapidly. On the contrary stereophony is widely used and many stereophonic recordings are available. Releasing the stereophonic listener from its static hearing position would be a decisive advantage.

2. 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 primarily increase localization blur will not be discussed here.

2.1. Adjustment of loudspeaker directivity

A number of papers exist which deal with broadening the area of stereophonic perception by adjusting loudspeaker characteristics. One of the early studies was published by Bauer [3]. He explains that the level difference between two loudspeakers in a listening point is dependent on 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 smaller than 3 dB. Bauer proposes a system, where the angel between the loudspeaker axes is approximately 120° to 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 Figure 1b.

An example of determining the optimal loudspeaker directivity using listening tests is Aarts [4]. Further publications dealing with loudspeaker directivity are [5, 6, 7].

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

2.2. Adjustment of loudspeaker signals

Another approach is the direct adjustment of the loudspeaker signals. Beside the level differences, the delay between the loudspeaker signals, due to different distances to the listener's position, is responsible for localization shift. The following methods try to adjust the delay for a specific listening position so that the speaker signals 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 [8] describes an interesting systems which uses groups of delayed directional loudspeakers. In all cases the systems are designed for static positions and do not respond to movement 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. The system however is not designed for stereophonic reproduction and continuous adjustment of the "sweet spot". For the first time Kyriakakis [10] described a system which uses head tracking and time delay between the loudspeaker signals to move the "sweet spot". Unfortunately, there are no publications about the usefulness of time delay and amplitude adjustment in off-center listening and emerging artifacts.

3. ADAPTIVE SIGNAL ADJUSTMENT

The present work presents a play back system which manipulates the loudspeaker signals depending on the listener's position in real-time. Therefore, the x-y position of the listener is tracked by a camera. 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 Figure 2 the signal paths from loudspeakers to ears become asymmetrical for off-center listening positions. For the situation in Figure 2 the signal at the left ear originating from the right loudspeaker (p_{RL}) will be more attenuated due to 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.

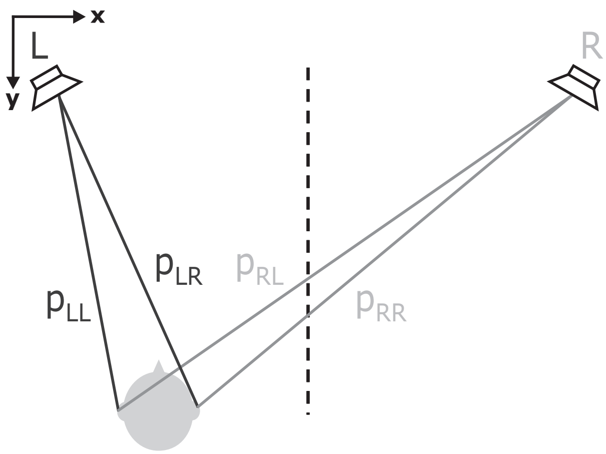


Fig. 2: Geometrical situation for excentric listening position.

The effect can be illustrated with an impulse phantom source positioned at the center between the loudspeakers. The two loudspeaker emit identical impulses, which are adjusted to reach the center of the head at the exact same time with the same amplitude. Figure 3 shows qualitative the resulting signals at the left and right ear for the off-center listening position as described in Figure 2. The dashed line represents the time, when both signals reach the center of the head in theory. 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, due to head shadowing. This leads to a phantom source shift to the right. From the listeners point of view, this is towards the center, where ideally the phantom source should be localized. This thought experiment helps understanding the usefulness of adaptive signal adjustment. The following section analyzes the effect from a quantitative point of view.

