Audio Engineering Society

Convention Paper

Presented at the 125th Convention
2008 October 2–5 San Francisco, CA, USA

A Method for Estimating Interaural Time Difference for Binaural Synthesis

Juhan Nam[1], Jonathan S. Abel[1], and Julius O. Smith III[1]

[1] *Center for Computer Research in Music and Acoustics (CCRMA), Stanford Univ., Stanford, CA, 94305, USA*

Correspondence should be addressed to Juhan Nam (juhan@ccrma.stanford.edu)

ABSTRACT

A method for estimating interaural time difference (ITD) from measured head-related transfer functions (HRTFs) is presented. The method forms ITD as the difference in left-ear and right-ear arrival times, estimated as the times of maximum cross-correlation between measured HRTFs and their minimum-phase counterparts. This arrival time estimate is related to a nonlinear least-squares fit to the measured excess phase, emphasizing those frequencies having large HRTF magnitude and deweighting large phase delay errors. As HRTFs are nearly minimum-phase, this method is robust compared to the conventional approach of cross-correlating left-ear and right-ear HRTFs, which can be very different. The method also performs slightly better than techniques averaging phase delay over a limited frequency range.

1. INTRODUCTION

For binaural synthesis, head-related transfer functions (HRTFs) are typically implemented as the cascade of a pure delay and a minimum-phase filter. The delay is selected according to the interaural time difference (ITD), the difference in sound arrival time between the listener's ears.

Nam et al. [1] and Kulkarni et al. [2, 3] examined the similarity between measured HRTFs and their minimum-phase counterparts using the maximum cross-coherence between them. The idea was that if the HRTF were a pure delay followed by a minimum-phase system, the maximum cross-coherence would be 1.0. Cross-coherence maxima close to 1.0 indicate a minimum-phase impulse response nearly identical in shape to a time-shifted version of the original impulse response. It was observed that the vast majority of maximum cross-coherences over azimuth and elevation exceed 0.9, implying that head-related impulse responses (HRIR) take on very similar shapes to their minimum-phase sequences. This work suggests our estimation of the time-shift be-

tween HRTFs and their minimum-phase counterparts from the time at which they are maximally correlated. This time-shift corresponds to the time of arrival of a source signal at a listener's ear. Accordingly, the ITD is the difference between the left-ear and right-ear time of arrival. This paper will explore this approach for estimating ITD.

2. ITD ESTIMATION METHODS

As described below, a number of methods have been proposed to estimate the ITD associated with a set of measured HRTFs from one subject.

2.1. Interaural Cross-Correlation

One conventional method to estimate ITD uses the interaural cross-correlation between the left-ear and the right-ear HRTFs for a given direction. Proposed by Kistler and Wightman in [4], this method determines the ITD as the time when the interaural cross-correlation is maximized. The drawback to this approach is that the left and right HRIRs are often quite different and do not correlate well. The estimated ITD becomes sensitive to the HRTF details, especially in the vicinity of the interaural axis where the contralateral HRTF is shadowed.

2.2. Threshold Detection

Another time-domain method examines the time at which the left and right HRIRs reach a threshold. This time determines the left-ear and right-ear times of arrival. The threshold is given as a percentage of the HRIR maximum value. Sandvad and Hammershoi [7] used 5% of the maximum value of HRIRs so as to detect their onset times. This method is sensitive to measurement noise, especially for the contralateral HRIRs which have relatively low signal to noise ratios, as it depends on a single data point. To avoid difficulties associated with additive measurement noise, Busson et al. [8] chose 50% of the maximum value for the threshold and other authors use the HRIR maximum. Doing so, however, introduces a bias, as the HRIRs change shape as a function of direction.

