PAN
ARCHIVES OF ACOUSTICS
Vol. **41**, No. 4, pp. 437–447 (2016)

Subjective Evaluation of Three Headphone-Based Virtual Sound Source Positioning Methods Including Differential Head-Related Transfer Function

Dominik STOREK, Frantisek RUND, Petr MARSALEK

Department of Radioelectronics, Czech Technical University in Prague
Technická 2, 166 27 Prague, Czech Republic; e-mail: storedom@fel.cvut.cz

(*received December 1, 2015; accepted March 2, 2016*)

This paper analyses the performance of Differential Head-Related Transfer Function (DHRTF), an alternative transfer function for headphone-based virtual sound source positioning within a horizontal plane. This experimental one-channel function is used to reduce processing and avoid timbre affection while preserving signal features important for sound localisation. The use of positioning algorithm employing the DHRTF is compared to two other common positioning methods: amplitude panning and HRTF processing. Results of theoretical comparison and quality assessment of the methods by subjective listening tests are presented. The tests focus on distinctive aspects of the positioning methods: spatial impression, timbre affection, and loudness fluctuations. The results show that the DHRTF positioning method is applicable with very promising performance; it avoids perceptible channel coloration that occurs within the HRTF method, and it delivers spatial impression more successfully than the simple amplitude panning method.

Keywords: virtual positioning; virtual reality; positioning method; positioning algorithm; head-related transfer function; amplitude panning.

Notations

AP – amplitude panning; panorama,
(D)HRIR – (differential) head-related impulse response,
(D)HRTF – (differential) head-related transfer function,
ILD – interaural level difference,
IPD – interaural phase difference,
ITD – interaural time difference,
ITF – interaural transfer function,
JND – just noticeable difference,
SL – sine law.

1. Introduction

For the purpose of auditory scene synthesis, multimedia applications require spatial separation of presented sound sources. The perception of spatial sound uses acoustical cues: delay times, sound level differences, and disparities due to the spectral characteristics of the outer ear (Blauert, 1997; 2013). Acoustical sound localisation cues arise from the geometrical and physical properties of sound wave propagation in the air (Xie, Zhang, 2010). The interaction between the sound wave and the listener's body can be described by *Head Related Transfer Functions* (HRTFs) Blauert (2013). Every individual has a unique HRTF, consequently an HRTF based on a prototypical listener's head can be used (Shinn-Cunningham *et al.*, 2000; Huang, Benesty, 2004; Yao, Chen, 2013). In virtual positioning it is necessary to simulate these acoustical features in both ear channels (Adams, Wakefield, 2008; Algazi, Duda, 2011; Rumsey, 2011). Naturally recorded or artificially generated auditory scenes can be reproduced with the use of headphones or via a set of spatially arranged loudspeakers. The former method employs processing of (usually) two separated channels, whilst the latter uses various sets of spatially separated channels (Zölzer, 2011). Two common methods employed to achieve the spatial illusion in headphone-based positioning include widely used *Amplitude Panning* (AP) (Pulkki, 2001) and filtering by HRTF (Blauert, 2013; Sodnik*et al.*, 2006). In this paper, the *Differential* HRTF (DHRTF) positioning method developed by the authors is compared to the AP and the HRTF in terms of quality of the rendered auditory space. Although it is very common to investigate primarily precision of a positioning method (Sodnik *et al.*, 2004; Pec *et al.*, 2007; Majdak *et al.*, 2010), this article focuses on particular aspects of perception of the virtual auditory environment: depth of the presented space, changes in timbre, and fluctuations in loudness.

In this paper, Sec. 2 *Binaural Cues* presents a brief basis of the sound source localisation in order to clarify several essential concepts. The principles of the introduced positioning algorithms are described in Sec. 3 *Positioning methods*. The next Sec. 4 *Objective comparison* reveals the objective differences between the particular positioning methods presenting their channel transfer function and position-dependent channel gain. The design and organisation of the listening tests for assessing the methods is introduced in Sec. 5 *Subjective Comparison*, while the consequent results are analysed and discussed in Secs. 6, 7, and 8.