4. 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 con-

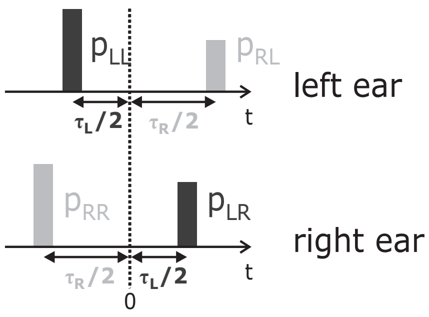


Fig. 3: Ear signals for transient stimuli. The loudspeaker signals are adjusted to reach the center of the head at the exact same time ($t = 0$).

verted into an Interaural Time Difference (ITD) and finally in a corresponding azimuth angle. In a more advanced approach, a binaural model after Braasch [12, 13] is used. This model simulates outer ear, inner ear and a decision device. The pathway to the ear is simulated by measured Head Related Transfer Functions (HRTF) 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 like bandpass noise.

4.1. 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 $\frac{L}{R}$ 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 using the described method of signal adjustment. The resulting signals from left and right loudspeaker at the left ear are:

$$\begin{aligned}\underline{p}_{LL} &= L \cdot \exp\left(\mathrm{j}\omega\frac{\tau_l+\tau_L}{2}\right)\\ \underline{p}_{RL} &= R \cdot \exp\left(-\mathrm{j}\omega\frac{\tau_l+\tau_R}{2}\right).\end{aligned}$$

They can be superimposed to the sum signal at the left ear:

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

The resulting signals from left and right loudspeaker at the right ear are:

$$\begin{aligned}\underline{p}_{LR} &= L \cdot \exp\left(\mathrm{j}\omega\frac{\tau_l-\tau_L}{2}\right)\\ \underline{p}_{RR} &= R \cdot \exp\left(-\mathrm{j}\omega\frac{\tau_l-\tau_R}{2}\right).\end{aligned}$$

They can be superimposed to the sum signal at the right ear:

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

The absolute value and phase of the interaural transfer function

$$\underline{A}(\omega) = \underline{p}_L/\underline{p}_R$$

are related to ILD and ITD, which can be transfered into a localization angle.

4.1.1. Modeling of a center source

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

$$|\underline{A}(\omega)| = 1$$

and the phase can be written as

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

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

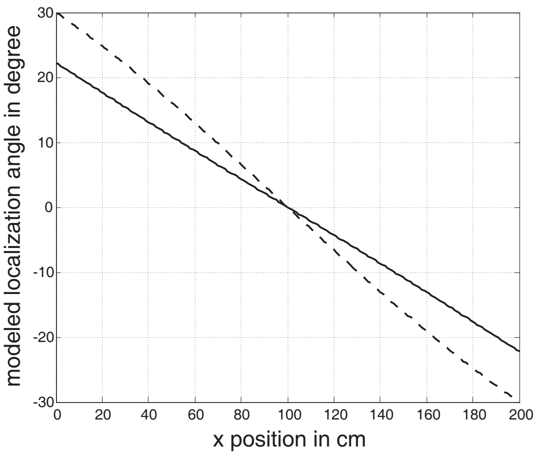


Fig. 4: Localization of a center phantom source with signal adjustment against the x position of the listener ($y = 1.73$ m, loudspeaker distance = 2 m). The plot shows the perceived angle to phantom source (solid), which should be at center, and actual angle to center (dashed). The perceived angle is modeled using the analytic approach.

the modeled localization angle Φ is compared with the actual target angle to center. A standard stereo setup with a loudspeaker distance of 2 m is used. The listening position is moved in x direction with constant $y = 1.73$ m. The orientation of the listener is straight ahead, as shown in Figure 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 loudspeaker ($x = 0$ cm and $x = 200$ cm). This means that the perceived position of the center phantom source remains almost constant, with a slight shift in the direction of listener movement.

4.2. Binaural Model

To investigate more complex signals and effects like 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 can be seen in Figure 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 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 and will be discussed in [14, 15].

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

Table 1: Used stimuli for binaural model. Bandpass white noise with lower (f_1) and upper (f_2) frequency limits. Stimulus 3 covers the relevant frequency range for speech and music.

