On the Extraction of Interaural Time Differences from Binaural Room Impulse Responses

Zur Extraktion von interauralen Laufzeitdifferenzen in binauralen Raumimpulsantworten

Studienarbeit

durchgeführt am:	Fachgebiet Audiokommunikation, Institut für Sprache und Kommunikation
vorgelegt von:	Jorgos Estrella
Studiengang:	Elektrotechnik
Matrikelnummer:	228620
Gutachter:	Prof. Dr. Stefan Weinzierl
	M.A Alexander Lindau
Abgabedatum:	13. September 2010

Eidesstattliche Erklärung

Ich versichere hiermit, dass ich meine Studienarbeit mit dem Thema:

> *On the Extraction of Interaural Time Differences from Binaural Room Impulse Responses*
> *Zur Extraktion von interauralen Laufzeitdifferenzen in binauralen Raumimpulsantworten*

selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt.

Berlin, den 13. September 2010

JORGOS ESTRELLA

Abstract

A common technique for binaural reproduction is to convolve anechoic audio with head related impulse responses (HRIR), thus providing the user with the spatial auditory cues required for a realistic listening experience. Although HRIRs depend on the physical structure of the subject's pinna, head and torso, which implies that they are different among individuals, it is in practice not feasible to conduct individual measurements. It is therefore necessary to re-use sets of so called non-individualized binaural transfer functions.

Degradation of the localization accuracy (i.e. constant localization offsets) and instability of the sound sources' location during head movements may occur as one consequence of using non-individualized binaural transfer functions, other kinds of degradation are related to spectral coloration. Nevertheless, relevant localization cues of the binaural dataset might be affected in order to improve plausibility of the auditory experience.

The ITD (interaural time difference) in pairs of head related transfer functions (HRTFs) is exploited for horizontal localization sound sources within the frequency range below approx. 1.5 kHz (Strutt 1907).

An approach to ITD individualization (see sec. 1.1) will be developed. In this context, the extraction of minimum-phase impulse responses out of HRIR has become an specially important topic on this work.

Contents

List of Figures

1. Motivation

Lord Rayleigh in it's Duplex Theory of human sound localization (Strutt 1907) stated that the most important cues for spatial hearing are the interaural level difference (ILD), caused by shadowing on the head, torso and pinnae and the interaural time difference (ITD) being the arrival time difference of a sound observable at left and right ears. Both cues are embedded in the binaural transfer function, because it is the complete description of the acoustic transfer paths from a sound source to the ears.

The ILD is related to localization in the median plane, and, for frequencies above ca. 1,5 kHz to localization in the horizontal plane. The ITD is more relevant in the horizontal plane and for frequencies below 1,5 KHz(Mills 1958). Above that frequency phase ambiguities disturb interpretation of arrival time differences in terms of a unique direction of incidence. For the binaural synthesis those transfer functions are convolved with anechoic audio in order to reproduce a realistic auditory experience at the listener's ear drums.

Since ITD and ILD are closely related to physiological characteristics, the use of non-individual HRTFs has a significant influence on the authenticity of the auralization. The use of non-individual ILD mostly affects the tone color (*timbre*) but as absolute memory for tone color is weak this arises as an issue typically only in direct comparison to real sound fields.

Opposing to that, a non-individual ITD produces the more obvious effect of instability of the sound sources and constant localization errors. In that case no adaption occurs and this artifact is also noticeable without direct comparison with real sound sources.

In head-tracked (dynamic) binaural synthesis a misaligned ITD causes a displacement of the sound sources in the same direction of the head's movement may be perceived, if the model's head for the data acquisition was smaller than the user's one, or in the opposite direction if the model's head was bigger (Algazi et al. 2001b).

The purpose of this work is to asses the behavior of empirical ITDs that occur when using a dummy-head based auralization approach. Research shall indicate solutions for the individual customization of the ITD, as this is expected to improve the auditory experience.

1.1. Intended solution for ITD individualization

An individualization approach based on the decomposition of the binaural room impulse responses (BRIRs) into minimum-phase impulse responses and and a variable delay line (VDL) - equivalent to the ITD - is proposed.Minimum phase impulse responses would replace the original HRIRs in the convolution process, while the interaural time differences are reinserted in form of a time delay between left and right ears, scaled by an individualized factor. Figure 1.1 shows a flow diagram of the model.

This approach entails several advantages which are shortly discussed in the following.

1.2. Latency reduction

Since the modified IR dataset has become shorter (the initial delay is now close to zero), there are less samples to process at the convolution stage, thus, resulting in reduction of latency and processing charge.

1.3. Artifact free cross-fading

The use of minimum phase IRs in the cross-fading stage avoids comb filtering (Wefers 2007) due to the addition of coherent time delayed signals, thus, improving the overall sound quality.

1.4. Separate processing of time and spectrum

Since interaural delay and binaural spectra are handled as separate processes they can work at different spatial resolutions. This would allow to record the head related impulse re-

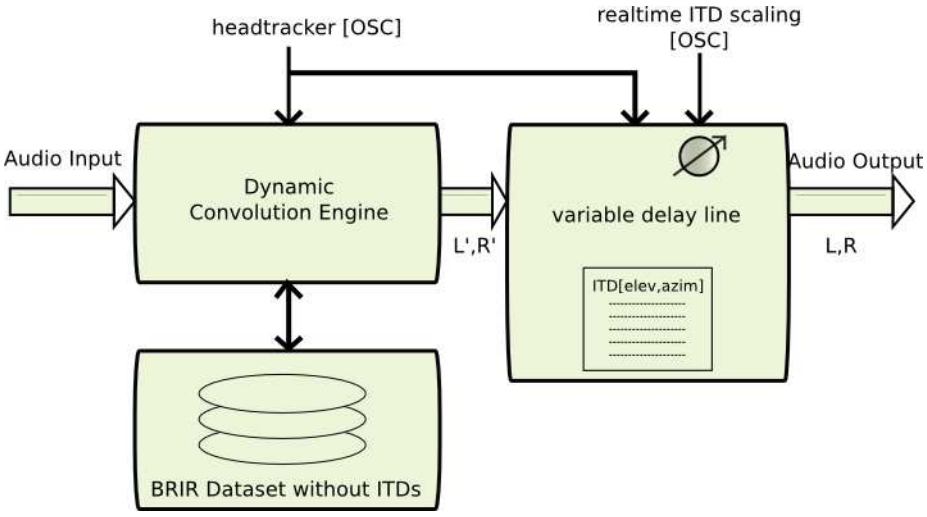

Figure 1.1.: Simplified schematic of the proposed individualization model

sponses with coarser resolution, while the temporal characteristics could be provided at a finer resolution (i.e. by means of interpolation). This could also reduce the memory requirements for the impulse responses.

1.5. Scope of this work

It has been explained that the use of non-individual HRIR/BRIRs can lead to artifacts in binaural reproduction systems and that the individualization of the ITD would solve the issue of sound source instability at head movements, thus, improving the listening experience. In sections 1.2 to 1.4 further advantages of the proposed individualization model were explained. Within this context, the present work is structured as follows:

- Chapter 2 reviews the methods for the separation of temporal and spectral characteristics on HRIRs and BRIRs from the system theory point of view.
- Chapter 3 reviews some individualization approaches in the literature for binaural synthesis using anthropometry and geometric head models. Those approaches are though not suitable for dataset-based auralization where the ITD can not be synthesized to fit a pre-defined source position but rather has to be estimated a-posteriori from data sets.
- In order to find a method, suitable for application in the model presented on section 1.1 several ITD estimation methods are covered on Chapter 4 .

- The proposed individualization model requires the extraction of minimum-phase impulse responses out of the original binaural dataset. In Chapter 5 two methods for the extraction of minimum-phase impulse responses are evaluated perceptually.
- Comparisons of FABIAN's (the head and torso simulator employed for the binaural dataset acquisition at the Audio Communication Institute of the TU-Berlin (Lindau 2006)) ITD to that of larger empirical samples are shown on appendix A.
- A comparison with between FABIAN's ITD and the synthetic ITD generated with geometrical models is presented on appendix B.
- Appendix C present Matlab's™ code for the ITD estimation method found to be most suitable for our ITD individualization model.
- Appendix D shows the user interface developed in order to perform the listening tests of Chapter 5.

2. Theoretical background

The individualization model of figure 1.1 requires that time and spectral components of an HRIR are treated separately. In this chapter, the theoretical background of this decomposition will be described.

2.1. Separation of binaural localization cues

Head-related transfer functions can be treated as linear time-invariant (LTI) systems. In LTI system theory the complex frequency response of a transfer function can also be expressed in terms of magnitude response and phase response. In the case of HRTFs, the phase can be split in a minimum phase component and an excess-phase component.

$$H(j\omega) = |H(\omega)| \cdot e^{j\varphi_{min}(\omega)} \cdot e^{j\varphi_{excess}(\omega)} \tag{2.1}$$

The frequency dependent excess-phase component can also be decomposed into linear-phase and all-pass components.

$$H(j\omega) = |H(\omega)| \cdot e^{j\varphi_{min}(\omega)} \cdot e^{j\varphi_{lin}(\omega)} \cdot e^{j\varphi_{allpass}(\omega)} \tag{2.2}$$

Since the sensitivity to phase spectra on humans low is (Preis 1982), the all-pass component can be neglected without disturbing the spatial perception (Minnaar et al. 1999) as has been shown for HRIRs that the contained all-pass component is inaudible for most directions of sound incidence.

$$H(j\omega) = |H(\omega)| \cdot e^{j\varphi_{min}(\omega)} \cdot e^{j\varphi_{lin}(\omega)} \tag{2.3}$$

Moreover, the linear-phase component on equation 2.3 can be replaced by a time delay without audible consequences as long as it adequately approximates the ITD (Kulkarni et al. 1999).

Figure 2.1 shows an example two IRs, both having the same frequency response but different phase responses.

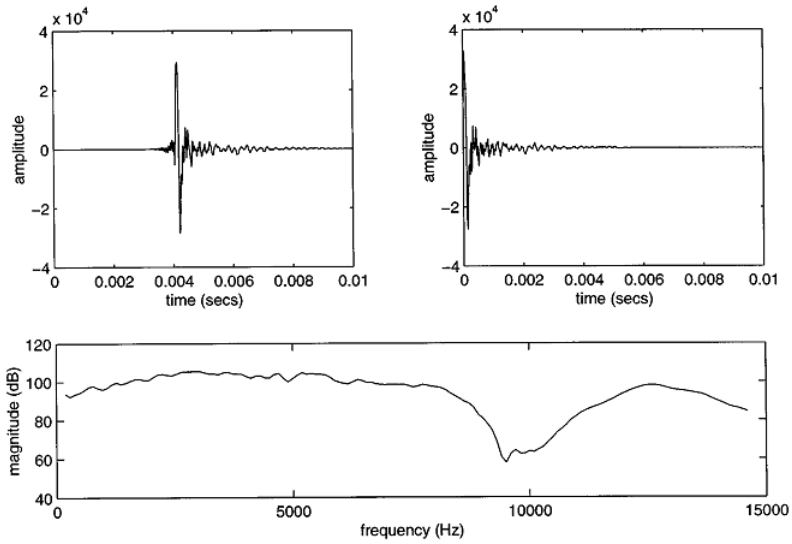

Figure 2.1.: Up: HRIRs with and without excess phase components. Below: frequency response of both HRIRs. From Kulkarni et al. (1999)

The application of this theory in our model requires effective methods for the extraction of the minimum phase impulse responses and the interaural time difference. Approaches discussed in Chapters 4 and 5 are mostly based on this foundations.

3. Overview of individualization approaches in binaural synthesis

3.1. The problem of using non-individualized HRIRs in binaural synthesis

The use of non individual impulse responses in binaural synthesis was widely discussed in the past years (Wenzel et al. 1988, 1993; Møller et al. 1996; Algazi et al. 1997). The most remarkable problems can be assigned to one of two categories:

- **Tone colour variation** given by non-individualized ILD having different spectral characteristics. As subjects may adapt to spectral coloration this issue might be less critical.
- **Localization errors** due to a non-individual ITD, are on the contrary more disturbing since they cause instability of the virtual sound sources on head tracked systems. Algazi et al. (2001b) mentions the annoying issue of sound sources slightly moving in the same direction as the listener's head if the artificial head used in the data acquisition had a smaller radius, or in the opposite direction if the artificial head had a bigger radius.

As mentioned on chapter 1, the manipulation of the ITD is used as framework for the analysis in this work.

3.2. Individualization using geometrical models

The need to affect the spatial cues has lead for many investigators to try to synthesize HRTFs and relate it's characteristics to anthropometric parameters in order to achieve individualization.

Woodworth et al. (1972) developed a formula for predicting the high frequency ITD based on just one anthropometric parameter, the head radius and the azimuthal position of the

sound source. This formula (eq. 3.1) takes account of the diffraction of a plane wave around the sphere:

$$ITD = \frac{a}{c}(\sin\theta + \theta) \tag{3.1}$$

a = head radius
c = speed of sound
θ = azimuth angle in [Rad] $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$

Larcher und Jot (1999) extended formula 3.1 to include the elevation dependency of the ITD and to cover the whole horizontal and frontal planes:

$$ITD = \frac{a}{c}(\arcsin(\cos\theta\sin\phi) + \cos\phi\sin\theta) \tag{3.2}$$

θ = azimuth angle in [Rad] $-\pi < \theta < \pi$
ϕ = elevation angle in [Rad] $-\frac{\pi}{2} < \phi < \frac{\pi}{2}$

Savioja et al. (1999) also extended Woodworth's formula to an equation better fitting their empirical data:

$$ITD = \frac{a}{c}(\sin\theta + \theta)\cos\phi \tag{3.3}$$

These approaches are though only applicable in dataset based binaural systems when the position of the sound sources is known. Appendix B analyzes and compares these methods.

3.3. Individualization aided by anthropometry

All mentioned geometrical models use the head radius as individualization parameter. In order to apply them, different methods have been proposed to derive a suitable estimation of the head radius.

Algazi developed an empirical formula to provide an optimal head radius for its use with Woodworth's ITD model (Algazi et al. 2001b). Its equation is based on three anthropometric measures: head width, head height and head depth (X1, X2 and X3 respectively in

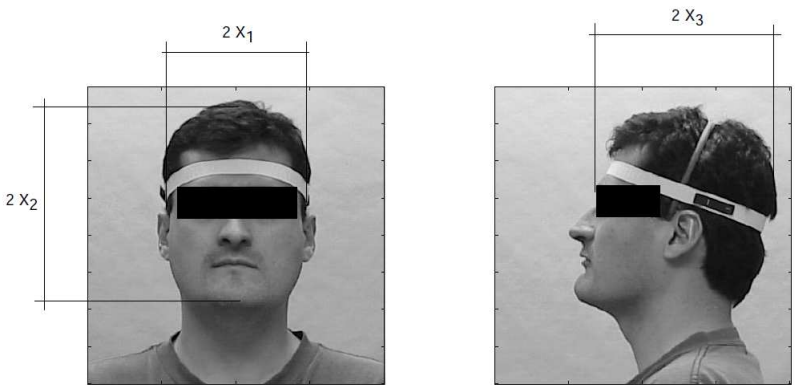

Figure 3.1.: Anthropometric measures used to find the optimal head radius in Algazi et al. (2001b).

figure 3.1).

