Audio Engineering Society

Conference Paper

Presented at the 21st Conference
2002 June 1–3 St. Petersburg, Russia

An Objective Model of Localisation in Binaural Sound Reproduction Systems

Alois Sontacchi[1], Markus Noisternig, Piotr Majdak and Robert Höldrich

Institute of Electronic Music and Acoustics, University of Music and Dramatic Arts, A-8010 Graz, AUSTRIA (http://iem.at)
[1]Correspondence should be addressed to alois.sontacchi@kug.ac.at

ABSTRACT

A mathematical model is presented to objectively derive sound localisation performance using HRIR (Head Related Impulse Response) based binaural sound reproduction systems. Rendering a sound source via panning methods causes artefacts that will lead to errors in localisation by human subjects. A localisation function and a localisation blur will be derived by comparing reference HRIRs with the distorted HRIRs, assuming that the cues specified by the reference HRIRs result in optimal localisations. Psychophysical effects will be incorporated as well. Studying the relationship between panning and perceived directions using listening tests entails an enormous effort of time. In addition, the presented mathematical model can be used to minimise the number of parameters which need to be evaluated by listening tests. Furthermore the localisation performance of several HRIR-based panning methods will also be evaluated.

1. INTRODUCTION

Convincing sound reproduction via headphones requires filtering of virtual sources with head related impulse responses (HRIRs) which describe signals at the two ear drums as a function of the incidence angle. The duplex theory of sound source localisation states that the two main cues are the interaural time difference (ITD) and the interaural level difference (ILD) which are caused by the wave propagation time difference and the shadowing effects of the head. Regarding hearing in natural sound fields, humans are able to improve source localisation capabilities due to small head movements [1], [2], [3]. In order to be able to benefit from this phenomenon in virtual reality (VR) applications, head tracking has to be incorporated in the time-varying binaural sound reproduction system (Fig. 1). Although simple in concept, this approach can be computationally expensive to implement. Typical systems accommodate head motion by high-quality time-varying interpolation between different HRIRs using various panning methods [4]. Hence audible artefacts may arise concerning localisation errors and localisation blurs.

In this paper a new mathematical model is proposed to evaluate the performance of binaural sound reproduction systems depending on various panning methods and used impulse responses. An error function based on the difference between the distorted HRIRs and the reference HRIRs will be introduced which accounts the psychophysical effects of binaural hearing as well. The basic assumption that guides the following approach is that the cues specified by the reference HRIRs result in optimal localisations. Localisation errors by the use of non-individualised HRIRs will not be examined more closely in the following paper. In a

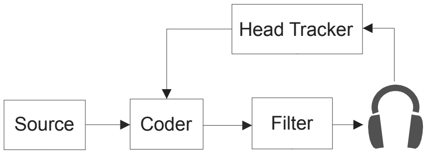

Figure 1: Binaural Reproduction System with Headtracking

companion paper [5] the mathematical model is verified by using listening tests.

2. THEORY

Sound localisation can be described by the following functions where Θ is the azimuth angle:

- The localisation function $L(\Theta)$
  that refers to the perceived direction of a sound source.
- The localisation blur $Bl(\Theta)$
  that refers to the width of the perceived sound source.

Due to the fact that the reference HRIRs are measured at N discrete positions over the azimuth angle, the localisation function and the localisation blur are discrete functions as well. Unfortunately one cannot usefully classify systems using the above functions. Therefore, a quality factor is developed that yields the following definitions:

- The average localisation error $\overline{L}$
  that refers to the root mean square of the difference between the localisation function and the target localisation.
- The average localisation blur $\overline{Bl}$
  that refers to the localisation blur averaged over the azimuth angle in multiples of the minimum audible angle (MAA) [1].

The duplex theory of sound localisation states that the two main cues of sound source localisation are the interaural time difference (ITD) and the interaural level difference (ILD) [1]. Higher level auditory cues like the interaural group delay (IGD) or monaurally effects are not considered in the following mathematical model.

2.1 Interaural Time Difference (ITD)

The ITD can be derived from the group delay time. Evaluating the group delay time using the negative derivation of the HRIRs phase yields to artefacts. Psychoacoustically, critical-band experiments show that the ear is usually not sensitive to relative timing or phase, as long as the signal components lie in different critical bands [1], [7]. Hence a group delay time $T_g(z,\Theta)$ over critical bands z is determined by filtering the impulse responses with zero-phase filters and calculating the energy gravity,

$$T_g(z,\Theta) = \frac{\sum\limits_n n \cdot h_F^2(n,z,\Theta)}{\sum\limits_n h_F^2(n,z,\Theta)} \tag{1}$$

where $h_F(n,z,\Theta)$ is the filtered impulse response. So the ITD follows from the difference of the group delay time between the ipsilateral and the contralateral eardrums.

