TIME DOMAIN ASPECTS OF ARTIFACT REDUCTION IN POSITIONING ALGORITHM USING DIFFERENTIAL HEAD-RELATED TRANSFER FUNCTION

Dominik Storek *

Department of Radioelectronics, Faculty of Electrical Engineering,
Czech Technical University in Prague
Prague, Czech Republic
storedom@fel.cvut.cz

ABSTRACT

This paper focuses on consequences of artifact reduction in virtual sound source positioning method based on Differential Head-Related Transfer Function (DHRTF). As resulted from previous experiments, spatial performance of this experimental method is very promising. However, under specific circumstances, artifacts may occur in the virtually positioned sound. Effective methods for artifact reduction were introduced before. This work discovers impact of the reducing algorithm in the time domain in order to understand phenomena occurring in the process. The cause of artifact presence results from narrow band peak(s) present in the DHRTF magnitude, which causes periodical character of the impulse response in the time domain.

1. INTRODUCTION

The HRTF contains localization cues for a human listener, i.e. Interaural Time Difference (ITD), Interaural Level Differences (ILD) and spectral cues [1, 2, 3]. To obtain virtually positioned sound by the HRTF method, a convolution of the original signal with appropriate HRIR pair is required. Many articles have already dealt with more effective measuring [4, 5] or rendering of the HRTFs [3]. However, simplifying the positioning process focused on reduction of the computational resources is not a well-explored issue yet. The effort was put on finding an algorithm with better spatial performance than simple amplitude panning and with less processing requirements than the HRTF method. This paper aims at time domain aspects of virtual sound source positioning method primarily focused on reducing the computational costs, called *Differential Head-Related Transfer Function* (DHRTF) [6, 7]. The essence of the DHRTF algorithm lies in introducing the frequency dependent ILD and ITD to the stereo signal not by separate filtering of both channels of the binaural signal (as is in the HRTF method [8]), but by filtering only one channel in a way that the same inter-channel differences will occur in the stereo sound, as when the HRTF method is applied. Therefore, only one channel is processed (always the contra-lateral), while the other one remains completely untouched. The DHRTF can be obtained from existing pair of HRTFs as

$$H_D(\vartheta,\omega) = \frac{|H_C(\vartheta,\omega)|}{|H_I(\vartheta,\omega)|}\cdot e^{j(\psi_C-\psi_I)} = \tag{1}$$

$$= \mathrm{A}_D(\vartheta,f)\cdot e^{j\cdot\Psi_D(\vartheta,f)},$$

where indices C and I stand for contra-lateral (farther) and ipsi-lateral (closer) ear. Equivalent of the DHRTF in time domain is called Differential Head-Related Impulse Response (dHRIR) and can be defined as

$$h_c(\vartheta,t) = \mathfrak{F}^{-1}\Big\{H_D(\vartheta,\omega)\Big\} \tag{2}$$

In some HRTF pairs, a specific phenomenon occurs. In unfavorable constellation the attenuation of the ipsi-lateral channel may be greater than in the contra-lateral (against expectation) for particular frequencies, having a character of a narrow-band notch. From the definition of the DHRTF, the same ILD is present. Therefore, the artifacts are caused by presence of sharp peak in the DHRTF (spike) exceeding the level of 0 dB. Since the property of the DHRTF lies in frequency-dependent attenuation and time delay of the contra-lateral channel, positive value of the DHRTF results in boosting the specific band in this channel.

The phenomenon was named Negative Interaural Level Difference (NILD) [9]. The NILD occurs especially within DHRTF derived from a real measured set. The origin of its occurrence results from a unique constellation of the HRTFs in a pair. As shown in Fig. 1 (a), the transfer function of the ipsi-lateral channel can have a spectral notch, which crosses the transfer function of the contra-lateral channel, i.e. the contra-lateral gain is lower than the ipsi-lateral for specific frequency band. This results in spectral spike presence in the DHRTF magnitude, as shown in Fig. 1 (b), black line. Panel (b) contains DHRTF derived from two different sets of HRTF corresponding to the same position. The black line refers to the inappropriate one. However, the presence of the spike in the DHRTF is neither determined for specific spectral bands, nor for specific positions and appears to occur chaotically. For instance, the gray line in the figure represents DHRTF for the same position obtained from another subject. No spectral spike is observable here. Its occurrence is different among various DHRTF sets [9].

