Individualization of dynamic binaural synthesis by real time manipulation of the ITD

J. Estrella, A. Lindau
Audio Communication Group TU Berlin
Email: jorgos.estrella@mailbox.tu-berlin.de

Introduction

Virtual acoustic environments (VAEs) are commonly realized via dynamic binaural synthesis. Therefore, anechoic audio is convolved in real time with head related impulse responses (HRIRs) or binaural room impulse responses (BRIRs). Binaural filters are exchanged inaudibly and in real time according to the listener's head movements, thus providing a realistic auditory experience. From Lord Rayleighs [1] duplex theory of hearing it is known, that interaural level (ILD) and time (ITD) differences are exploited for auditory localization. ITDs are being evaluated mainly below 1500 Hz, whereas above, envelope delays and ILDs determine localization. If the signal contains ambiguous temporal and spectral localization cues, the ITD tends to dominate the perceived sound source direction [2]. As it is mostly unfeasible to conduct individual binaural measurements, binaural data sets of dummy heads are commonly used. ILDs and ITDs of the binaural data sets therefore reflect the physical structure of a certain individual's pinna, head and torso. ITDs differ individually due to varying head geometries. As a result, when listening to a VAE with non-individualized binaural data sets, degradation of localization accuracy and the perceived stability of the sound sources occur: If the head in the original data sets was smaller than the listeners head, a displacement of the sound sources in the same direction as the heads' movement may be perceived during head movements; if on the other hand, the head was larger, an apparent movement in the opposite direction occurs [3]. This effect can be very annoying, especially as adaptation does not seem to occur. Therefore, a method for the customization of binaural synthesis by means of real time manipulation of the ITD contained within binaural impulse response data sets is being proposed. Additionally, based on the listener's individual anthropometry and listening tests, a predictive model for adequate individual manipulation will be presented.

Method

Binaural impulse response sets of our head and torso simulator (HATS) FABIAN [4] where considered for the development of the individualization approach. In order to be able to individually manipulate the interaural time differences of binaural impulse response data sets, at first, time delays had to be extracted and removed from the datasets in way that the BRIR can later be reconstructed without artifacts. In [5] several methods for the detection and extraction of the ITD were reviewed and evaluated. Because of its applicability on empirical binaural room impulse response data sets, its robustness, and its subjectively artifact-free results, the onset detection method was selected (see Figure 1). To obtain ITD values with subliminal precision, preprocessing takes place in the ten times upsampled domain. The quasi-minimum phase impulse responses and ITD values are stored for further use.

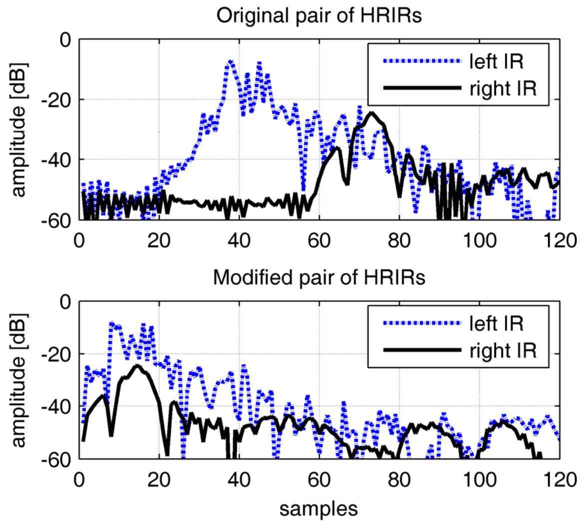

Figure 1: Extraction of ITD from HRIRs via onset detection.

The individualized re-synthesis process is an extension of the time-variable fast convolution algorithm conventionally used for dynamic binaural synthesis. Figure 2 shows a schematic depiction of the implementation. The conventional convolution process is split up in order to independently process magnitude and phase information. Thus, the conventional dynamic convolution is conducted as before, but using binaural datasets without ITD. Because cross fading is now conducted on time-aligned signals, characteristic comb filter artifacts are largely diminished. This results in a clearly audible improvement of the binaural rendering. The output from the fast convolution (signals L' and R' in Fig. 2) is then subjected to a variable delay line (VDL). The previously extracted time delay is re-inserted with subsample accuracy according to the current head position thereby re-establishing the ITD between left and right ear signals. Subsample accuracy is guaranteed by means of a band limited interpolation method [6]. For implementation an open source sample rate converter library [7] was used. This approach allows for glitch-free time stretching while maintaining a bandwidth of 97% and a signal to noise ratio of 97dB. Moreover, as changes of the ITD are realized by fractional sample rate conversion the Doppler effect is correctly imitated for the direct sound. Individual customization of the ITD can be accomplished by simply scaling the re-inserted ITD by a constant, frequency independent factor. This approach inherently assumes that any person's ITD can be correctly obtained from that of any other person by simple scaling. This would only be true if all subjects' heads were only scaled versions of each other. However, for now this simplified assumption is justified by the perceptual improvements that can already be obtained.

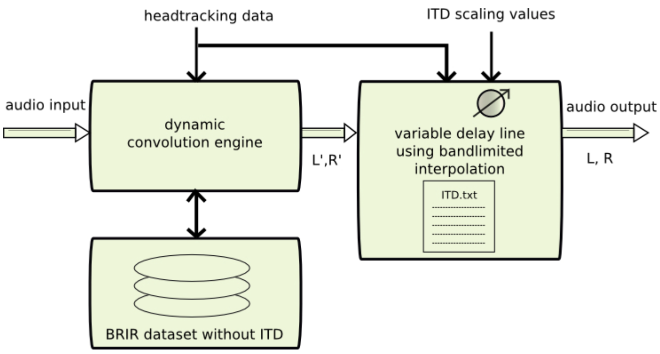

Figure 2: Schematic depiction of dynamic binaural synthesis with ITD individualization.