HRTF recordings conducted for 25 subjects male and female, Caucasian and Asian were used in this method. Least squares fitting between the measured ITD[1] and the ITD produced by the Woodworth's formula was applied delivering a model of the optimal head radius for each subject. For predicting this optimal head radius for a subject whose HRIRs are unknown, a three-parameter linear model was considered for regression:

$$a_{opt} = W_1X_1 + W_2X_2 + W_3X_3 + b[cm] \tag{3.4}$$

With:
X_1= head width/2
X_2= head height/2
X_3= head depth/2

By means of multiple linear regression of the individual optimal head radii on the 25 subjects' head-dimensions an empirical formula for predicting a generic optimal head radius was achieved.

$$a_{opt} = 0.51X_1 + 0.019X_2 + 0.18X_3 + 3.2[cm] \tag{3.5}$$

[1]The method for the ITD extraction used by Algazi's research team was the onset detection. For more details on this method see chapter 4 section 4.3

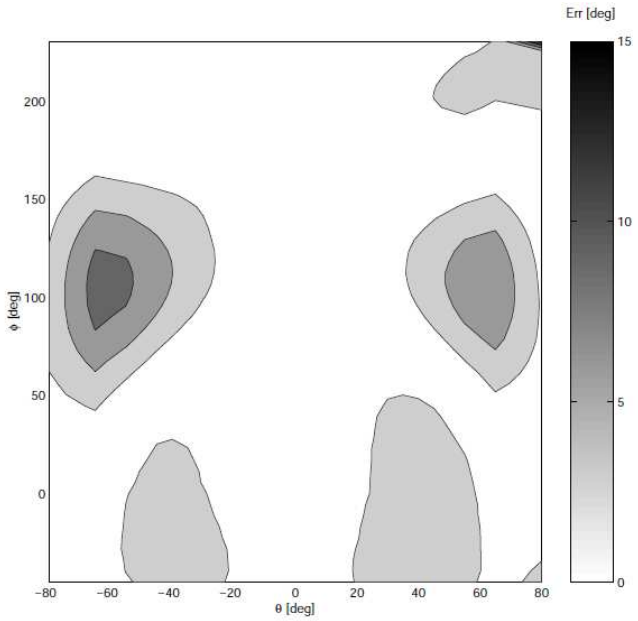

Figure 3.2.: Mapping of the average angular error of the optimal head radius. From: Algazi et al. (2001b).

Figure 3.2 shows the average angular error of the optimal head radius using the Woodworth-Schlosberg formula. It should also be mentioned that no perceptual evaluation validating this approach was presented by the researchers. Though, it can be found in the work of Busson (Busson et al. 2005) (see figure 3.3), where the method was shown to underestimate the perceptual ITD.

3.4. Chapter's resume

In this chapter the problem of using non-individualized binaural cues was discussed. ITD individualization approaches based on geometrical models were also reviewed.

It has been explained that the geometrical models, derived from the Woodworth-Schlosberg formula, require the head-radius as individualization parameter as well as the position of the sound sources for generating the ITD. Algazi's anthropometric method for finding an optimal head-radius represent an enhancement on the applicability of the geometric models and at the same time an interesting approach for relating human-head's dimensions to the individualized ITD.

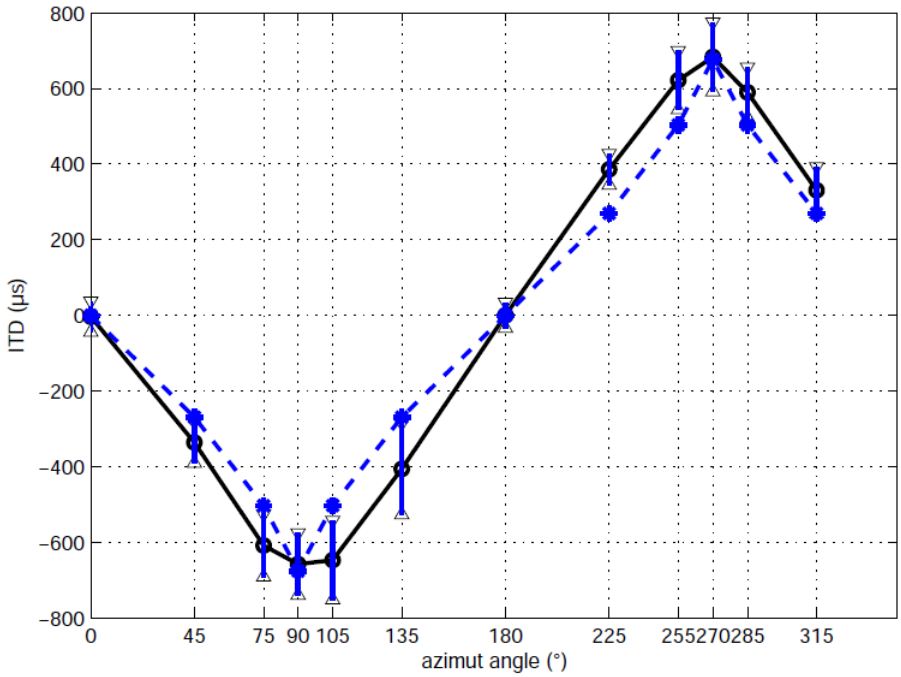

Figure 3.3.: ITD comparison: Perceptually retrieved ITD vs. ITD estimated with the Woodworth-Schlosberg method and Algazi's optimal head radius. Means and standard deviations are plotted with solid lines. Dotted line represent the ITD estimation method. Only the horizontal plane is considered. From: Busson et al. (2005).

It is important to remember that the position of the sound sources has to be known in order to apply the individualization models. However, in data-based auralization the position of the sound sources is mostly unknown. Thus, these methods are not suitable for our purposes; but the procedure of relating anthropometric head measures to the interaural time difference serve as inspiration for our new individualization method.

4. Evaluation of ITD estimation methods

4.1. Introduction

For the system proposed on Fig. 1.1 to be realized, it is important to find reliable and perceptually correct **ITD estimation** and **IR decomposition** methods. This chapter evaluates several ITD estimation methods while Chapter 5 analyzes IR decomposition methods.

Within this scope, the work of Minnaar (Minnaar et al. 2000) is a good starting point as it references several of the currently existing methods, provides graphic comparisons and gives some insights on the applicability. On the other hand, the work of Busson (Busson et al. 2005) presents a subjective evaluation of some of the ITD estimation methods.

Almost all methods treated in this chapter are explained in those papers. The estimation methods considered are grouped in three categories as in Busson et al. (2005):

1. **Cross-correlation methods (CC)** Two methods are considered in this category:
   a) Maximum of the interaural CC between left and right ears.
   b) CC between HRIRs and their minimum phase representations.
2. **Threshold method** A time domain method based on the onset detection in the impulse responses.
3. **Phase methods** Two methods are treated in this category:
   a) Interaural group delay difference at 0 Hz.
   b) Linear phase fitting.

The interaural time differences of data sets recorded with the manikin FABIAN (Lindau 2006) will be used for this analysis.

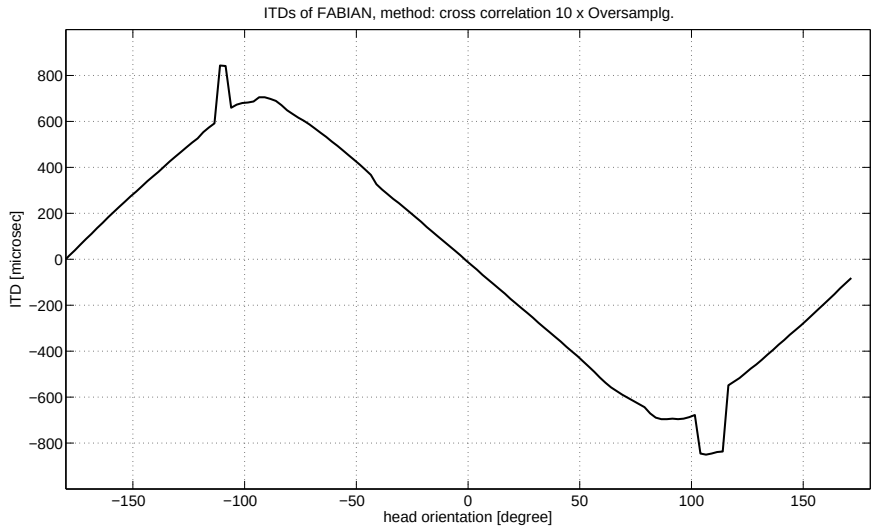

Figure 4.1.: ITD extracted using the maximum of the cross-correlation method with 10x up-sampling. Note the discontinuities at +-110°. Data set: FABIAN's HRTFs

4.2. Cross-correlation methods

4.2.1. Maximum of the interaural cross-correlation (MIACC)

This method consists of cross correlating the impulse responses of left and right ears with each other and measure the time to it's maximum. According to Mills (1958) the threshold for detection of ITD changes is approx. 10μs when the conditions are optimal. For 44100 Hz samplerate, the time difference between one sample to another is already 22μs. Therefore, for appropriate accuracy, the HRIRs should be first up-sampled.

Figure 4.1 shows the results of this method for 10x up-sampling using HRIRs corresponding to the horizontal plane. At some points near -+110° the cross-correlation method seems to give erratic ITD values. This are most probably due to the minor SNR of the contralateral IR and the lack of coherence between ipsilateral and contralateral IR at those angles (Busson et al. 2005).

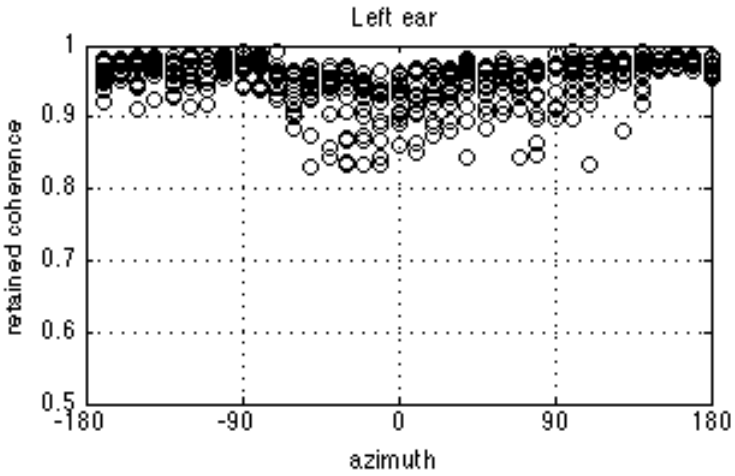

Figure 4.2.: High degree of coherence of an HRIR with its minimum phase version (left ear). From Nam et al. (2008)

4.2.2. Cross Correlation with minimum phase impulse responses

Nam et al. (2008) showed that for the vast majority of HRIRs, the correlation between an impulse response and its minimum phase representation is over 0.9 (see Fig. 4.2). Thus, finding the times until maximum of this type of cross-correlation for left and right HRTFs and subtracting them from each other gives us the ITD.

Figure 4.3 shows an ITD estimation example.

The method presents also discontinuities (around +- 50° to 130°). It also requires a lot of processing time because the extraction of the minimum phase impulse responses and the cross-correlation, are both realized with up-sampled IRs. BRIRs of large rooms which already consist on large vectors become problematic in this sense.

Subjective evaluation of the MAICC when applied on HRIRs can be found in the work of Busson et al. [2] Figure 4.4 shows that this ITD estimation method fits between the standard deviations of the subjective ITD. Thus, it might be perceptually appropriate.

[2] On that listening test, an auralization unit consisting on a minimum phase IR and a pure delay was employed to generate the perceptual ITD. On the experiment the subjects had to match the delay (with 22μs of resolution) until it resembles the reference (auralization using the own raw HRIR). At the end the estimations methods were compared to the generated perceptual ITD. For more details see Busson et al. (2005).

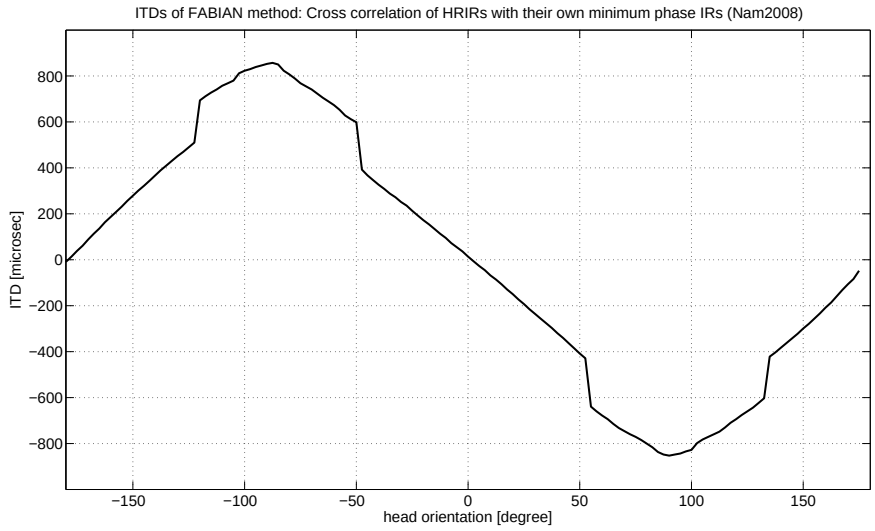

Figure 4.3.: ITD estimation by cross-correlation of HRIRs with their minimum phase versions. Note the discontinuities around the ipsilateral and contralateral azimuth angles. Data set: HRTFs FABIAN

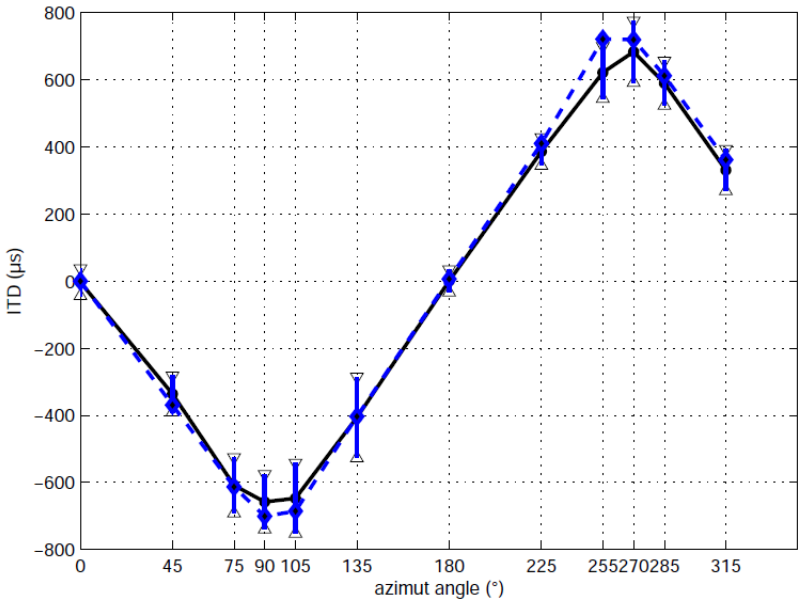

Figure 4.4.: Subjective ITD vs. ITD extracted with the IACC method. Means of subjects answers and standard deviations are plotted with continuous lines. Dotted line represent the ITD estimation method. Only the horizontal plane is considered. From Busson et al. (2005).

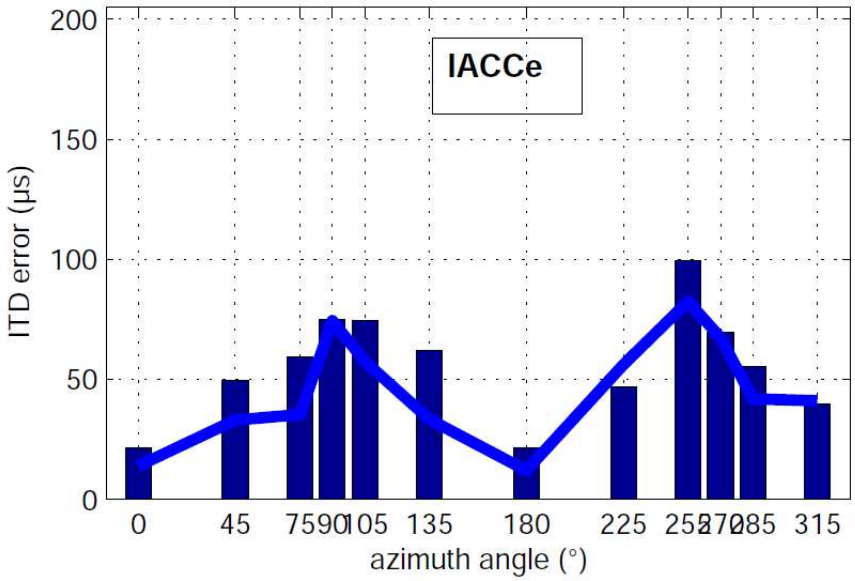

Figure 4.5.: Means of absolute errors between subjective ITD and ITD extracted with the IACC method as a function of azimuth angle. From Busson et al. (2005).

The estimation error in [μs], as a function of azimuth angle computed according equation 4.1 is plotted on figure 4.5. The maximum (100μs) takes place at 255°.

$$EC(\theta) = \frac{1}{N}\sum_{i=1}^{N} |ITD_{psych}(\theta,i) - \widehat{ITD}(\theta,i)| \tag{4.1}$$

where:
ITD_{psych} is the psychoacoustic ITD and $\widehat{ITD}$ the estimated ITD.
θ is the azimuth angle and i, N are subject number and amount of subjects ($N = 11$)respectively.