$$ITD(z,\Theta) = T_{g,Left}(z,\Theta) - T_{g,Right}(z,\Theta) \tag{2}$$

Localisation as a result of ITD:

The ITD causes localisation cues primarily below 1.5 kHz [1]. A general approach is that the position of a sound source can be evaluated from the ITD between the reference HRIRs representing the left and right ear. Consequently a distortion of the ITD yields to dislocalisation. The perceived azimuth angle from a sound source at position Θ_0, the frequency dependent ITD angle (ITDA) $\Theta_{ITD}(z,\Theta_0)$, can be evaluated by comparing the reference HRIRs with the distorted HRIRs. An efficient way to calculate the ITDA is to determine the angle where the distorted ITD matches the ITD caused by the reference HRIRs.

$$\Theta_{ITD}(z,\Theta_0) = \arg\min_{\Theta} \left| ITD_{REF}(z,\Theta) - ITD_{DIST}(z,\Theta_0) \right| \tag{3}$$

It is important to restrict the area of searching, thus the scope is limited to

$$\forall \Theta \in \left[\Theta_0 - \frac{\Theta_{WND}}{2} ; \Theta_0 + \frac{\Theta_{WND}}{2} \right] \tag{4}$$

where Θ_{WND} terms the size of the window.

If at any frequency band the value of a distorted ITD is larger than the reference ITD values at the corresponding frequency band

$$\forall \Theta : ITD_{REF}(z,\Theta) < ITD_{DIST}(z,\Theta_0) \tag{6}$$

it must be taken into account that the search process will not be determined. In this case the ITDA is set to ±90° regarding psychophysical effects [6].

$$\Theta_{ITD}(z,\Theta_0) = \left\{ \begin{array}{lll} +90° & \mathit{for} & \Theta_0 \in [0°;180°] \\ -90° & \mathit{for} & \Theta_0 \in (0°;180°) \end{array} \right\} \tag{7}$$

2.2 Interaural Level Difference (ILD)

The ILD represents the level difference between the two ears and is calculated in the frequency domain. The ILD causes localisation cues primarily above 1.5 kHz. Incorporating psychophysical aspects, the level difference is calculated over frequency bands as well.

$$ILD(z,\Theta) = 10\log \sum_{f(z-1)}^{f(z)-1} \left[FFT\ h_L(n,\Theta)^2 \right]^2 - 10\log \sum_{f(z-1)}^{f(z)-1} \left[FFT\ h_R(n,\Theta)^2 \right]^2 \tag{8}$$

$h_i(n,\Theta)$ is the HRIR respective to the left and right ear.

Localisation as a result of ILD:

In spatial hearing research the frequency dependant ILD angle (ILDA) $\Theta_{ILD}(z,\Theta_0)$ describes the perceived azimuth angle caused by the ILD from a sound source of certain direction Θ_0. An auditory cue is transformed to the appropriate direction angle by comparing the ILD caused by the reference HRIR with the distorted ILD.

$$\Theta_{ILD}(z,\Theta_0) = \arg\min_{\Theta} \left| ILD_{REF}(z,\Theta) - ILD_{DIST}(z,\Theta_0) \right| \tag{9}$$

However, it is important to restrict the search area as well. Like calculating the ITDA the ILDA is set to ±90° in case the value of the distorted ILD is larger than the reference ILDs value.

2.3 Merging of ILD and ITD

Both the localisation based on ITD and on ILD are important for the detection of the position of a virtual sound source. While human listeners are sensitive to interaural timing information in various forms over a wide range of frequencies, their use as a cue to sound location seems to be restricted to the lower range of frequencies. Human subjects are uniformly sensitive to ILDs across the audiometric range. Because of diffraction, significant location dependent differences in sound level at each ear only occur for frequencies in the mid to high frequency range. Merging the ITD and ILD the data will be prepared first in the following way:

- Because of the systems critical frequency the ITD as well as the ILD are faded out in the first critical band.
- For frequencies above 800 Hz (z = 7 bark) the ITD is faded out.

The result yields the localisation function averaged over critical bands where $w_i(z)$ are the weightings for superposition referring to ITD and ILD

$$L(\Theta)=\frac{1}{2}\cdot\left(\frac{1}{\sum w_{ITD}(z)}\cdot\sum_{z=1}^{24} w_{ITD}(z)\cdot\Theta_{ITD}(z,\Theta)\right)+ \frac{1}{2}\cdot\left(\frac{1}{\sum w_{ILD}(z)}\cdot\sum_{z=1}^{24} w_{ILD}(z)\cdot\Theta_{ILD}(z,\Theta)\right) \qquad (10)$$

For a given azimuth angle, the localisation blur refers to the standard deviation of the localisation function over critical bands.