Figure 2 shows the extracted NILD spikes in one DHRTF set constructed from a measured set of HRTF [10]. This set of DHRTF is heavily distorted by the NILD spectral features. As can be observed, the NILD occurs mostly in form of narrow spikes. However, even wider frequency bands can appear in particular DHRTFs, especially around frontal axis, i.e. positions $\vartheta = 0^\circ$ and $\vartheta = 180^\circ$. Due to the principle of the method, the artifacts are generally likely to occur around these positions. The Negative ILD may generate noticeable disturbing artifacts of the positioned sound. Perception effect of the mentioned spectral spike is a pure tone character disturbance in the contra-lateral channel. This undesirable phenomenon can be avoided either by selection of appropriate DHRTF set (which actually limits personalization) or by adjusting of the selected individual set of the DHRTF. This adjustment is

* This work was supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS14/ 204/OHK3/ 3T/ 13.

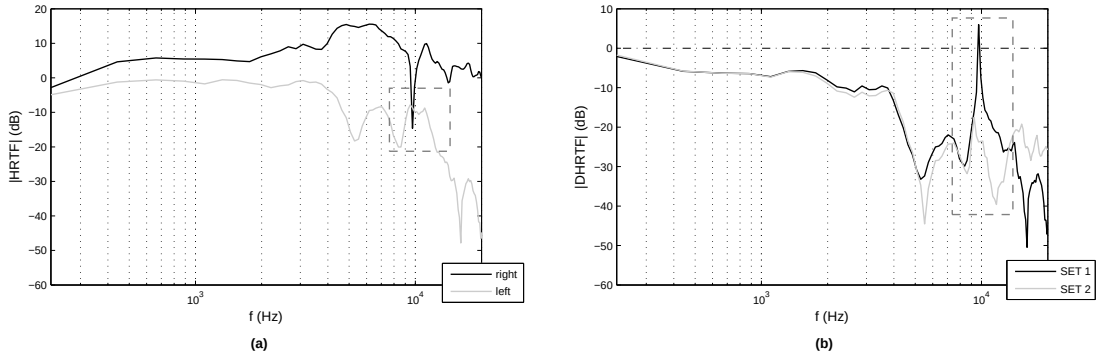

Figure 1: **Artifact origin.** Magnitude of left and right HRTF for $\vartheta = 70^\circ$ and $\vartheta = 0^\circ$ (a). When a notch attenuation of the HRTF for the ipsi-lateral ear crosses HRTF of the contra-lateral, a spike-like peak in DHRTF occurs (b). Spectral peaks and notches vary uniquely in accordance with source position and selected HRTF set.

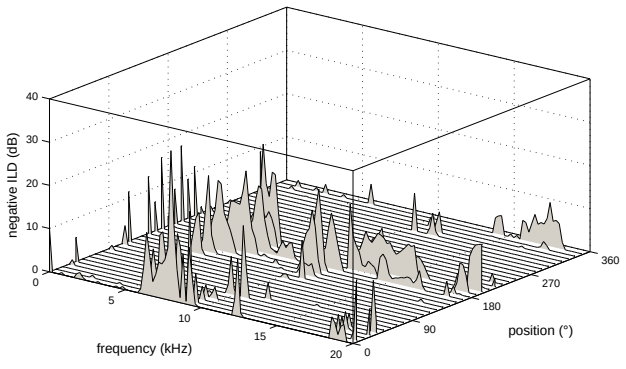

Figure 2: **Negative ILD in the DHRTF.** Extracted spectral spikes of the Negative ILD ($|H_d[\Omega]| > 0$) extracted from DHRTF. Notice highest spike occurrence around positions $\vartheta = 0^\circ$ and $\vartheta = 180^\circ$.

considered to be performed prior to practical use. Previous work published in [9] focused primarily on design of a simple artifact reduction method and verification of its efficiency by subjective listening tests. This paper extends [9] by presenting time domain aspects of application of the artifact reduction method that were investigated thereafter.