2.3. Linear Phase Fit

Jot et al. [9] presented fitting a linear characteristic to the excess phase over a specified frequency range as a means of estimating HRIR time of arrival. A frequency range between 1 kHz and 5 kHz was chosen, while Huopaniemi and Smith [10] suggested 500 Hz to 2 kHz, claiming that the ear is sensitive to low-frequency phase below approximately 1.5 kHz. These methods provide good results, but the frequency range needs to be carefully selected to avoid the effect of any non-minimum-phase zeros present. In addition, averaging over a narrow bandwidth reduces statistical leverage.

2.4. Interaural DC Group Delay

Minnaar et al. [5] proposed another method which computes the ITD as the interaural group delay at DC. Psychoacoustic experiment results are presented that support replacing the ITD with the DC group delay [6]. This method is expected to produce similar results to linear phase fit due to the scarcity of non-minimum-phase zeros at low frequencies. However, the DC group delay is difficult to measure due to customary DC removal in recording systems.

3. PROPOSED ITD ESTIMATOR

3.1. Estimator Description

The interaural time difference δ is estimated as the difference between estimated left and right HRIR arrival times,

$$\hat{\delta} = \hat{\tau_L} - \hat{\tau_R}. \tag{1}$$

As described above, the HRIR arrival times are estimated as the times of maximum HRIR correlation with their minimum-phase versions,

$$\hat{\tau} = \underset{\tau}{\operatorname{argmax}} \left\{ \sum_n h(n-\tau)\, h_{\mathrm{mp}}(n) \right\} \tag{2}$$

where $h(n)$ is the HRIR and $h_{\mathrm{mp}}(n)$ is its minimum-phase sequence. Note that since the HRIRs are available in discrete time, the arrival time can be more precisely estimated by quadratically interpolating the correlation maximum [12].

3.2. Estimator Interpretation

The expectation that the estimator (1) and (2) performs well flows from the fact that the HRIRs are substantially pure delays followed by minimum-phase systems. HRIRs therefore correlate well with their minimum-phase versions, with the correlations peaking at the unknown pure delay.

An HRTF $H(\omega)$ can be written as

$$H(\omega) = |H(\omega)| e^{j\{\mu(\omega)+\eta(\omega)\}} \tag{3}$$

where $|H(\omega)|$ is the magnitude response, and the phase response $\mu(\omega) + \eta(\omega)$ is the sum of the minimum-phase $\mu(\omega)$ and excess-phase $\eta(\omega)$ components. The minimum-phase frequency response is

$$H_{\rm mp}(\omega) = |H(\omega)|e^{j\mu(\omega)}, \quad (4)$$

and therefore (3) can be written as the product of the minimum-phase and excess-phase responses,

$$H(\omega) = H_{\rm mp}(\omega)e^{j\eta(\omega)}. \quad (5)$$

The estimated arrival time is then

$$\hat{\tau} = \underset{\tau}{\rm argmax} \left[\mathcal{F}^{-1}\left\{(H(\omega)H^*_{\rm mp}(\omega)\right\}\right], \quad (6)$$

where $\mathcal{F}^{-1}$ is the inverse discrete-time Fourier transform. The arrival time is then

$$\hat{\tau} = \underset{\tau}{\rm argmax} \left[\sum_{\omega} |H(\omega)|^2 \cos(\omega\tau - \eta(\omega))\right]. \quad (7)$$

Finally, $\hat{\tau}$ may be expressed as the result of a minimization,

$$\hat{\tau} = \underset{\tau}{\rm argmin} \left[\sum_{\omega} |H(\omega)|^2 \left\{1 - \cos(\omega\tau - \eta(\omega))\right\}\right]. \quad (8)$$