4.2.1. Modeling of a center source

As already 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 in comparison with the actual target angle to center can be seen in Figure 6. Again the standard setup is used as described above. Figure 6a 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 $x = 0$ cm and $x = 200$ cm). The localization error (difference between perceived angle to phantom source and actual angle to center) remains constant, because the absolute value of the reference angle to center increases. In Figure 6b 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 x-direction (e.g. from left to right). Although the same image shift in the direction of movement can be seen. Again a small error

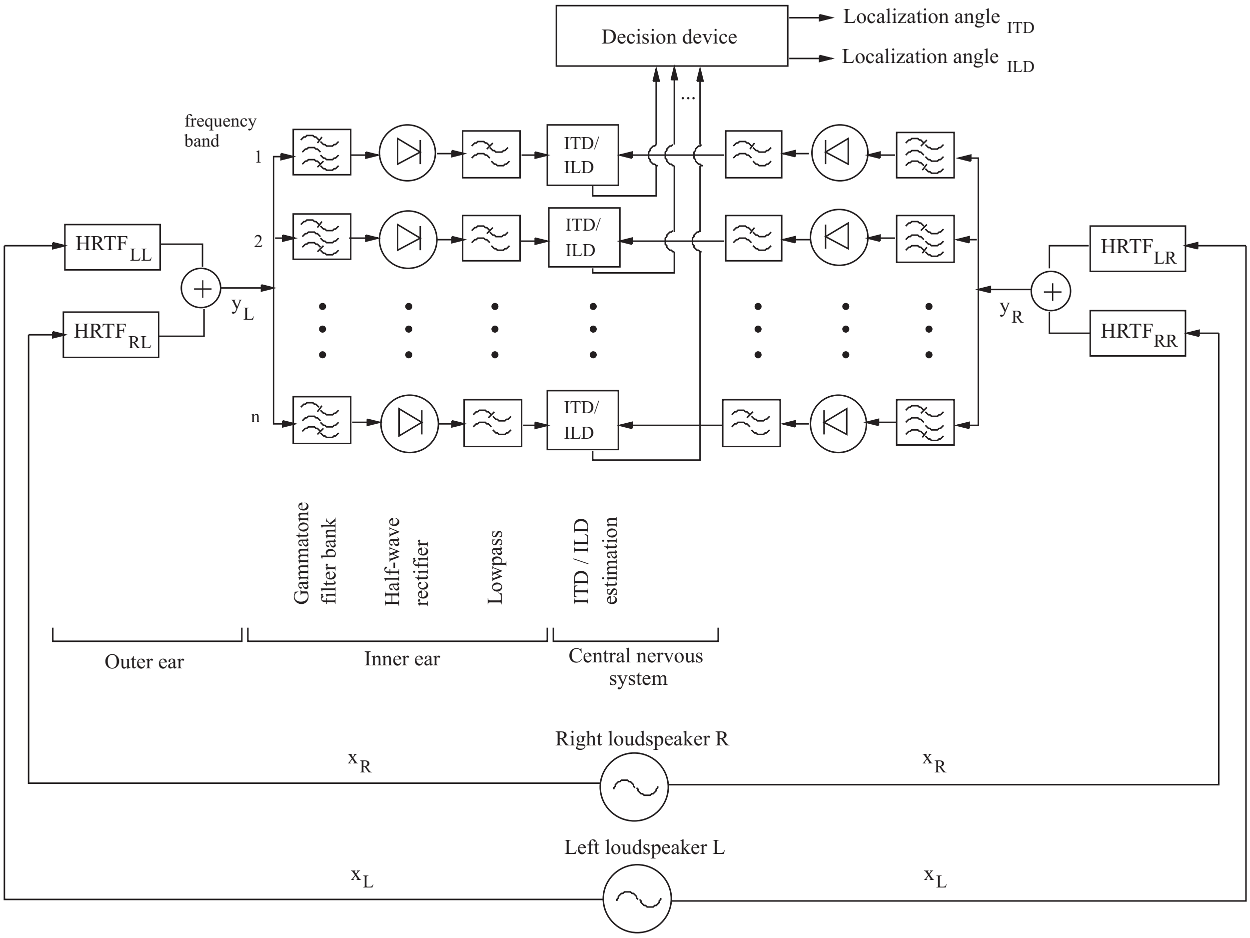