4.3. Onset detection

This method, also known as edge detection, measures the time in samples up to a given threshold in the left and right onsets of the binaural IRs (ie. 10% of the peak in Minnaar et al. (2000)). The ITD equals the difference between the times found.

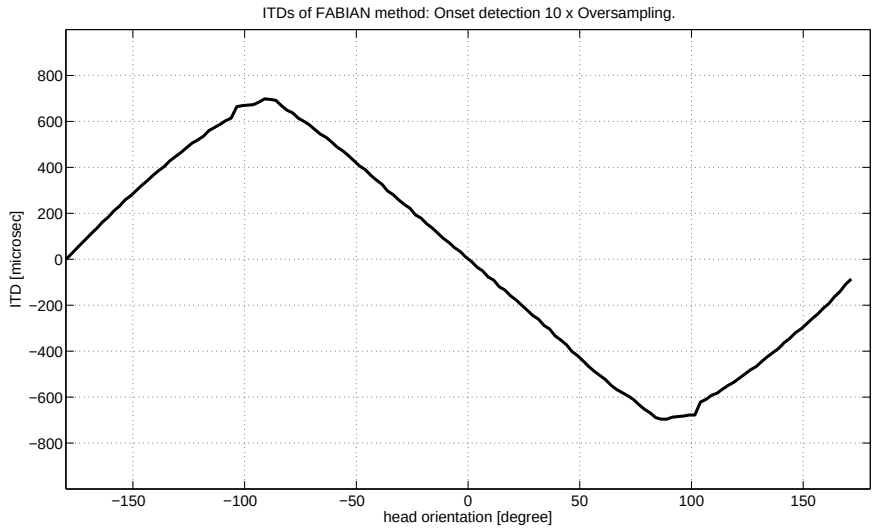

Figure 4.6.: ITD extracted using the onset detection method with 10x up-sampling, threshold -3dB. Data set: FABIAN's HRIRs

Figure 4.6 shows an ITD estimation example. The IRs should be up-sampled for appropriate accuracy.

Visual inspection of the data set should help finding an appropriate threshold. The ITD in BRIRs is reliably detected when using thresholds of -20 to -40 dB of the maximum peak.[3] In figure 4.7 the onsets of two HRIRs data sets are plotted. Note the different onset characteristics.

This estimation method performs quite fast and robust but it depends on the chosen threshold, thus not all all-pass components can be extracted with it.[4].

The performance of this method compared to the perceptual ITD can be seen on figure 4.8. As for the previous method the estimation fits best to the perceptual ITD. This method might be suitable for our individualization model too.

The estimate error as a function of the azimuth angle (equation 4.1) can be read in figure 4.9. At 105° the error reaches 80μs, its maximum value.

[3] In appendix C the Matlab™ code for computing the ITD with the edge detection method can be found

[4] Minnaar et al. (2000) mentions that all-pass components might be audible if they are larger than 30μs. On Chapter 5 the subjective performance of this method will be analyzed.

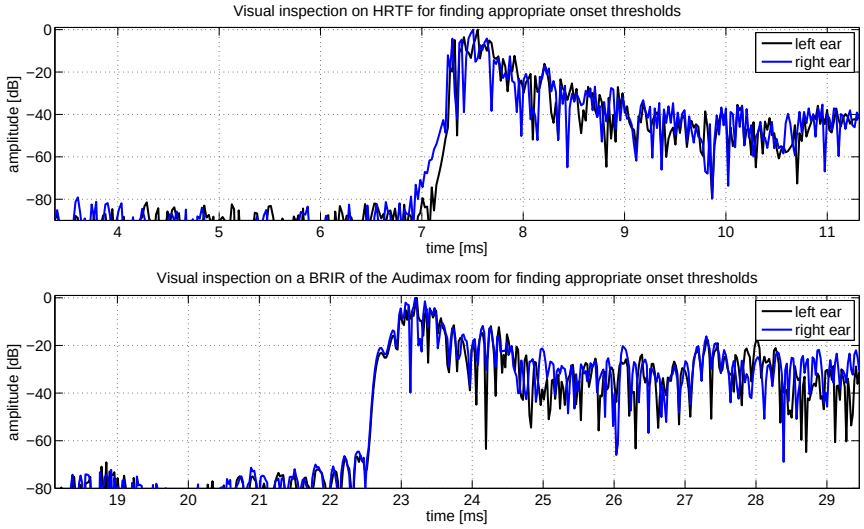

Figure 4.7.: Visual inspection required in the onset detection method. Note the different rise-up characteristics and noise levels on the onsets. Data set: FABIAN's HRTFs recorded at the anechoic room of the TU-Berlin and BRIRs recorded at the Audimax hall of the TU-Berlin.

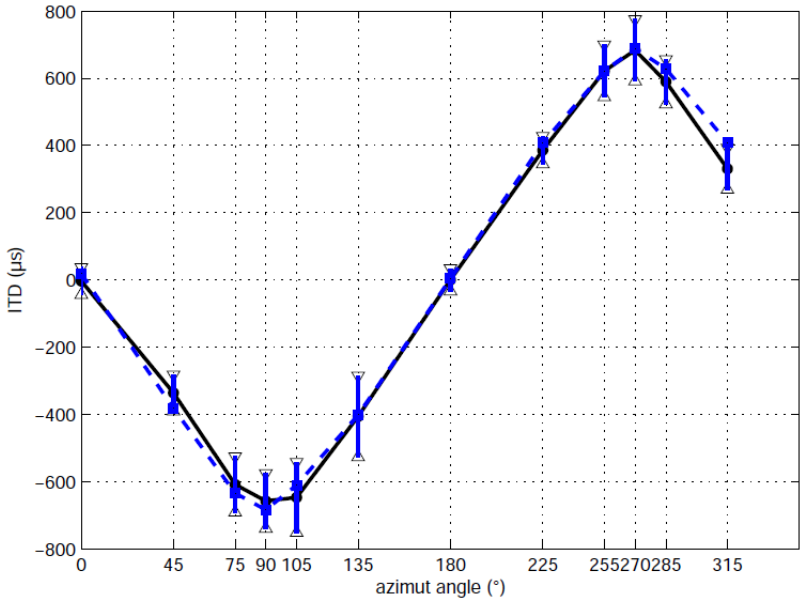

Figure 4.8.: Subjective ITD vs. ITD extracted with the Edge Detection method. Means of subjects answers and standard deviations are plotted with continuous lines. Dotted line represent the ITD estimation method. Only the horizontal plane is considered. From Busson et al. (2005).

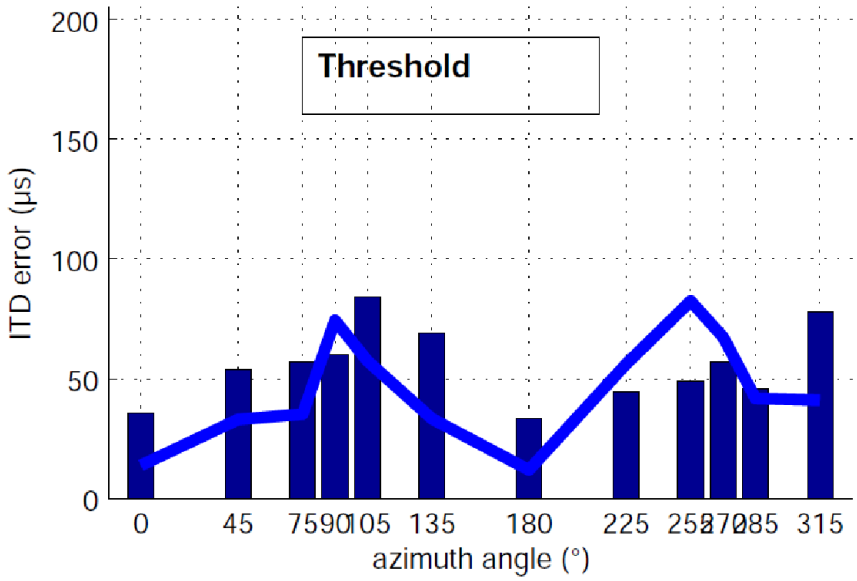

Figure 4.9.: Means of absolute errors between subjective ITD and ITD extracted with the edge detection method as a function of azimuth angle. From Busson et al. (2005).

4.4. Phase methods

4.4.1. Interaural group delay difference at 0Hz, (IGD_0)

As explained in Chapter 2, a HRTF can be decomposed in minimum-phase and excess phase component. Here, the ITD is the interaural group delay difference of the excess phase components evaluated at 0Hz.

In the work of Minnaar et al. (2000) four methods for achieving this task are briefly described. The method we have chosen is based on the following steps:

a) Calculate the group delay of an HRTF pair and the group delay of it's minimum phase representation[5].
b) Subtract them from each other to obtain the group delay of the excess phase components.
c) The difference between the values obtained for left and right ears (the interaural group

[5] Minnaar et al. (2000) computes first the unwrapped phase response of the original impulse response and subtracts from it the unwrapped phase of the minimum phase impulse response. Applying the derivative (gradient), the group delay of the excess component is found. In our approach the group delays are computed using Matlab's `grpdelay` function.

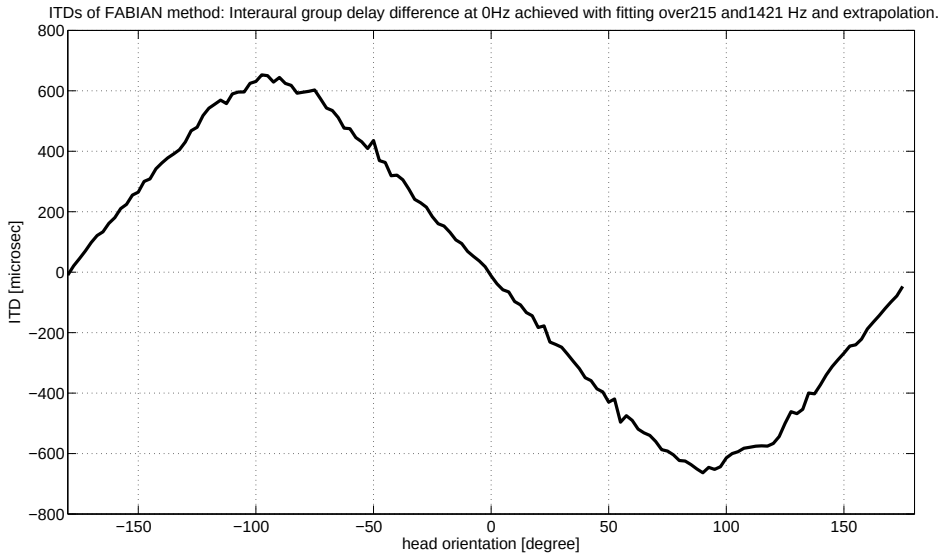

Figure 4.10.: ITD estimation using the interaural group delay difference at 0 Hz. Data between 215 and 1421 Hz used for extrapolation. Data set: FABIAN's HRTFs

delay difference) evaluated at 0 Hz is the ITD.

However as binaural data sets are recorded using real electro acoustical transducers (loudspeakers and microphones) as well as AD converters utilizing DC-blockage, thus, not providing any useful information at 0 Hz (DC). One approach to overcome this problem is to employ extrapolation using as reference data of a frequency range below 1,5 kHz, where according to Minnaar et al. (2000) the group delay should be almost constant.

Fig. 4.10 shows an ITD estimation example where a frequency range of 215 Hz to 1421 Hz is used for the extrapolation. This method has the disadvantage of being highly dependent on the frequency range chosen and requires a lot of computation time with longer impulse responses.

4.4.2. Phase delay fitting

This method was first proposed on Jot et al. (1995), it assumes that the excess phase of an HRTF is a linear function of frequency until 8 to 10 kHz. Since the all-pass components on a HRTF can be replaced with a pure delay, this delay can be calculated by fitting a

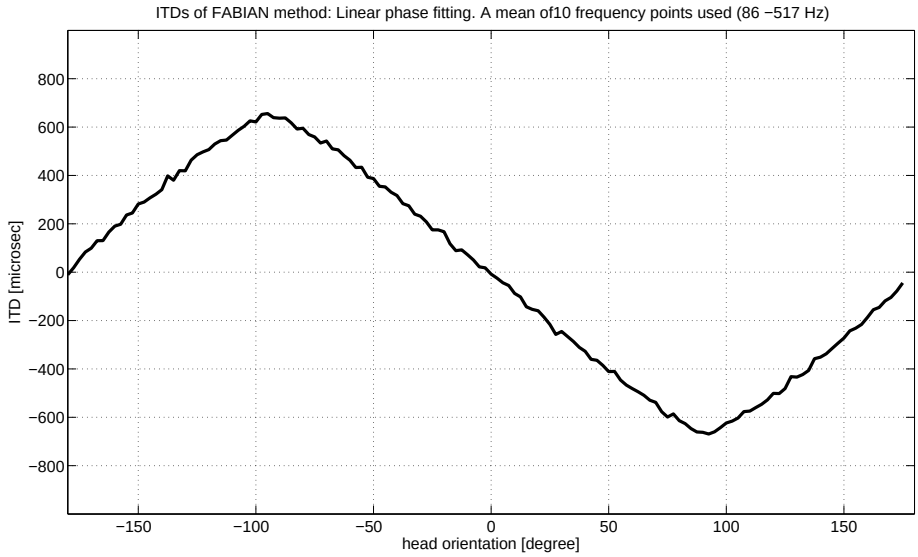

Figure 4.11.: ITD estimation using phase delay fitting. Data between 83 and 500 Hz was used for fitting. Data set: FABIAN's HRTFs

linear curve on the excess-phase response between 1 kHz and 5 kHz for left and right ears and computing the difference. Huopaniemi und Smith (1999) proposed another frequency range, 500 Hz to 2 KHz. While Minnaar et al. (2000) states that the phase can only be linear as a function of frequency for frequencies below 1.5 KHz.

Figure 4.11 shows an ITD estimation example.

The perceptual performance of Jot's method according to Busson et al. (2005) tells us that the estimation fits well at almost all frequencies except at lateral locations (see figure 4.12), where it departs from the subjective values more strongly than the IACC and Edge Detection methods.

In figure 4.13 this aspect can clearly be seen. The error as a function of azimuth reaches as much as 200μs for 105° and 255°.

On figure 4.14 the group delays of left and right HRTFs are plotted as an example of the critical role on the frequency range selection of the phase methods discussed in this section. Note the non-constant characteristic of the group delays.

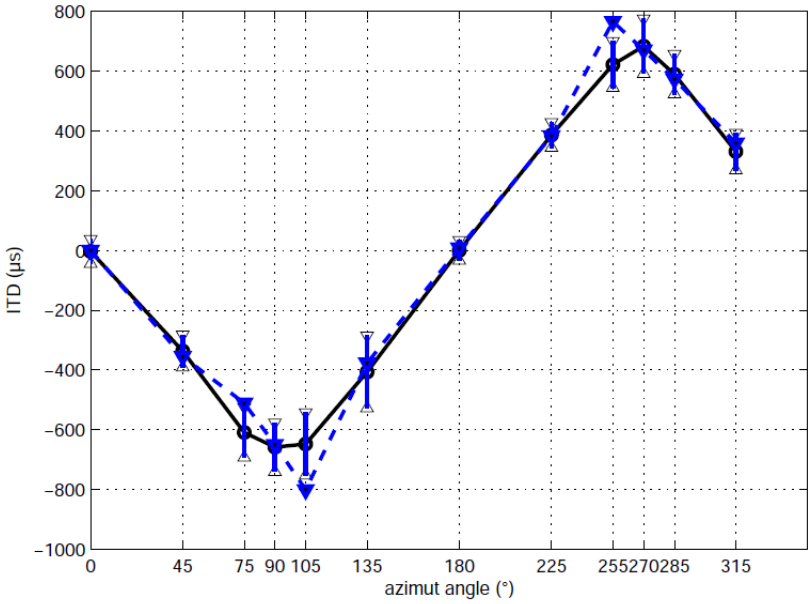

Figure 4.12.: Subjective ITD vs. ITD extracted with the Linear Phase Fitting method. Means of subjects answers and standard deviations are plotted with continuous lines. Dotted line represent the ITD estimation method. Only the horizontal plane is considered. From Busson et al. (2005).