$$Bl(\Theta)=\sqrt{\frac{1}{2}\cdot\sum_{i=ILD}^{ITD}\frac{1}{\sum w_i(z)}\cdot\sum_{z=1}^{24} w_i(z)\cdot\left[\Theta_i(z,\Theta)-L(\Theta)\right]^2} \qquad (11)$$

In addition, both an average error of localisation (Equ. 12) and an average localisation blur (Equ. 13) can also be derived, where N depends on the number of measured reference HRIRs.

$$\overline{L}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[L(\Theta_i)-\Theta_i\right]^2} \qquad (12)$$

Blauert showed in [1] that the resolution of human directional hearing is much more better in frontal direction than in lateral direction. Taking this fact into account, the localisation blur is referred to the minimum audible angle (MAA).

$$\overline{Bl}=\frac{1}{N}\cdot\sum_{\Theta=0}^{360^\circ}\frac{Bl(\Theta)}{MAA(\Theta)} \qquad (13)$$

2.4 Examples

In the following examples the calculation was performed for different panning methods (setup A, setup B) using several reference HRIRs (1, 2). Fig. 2 shows the localisation error in degrees, Fig. 3 shows the localisation blur in multiples of the MAA.

Result 1: Using shorter filters, the localisation error as well as the localisation blur decreases. This can be traced back to the fact, that shorter HRIRs include less azimuth dependent information.

Result 2: It is possible to map dislocalisation caused by the localisation error. There is no possibility to equalise the localisation blur. Although setup 2A causes a better performance in considering the localisation error, setup 2B yields to a better performance because the localisation blur is smaller than in setup 2A.

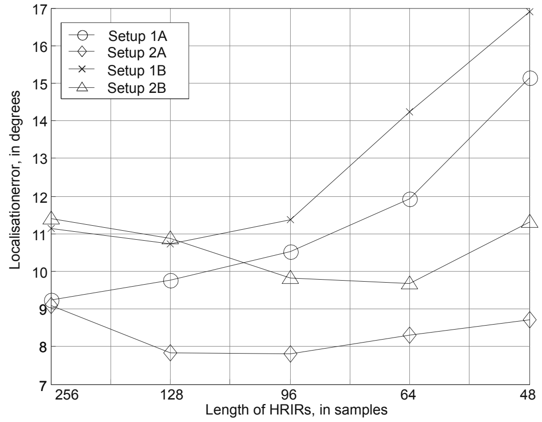

Figure 3: Localisation error in degrees for systems using different reference HRIRs (1,2) and different interpolation techniques (A,B)

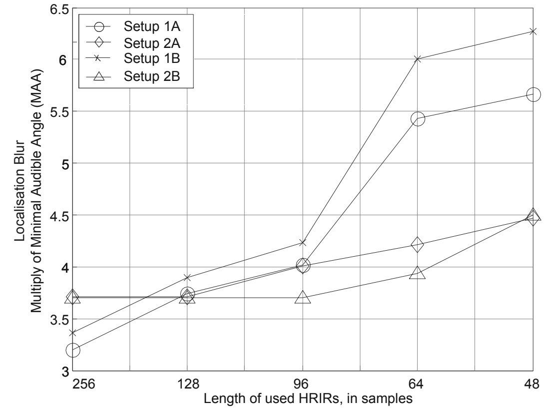

Figure 2: Localisation blur in multiples of the MAA for systems using different reference HRIRs (1,2) and different interpolation techniques (A,B)

3. CONCLUSION

The proposed mathematical model enables an objective classification of systems for binaural reproduction of virtual sound sources regarding to the localisation error and the localisation blur.

4. REFERENCES

[1] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localisation*, rev. ed. (MIT-Press, Cambridge, MA, 1997)

[2] F. L. Wightman and D. J. Kistler, *The Importance of Head Movements for Localizing Virtual Auditory Display Objects*, in Proc. 1994 Conference on Auditory Displays

[3] F. L. Wightman and D. J. Kistler, *Resolution of Front-Back Ambiguity in Spatial Hearing by Listener and Source Movement*, J. Acoust. Soc. Am, vol. 105, pp. 2841 – 2853 (1999)

[4] A. Sontacchi and R. Höldrich, *Kopfhörer und Lautsprecher-konzepte für eine virtuelle Akustik*, 46. Internationales Kolloquium, Technische Universität Ilmenau, (2001)

[5] A. Sontacchi et al., *Subjective Validation of an Objective Localisation Model*, 21st International Conference of the Audio Engineering Society, St. Petersburg, Russia (2002)

[6] D. Griesinger, *The Psychacoustics of Apparent Source Width, Spaciousness and Envelopment in Performance Spaces*, Acustica Vol. 83, pp. 721-731, (1997)

[7] V. R. Algazy et al., *Dependence of Subject and Measurement Position in Binaural Signal Acquisition*, J. Audio Eng. Soc., Vol. 47, pp. 937-947, (1999 Nov.)