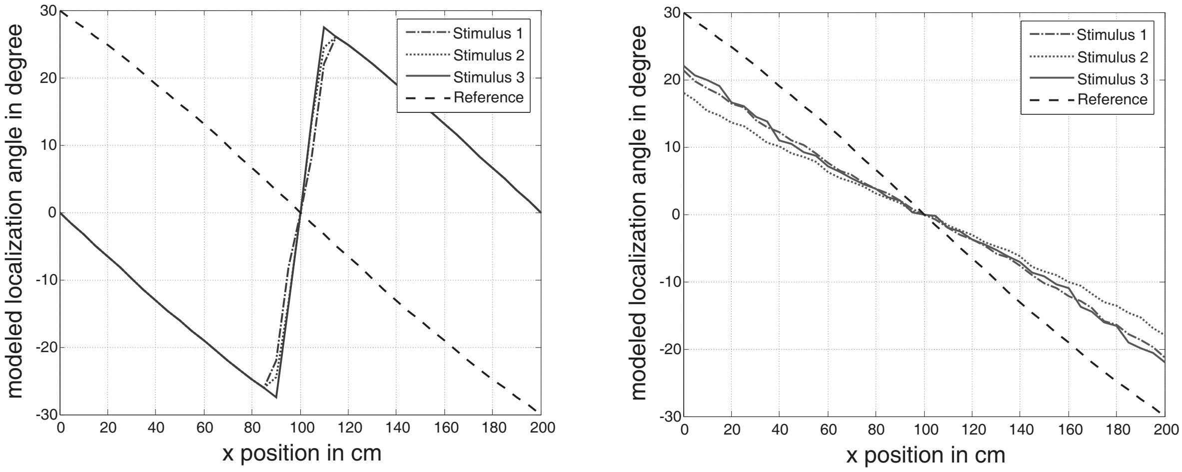
Fig. 5: Adapted binaural model for phantom source localization in a stereophonic setup after Braasch [12, 13].

of approximately 10° is found for extreme listening positions.

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

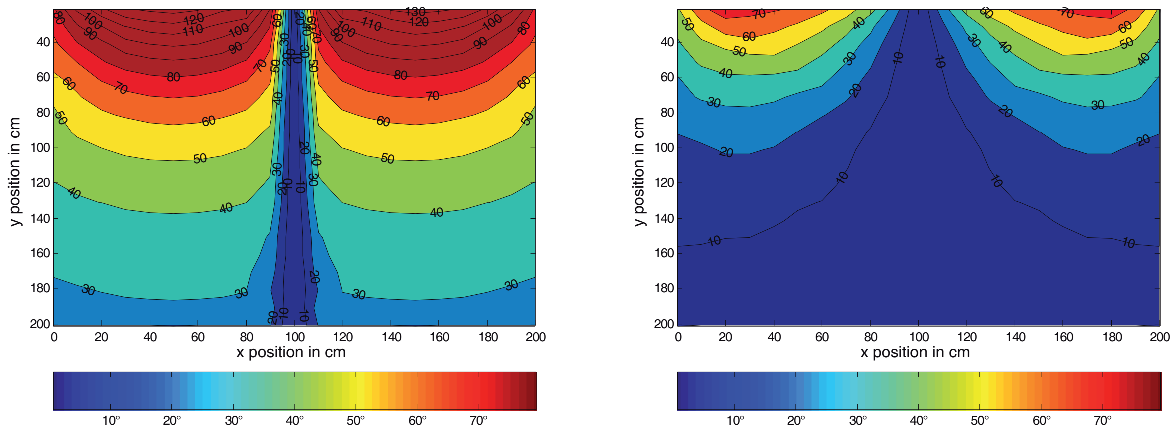
Even more intuitive is the visualization of the modeled localization angle using "vector maps". In Figure 8 the modeled azimuth angle is plotted as a direction vector for a center source without and with signal adjustment. Figure 8a shows the varying direction of source localization depending on listener position. If signals are adjusted, the perceived source position remains almost stable as can be seen in Figure 8b.