Note that $\eta(\omega)$ consists of linear phase (pure delay) and nonlinear phase allpass components. We see that the estimated arrival time $\hat{\tau}$ is the one minimizing the weighted sum of errors resulting from fitting a linear phase characteristic $\omega\tau$ to the measured excess phase $\eta(\omega)$. The weighting, $|H(\omega)|^2$, is the spectral energy, emphasizing phase errors in spectral regions having high energy and de-emphasizing phase errors associated with spectral nulls. Note that in the presence of small phase error, $|\omega\tau - \eta(\omega)| \ll 1$, the term $1 - \cos(\omega\tau - \eta(\omega))$ is approximately

$$1 - \cos(\omega\tau - \eta(\omega)) \approx \left[\omega\tau - \eta(\omega)\right]^2/2, \quad (9)$$

and the time of arrival estimate is the minimizer of a weighted sum of squared errors,

$$\hat{\tau}_{ls} = \underset{\tau}{\rm argmin} \left[\sum_{\omega} |H(\omega)|^2 \left[\omega\tau - \eta(\omega)\right]^2/2\right]. \quad (10)$$

Note that this *least squares* estimator $\hat{\tau}_{ls}$ is a weighted version of that proposed by Jot et al. [9] and Huopaniemi and Smith [10]. Compared to the cross-correlation maximizer proposed here (2), the least-squares estimator gives more weight to large errors, Fig. 1. The cross-correlation maximizer is expected therefore to be less sensitive to large phase errors, for instance caused by non-minimum-phase zeros near the frequency axis.

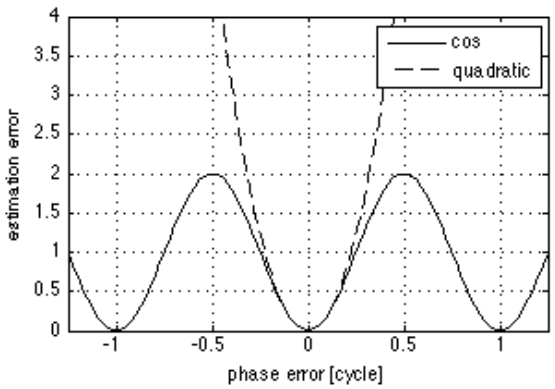


Fig. 1: Cosine and Quadratic Phase Errors

In fact, the phase error arises from non-minimum-phase HRTF zeros. Since it follows the allpass phase response, the estimation error is accumulated sweeping over one cycle as the frequency passes that of each non-minimum-phase zero, for example, moving from 0 to 1 in Fig. 1. However, in the proposed estimator, the phase error is suppressed at spectral nulls induced by zeros. Fig. 2 shows a typical HRTF excess group delay and its corresponding magnitude response. The excess group delay is

$$\gamma(\omega) = -\frac{d}{d\omega}\eta(\omega). \quad (11)$$

The phase error occurs in the non-constant group delay regions shown as peaks. The group delay peaks are associated with notches in the magnitude response. As a result, the proposed estimator discounts these regions of excess phase.

Consequently, both the nonlinear nature of the phase error and the spectral weighting are seen to provide advantages to the proposed estimation method (1) and (2).

3.3. Alternative Methods

The weighted least squares estimation above, (10), suggests an alternative arrival time estimator: the

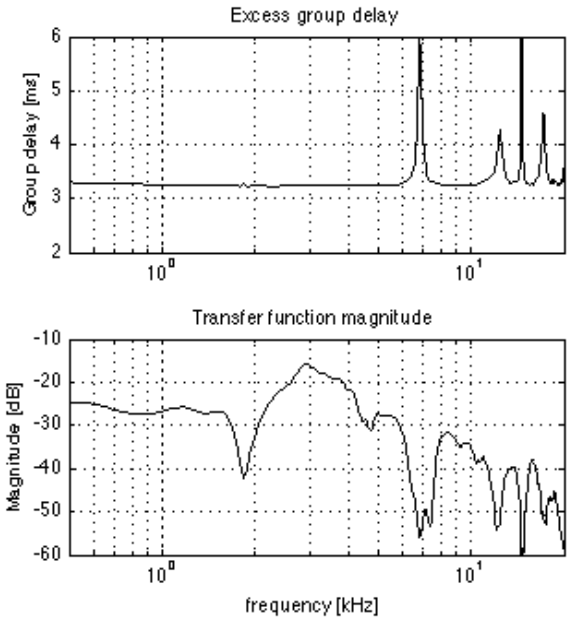


Fig. 2: Top: Excess group delay of an HRTF. Bottom: Corresponding magnitude response.