2. Binaural cues

For a real sound source placed (out of the ears axis) within the horizontal plane, the incident sound wave reaches the farther (*contra-lateral*) ear with time delay corresponding to its longer pathway, given by the speed of sound in the air. Figure 1 shows the top view of a head with the sound source located at azimuth ϑ. The difference in delay times is referred to as the *Interaural Time Difference* (ITD). For a harmonic signal it can be expressed by *Interaural Phase Difference* (IPD, further denoted Ψ_{IPD}). Inter-channel attenuation is known as the *Interaural Level Difference* (ILD, expressed in dB as L_{ILD} and in linear scale as A_{ILD}). This attenuation is caused primarily by the head shadowing on particular wavelengths (Sodnik *et al.*, 2004). In a natural listening environment, both ITD and ILD are frequency dependent (Hartmann, Rakerd, 1989). In binaural hearing, the border between low and high frequencies is approximately 1.5 kHz (with respect to anthropometrical parameters). The ITD effects occur at lower frequencies and the effects of the ILD are present in the high frequency range (Blauert, 2013). This is primarily determined by the mechanisms of signal coding in the inner ear (Algazi, Duda. 2011) and in subsequent neurons of the auditory pathway (Marsalek, 2001; Marsalek, Kofranek, 2004; Kostal, Marsalek, 2010; Sanda, Marsalek, 2012). The elevation φ of the sound source measured as an angle within the sagittal plane is perceived due to the propagation of high frequency sound and its reflection within the outer ear (ear canal and pinna). These high frequency spectral components (monaural cues) can be observed above approximately $f = 6$ kHz (Blauert, 2013) and their character varies according to their vertical position (Blanco-Martin *et al.*, 2011; Malinina, Andreeva, 2010).

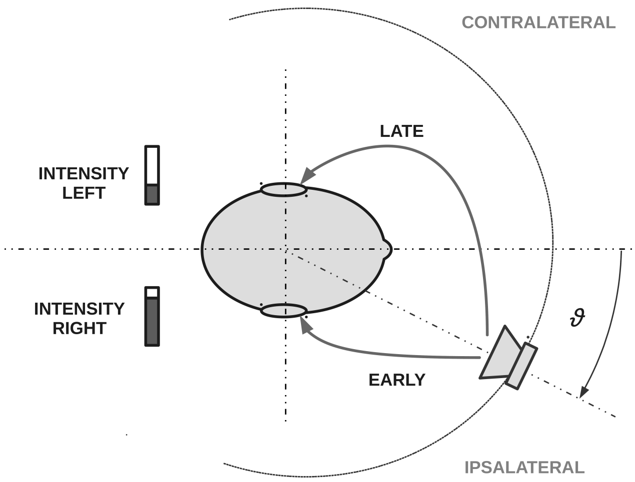

Fig. 1. ILD, ITD, and azimuth. A schematic top view of the spatial arrangement of the listener and sound source. Azimuth ϑ is measured from the eye-view. Higher sound intensity and early sound arrival are on the right side.

3. Positioning methods

Employing of the particular amplitude and time features is crucial for the virtual sound positioning. Since the DHRTF method is intended for positioning only in the horizontal plane, the following description considers that.

3.1. Amplitude panning

The simplest method to implement is *amplitude panning* (AP, panorama). This method puts into relationship the position of the source in the horizontal plane and the corresponding (frequency independent) gains of the left and right channels. Simplification of the geometry of the head is known as *sine law* (SL) formula (Pulkki, 2001; Zölzer, 2011), expressing the signal *amplitude* difference $L_{\mathrm{ILD}}(\vartheta)$ by linear-scaled gain for each channel. In the *sine law* formula, $g_{\mathrm{L}}(\vartheta)$ and $g_{\mathrm{R}}(\vartheta)$ refer to the respective channel gains and ϑ corresponds to the source angle position in the horizontal plane. The left and right channels are obtained as:

$$\begin{aligned} g_{\mathrm{L}}(\vartheta) &= \frac{1-\sin(\vartheta)}{\sqrt{2\left(1+\sin^2(\vartheta)\right)}}, \\ g_{\mathrm{R}}(\vartheta) &= \frac{1+\sin(\vartheta)}{\sqrt{2\left(1+\sin^2(\vartheta)\right)}}. \end{aligned} \tag{1}$$

The left and right channel amplitudes are multiplied by $1 \pm \sin\vartheta$, thus there are singular directions where one of the gains is set to 0, an occurrence which is not realistic. There are several other descriptions of the $L_{\mathrm{ILD}}(\vartheta)$ dependence on azimuth (e.g. tangential law or methods for bias reduction); however, for the purpose of this study, the SL primarily represents positioning approach of frequency independent gain modification, thus there is no need to discuss other geometric simplifications.

The amplitude panning is widely used in various multimedia applications with no special requirement for spatial fidelity such as simple PC games, music industry, film production, etc. This method is not designated for positioning in the sagittal plane.

3.2. Head-Related Transfer Function

The more elaborated method used for virtual sound source positioning is aimed at more precise description of the 3D head shape and corresponding sound interaction by utilising the HRTF (Blauert, 2013; Oreinos, Buchholz, 2013). The equivalent of the HRTF in the time domain is *Head-Related Impulse Response* (HRIR). HRTF can be considered as a pair of direction-dependent filters (Otcenasek, 2008) and usually defined and written as

$$H_{\mathrm{X}}(\vartheta, \varphi, \omega) = \frac{p_{\mathrm{X}}(\vartheta, \varphi, \omega)}{p_{\mathrm{S}}(\vartheta, \varphi, \omega)}, \qquad (2)$$

where $p_{\mathrm{X}}(\omega)$ represents the sound pressure in frequency domain at the position of the left or right ear canal entrance. Based on the context, X denotes the left or right side ($\mathrm{X} = L, R$), p_{S} corresponds to the sound pressure at the place of the sound source S at azimuth ϑ and elevation φ. Examples of HRTF filter curves that correspond to a spatial arrangement similar to that in Fig. 1 are shown in Fig. 2. Implementation of the positioning algorithm consists of double-channel filtering by the pair of transfer functions (or HRIRs). The HRTF is a well-known method commonly used in headphone based applications, where high-fidelity reproduction is required, e.g. virtual reality, simulators, advanced gaming.