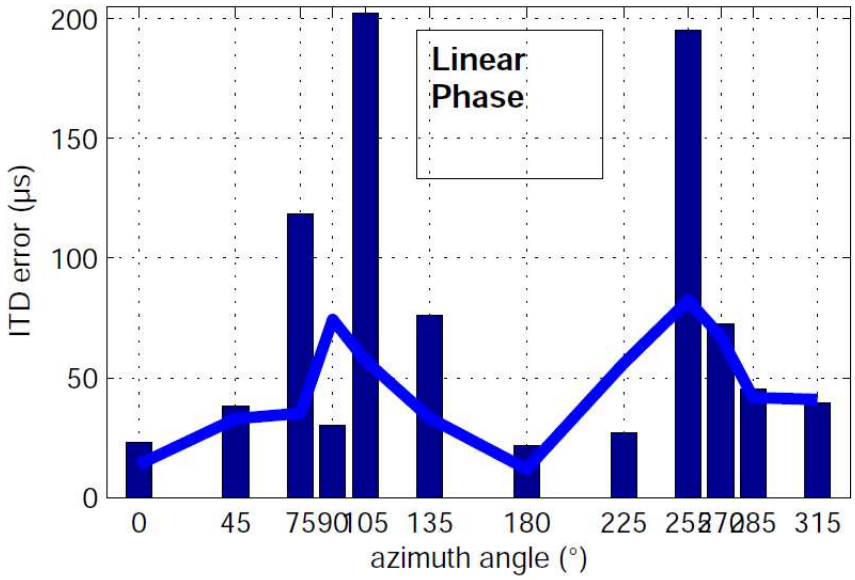

Figure 4.13.: Means of absolute errors between subjective ITD and ITD extracted with the linear phase fitting method as a function of azimuth angle. From Busson et al. (2005).

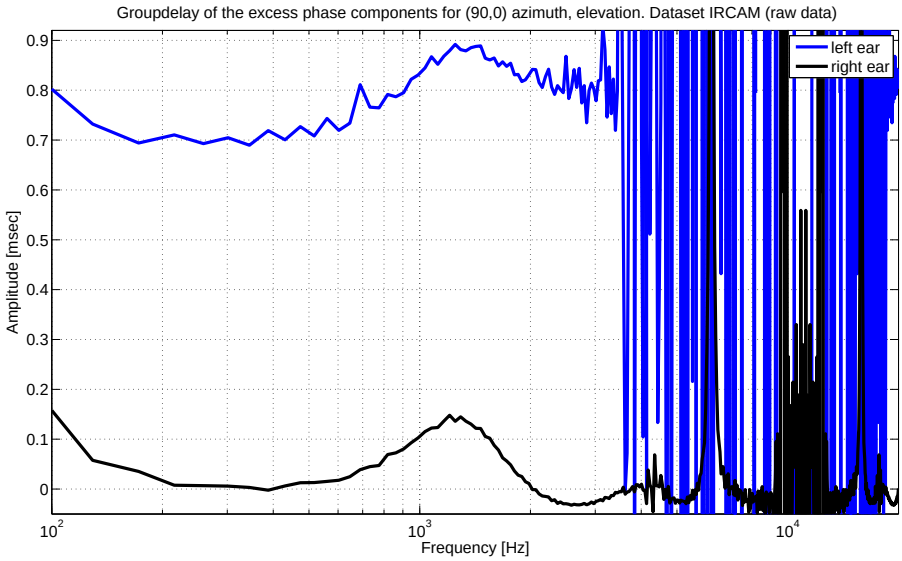

Figure 4.14.: Groupdelay of the excess phase components from an HRTF pair.Data set IRCAM (90,0) azimuth elevation.

Figure 4.15 shows a fitting example in a frequency range of 500 Hz to 1500 Hz, using an HRTF pair of a subject from the IRCAM's public database. Note that the lines are not exactly parallel.

The frequency dependency on figure 4.15 can also clearly be seen, meaning that the ITD obtained by this methods varies according to the frequency evaluation range. In the work of Algazi et al. (2001a), it is also mentioned that phase related methods are problematic because of reflections and resonances of torso and pinnae causing unpredictable phase responses on the HRTFs.

4.5. Chapter's resume

ITD estimation methods according to three categories (cross-correlation, threshold detection, phase difference) were assessed in this chapter using this validation criteria:

- Reliability and applicability in the proposed binaural individualization model.
- Perceptual correctness (according to results from literature).

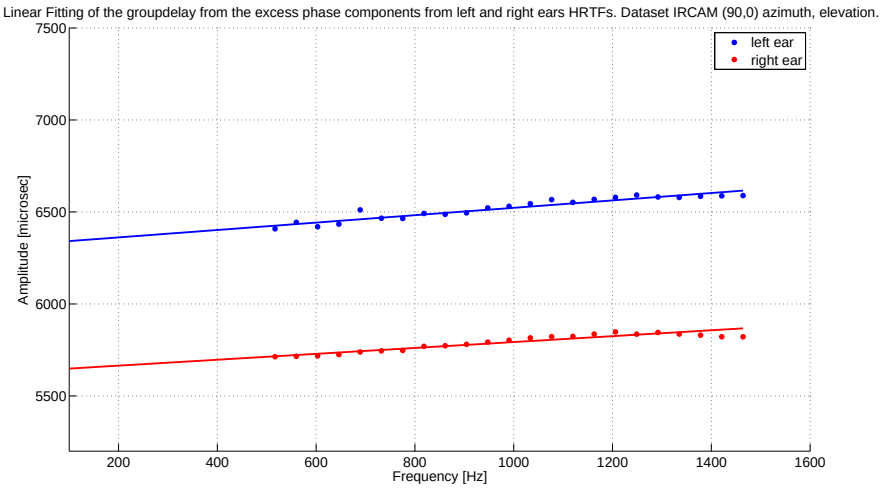

Figure 4.15.: Linear fitting of the group delays from the excess phase components of an HRTF pair. Note that the fitted lines are not parallel. Data set IRCAM, subject 38, 90° azimuth, 0° elevation.

Out of the methods analyzed in this chapter the threshold detection method seems to be the most appropriate since it delivers estimations of ITD which are continuous functions of the angle with most kinds of data sets, the method is computationally fast and delivers values which are most similar to the perceptually correct ones. The cross-correlation methods seem to also match well with the perceptual ITD, but the estimation is problematic at lateral locations where it presents some discontinuities. The phase methods are computationally more expensive, not providing better performance as the other methods. Moreover, they appear to lack perceptual fit at positions (around +-110°).

5. Perceptual evaluation of HRIR decomposition methods

The individualization model introduced on Chapter 1 is based on the decomposition of the HRIRs into **a variable delay line simulating the ITD** and two **minimum phase IRs** replacing the left and right ears BRIRs.

A suitable ITD estimation method was discussed already in Chapter 4.

In this chapter two methods for the extraction of minimum phase impulse responses[6] from HRIRs will be analyzed perceptually. According to Julius O. Smith, Hilbert minimum-phase filters compared to causal signals having the same amplitude response, have faster decay as their energy is maximally concentrated towards the beginning (time $\rightarrow$ zero). The perceptual result of this aspect is an important topic in this analysis.

The methods for HRIR decomposition assessed in this chapter are:

- Hilbert transformation based method (Oppenheim et al. 1999) also known as Kolmogorov method of spectral factorization. Obtained using Matlab's™ `rceps` function. Matlab's™ algorithm finds first the real cepstrum of the input signal as:
  `y = real(ifft(log(abs(fft(x)))));`
  The minimum phase impulse response is computed after windowing in the cepstral domain.
  `window = [1;2*ones(n/2-1,1);ones(1-rem(n,2),1);zeros(n/2-1,1)];`
  `min_phase = real(ifft(exp(fft(window.*y))));`

- Threshold method, consists in extracting the impulse response starting at the ITD detection spot [7]. For better accuracy the ITD estimation and the decomposition are realized with 10x up-sampled HRIRs. Note that this method does not extract all all-pass components, thus, the extracted IRs are indeed quasi minimum phase impulse responses. This aspect is though not critical as long as the remaining all-pass components are kept mostly below 30μs (Minnaar et al. 2000).

[6] A minimum-phase filter is a filter that contains all it's poles and zeros inside the unit circle $|z| = 1$ (Oppenheim et al. 1999, on pg. 281).
[7] Using the onset detection as ITD estimation method

5.1. Comparison of ear-weighted minimum phase impulse responses

In order to visually assess perceptual differences between Hilbert minimum-phase IRs and onset minimum-phase IRs, comparisons in different room sizes and IR lengths were realized after applying a weighting on the time signal simulating the inertial behavior of the ear (see Weinzierl 2008, chap. 5). (using 25 ms integration window).

The Hilbert minimum-phase IRs were first zero padded to double length to avoid circular convolution artifacts.

Three rooms of either big, medium and small volumes were considered:

- Audimax hall at the TU-Berlin (volume 8500 m^3).
- TU - Berlin lecture hall H104 (volume 3000 m^3).
- TU - Berlin small Electronic Studio (volume 230 m^3).

The results are displayed on figures 5.1 to 5.3 in form of amplitude plots, and energy time curves.

Besides for room H104, only minor differences among the two methods are visible using this approach. Note that in this section the temporal behavior of the two methods were compared.

5.2. ABX listening test: Minimum-phase impulse responses (Hilbert method) vs original impulse responses

In order to assess if perceptual differences between the Hilbert minimum phase IRs and the original HRIRs can be detected, an ABX listening test was conducted. ABX tests allow to assess whether discrimination between two samples is possible (performance better than chance).

In this listening test the hypothesis ("no audible difference existing") was our H_0 research hypothesis. As the H_0 cannot be proved directly in inferential statistic tests, instead, on tries to neglect a rather small-effect-size H_1, indirectly supporting the H_0 if a small effect can be shown to be absent (Leventhal 1986).

10 subjects participated on the test. Each of them had to listen 14 times to each stimulus, resulting in 42 decisions per subject.

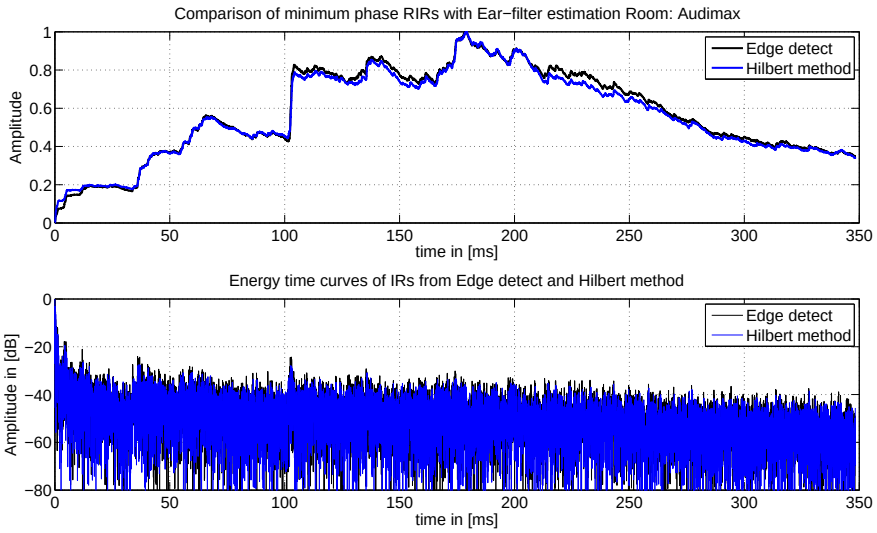

Figure 5.1.: Ear-weighted minimum-phase impulse responses: onset detection vs. Hilbert-transformation method. Room: Audimax hall TU-Berlin

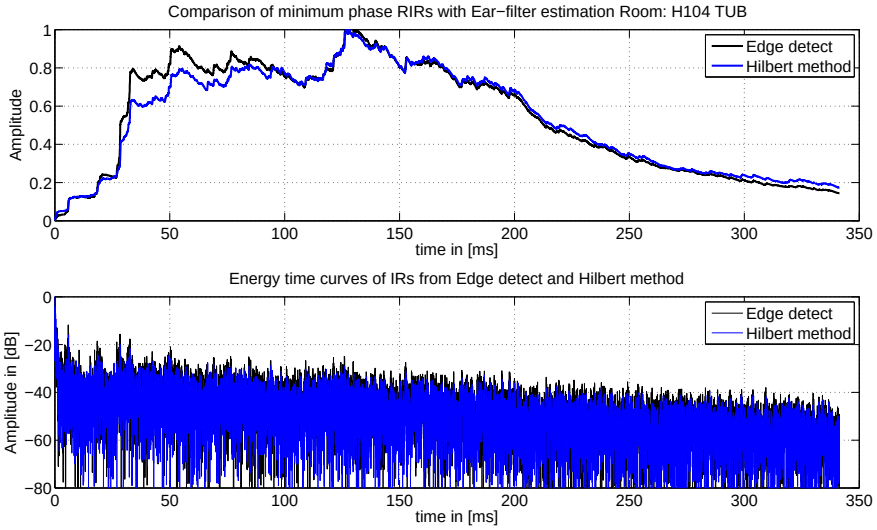

Figure 5.2.: Ear-weighted minimum-phase impulse responses: onset detection vs. Hilbert-transformation method. Room: lecture hall H104 TU Berlin

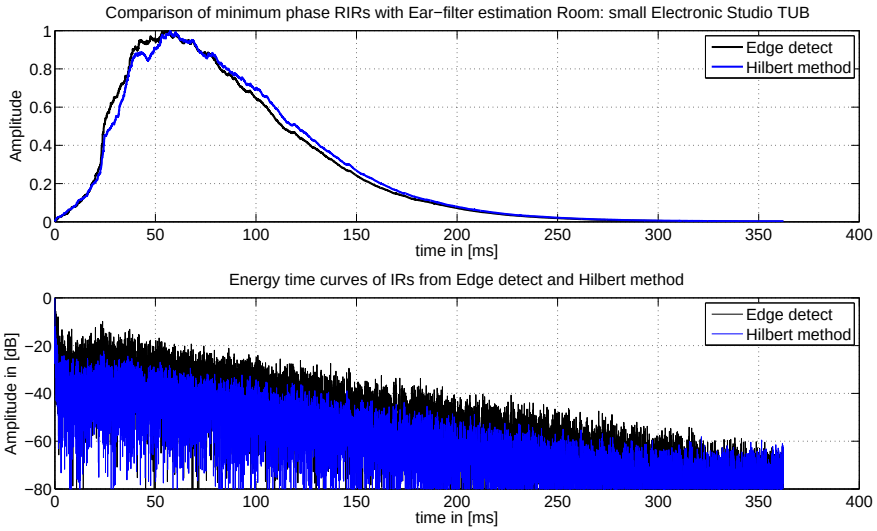

Figure 5.3.: Ear-weighted minimum-phase impulse responses: onset detection vs. Hilbert-transformation method. Room: Small electronic Studio - Tu Berlin

Being limited to this reasonable sample size we could test the existence of at least 75% detection rate per individual on 5% significance level with a test power of 95% [8].

The left and right BRIRs of three rooms: large, medium and small with respectively 1.2, 1.8 and 2 seconds reverberation time, recorded with the HATS FABIAN at 0° azimuth, 0° elevation were used for this ABX-test[9].

The minimum phase impulse responses were extracted using the Hilbert-transformation-based algorithm of Matlab's™ `rceps` function.The test consisted on the comparison between the original BRIRs and the Hilbert minimum-phase BRIRs convolved with a short piece of drum solo as content. This stimulus was chosen because it contains many transients, which are supposed to ease the detection of time domain alterations.

The hypothesis H_0 would be rejected if at least 27 of the 42 decisions were correct. Figure 5.4 shows the results of this test. It can clearly be seen that all participants could easily recognize the Hilbert minimum phase IRs from the original. For half of the subjects the detection rate was above 97% and never sank below 78.4%.

The cues leading to perceptual distinction mentioned mostly were, in order:

- Sound source distance alteration,

[8] The test characteristics were computed using Burtein's approximation formulas (Burstein 1988).
[9] The user interface used can be seen on appendix D.