weighted mean of the excess group delay,

$$\hat{\tau}_{gd} = \sum_{\omega} |H(\omega)|^2 \gamma(\omega) \qquad (12)$$

where $|H(\omega)|^2$ is the spectral power and $\gamma(\omega)$ is the measured excess group delay. As illustrated in Fig. 2, the HRTF excess group delay is nearly constant with isolated peaks associated with spectral regions weighted by less power. As a result, the estimation (12) is expected to perform well.

Consider that averaging the group delay over the entire frequency band can introduce a bias, as the non-pure-delay, non-minimum-phase components produce positive group delay peaks. In order to avoid such effects, two approaches are suggested: 1) forming a trimmed mean by sorting the excess group delays and excluding the highest, say, 30%, 2) selecting specific frequency ranges, for instance, 500Hz to 2kHz. The first is intended to remove the group delay peaks and the second is to follow the frequency range suggested by [10].

4. ESTIMATOR PERFORMANCE

To explore estimator performance, HRIRs were measured [1] for 8 subjects outfitted with blocked meatus microphones in the CCRMA recording studio, using the method described in [11]. The test signal used to measure impulse responses was an exponentially swept sinusoid. The directional sampling interval was 15 degrees in azimuth and 10 degrees in elevation. The azimuth increases counterclockwise and the elevation is limited to $[-40°, 40°]$, with reference to 0° for the horizontal plane. In addition, a set of transfer functions for a microphone mounted on a rigid sphere were measured at 15° increments in the horizontal plane.

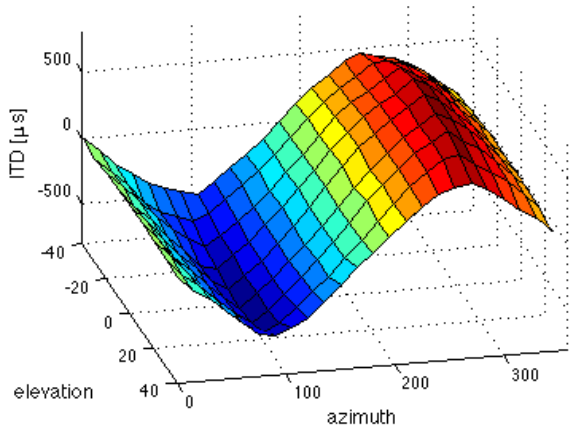
4.1. ITD Estimation

Here the proposed method (1) and (2), the interaural cross-correlation method, and the linear-phase fitting method are compared first, followed by the weighted excess group delay method. For the linear-phase fitting method, we used the 500 Hz to 2 kHz frequency range suggested by Huopaniemi and Smith [10], because it outperformed the 1 kHz to 5 kHz range suggested by Jot et al. [9]. The interaural DC group delay method by Minnaar et al. [5] was not computed due to DC blocking in the converters.

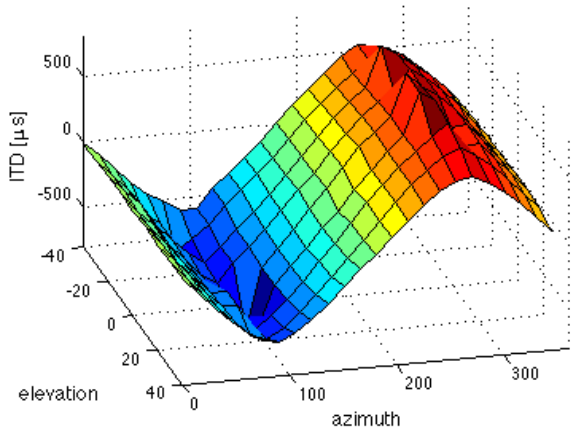
Fig. 3 shows the ITDs of a representative HRTF computed by the proposed method and the two prior methods. They produced quite similar ITD estimates overall. The interaural cross-correlation method, however, shows clear disturbances around the interaural axis near the horizontal plane, corresponding to 105° and 255° azimuth and -20°, 0°, and 10° elevation. This may be ascribed to the lack of coherence between left and right HRTFs in these directions.