2. ARTIFACT REDUCTION

Several artifact reduction approaches were introduced in [9]. According to performed subjective tests, the most effective algorithm contains the following steps.

2.1. Spectral amplitude limiting

The first step performs hard limitation of the DHRTF magnitude curve, where every spectral sample exceeding the threshold (i.e. the negative ILD spike) is reduced to this reference level. This algorithm can be written as

$$\left|\widehat{H}_D^{\vartheta}[\Omega]\right| = \begin{cases} \varepsilon & \text{for } \left|H_D^{\vartheta}[\Omega]\right| > \varepsilon \\ \left|H_D^{\vartheta}[\Omega]\right| & \text{elsewhere,} \end{cases} \tag{3}$$

where Ω denotes spectrum of a discrete signal, $\widehat{H}_D^{\vartheta}[\Omega]$ represents adjusted DHRTF for particular position ϑ, $H_D^{\vartheta}[\Omega]$ represents initial DHRTF, and ε stands for the threshold level. The reference level was considered 0 dB in the logarithmic scale, which is equivalent to 1 in the linear scale. The tests showed that the simple *spectral limiting* may sometimes appear insufficient due to remaining local maximums in the transfer function *under* the reference level ε. Therefore, the algorithm shall be extended by smoothening of the DHRTF magnitude while preserving ILD localization cues.

2.2. Spectral smoothing

This step of smoothing the DHRTF curve by moving average, when each sample of the transfer function is obtained as an average of its neighbors. This process can be also interpreted as convolution of the DHRTF with convolution kernel $M_A[\Omega]$, as described below.

$$\left|\bar{H}_D^{\vartheta}[\Omega]\right| = \left|\widehat{H}_D^{\vartheta}[\Omega]\right| * M_A[\Omega] \tag{4}$$

The $M_A[\Omega]$ is defined as a series of uniformly weighed coefficients of length K with amplitude of $1/K$, as shown here:

$$M_A[\Omega] = \begin{cases} 1/K & \text{for } \Omega \in (0, K) \\ 0 & \text{elsewhere.} \end{cases} \tag{5}$$

The $M_A[\Omega]$ can be comprehended as coefficients of impulse response of a low-pass FIR filter. This filtering procedure ensures that the magnitude of the DHRTF will not cross the reference level ε again. Block diagram representing the signal flow in the artifact reduction method and obtaining the pre-processed dHRIR ready for virtual sound source positioning is summarized in Fig. 3. As can be seen in the diagram, all the processing is done within the DHRTF magnitude. In the first step, the phase of the DHRTF is extracted as

$$\Psi_c^{\vartheta}[\Omega] = arg\Big(H_D^{\vartheta}[\Omega]\Big) \tag{6}$$

and after the processing algorithms, the original phase is returned to the signal.

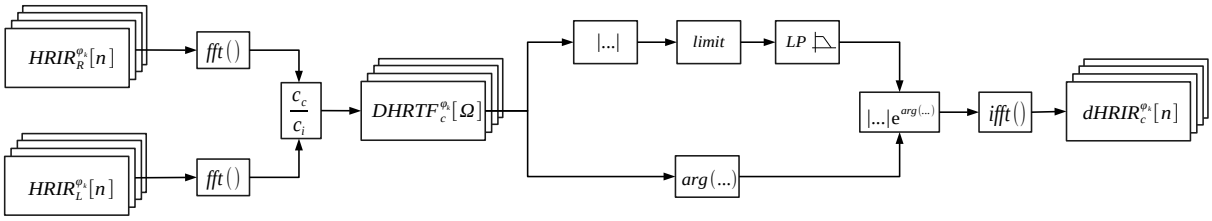

Figure 3: **Processing diagram.** A block scheme of constructing dHRIR with recommended pre-processing for reduction of the artifact occurrence. Once the DHRTF is obtained from the HRTF pair, the phase is extracted, the module is finally limited and smoothed, and merged with the original phase. Inverse Fourier transform is the last step.