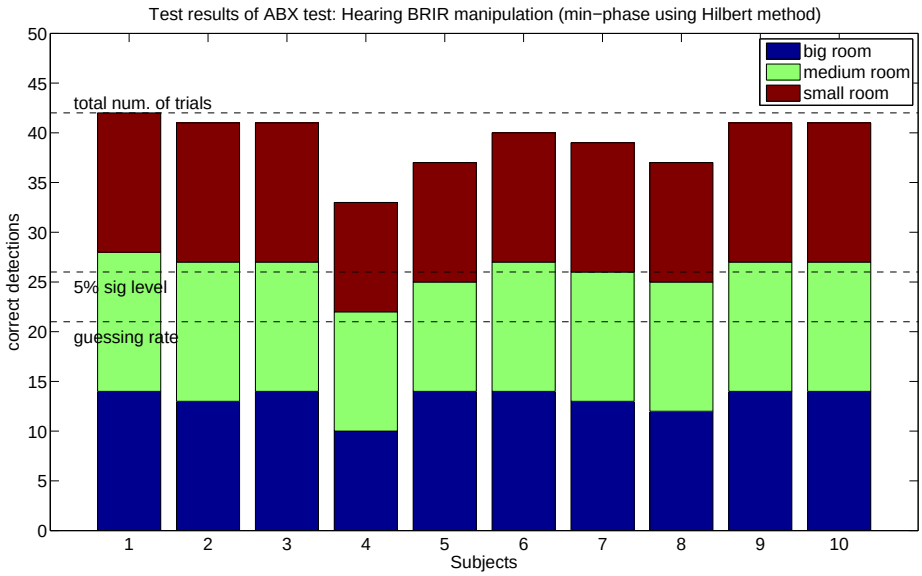

Figure 5.4.: Results of ABX hearing test of minimum-phase IRs (Hilbert method) vs. original impulse responses.

- Compression-like effect and
- Different tone color.

5.3. ABX listening test: Minimum phase impulse responses (onset method) vs original impulse responses

As already mentioned on section 4.3 not all excess phase components can be extracted with the onset method and this could be inaudible as long as they represent less than 30μs. In order to provide information about the detectability of the impulse response manipulation with this method another ABX listening test was conducted. The research hypothesis H_0 ("not hearing any difference") was again tested, via trying to neglect a small-effect-size H_1, using the same effect sizes and significance level as for the previous test.

The hypothesis H_0 would have to be rejected if at least 31 of the 48 decisions were correct. The same 10 subjects of section 5.2 participated in this ABX listening test. Two contents: male speaker and noise bursts, were convolved with the original and the manipulated HRIRs. The task consisted of identifying whether the reference corresponded either to the auralization using the original HRIR or the manipulated HRIR on 48 decisions (8

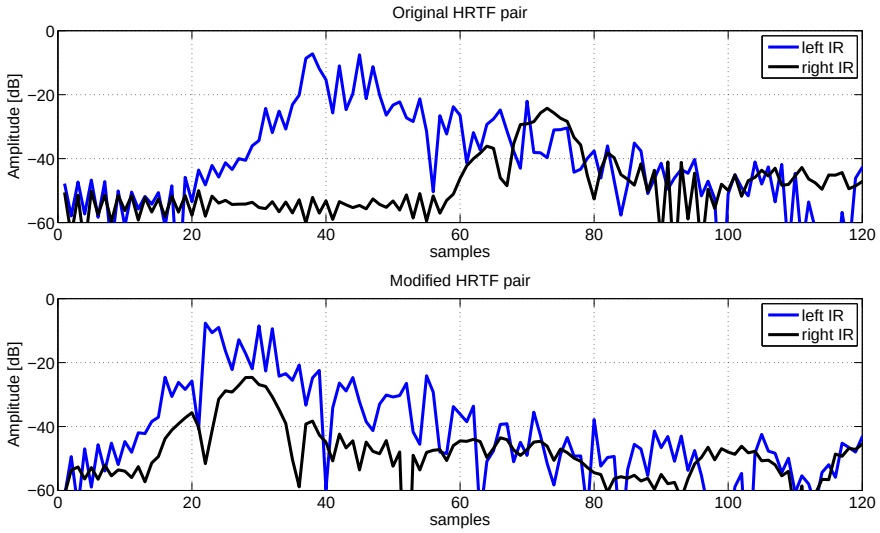

Figure 5.5.: Extraction of quasi minimum-phase impulse responses with the onset detection method. Note that the envelope has slightly changed due to manipulation. It were these kind of differences that were assessed for audibility in the listening test.

auralization directions x 2 contents x 3 runs). For every decision the direction of sound incidence and the audio content were randomized.

To test the performance of the onset method at large ITD differences, HRIRs of IRCAM's public database at 8 auralization directions $[90°, 0°]$, $[90°, 45°]$, $[90°, -45°]$, $[-90°, 0°]$, $[-90°, 45°]$, $[-90°, -45°]$, $[45°, 45°]$, $[-45°, -45°]$, were selected and manipulated as follows:

- Upsampling using a factor of 10.
- ITD detection and extraction (shortening of the impulse response).
- Downsampling to the original samplerate.
- Convolution with audio content.
- Upsampling with a factor of 10.
- Zero padding on one of the IRs to an equivalent ITD.
- Downsampling to the original samplerate.

Figure 5.5 shows the extraction of the quasi-minimum-phase impulse response on a selected contralateral HRTF.

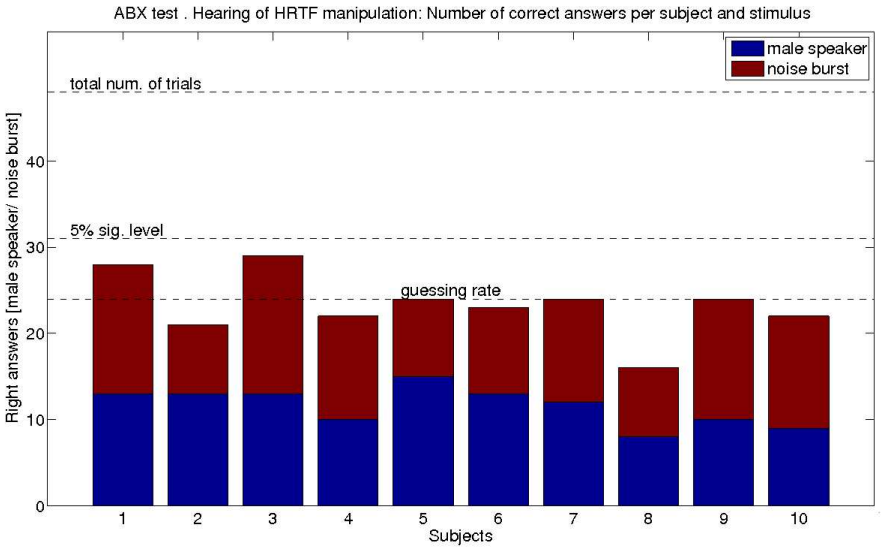

Figure 5.6.: Results of ABX hearing test of minimum-phase IRs (extracted with the onset detection method) vs. original impulse responses.

Figure 5.6 shows the results of this test. None of the subjects was able to reach the 31 correct decisions. This approach can practically be considered as not having obvious audible consequences.

5.4. Chapter's Resume

Two methods for decomposition of HRIRs into minimum-phase impulse responses were quantitatively and perceptually analyzed on this chapter: The onset detection method and the Hilbert transformation based approach.

It has been found that the onset detection offers the best results since it does not introduce artifacts to the impulse responses and the proposed manipulation can be considered as not audible.

The Hilbert transform method on the contrary introduces very noticeable artifacts to the impulse responses and convolved audio-content. Moreover, this method should be not used for auralization purposes since it might degrade the plausibility of the virtual acoustical environment by distorting the original spatial dimensions.

Bibliography

Algazi et al. 1997

ALGAZI, V. R. ; DIVENYI, P.L. ; MARTINEZ, V.A. ; DUDA, R. O.: "Subject Dependent Transfer Functions in Spatial Hearing". In: *IEEE, Proc. of the 40th Midwest Symposium, Sacramento, CA, Aug. 3-6 1997* Bd. 2, 1997, S. 877–880

Algazi et al. 2001a

ALGAZI, V. R. ; DUDA, R. O. ; THOMPSON, D. M. ; AVENDANO, C.: "The CIPIC HRTF Database". In: *IEEE, Proc. of the Workshop on Applications of Signal Processing to Audio and Acoustics, Mohonk Mountain House, New Paltz, NY, Oct.21-24 2001*, 2001, S. 99–102

Algazi et al. 2001b

ALGAZI, V. R. ; AVENDANO, C. ; DUDA, R. O.: "Estimation of a Spherical-Head Model from Anthropometry". In: *J. Audio Eng. Soc* 49 (2001), Nr. 6, S. 472–479

Algazi et al. 1999

ALGAZI, V. R. ; AVENDANO, C. ; THOMPSON, D.: "Dependence of Subject and Measurement Position in Binaural Signal Acquisition". In: *J. Audio Eng. Soc* 47 (1999), S. 937–947

Burstein 1988

BURSTEIN, H.: "Approximation Formulas for Error Risk and Sample Size in ABX Testing". In: *J. Audio Eng. Soc* 36 (1988), Nr. 11, S. 879–883

Busson et al. 2005

BUSSON, S. ; KATZ, B. ; NICOL, R.: "Subjective Investigations of the Interaural Time Difference in the Horizontal Plane.". In: *Proc. of the 118th Convention of the Audio Eng. Soc., Barcelona Spain. Preprint 6324*, 2005

Huopaniemi und Smith 1999

HUOPANIEMI, J. ; SMITH, J. 0.: "Spectral and Time-Domain Preprocessing and the Choice of Modeling Error Criteria for Binaural Digital Filters". In: *Proc. of the 16th International Conference on Spatial Sound Reproduction of the Audio Eng Soc., Rovaniemi, Finland*, 1999, S. 301–312

IRCAM

IRCAM: *Web page of the IRCAM's public HRTF database*. http://recherche.ircam.fr/equipes/salles/listen/system_protocol.html,

Jot et al. 1995

JOT, J. M. ; LARCHER, V. ; WARUSFEL, O.: "Digital Signal Processing Issues in the Context of Binaural and Transaural Stereophony". In: *Proc. of the 98th Convention of the Audio Eng. Soc.* Paris, France, Februar 1995

Kulkarni et al. 1999

KULKARNI, A. ; ISABELLE, S. K. ; COLBURN, H. S.: "Sensitivity of Human Subjects to Head-Related Transfer-Function Phase spectra". In: *J. Ac. Soc. Am.* 105 (1999), Nr. 5, S. 2821–2840

Larcher und Jot 1999

LARCHER, V. ; JOT, J. M.: "Techniques D'Interpolation de filtres Audio-numériques : Application à la Reproduction Spatiale des sons sur Écouteurs". In: *Congrés Français D'Acoustique, Marseille, France*, 1999

Leventhal 1986

LEVENTHAL, L.: "Type 1 and Type 2 Errors in the Statistical Analysis of Listening Tests". In: *J. Audio Eng. Soc* 34 (1986), Nr. 6, S. 437–453

Lindau 2006

LINDAU, A.: *"Ein Instrument zur softwaregestützten Messung binauraler Raumimpulsantworten in mehreren Freiheitsgraden."*. Magister Arbeit, Technische Universität Berlin, 2006

Mills 1958

MILLS, A. W.: "On the Minimum Audible Angle". In: *J. Ac. Soc. Am.* 30 (1958), April, Nr. 4, S. 237–246

Minnaar et al. 1999

MINNAAR, P. ; PLOGSTIES, J. ; CHRISTENSEN, F. ; MØOLLER, H. ; OLESEN, S. K.: "The Audibility of All-Pass Components in Binaural Synthesis". In: *Proc. of the 106th Audio Eng. Soc. Convention, Munich, Germany*, 1999 (4911)

Minnaar et al. 2000

MINNAAR, P. ; PLOGSTIES, J. ; OLESEN, S. K. ; CHRISTENSEN, F. ; MØOLLER, H.: "The Interaural Time Difference in Binaural Synthesis". In: *Proc. of the 108th Audio Eng. Soc. Convention, Paris, France*, 2000 (Preprint 5133)

Moldrzyk et al. 2004

MOLDRZYK, C ; AHNERT, W. ; FEISTEL, S. ; LENTZ, T. ; WEINZIERL, S.: "Head-Tracked Auralization of Acoustical Simulation". In: *Proc. of the 117th Audio Eng. Soc. Convention, San Francisco Ca., U.S.A.*, 2004

Møller et al. 1996

MØLLER, H. ; SORENSEN, F. M. ; JENSEN, B.C. ; HAMMERSHOI, D.: "Do We Need Individual Recordings? ". In: *J. Audio Eng. Soc* 44 Issue 6 (1996), June, S. 451–469

Nagoya

NAGOYA, University: *Web page of the Nagoya University's public HRTF database*. http://www.sp.m.is.nagoya-u.ac.jp/HRTF/,

Nam et al. 2008

NAM, J. ; ABEL, J. S. ; III, J. O. S.: "A Method for Estimating Interaural Time Difference for Binaural Synthesis". In: *Proc of the 125th Audio Eng. Soc Convention, San Francisco Ca., U.S.A.*, 2008

Oppenheim et al. 1999

In: OPPENHEIM, A. V. ; SCHAFER, R. W. ; BUCK, J. R.: *Discrete-Time Signal Processing (2nd Edition) (Prentice-Hall Signal Processing Series)*. 2. Prentice Hall, 1999. – ISBN 0137549202, S. 775–802

Preis 1982

PREIS, D.: "Phase Distortion and Phase Equalization in Audio Signal Processing. A Tutorial Review". In: *J. Audio Eng. Soc* 30 (1982), S. 774–794

Savioja et al. 1999

SAVIOJA, L. ; HUOPANIEMI, J. ; LOKKI, T. ; VNNEN, R.: "Creating Interactive Virtual Acoustic Environments". In: *J. Audio Eng. Soc.* 47 (1999), S. 675–705

Smith

SMITH, J. O.: *"Introduction to Digital Filters with Audio Applications"*. https://ccrma.stanford.edu/~jos/filters/Minimum_Phase_Means_Fastest.html

Strutt 1907

STRUTT, J. W.: "On our Perception of Sound Direction". In: *Philos* 13 (1907), S. 214 –232

Wefers 2007

WEFERS, F.: *"Optimizing Segmented Realtime Convolution"*. Diploma thesis at the RWTH Aachen University, September 2007

Weinzierl 2008

Kapitel 5. In: WEINZIERL, S.: *"Handbuch der Audiotechnik"*. 1. ed., Springer, Berlin. 2008. – ISBN 978–3540343004, S. 187

Wenzel et al. 1988

WENZEL, E. ; WIGHTMAN, F. ; KISTLER, D. ; FOSTER, S.: "Acoustic Origins of Individual Differences in Sound Localization Behavior". In: *J. Ac. Soc. Am.* 84 (1988), Nr. S1, S. S79–S79

Wenzel et al. 1993

WENZEL, E. M. ; ARRUDA, M. ; KISTLER, D. J. ; WIGHTMAN, F. L.: "Localization using nonindividualized head-related transfer functions". In: *J. Ac. Soc. Am.* 94 (1993), Nr. 1, S. 111–123

Woodworth et al. 1972

WOODWORTH, R. S. ; SCHLOSBERG, H. ; KLING, J. W. ; RIGGS, L. A.: *"Woodworth and Schlosberg's Experimental psychology"*. 3d ed. by J. W. Kling and Lorrin A. Riggs and seventeen contributors. Methuen, London„ 1972. – xv, 1279 p. S. – ISBN 0416674607

A. Comparison of FABIAN's ITD with ITDs from public HRTF databases

In this appendix results of comparisons between the ITD of the head and torso simulator (HATS) FABIAN[10] and the ITD values extracted of public HRTF databases are presented. As in earlier investigations using the FABIAN HATS' BRIRs there was a tendency to report artifacts related to the ITD being too large (source movement opposed to head movement on head-tracked systems) it is the aim of this comparison to assess whether there are systematic differences in size of the ITD of FABIAN with respect to the average ITD of public databases.

The method used for the ITD detection was onset detection (see sec. 4.3) as this method provided the best performance. The following HRTF databases were taken into account:

CIPIC from the CIPIC Interface Laboratory from the University of California Davis, U.S.A. The database includes 1250 measurements of head-related impulse response pairs for each of 43 subjects (27 male, 16 female). These measurements were recorded at 25 different azimuths and 50 different elevations (Algazi et al. 2001a).

IRCAM from the Institut de Recherche et Coordination Acoustique/Musique Paris, France. Database with HRTFs from 52 subjects, males and females measured on 24 azimuths and 10 elevations (IRCAM).

AALBORG from the Acoustics Laboratory Aalborg, Denmark. This database is not available for the public, but the mean ITD of 70 subjects among 16 azimuths can be read from plots published on Minnaar et al. (2000).

NAGOYA from the Nagoya University, Japan. This database has HRTFs of 100 subjects males and females measured on 72 azimuths (Elevation = 0°) (Nagoya)

[10] FABIAN's HRTF dataset was recorded at the anechoic chamber of the TU-Berlin. Only horizontal plane was considered (Elevation = 0°).

