Adaptive Adjustment of the "Sweet Spot" to the Listener's Position in a Stereophonic Play Back System - Part 2

S. Groth[1], S. Merchel[2]

Chair of Communication Acoustic, Dresden University of Technology, Germany,
[1] *stephan.groth@gmx.net* [2] *sebastian.merchel@ias.et.tu-dresden.de*

Introduction

One major disadvantage of stereophonic play back systems is the narrow "sweet spot" in which correct audio localization is possible (see Figure 1). Due to this limitation, the listener's freedom of movement is drastically restricted. The present paper focuses on a system designed to adaptively adjust the "sweet spot" to the listener's position. The primary concern is the analysis of sound localization during the adjustment of the "sweet spot" via time delay and level differences. Therefore, different models of binaural hearing are used to validate the utility of the system. Furthermore, different stereo recording techniques are evaluated regarding their suitability for play back systems with signal adjustment. The theoretical results are examined using the C++ program "Sweetspotter", which adjusts the loudspeaker signals according to the listener's position in real-time.

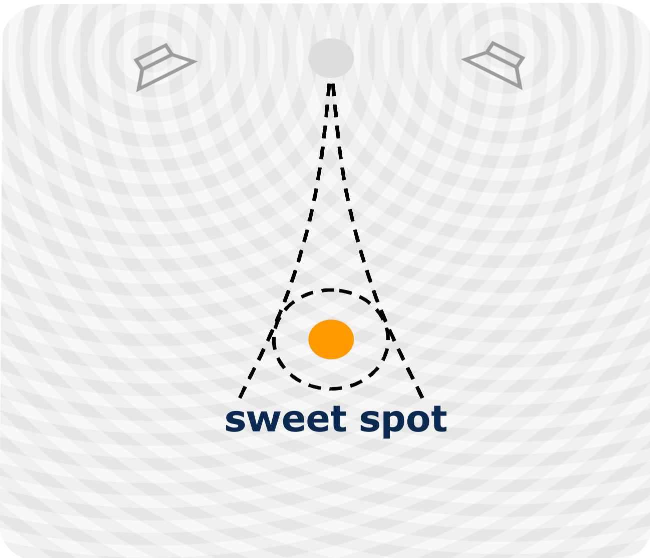

Figure 1: Schematic figure of the "sweet spot"

Literature Survey

Present methods to broaden the "sweet spot" can be separated into two groups: Those who attempt to adjust the radiation pattern of the loudspeakers and those who adjust the signals of the loudspeakers directly. Since all methods dealing with loudspeaker characteristics only affect the level differences, the precedence effect can not be prevented which leads to localization blur and image broadening at off-center listening points. Therefore, only methods that adjust the loudspeaker signals directly will be discussed in this paper. For example, Kyriakakis [1] described a system which uses head tracking and time delay between the loudspeaker signals to adjust the "sweet spot" while the listener is moving. Unfortunately, there are no publications about the usefulness of time delay and amplitude adjustment in off-center listening and emerging artifacts. Although several researchers have analyzed the human localization by using analytical (i.e. Lipshitz [2]) and numerical models (i.e. Stern [3], Braasch [4, 5]), none of them have ever described the localization in a stereo system with signal adjustment.

Adaptive Signal Adjustment

The image shift to the nearer loudspeaker in eccentric stereo hearing situations results from the occurring Interaural Time and Level Differences (ITD and ILD). For that reason, the loudspeaker signals have to be adjusted according to the x-y position of the listener. 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 (corresponding to the "sweet spot"). Additionally, 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 the loudspeakers to the ears become asymmetrical for off-center listening positions due to head shadowing and travel distance. This asymmetry is important for correct off-center localization.