Both models lead to the same conclusion: They show that an adjusted "sweet spot" using interchannel time delay and amplitude adjustment improves the



(a) (b)

Fig. 6: Localization of a center phantom source against the x position of the listener ($y = 1.73$ m, loudspeaker distance $= 2$ m). The perceived angle is estimated using the binaural model. The plot shows the perceived angle to phantom source for three different stimuli and the actual angle to center (dashed) **(a)** without signal adjustment, **(b)** with signal adjustment.



(a) (b)

Fig. 7: "Quality maps" showing the absolute value of the 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.

localization over the whole off-center listening area. The asymmetric signal paths for excentric listening positions are important to maintain correct stereophonic localization. Although some localization error remains. It can be estimated and compensated using a binaural model.

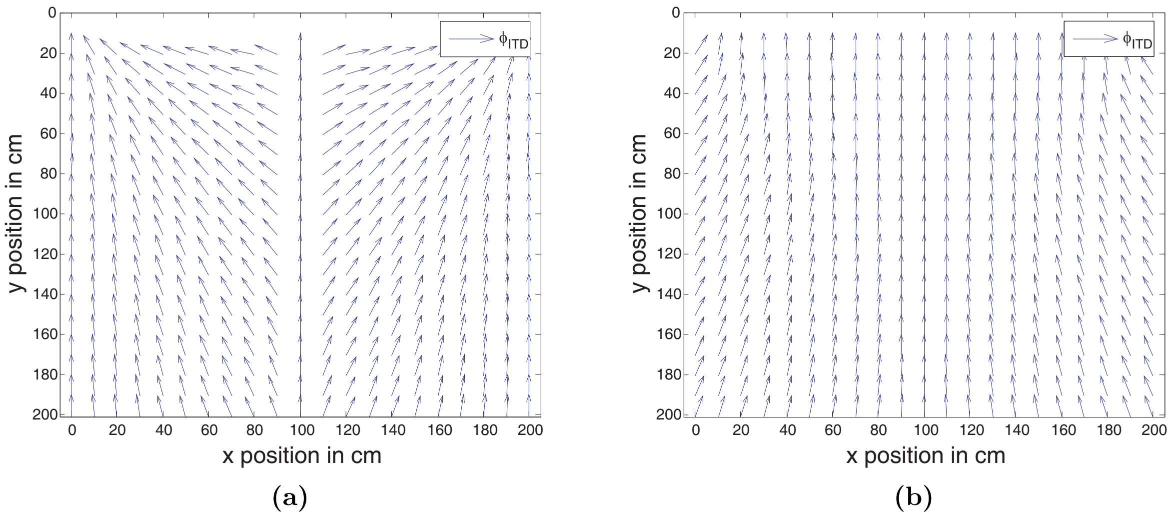


Fig. 8: "Vector maps" showing the modeled localization angle of a center phantom source. **(a)** Without signal adjustment the precedence effect distorts correct phantom source localization. **(b)** With signal adjustment the localization is maintained over a wide listening area.

5. IMPLEMENTATION

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 Figure 9). 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 more a small area than a point. To be suitable for daily use the tracking system must not use markers. Therefore a camera based head tracker was developed in cooperation with the faculty of computer science at TU Dresden. Our current system runs in real time on a single laptop using the integrated camera. The practicability of the system was confirmed through informal listening tests.

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

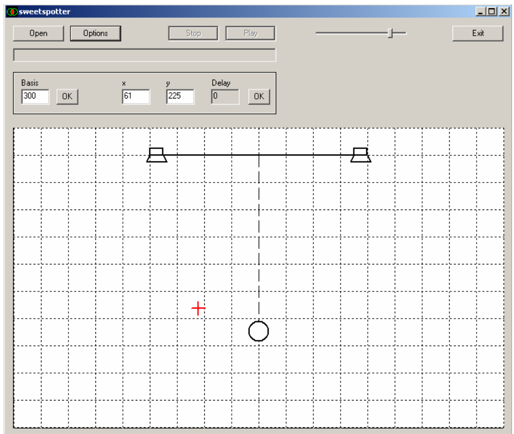


Fig. 9: "Sweetspotter" is a real time C++ implementation running on Windows. The listener position is tracked using a camera and a face recognition algorithm. Delay and level are adjusted accordingly.