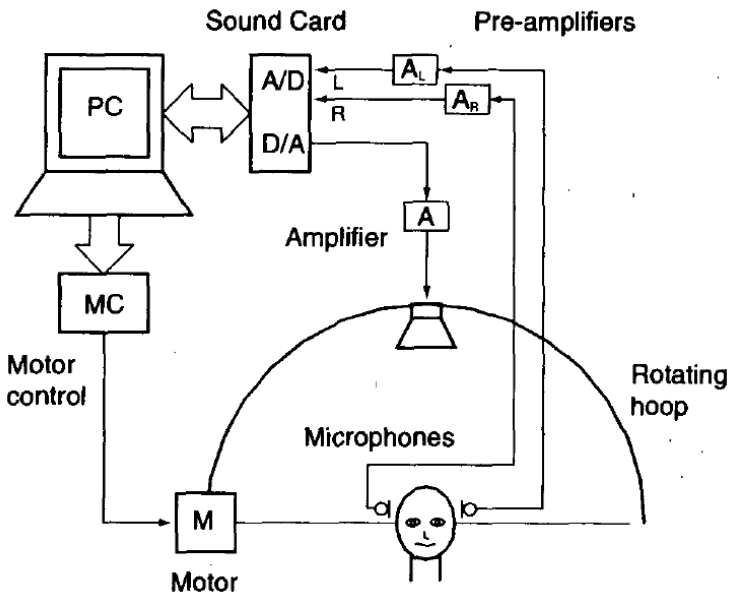

Figure A.1.: Experimental setup for the HRTF acquisition at CIPIC. Source Algazi et al. (1999)

A.1. FABIAN vs. CIPIC HRTF database.

A.1.1. Experimental setup at CIPIC

For the dataset acquisition at CIPIC the subject was seated in the center of a 1 m radius hoop whose center were aligned with the subject's interaural axis. A Bose Acoustimass loudspeaker with 5.8 cm cone radius was situated at various positions along the hoop. (see Fig. A.1).

The subjects head movements were not restricted. Datasets of subjects containing abrupt changes in ITD due to small head movements were excluded. The subjects ear canals were blocked and Etymotic Research ER-7C probe microphones were used to pick up Golay code sequences for impulse response measurement.

The samplerate used was 44100 with 16 bits quantization. A modified Hanning window was applied to the raw HRIR measurements to remove room reflections, and the results were free-field compensated to correct for the spectral characteristics of the transducers. The length of each HRIR is 200 samples.

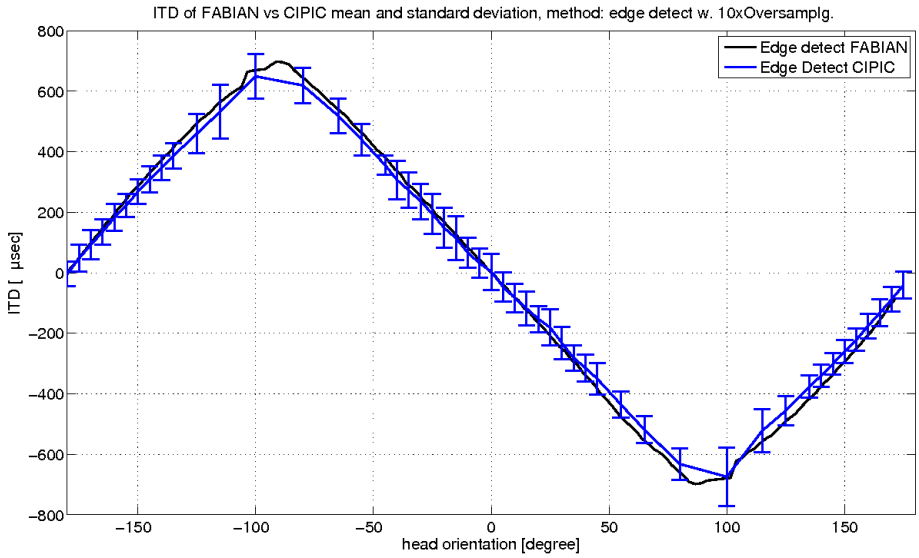

Figure A.2.: ITD of FABIAN vs. mean and standard deviation of the CIPIC database.

A.1.2. Results

Fig. A.2 shows the mean and the standard deviation of the CIPIC database. It can clearly be seen that FABIAN has a slightly bigger ITD as CIPIC's mean, this is though not surprising because the CIPIC database contains data of male (60%) and female (40%) subjects where the latter, on average, exhibit a smaller head size.

Near the +/- 90° region the standard deviation increases, possibly due to the small amplitude of the contra-lateral impulse response making it harder to find an appropriate ITD[11]. However, FABIAN's ITD remain within the standard deviation ranges at all angles.

A.2. FABIAN vs. IRCAM's HRTF database

A.2.1. Experimental setup at IRCAM

The IRCAM measurements were realized in an anechoic room (8.1 x 6.2 x 6.45 = 324m^3). The walls of the room were covered with 1.1 m glass wool wedges absorbing sound waves above 75 Hz (see Fig. A.4). The loudspeaker was attached to a crane whose position was

[11] See section 4.3

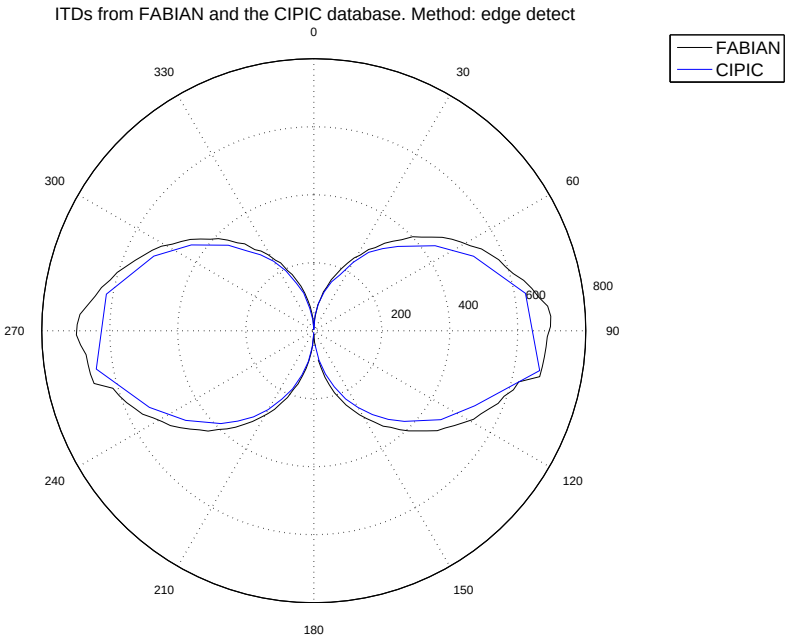

Figure A.3.: ITD of FABIAN vs. mean of the CIPIC database. Extraction method: edge detection.Notice the bigger ITDs on FABIAN's dataset.

controlled by step-by-step motors. An angular sensor sent feedback to the computer regarding the crane's elevation.

The subject to be measured was sitting on a remotely controlled rotating chair. The chair was adjustable in height and an aluminium mount was attached in the chair's back rest, with headrest for helping subjects to keep the head straight ahead. Validation of head position was performed by a head tracking system linked to the measurement software. This allowed triggering the measurement signal only when subject's head was at the correct position.

A.2.2. Results

As can be seen on figure A.5 the standard deviation of the ITDs from this database is very small. This may be a consequence of using a head tracked system for starting the recordings, which helped minimizing errors due to small head movements.

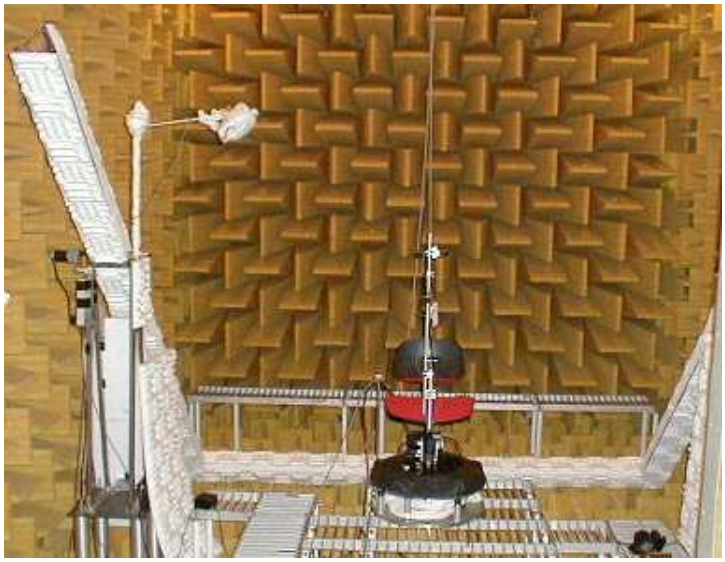

Figure A.4.: Experimental setup for the dataset acquisition at IRCAM. Source IRCAM

Here again the ITD of FABIAN fits in the standard deviation ranges at all angles. The mean of the database seems be almost identical FABIAN's ITD for the azimuth range of $-80°$ to $+80°$ and slightly bigger for other angles. This is also not surprising because this database has 37% female subjects.

On the polar plot of fig. A.6 we see the biggest differences between $80°$ to $120°$ and $-80°$ to $-120°$.

A.3. FABIAN vs. Alborg's HRTF database

The ITD mean values of the 70 subjects [12] of this database were extracted from a publication of the Aalborg Institute (see Minnaar et al. 2000, page 13). 30 datasets were recorded while the subjects were seated, the remaining 40 while they were standing. The spatial resolution was $22.5°$. There is no published information about further details of the data acquisition.

[12] The amount of male and female subjects was not published

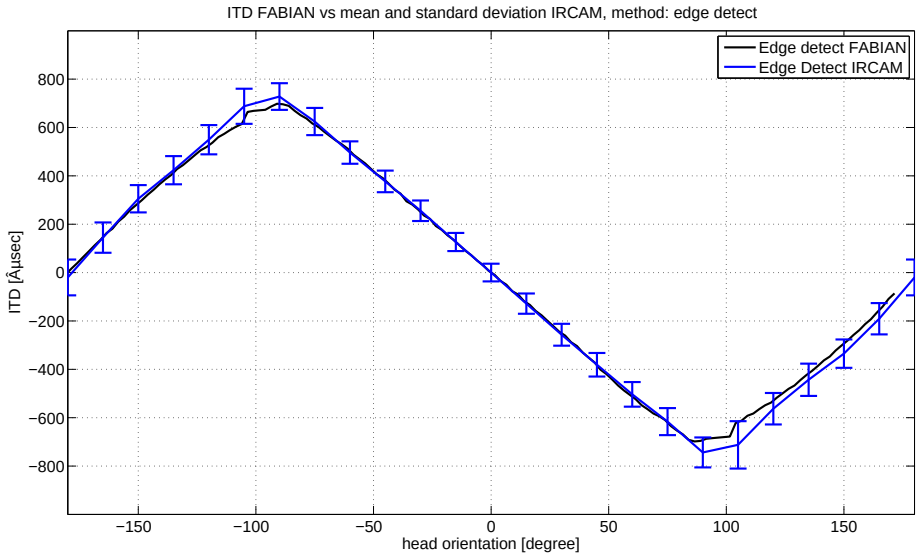

Figure A.5.: ITD of FABIAN vs. the mean and standard deviation of the IRCAM HRTF database. Note that the ITD of FABIAN fits inside the standard deviations at all angles.

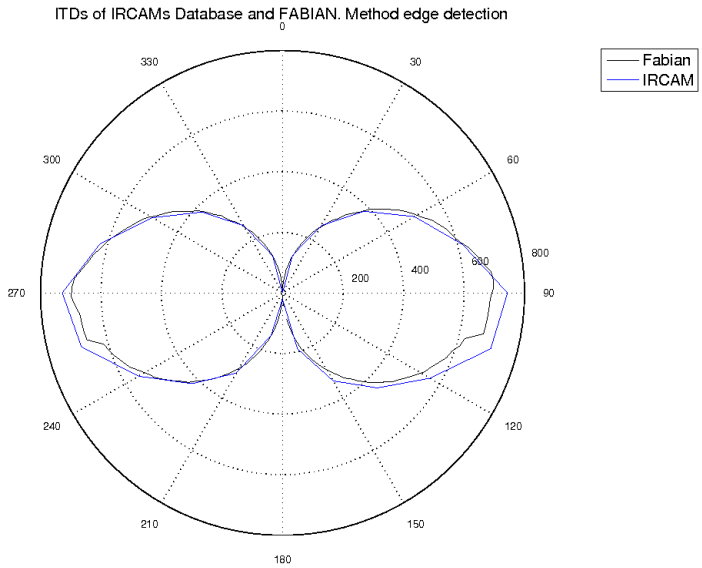

Figure A.6.: ITD of FABIAN vs. mean of the IRCAM HRTF database. Extraction method: edge detection. Note the improved symmetry of the mean ITD of this public database compared to CIPIC. (fig. A.3).

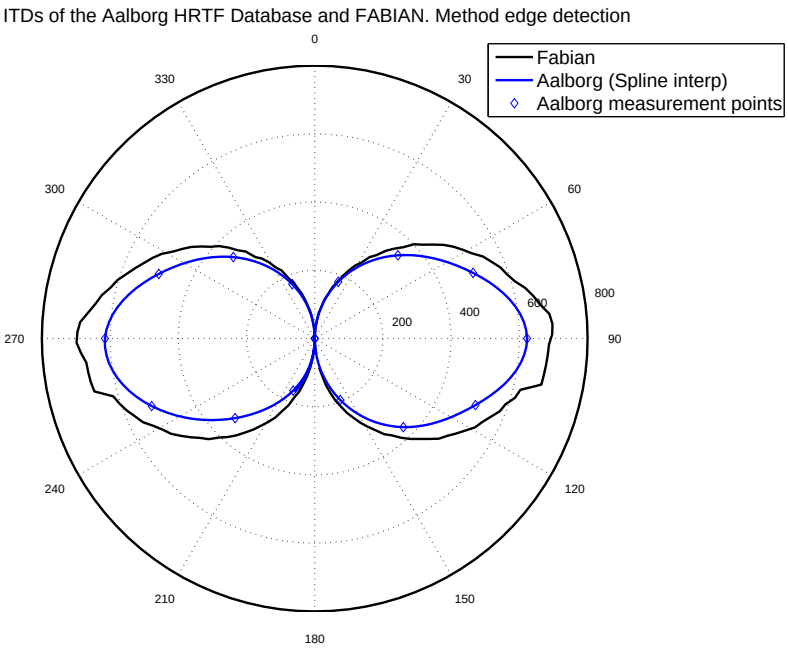

Figure A.7.: ITD of FABIAN vs mean of the Aalborg HRTF database. Extraction method for FABIAN: edge detection

A.3.1. Results

Figure A.7 shows a the ITD of FABIAN being only approximately the same for $-30°$ to $30°$ for all other elevations the mean ITD of the Aalborg database is smaller. Since the percentage of female subjects was not published we can't exclude the possibility that the differences are in part related to this aspect.