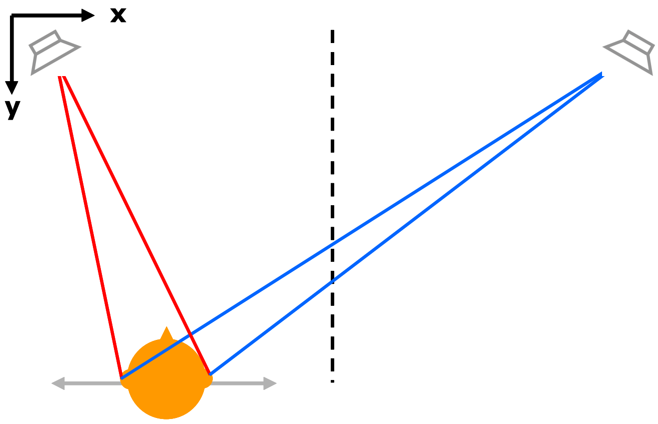

Figure 2: Asymmetric signal paths between loudspeakers and ears for an eccentric listening position

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 [2] is adapted for off-center stereo with signal adjustment. This approach analyzes the superimposed signals at the listener's ears, if two sources emit low-frequency sine waves. The emerged phase difference is converted into an Interaural Time Difference (ITD) and finally in a corresponding azimuth angle. It can be shown that a phantom source remains stable for a large off-center listening area with small remaining errors at extreme positions.

To investigate more complex signals and effects like head shadowing and reflections on the head and the torso, a binaural model after Braasch [4, 5] is adapted and implemented in MATLAB. The general model structure can be seen in Figure 3. The input consists of two sources which generate the left and right channel of a stereophonic system. The pathways to the ear canals are modeled using angle dependent measured HRTFs. This outer ear model is followed by a model of the inner ear behavior. In a cue estimation stage, binaural localization cues (ITDs and ILDs) are estimated and analyzed using a cross-correlation algorithm with specialized weighting functions. The corresponding localization angle is found with a remapping algorithm using measured HRTFs. A detailed description of the system can be found in Groth [6]. Since the interaction of ITDs and ILDs in the localization process is still object of research, resulting localization angles will be examined separately.

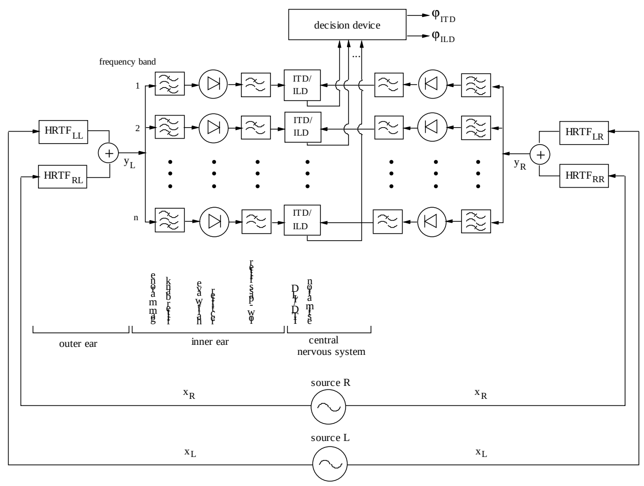

Figure 3: Adapted binaural model for phantom source localization in a stereophonic setup (after Braasch [4, 5])

By means of the described model, different play back and hearing situations with or without signal adjustment are simulated in a stereo reference setup with a base width of 2 meters. Three different noise stimuli are used with bandwidth limitation as shown in Table 1.

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

First, both loudspeakers are driven with identical signals to produce a center source. The improvement of the system with signal adjustment can be illustrated by using "quality maps" which show the localization error (absolute value of the difference between target angle to phantom source and modeled localization angle) over the whole listening area. The upper row of Figure 4 shows the resulting error for a center source (a) without and (b) with signal adjustment considering only ITD cues. Without signal adjustment only a narrow area with a minor localization error emerges – the "sweet spot". When leaving this area, the error increases rapidly until the precedence effect appears. With combined time and level adjustment the area with a minor localization error can be considerably broadened, except for extreme hearing positions near the loudspeakers. This can be shown even more intuitively by using "vector maps". In the lower row of Figure 4 the modeled azimuth angle is plotted as a direction vector over the whole listening area for a center source without and with signal adjustment. As can be seen, the localization vector rapidly points to the nearer loudspeaker while leaving the "sweet spot". With signal adjustment, the position of the perceived phantom source almost remains stable at the center.