Both the proposed method and the linear-phase fitting method show very stable shapes. But, it is observed that the linear-phase fit is slightly more irregular than that of the proposed method. The subtle difference is thought to arise from the limited frequency range taken into account by the linear-phase fitting method. In addition, in results not presented here, the linear-phase fitting occasionally produced spurious ITD estimates. In these cases, a non-minimum-phase zero was located within the 500 Hz to 2 kHz range.

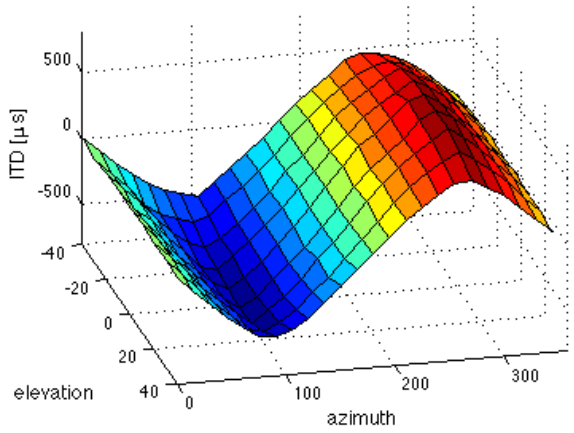
Fig. 4 shows the ITD computed by the weighted excess group delay—first, using the entire audio band, second, using the trimmed mean, and third, using