A.4. FABIAN vs. Nagoya's HRTF database

A.4.1. Experimental setup at the Nagoya university

For the HRTF data acquisition of the 100 male and female subjects[13] the set-up of figure A.8 was used.
The sampling frequency of the HRTFs is 48000 Hz. Every impulse response has a length of 512 samples. During the measurements there was a noise level of 33.9 dB A. The data is

[13]The amount of male and female subjects was not available

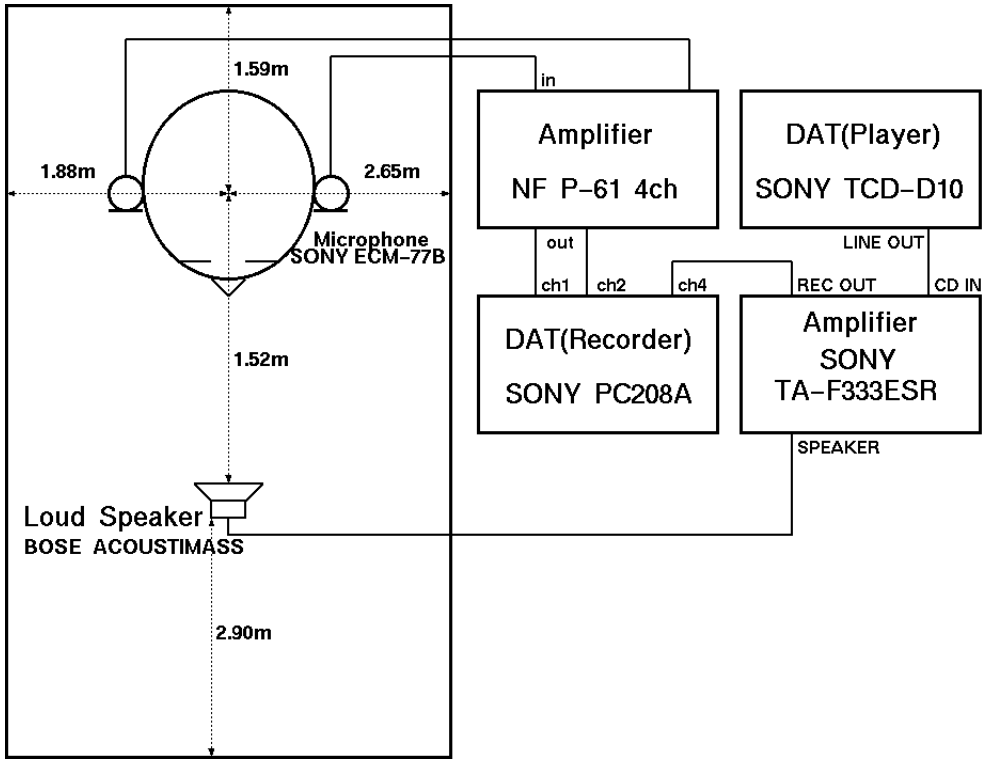

Figure A.8.: Experimental setup for the database acquisition at the Nagoya University. Source (Nagoya)

available as double precision numbers in ASCII format (.dat files).

The spatial resolution of the database was 5° covering 360° azimuth. Elevation data is not available for this amount of subjects.

A.5. Results

The big standard deviations on the database ITDs is an indication for imprecise recordings (see figure A.9). Especially close to the left and right 90° angles, the deviation were larger than 200 μsec.
The ITD of FABIAN is bigger as the Nagoya's mean, but is still within the standard deviations at all azimuth angles except for 105° and −105°. The polar plot of the mean ITDs (fig. A.10) reveals also this aspect and a counter-clockwise rotational offset.

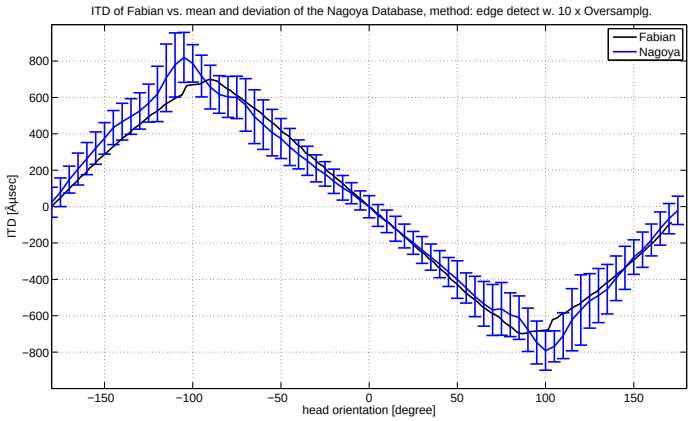

Figure A.9.: ITD of FABIAN vs. the mean and standard deviation of the Nagoya HRTF database

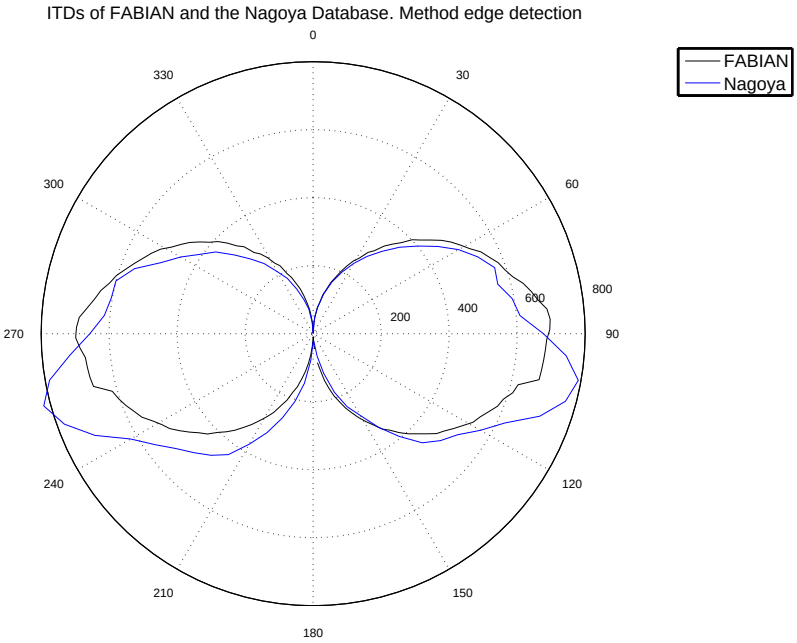

Figure A.10.: ITD of FABIAN vs. mean of the Nagoya HRTF database. Extraction method: edge detection

A.6. Chapter's Resume

The interaural time differences of 4 HRTF databases were extracted using the edge detection method. The mean of the extracted values were compared with FABIAN's ITD.

The artificial head's ITD was shown to be within the standard deviations at almost all angles except for the Aalborg dataset where the HRTFs are not available to the public and the ITD were read from a publication's plot, thus, no standard deviations are available. The HRTFs from the Nagoya University, shown to contain deviations/errors which probably can be assigned to measurements errors, thus preventing us from making valid judgments based on these data.

IRCAM's HRTF database seem to be have the most consistent data, possibly because of:

- Using a head tracking system for triggering the recordings, thus recording only when the subject head's stays at the correct position.
- Controlling the step by step motors with a sensory aided feedback system. Such a highly motion-controlled system seems recommendable when measuring HRTFs.

B. Comparison of the ITD synthesized from geometrical models

The previous appendix compared FABIAN's ITD with the ITD of empirical datasets. In this chapter FABIAN's ITD will be compared to ITD's generated with the help of geometrical models.

This analysis can be of interest for this reasons:

- The use of anthropometry for the individualized ITD generation can be compared with FABIAN's estimated ITD following the approach of Algazi et al. (2001b) in order to prove the suitability of a regression model to be used for ITD prediction/individualization.
- With the help of geometric models the elevation dependency of the ITD can be easily assessed visually.
- In the individualization model discussed in chapter 1 (section 1.1 fig. 1.1) the $ITD(\theta,\phi)$ is a function of the head's position given by azimuth and elevation. In the special case of sound sources with known position, the ITD could be easily synthesized using a geometric model.[14]

In section 3.3 an individualization method using an optimal head radius was discussed. Applying FABIAN's head dimensions on equation 3.5 the optimal head radius a_{opt} for the HATS FABIAN was determined:

$$a_{opt} = 0.51 \cdot (0.0790m) + 0.019 \cdot (0.1245m) + 0.18 \cdot (0.0995m) + 0.032 = 0.0926[m] \quad \text{(B.1)}$$

In this chapter FABIAN's a_{opt} will be used as head radius.

[14] With empiric BRIR datasets the position is mostly unknown

B.1. Extracted ITD vs. Woodworth- Schlosberg 's geometric model

The Woodworth-Schlosberg Formula (eq. 3.1) applied to FABIAN's optimal head radius a_{opt} gives the ITD of figure B.1. Both ITDs seem to be very close to each other. Note that the Woodworth-Schlosberg equation is defined only for an azimuth range of $-90°$ to $90°$. As this model relies on a spherical head model symmetry is assumed.

In order to quantify the perceptual performance, the absolute ITD difference was calculated.
According to Minnaar et al. (2000) ITD differences start being audible at around 30μ seconds. Figure B.2 shows that the maximum difference between FABIAN's ITD and the geometric model ITD reaches values of more than 35μs. Therefore the model could still replace the ITD extracted with the onset detection method without audible consequences. However as already discussed on Chapter 3 the subjective study of (Busson et al. 2005) mentions that Algazi's approach underestimates the subjective ITD although the onset extracted ITD (percepetively best ITD estimator in Busson) and the Algazi-model are very close in the case of FABIAN.

Another interesting aspect on figure B.2 is that the absolute ITD difference is not symmetrical. This could be related to a systematic error at the dataset acquisition, the extraction method, and/or asymmetries in the artificial head's dimensions.[15]

B.2. Modelling the influence of distance and source elevation on the ITD

In order to assess the influence of the distance to the ITD, the results of a simple calculation of the time arrival difference between two points for -60° to 90° elevations and 0.5 to 100 m distances are displayed in figure B.3.

Every sinusoid curve generated this way represent a given elevation at 200 steps of distance. It can clearly be seen that above 0.5m the distance of the source has no influence on the time arrival difference. It has to be stated though that this model does not take into account a real head's frequency-dependent diffraction it shall only serve to demonstrate, that a) distance

[15] On (Busson et al. 2005) the subjective ITD is never symmetric.

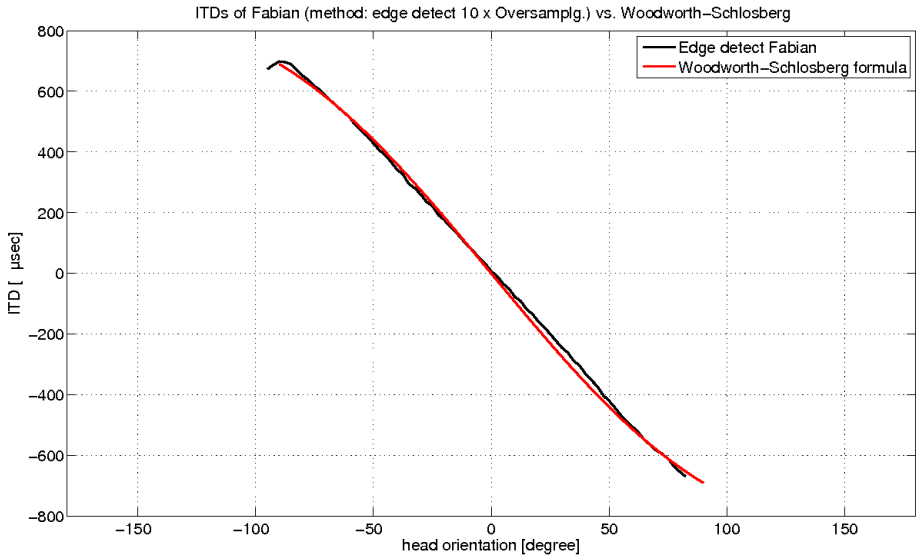

Figure B.1.: ITD of FABIAN compared to the ITD generated by the Woodworth-Schlosberg formula

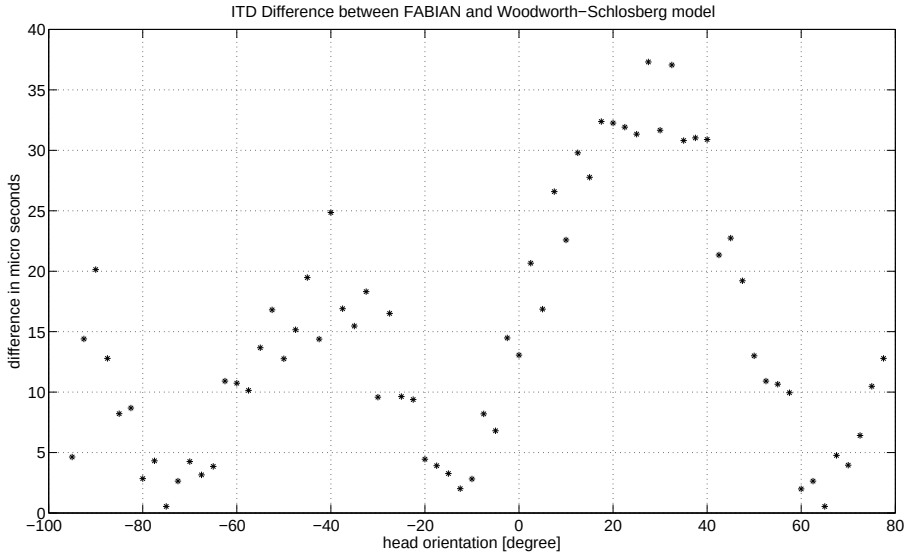

Figure B.2.: Absolute difference between the extracted ITD of FABIAN (method: edge detection w. oversampling) and Woodworth-Schlosberg's geometric model. Only horizontal plane.

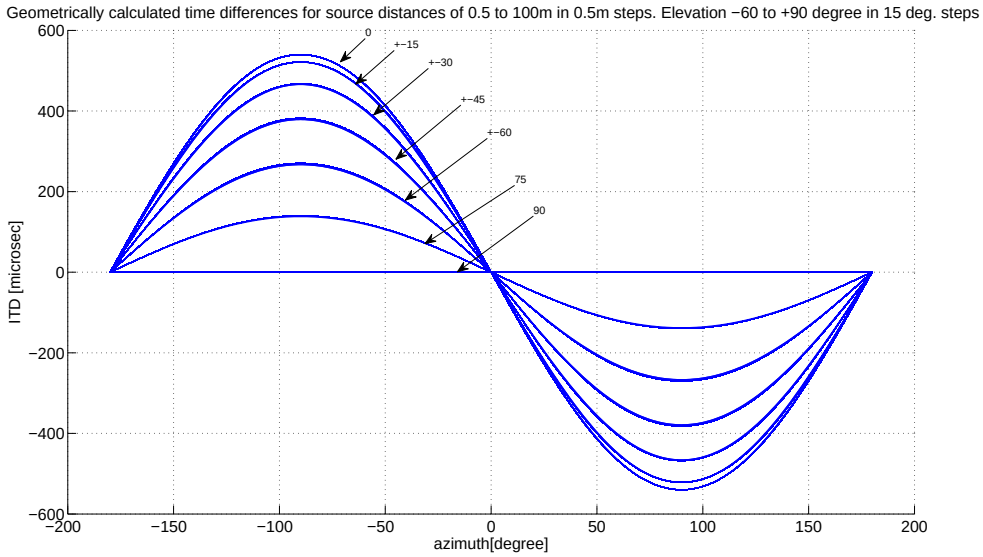

Figure B.3.: Arrival time difference at two receivers for different distances and elevations.

nearly plays no role at above 1 m and that b) elevation can not be neglected in characterizing the ITD related to a certain sound source.

B.3. Performance of the geometric ITD models regarding elevation

The Algazi model does not consider elevation. But as the influence of elevation could be clearly shown in the last section, improved versions of the Woodworth-Schlosberg formula were assessed and compared to ITD derived from FABIAN HRTFs from different elevations.

The optimal head radius according Algazi (eq. 3.4) was used in Larcher's (eq. 3.2) and Savioja's (eq. 3.3) equations to synthesize the ITD for various elevations. To counteract the current unavailability of 360° azimuth HRTFs at different elevations of the HATS FABIAN, a dataset with the same dummy-head was used (see figure B.4).

This dataset (Moldrzyk et al. 2004) was recorded with 0.5° azimuth resolution and 5° elevation at the Institute of Technical Acoustics of the RWTH Aachen.

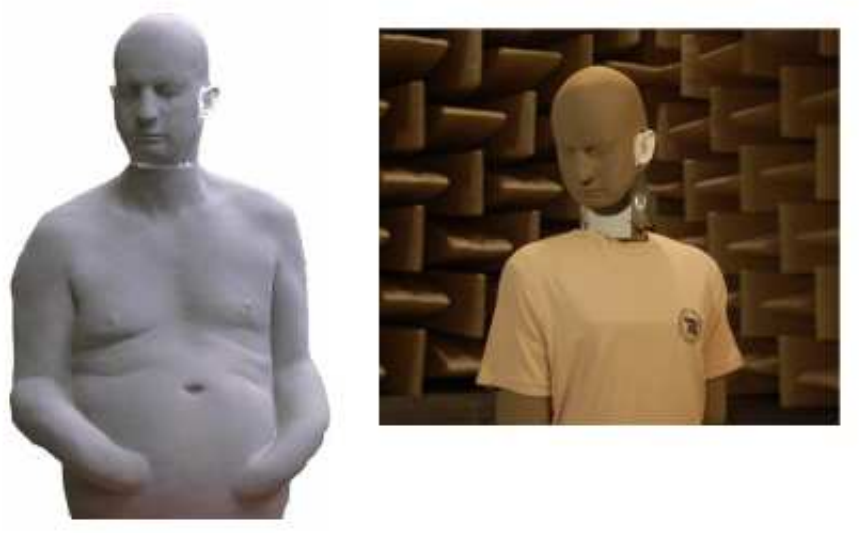

Figure B.4.: Moldzdryk's dummy head (Moldrzyk et al. 2004) and FABIAN (Lindau 2006). Both artificial heads were molded from the same individual's head.