3. TIME ANALYSIS

The impact of the introduced algorithms in spectral domain has been already presented in details in [9]. This section focuses on the final effect of the artifact reduction in the time domain. The phase characteristics contains the information about the ITD. For demonstration, see actual values of both HRIR's onsets and resulting ITD obtained from a set of HRIRs measured on acoustic manikin available in [10], which are shown in Fig. 4. The black line denotes onset of the left channel response, the red line denotes onset of the right channel response, and the dashed line shows resulting inter-channel ITD in regular shape of two triangles. The onset time was obtained by thresholding of the impulse response energy. The time information was quantized to particular samples in standard sampling rate of 44.1 kHz. The maximum ITD usually corresponds to approximately 29-33 samples in dependence on the head proportions. The ITD is symmetrical with almost linear character. Figure 5 shows a pure set of dHRIRs for an azimuth in range of $\vartheta \in (0, 360)$ with step of 5 degrees. The dHRIR data were constructed from the author's own measured HRTF set. The particular spike-like spectral features similar to the one demonstrated in Fig. 1 results in pseudo-periodical character of the dHRIR. Several selected responses corresponding to heavily distorted DHRTFs are marked in red. The biggest distortion is observable for positions close to frontal axis, i.e. near $\vartheta \in \{0, 180\}$.

For demonstration of the final effect of the artifact reduction algorithm on the dHRIR response, see Fig. 6. After application of the reduction algorithm, the pseudo-periodicity of the particular responses is suppressed as the spike-like spectral features in the DHRTF magnitude are removed. Notice also decreased noise level in the dHRIRs. Application of the low-pass filter performed by moving average also reduces the *noisy tale* of the response. Another important feature in the dHRIR set is a pair of clearly visible triangular shapes in the horizontal plane indicating the onset of the dHRIR. The shift corresponds to the ITD and the triangle profile is the same as presented in Fig. 4.

As stated above, the dHRIR (DHRTF) contains information about relative time shift and relative attenuation of the contralateral (farther) channel in relation to the ipsi-lateral (closer). The ITD is visible through the onset and the attenuation is observable through energy of the response. For positions close to the axis of $\vartheta \in \{0, 180\}$, the response energy is much higher than for the side positions $\vartheta \in \{90, 270\}$, where the attenuation of the contralateral channel is the highest.

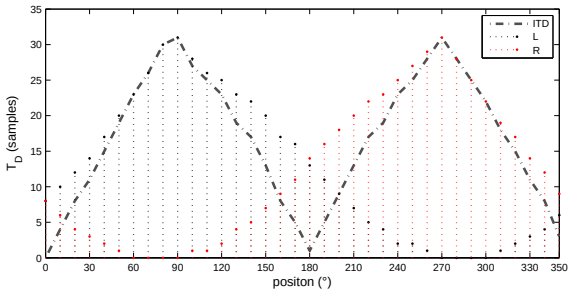

Figure 4: **ITD from HRTF set.** Interaural Time Delay extracted form a real HRTF set by detecting signal energy onset of each impulse response. The dotted black and red lines corresponds to the time shift of left and right HRIRs, respectively. Dashed triangles represent resulting ITD.

4. CONCLUSIONS

This paper presents time domain scope of the effect of artifact reduction algorithm for DHRTF-based positioning and extends previous work published in [9], where also perceptual analysis by listening tests is present. The basis of the algorithm is hard limiting of the DHRTF magnitude and smoothening of the magnitude by low pass filtering performed by moving average convolution kernel. The most affected dHRIRs are around the front-back axis. As the artifact reduction algorithm removes local maxima (peaks) in the DHRTF magnitude, the equivalent in the time domain is removing the periodicity of the impulse response. The algorithm preserves identical time domain localization cues (ITD). The ILD information is modified; however, the main features in the frequency domain are still preserved. Future work will be focused on the origin of the Negative ILD in the data sets as it is present primarily in the measured HRTF sets.

5. ACKNOWLEDGMENTS

This work was supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS14/ 204/OHK3/ 3T/ 13. Thanks to C. Fialova for fruitful discussions.