(a) Proposed method

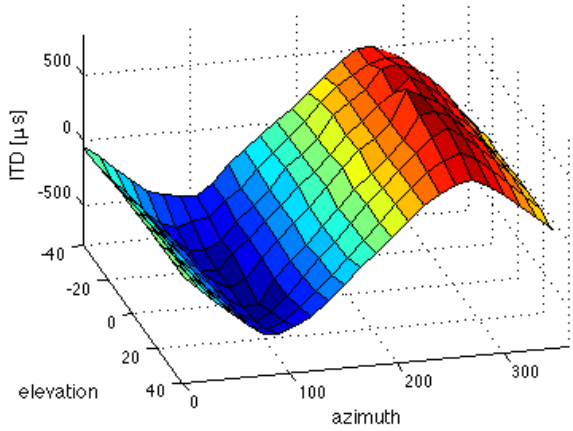


(b) Interaural cross-correlation

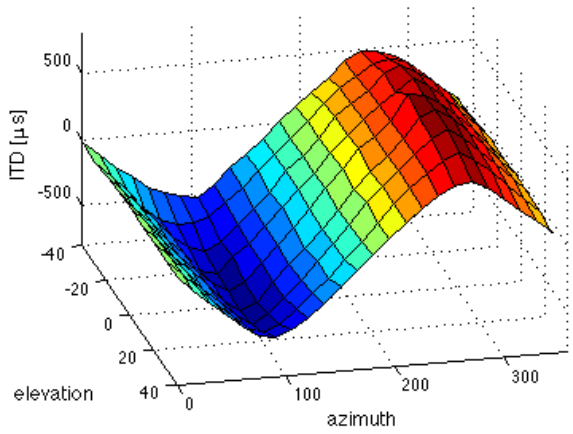


(c) Linear phase fit to 500 Hz – 2 kHz

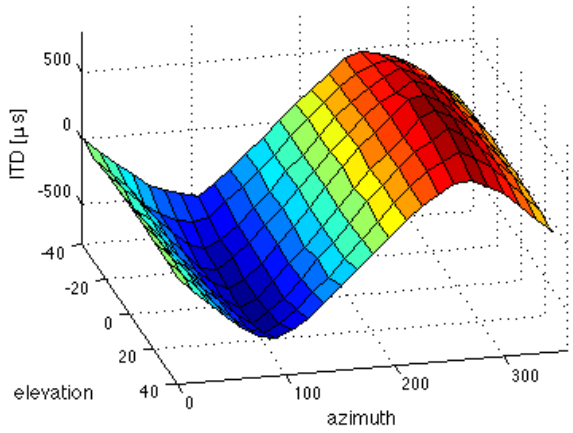
Fig. 3: Estimated ITD from a set of HRTFs



(a) Weighted excess group delay over the entire audio range



(b) Weighted excess group delay with trimming (30%)



(c) Weighted excess group delay with selection (500 Hz to 2 kHz)

Fig. 4: Estimated ITD using weighted excess group delay

the limited bandwidth suggested by Huopaniemi and Smith [10]. The three cases seem to produce better estimates than the interaural cross-correlation method while they are distinguished by the amount of disturbance around the interaural axis near the horizontal plane. The estimation using the entire frequency range is worst, probably because of the extreme peaks in the group delay. The trimmed mean is slightly improved, but the disturbances are easily detectable. The last case, the weighted excess group delay between 500 Hz to 2 kHz, shows the best result. Note that the shape of the ITD is almost identical to the result of the linear-phase fitting method. This is likely due to the linearity of HRTF phase in the frequency range used. However, the weighted excess group delay is considered to slightly outperform linear phase fit because the weighting reduces the effect of non-minimum-phase zeros near the frequency axis.

4.2. ITD estimation for a rigid sphere

Woodworth and Schlosberg presented a theoretical formula to calculate the ITD of a rigid sphere for diametrically opposed ears [13]. The formula is given as

$$\mathrm{ITD} = \frac{r}{c}(\sin\theta + \theta) \tag{13}$$

where r is the radius of the sphere, c is the speed of sound and θ is the azimuth angle in radians. To verify the result of the ITD estimators above, "HRTFs" of a rigid sphere were measured in the horizontal plane. The measured impulse responses had an 80 dB impulse level to mean noise floor SNR. The proposed method, the interaural cross-correlation method, and the linear-phase fitting method were compared. As seen in Fig. 5, the ITDs are very similar to each other, and to the theoretical ITD. As a way of examining the slight differences among them, the RMS ITD error was computed. We found RMS errors of 0.49, 0.53, and 0.82 samples for the proposed method, the interaural cross-correlation method, and the linear phase fitting method, respectively. Though these errors are insignificant, the proposed method is more accurate than the others.

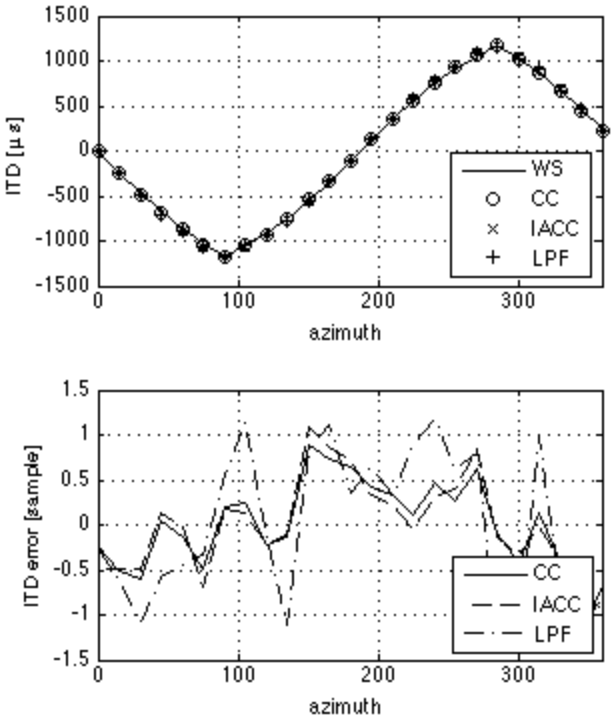


Fig. 5: Top: ITD estimation from measurement of a rigid sphere. Bottom: Estimation errors: WS - theoretical formula by Woodworth and Schlosberg, CC - proposed method, IACC - interaural cross-correlation, LPF - linear phase fit