B.3.1. Larcher's geometric model

Equation 3.2 was applied over a range of 360° in azimuth and 30°, 60° and 90° elevation on figure B.5 and −60°, −30° and 0° elevation on figure B.6.

The closest similarities are found at 60°. The absolute ITD difference is shown on figure B.7. Once again the differences are far beyond Mills' 10μs jnd and Minnaar's 30μs jnd, but despite that the model is able to synthesize the overall ITD variation fairly well.

B.3.2. Savioja's geometric model

Equation 3.3 was applied in the same ranges as for Larcher's equation (see figs. B.8, B.9). Comparing to Larcher's equation the performance of Savioja's geometric model was slightly inferior. Discrepancies between extracted vs. synthetic ITD can clearly be seen on figure B.10, where the absolute ITD difference reaches values of more than 120μs.

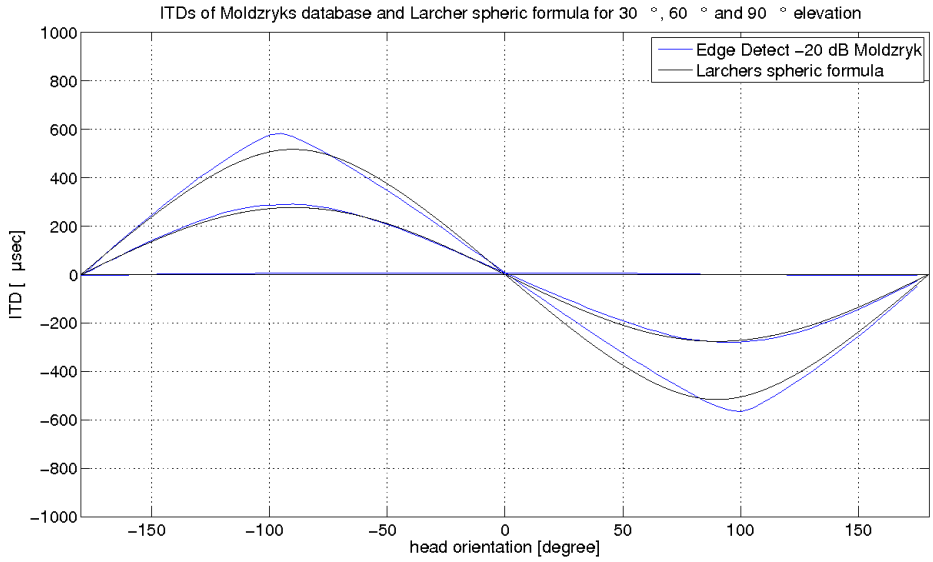

Figure B.5.: ITD of Moldzryk dataset compared to the ITD generated by the Larcher formula for 30°, 60° and 90° elevation

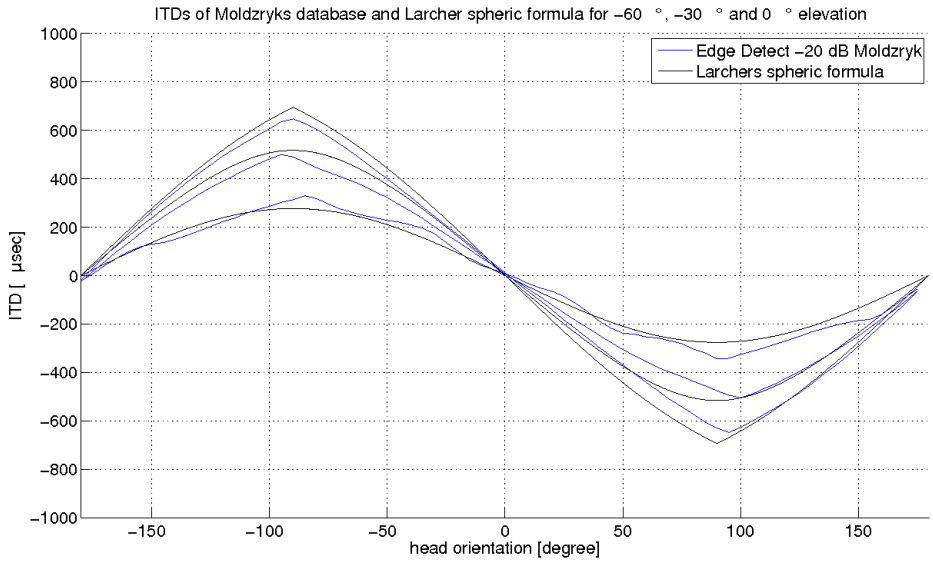

Figure B.6.: ITD of Moldzryk dataset compared to the ITD generated by the Larcher formula for −60°, −30° and 0° elevations

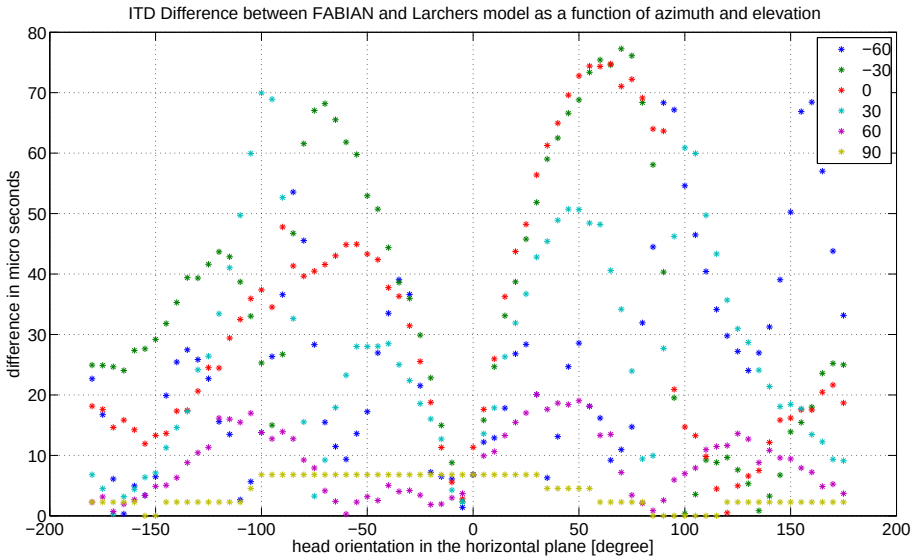

Figure B.7.: Absolute ITD difference between Moldzryk dataset and the ITD generated by the Larcher formula for different elevations (-60° to 90°) and azimuth angles (-180° to 180°)

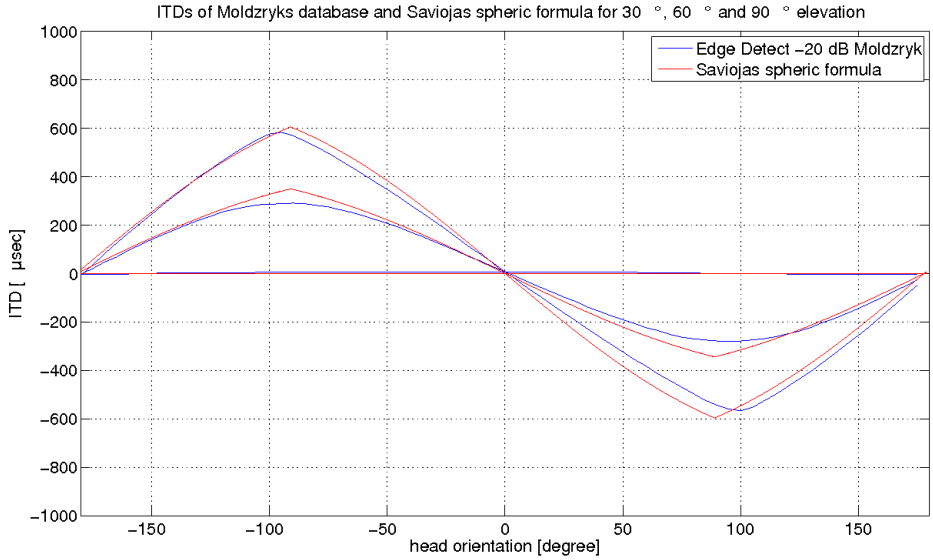

Figure B.8.: ITD of Moldzryk dataset compared to the ITD generated by the Savioja formula for 30°, 60° and 90° elevation

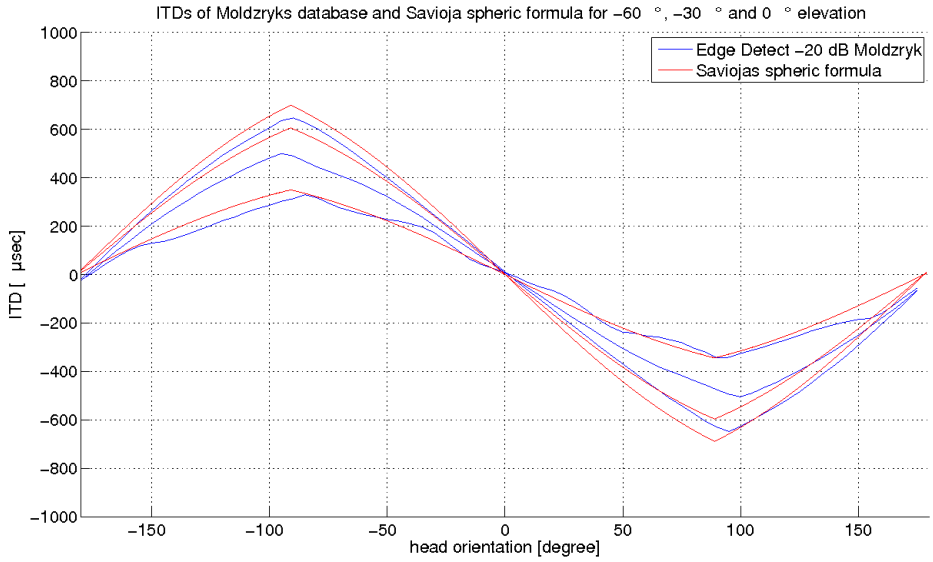

Figure B.9.: ITD of Moldzryk dataset compared to the ITD generated by the Savioja formula for $-60°$, $-30°$ and $0°$ elevation

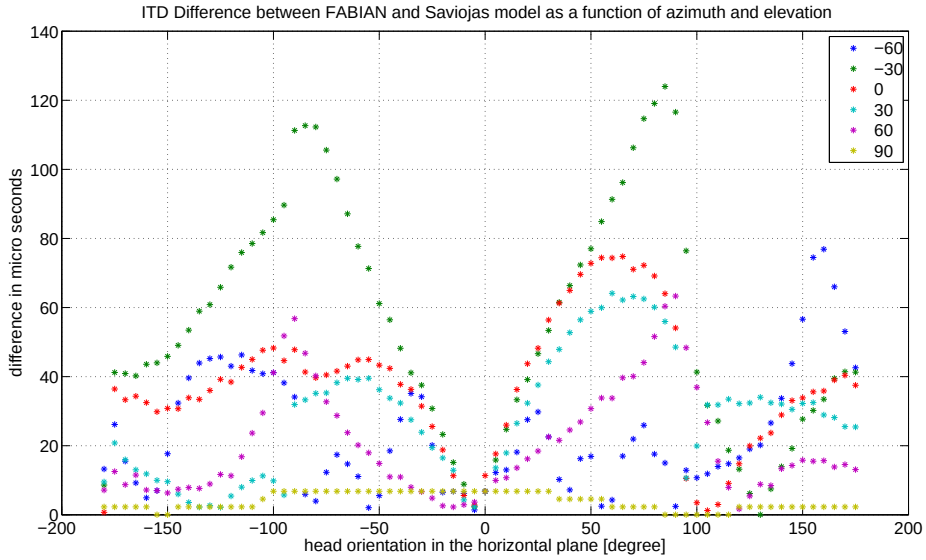

Figure B.10.: Absolute ITD difference between Moldzryk dataset and the ITD generated by the Savioja formula for different elevations (-60° to 90°) and azimuth angles (-180° to 180°)

B.4. Chapter's Resume

In this chapter the performance of the geometric models for ITD synthesis were analyzed and compared with the extracted ITD from datasets using onset detection as estimation method and the optimal head radius a_{opt} described on Algazi et al. (2001b).

The spherical head models seem to provide a fairly good approximation of the ITD. The equations including elevation showed a similar fit for positive elevation angles, while negative elevation was worse for both formulas.

The influence of the source distance on the arrival time difference between two points was also found to be irrelevant above a distance of 0.5m while neglecting the head diffraction. It has been clearly verified that elevation plays a role in the ITD. The "azimuthal-only" Woodworth-Schlosberg model was shown to be insufficient for full sphere ITD synthesis.

C. Matlab code for extracting the ITD with the onset detection method

```
function [ itd ] = OnSetItd( left , right , onset_threshold_dB , fs , up)
% Funtion to calculate the ITD with detection of Onsets
% Input parameters are IR vectors left and right , the onset threshold in
% dB, the sample frequency and the upsampling factor that the IRs have.
%
tauUp = 1/( up*fs);
% calculate linear onset threshold from dB value
 onset_threshold =10^(onset_threshold_dB /20);

% find peaks and compute the sample position : Left
[maxLeft, iLeft ] = max( left );
kL = 0;

while kL <= iLeft
    kL = kL +1;
     if abs( left (kL)) > abs(maxLeft*onset_threshold )
         break;
    end;

end
 if kL == 0,
        fprintf ('Error #1 Left: Problem finding the onset\n');
       kL = 1;
end

% find peaks and compute the sample position : Right
[maxRight,iRight] = max(right);
kR = 0;
```

```
29
30 while kR <= iRight
31     kR = kR + 1;
32     if abs(right(kR)) > abs(maxRight*onset_threshold)
33         break;
34     end;
35 end
36 if kR == 0,
37         fprintf('Error #1 Right: Problem finding the onset\n');
38         kR = 1;
39 end
40
41 % calculate the ITD in seconds instead of samples
42 itd = (kL-kR)*tauUp;
```

D. Screenshots of the ABX software

In order to conduct the listening tests explained on Chapter 5, an ABX-test software was developed as a standalone C++ application. This software was able to fulfill all test requirements with no constraints.

Figure D.1 shows the graphical user interface that the subjects operate.

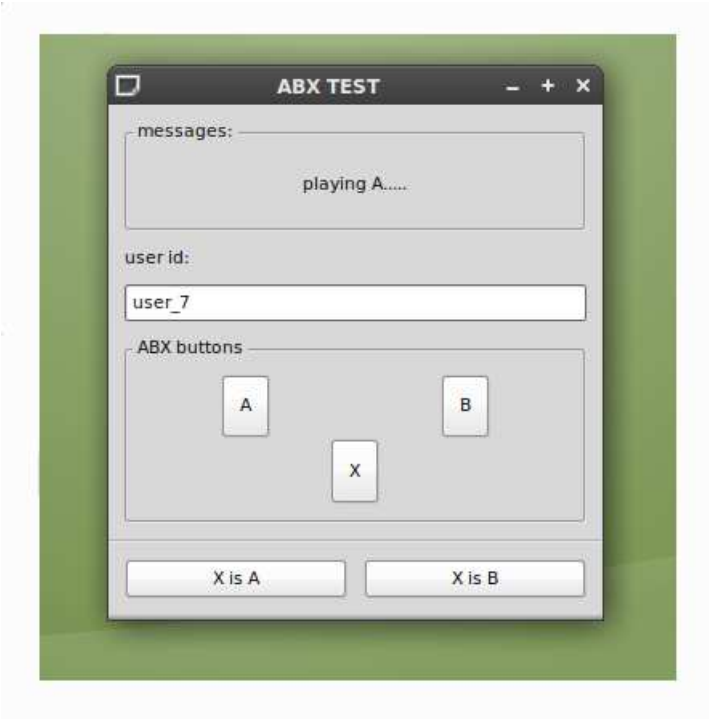

Figure D.1.: Screenshot of the user interface of the ABX-test software especially developed for the listening tests of Chapter 5