Apart from a center source, different stereo recording techniques are simulated to investigate the stability of a shifted source. In "intensity" stereo a shifted source is produced by applying different levels to the right and the left channel. Figure 5 (top) shows the "quality maps" that are based on (a) ITD and (b) ILD cues for a phantom source which is shifted about 10° to the left (relative to the "sweet spot"). This relates to a level difference of ΔL = 6 dB between the left and the right loudspeaker. The figure also shows a broadening of the listening area for both ITD and ILD cues, but in this case the area becomes asymmetric. The "vector maps" in Figure 5 (bottom) show an almost stable localization angle to the shifted source; ITD and ILD cues lead to the same direction. Therefore, a sharp image is produced.

However, there is a different situation when using time stereo. In time stereo, the shifted source is produced via time delay between the channels. It can be shown that time delay between the loudspeaker signals lead to unnatural level differences at the listener's ears (see also Lipshitz [2] and Groth [6]). This can be seen in the "quality maps" in Figure 6 (top) where a shifted source is simulated throughout a time difference of Δt = 0.14 ms. For ITD cues the listening area is broadened similar to "intensity" stereo, but for ILD cues the error becomes larger. In the "vector maps" in Figure 6, one can see that this leads to contradicting localization cues and thus to localization blur and image broadening.

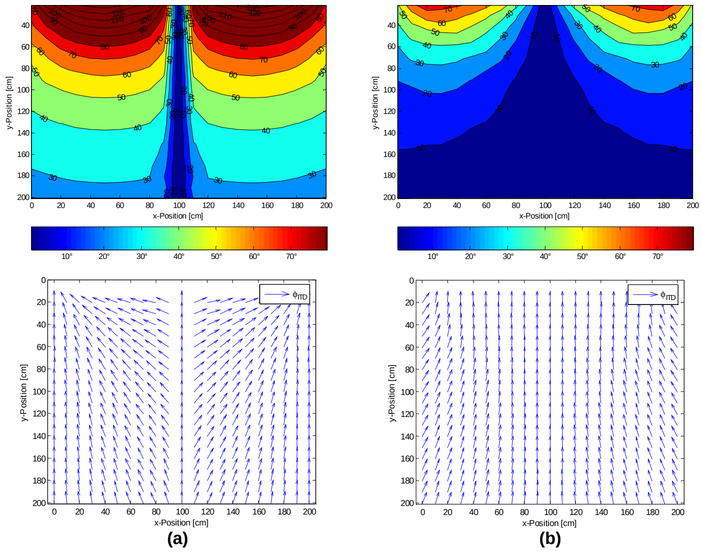

Figure 4: “Quality maps” (top) and “vector maps” (bottom) for a simulated center phantom source based on ITD cues; **(a)** without signal adjustment; **(b)** with signal adjustment

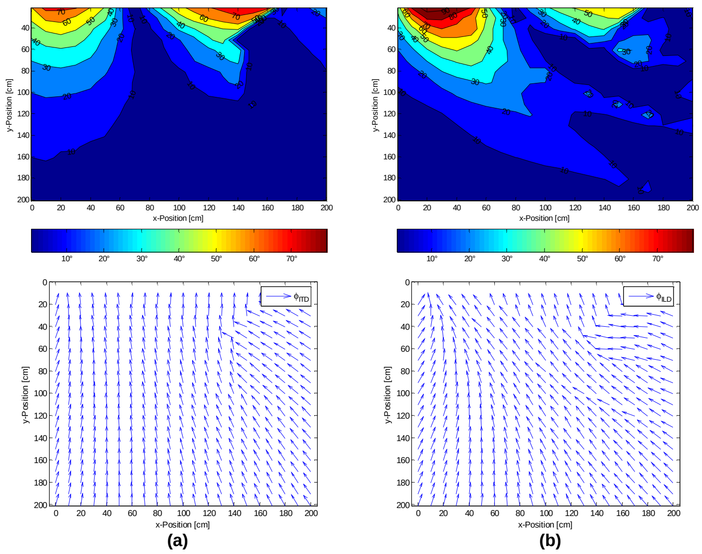

Figure 5: “Quality maps” (top) and “vector maps” (bottom) for “intensity” stereo (ΔL = 6 dB) with signal adjustment; **(a)** based on ITD cues; **(b)** based on ILD cues