5. CONCLUSION

A method for estimating ITD from measured HRTFs was suggested. The method computes the ITD as the difference between left-ear and right-ear HRIR

arrival times. The arrival time is estimated as the time of maximum cross-correlation between HRIRs and their minimum-phase versions. The maximum cross-correlation is interpreted as a nonlinear minimizer of estimation error with spectral weighting. Based on the analysis, the weighted excess group delay was presented as an alternative method. We showed that the proposed method outperforms compared methods—the interaural cross-correlation, the linear phase fit and the weighted excess group delay.

6. ACKNOWLEDGEMENT

We would like to thank CCRMA students and staff for their assistance and participation in the HRTF measurements, and Miriam Kolar for measuring the transfer functions of a rigid sphere.

7. REFERENCES

[1] J. Nam, M. A. Kolar, J. S. Abel, "On the minimum-phase nature of head-related transfer functions," presented at the 125th Audio Engineering Society Convention, October 2008.

[2] A. Kulkarni, S. K. Isabelle, H. S. Colburn, "On the minimum-phase approximation of head-related transfer functions," IEEE ASSP Workshop on Application of Signal Processing to Audio and Acoustics, 1995.

[3] A. Kulkarni, S. K. Isabelle, H. S. Colburn, "Sensitivity of human subjects to head-related transfer-function phase spectra," J. Acoust. Soc. Am. 105 (5), May 1999.

[4] D. J. Kistler, F. L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," J. Acoust. Sot. Am., Vol. 91, pp. 1637-1647, 1992.

[5] P. Minnaar, J. Plogisties, S. K. Olesen, F. Christensen, H. Moller, "The interaural time difference in binaural synthesis," presented at the 108th Audio Engineering Society Convention, February 2000.

[6] J. Plogsties, P. J. Minnaar, S. K. Olesen, F. Christensen, H. Moller, "Audibility of allpass components in head-related transfer functions," Presented at the 108th Convention of the Audio Engineering Society, Paris, France, February 2000.

[7] J. Sandvad, D. Hammershoi, "Binaural auralization. Comparison of FIR and IIR Filter representation of HIRs," presented at the 96th Convention of the Audio Engineering society, Amsterdam, The Netherlands, February 1994.

[8] S. Busson, R. Nicol, B.F.G. Katz, "Subjective investigations of the interaural time difference in the horizontal plane," presented at the 118th Audio Engineering Society Convention, May 2005.

[9] J.-M. Jot, V. Larcher, O Warusfel, "Digital signal processing issues in the context of binaural and transaural stereophony," presented at the 98th Audio Engineering Society Convention, February 1995.

[10] J. Huopaniemi, J. O. Smith, "Spectral and time-domain preprocessing and the choice of modeling error criteria for binaural digital filters", presented at the 16th International Conference of the Audio Engineering Society, Rovaniemi, Finland, April 1999.

[11] J. S. Abel, S. H. Foster, "Measuring HRTFs in a Reflective Environment," International Community for Auditory Display, November 1994.

[12] J. O. Smith, "Spectral audio signal processing", March 2007 Draft, `//ccrma.stanford.edu/~jos/sasp/`, online book, accessed July 2008.

[13] R. S. Woodworth, G. Schlosberg, "Experimental psychology", Holt, Rinehard and Winston, New York, pp.349-361, 1962.