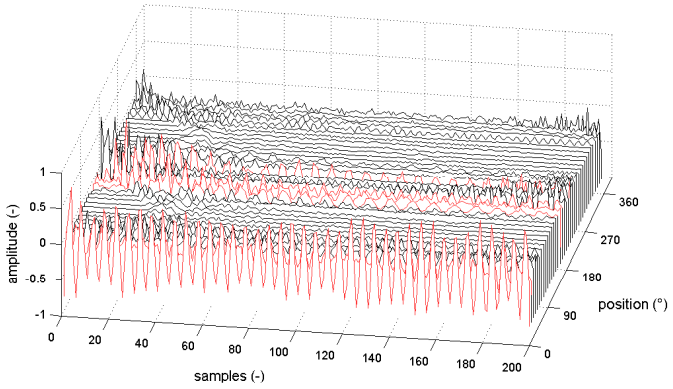

Figure 5: **An original dHRIR set.** A set of $dHRIR[n]$ derived from author's own set of $HRTF$. Particular responses affected subjected to artifact occurrence resulting from spectral negative ILD spikes are highlighted by red color.

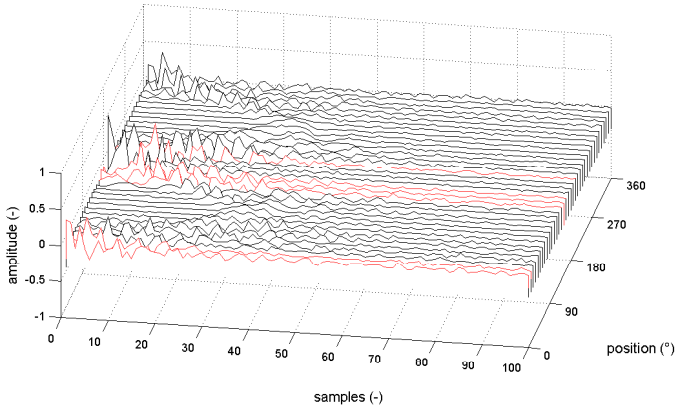

Figure 6: **A processed dHRIR set.** The effect of processing focused on artifact reduction is well observable. The problematic responses highlighted in red looses their pseudo-periodical character and extensive amplitude.

6. REFERENCES

[1] Jens Blauert, *Spatial hearing: the psychophysics of human sound localization*, MIT press, 1997.

[2] Jens Blauert, *The technology of binaural listening*, Springer Verlag, Berlin, 2013, eBook.

[3] Robert Baumgartner, Piotr Majdak, and Bernhard Laback, "Modeling sound-source localization in sagittal planes for human listeners," *The Journal of the Acoustical Society of America*, vol. 136, no. 2, pp. 791–802, 2014.

[4] Areti Andreopoulou, Agnieszka Rogińska, and Hariharan Mohanraj, "A database of repeated head-related transfer function measurements," 2013.

[5] Chris Oreinos and Jörg M Buchholz, "Measurement of a full 3D set of HRTFs for in-ear and hearing aid microphones on a head and torso simulator (HATS)," *Acta Acust. United Ac.*, vol. 99, no. 5, pp. 836–844, 2013.

[6] Dominik Storek and Frantisek Rund, "Differential head related transfer function as a new approach to virtual sound source positioning," in *Radioelektronika, 2012 22nd International Conference*. IEEE, 2012, pp. 1–4.

[7] Dominik Storek, "Virtual sound source positioning by differential head related transfer function," in *Audio Engineering Society Conference: 49th International Conference: Audio for Games*. Audio Engineering Society, 2013.

[8] Udo Zölzer (ed.), *DAFX: Digital Audio Effects*, vol. 1, John Wiley & Sons, 2002.

[9] Dominik Storek, Jaroslav Bouse, Frantisek Rund, and Petr Marsalek, "Artifact reduction in positioning algorithm using differential HRTF," *Journal of the Audio Engineering Society*, vol. 64, no. 4, pp. 208–217, 2016.

[10] V Ralph Algazi, Richard O Duda, Dennis M Thompson, and Carlos Avendano, "The CIPIC HRTF database," in *IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, 2001, pp. 99–102.