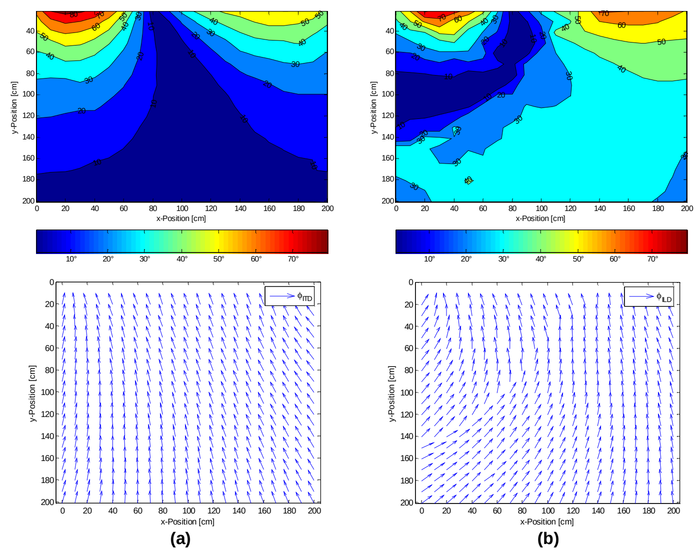

Figure 6: “Quality maps” (top) and “vector maps” (bottom) for time stereo (Δt = 0.14 ms) with signal adjustment; **(a)** based on ITD cues; **(b)** based on ILD cues

Implementation

In order to examine the theoretic results, a test system named “Sweetspotter” was implemented on a PC using C++. The system is able to adaptively adjust the loudspeaker signals to the listener’s position. 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. Figure 7 presents the graphical user interface of “Sweetspotter”. The practicability of the system and the theoretic predictions have been confirmed through first informal listening tests.

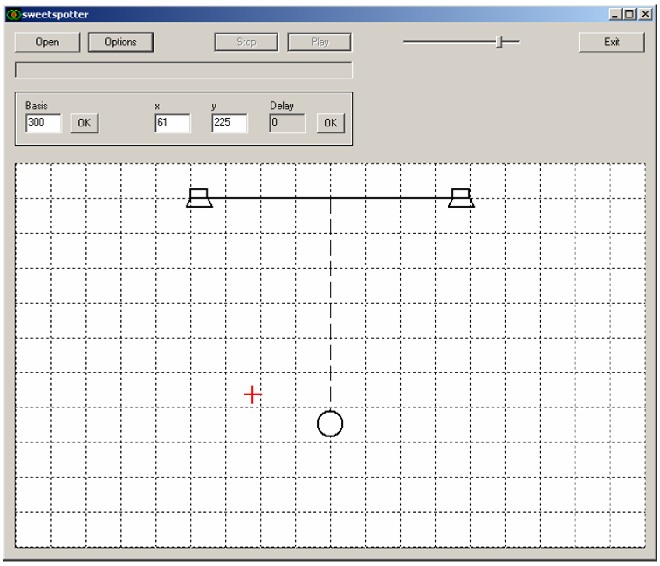

Figure 7: “sweetspotter” GUI

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 web cams. The direct integration in loudspeaker or hi-fi systems is also possible. Because the system only works for a single person, it has to automatically switch off if multiple persons are entering the listening area.

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 indicated that an adaptive adjustment of the signals relative to the center of the listener’s head improves the localization over the whole listening area. Moreover, the best results in localization stability can be achieved for “intensity” stereo with combined time and level adjustment. However, adaptive signal adjustment in reference to the center of the listener’s head still results in a small localization error, which could be estimated and compensated for using a binaural model.

Acknowledgment

The authors want 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.

References

[1] 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

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

[3] Stern, R. M. et al.: *Lateralization of complex binaural stimuli: A weighted image model*, J. Acoust. Soc. Amer. 84, 1988, S. 156-165

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

[5] Braasch, J.: *Modelling of Binaural Hearing. In J. Blauert*, editor, Communication Acoustics, chapter 4, pages 75–108, Berlin: Springer-Verlag, 2005

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