6. CONCLUSION AND OUTLOOK

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 listeners head improves the localization over the whole listening area. The remaining error can be estimated using the described models. The major advantage of adaptive signal adjustment is the compatibility with existing source material and equipment. Stereophonic multichannel reproduction is still an interesting topic and can be further improved using the described methods.

The influence of head orientation and different stereophonic recording techniques on the utility of the system is an important factor. A first study is discussed in [14, 15]. To further evaluate the system, comprehensive listening tests are necessary. The ability to compensate for elevation effects or coloration in reverberant environments will be investigated.

7. ACKNOWLEDGEMENT

The authors wants to thank Prof. U. Jekosch, Prof. J. Blauert and Dr. E. Altinsoy for supervision and informative discussions. Thanks to Tobias Pietzsch for support with the head tracking system.

8. REFERENCES

[1] Keibs, L., *Perspektiven für eine raumbezogene Rundfunkübertragung*, Gravesaner Blätter, Nr. 22, 1961, S. 2-40.

[2] Gaál, D., *Calculation of the Stereophonical Localization Area*, preprint of the 53th Convention Zürich 1976, Audio Engineering Society, S. 1-7.

[3] Bauer, B., *Broadening the Area of Stereophonic Perception*, Journal of the Audio Engineering Society, Vol. 8, No. 2, 1960, S. 91-94.

[4] Aarts, R. M., *Enlarging the Sweet Spot for Stereophony by Time/Intensity Trading*, preprint of the 94th Convention Berlin 1993, Audio Engineering Society, S. 1-16.

[5] Kates, J. M., *Optimum Loudspeaker Directional Patterns*, Journal of the Audio Engineering Society, Vol. 28, No. 11, 1980, S. 787-794.

[6] Keele, D.B., *A Loudspeaker Horn that Covers a Flat Rectangular Area from an Oblique Angle*, preprint of the 74th Convention New York 1983, Audio Engineering Society, S. 1-22.

[7] Davis, M. F., *Loudspeaker Systems with Optimized Wide-Listening-Area Imaging*, Journal of the Audio Engineering Society, Vol. 35, No. 11, 1987, S. 888-896.

[8] Aoki, S. et al., *Stereo Reproduction with Good Localization over a Wide Listening Area*, Journal of the Audio Engineering Society, Vol. 38, No. 6, 1990, S. 433-439.

[9] Kim, S. et al., *Adaptive Virtual Surround Sound Rendering System for an Arbitrary Listening Position*, Journal of the Audio Engineering Society, Vol. 56, No. 4, 2008, S. 243-254.

[10] Kyriakakis, C. et al., *Signal Processing, Acoustics, and Psychoacoustics for High Quality Desktop Audio*, Journal of Visual Communication and Image Representation, Vol. 9 No. 1, 1998, S. 51-61.

[11] Lipshitz, S. P., *Stereo Microphone Techniques: Are the Purists Wrong?*, Journal of the Audio Engineering Society, Vol. 34, No. 9, 1986, S. 716-744.

[12] Braasch, J., *Modelling of binaural hearing*, Communication Acoustics (J. Blauert ed.), Springer Verlag, 2005, S. 75-108.

[13] Braasch, J., *Auditive Lokalisation und Detektion in Mehrschallquellen-Situationen*, Dissertation, Düsseldorf: VDI Verlag, 2002.

[14] Merchel, S., Groth, S., *Adaptive adjustment of the "sweet spot" to the listener's position in a stereophonic play back system*, Proceedings of NAG/DAGA, Rotterdam, 2009.

[15] Groth, S., *Untersuchung eines Stereo-Systems mit Signalanpassung an die Hörposition*, Diploma Thesis, Institute of Acoustics and Speech Communication, TU Dresden, Dresden, 2008