Listening test

For determining individual scaling factors of a foreign datasets' ITD for different subjects, the developed binaural rendering method was used in a listening test. A dataset of BRIRs was measured in an acoustically damped room (V = 155 m³, RT = 0.47 s) using the HATS FABIAN. A Genelec 1030A loudspeaker was positioned frontally in a distance of 2 m. BRIRs were measured for horizontal head rotations within ±90° in angular steps of 1°. To allow for direct comparison between real sound source and the individually adjustable auralization in listening test, the HATS wore acoustically transparent headphones during the measurements. In the listening test subjects were the seated at the former position of the HATS. Using the method of adjustment, the subjects' task was – while instantly switching between simulation and reality – to adjust the foreign ITD by changing the scaling factor until localization and source stability was perceived to be similar in reality and simulation. During training subjects were instructed to rotate the head widely, in order to maximize audible artifacts of misaligned interaural delay. The ITD could be modified in a range of 0-200% at a resolution of 1% using up-down pushbuttons as interface. To minimize the impact of concurrent ILD, low-pass filtered white noise bursts were used as stimulus (f_{stop} = 1.5 kHz). Ten repeated runs were conducted per subject while starting from randomized ITD scaling-factors. Eleven subjects took part in the test.

Results

Despite training the task was difficult for some subject. This is also reflected by the rather large confidence intervals in Fig. 3. By means of residual analysis and outlier tests, 2 of the 11 subjects were excluded from the final analyses.

When using ITD manipulation in the future, individually correct scaling factors would have to be established by lengthy manual adjustment of the ITD. A prediction formula would thus be more convenient. In order establish a functional relation between the head's proportions and the ITD scaling factor, four anthropometric measures were taken from each subject: width, height and depth of head and the *intertragus* distance, which is the distance between both ears' *incisura anterior* marking the tragus' upper end. This measure was chosen due to the tragus' proximity to the ear channel and its simple and reliable determination. Individual scaling values were predicted from these anatomical measures by means of multiple regression analysis. The explained variance (adj. R^2) revealed that a single predictor – the *intertragus* distance – is sufficient for predicting the individual ITD scaling factors. In this case the explained variance was 70%. Figure 3 depicts all average individual scaling factors together with 95% CIs, and the linear regression model with its 95% CIs. The linear model could possibly be generalized to arbitrary binaural data sets by scaling the predicted values by a certain ratio of the *intertragus* distances of a foreign and our artificial head. Besides, this method has not been evaluated so far. A formal evaluation of the achieved localization improvement is subject to future work, too.

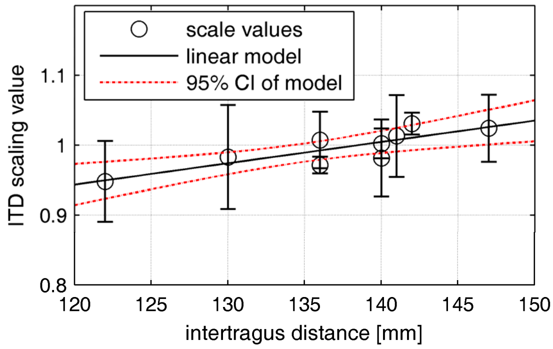

Figure 3: Mean individual ITD scale values plotted over intertragus distance (with 95% CIs). Linear model is shown with 95% CIs.

Conclusions

A method for customizing dynamic binaural reproduction by means of real time manipulation of the ITD was proposed. Furthermore, an anthropometry-based prediction model for an individual ITD correction factor was discussed. Further advantages of the proposed approach are: the elimination of cross fade comb filtering by using minimum phase audio signals, the possibility to use different resolution and interpolation methods for both temporal and spectral cues, and a correct simulation of the Doppler effect for the direct sound by means of sample rate conversion. Overall subjective quality of dynamic binaural synthesis could thus be noticeably improved.

Alexander Lindau was supported by a grant from the Deutsche Telekom Laboratories.

References

[1]Strutt, J. W. (Lord Rayleigh) (1907): "On Our Perception of Sound Direction.", In: *Philosophy Magazine*, Vol. 13, pp. 214–232
[2]Wightman, F.L.; Kistler, D.J. (1992): "The dominant role ....", In: *J. Acoust. Soc. Am.*, **91**(3), pp. 1648-1661
[3]Algazi, V. R. et al (2001): "Estimation of a spherical-head model from anthropometry.", In: *J. Aud. Eng. Soc*, **49**(6): 472-479
[4]Lindau, A.; Weinzierl, S. (2006): "FABIAN - An instrument …", In: VDT (Hrsg.): *Bericht der 24. Tonmeistertagung*. Leipzig
[5]Minaar, P. et al. (2000): "The Interaural time difference in binaural synthesis. ", In: *Proc. of the. 108th AES Conv.*, preprint no. 5133
[6]Digital Audio Resampling Home Page: https://ccrma.stanford.edu /~jos/resample/resample.html
[7]Libsamplerate Homepage: http://www.mega-nerd.com/SRC/index.html