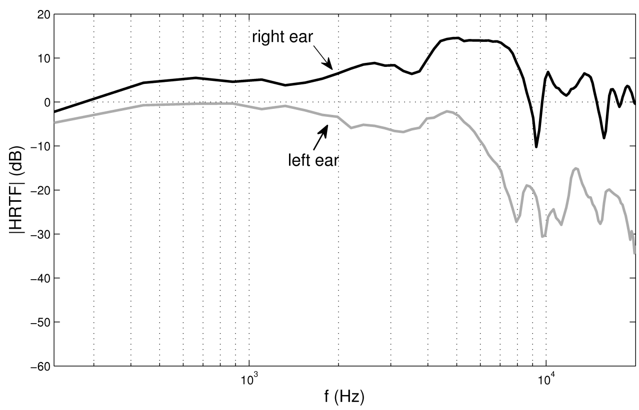

Fig. 2. HRTF magnitude. The magnitude of the HRTF of an artificial head (Algazi *et al.*, 2001) for $\vartheta = 70°$ and $\varphi = 0°$. Spectral peaks and notches vary uniquely in accordance with the source position.

3.3. Differential HRTF

Even though the AP is very simple and computationally almost trivial, its performance does not correspond to the real signal perception of either ear. The ILD does not take into consideration the frequency dependence, and the time shift between both of the channels is omitted. Hence, poor illusion of a virtual space with lateralisation of the sources is obtained when the sound is presented by headphones. In contrast, HRTF-based positioning provides more realistic spatial effect, ensuring the sound source is ideally perceived out of the head when listened via headphones allowing front-back resolution in the horizontal plane (Suzuki *et al.*, 2008; Ortega-González *et al.*, 2010; Zhang, Hartmann, 2010; Wersenyi, 2009). This section introduces the basis of the experimental method called *Differential Head-Related Transfer Function* (DHRTF). The first pilot study by the authors was presented in (Storek, 2013). Assume that the common AP processing changes the amplitude ratio in both channels. The final perceived in-head position does not depend on the absolute amplitude of both signals, but on their difference expressed by the ILD. It can be also assumed that both signals are not approaching extreme high or low levels within the hearing dynamic range. When the HRTF positioning method is applied, separate HRTF filtering results in mutual differences in both channels and frequency-dependent ILD and ITD emerge. The principle of the Differential HRTF lies in introduction of the frequency dependent ILD and ITD to the stereo signal. Therefore, filtering by a pair of HRTFs is reduced to a one-channel filtering, where the same inter-channel differences occurs in the positioned sound as when filtered by HRTF. Only one channel is processed while the other one remains completely untouched. The concept is demonstrated in Fig. 3. However, this procedure heavily distorts the monaural spectral cues that are essential for sound localisation in sagittal planes (Baumgartner *et al.*, 2014; Langendijk, Bronkhorst, 2002), thus the method is intended to be used only within the horizontal plane (as AP is). The Differential HRTF can be defined as the ratio of contra-lateral (farther) and ipsi-lateral (closer) HRTFs. This can be expressed as (Storek, 2013):

$$\begin{aligned} D_{\mathrm{HRTF}}(\vartheta, \omega) &= \frac{H_{\mathrm{C}}(\vartheta, \omega)}{H_{\mathrm{I}}(\vartheta, \omega)} \\ &= \frac{|p_{\mathrm{C}}(\vartheta, \omega)|}{|p_{\mathrm{I}}(\vartheta, \omega)|} \exp\left(j\psi_{\mathrm{C}}(\vartheta, \omega) - j\psi_{\mathrm{I}}(\vartheta, \omega)\right) \\ &= A_{\mathrm{ILD}}(\vartheta, \omega) \exp(j\Psi_{\mathrm{IPD}}(\vartheta, \omega)), \end{aligned} \qquad (3)$$

where Ψ_{C} and Ψ_{I} denote the phase of the particular channel. The equation expresses the DHRTF as the ratio of the sound pressures in frequency domain at both sides. Therefore, the ILD cue is coded in the magnitude of the DHRTF and the IPD (ITD) cue in the phase of the DHRTF. In the definition, sides are denoted as Ipsi-lateral and Contra-lateral (I, C), hence calculation of the DHRTF does not depend on the choice

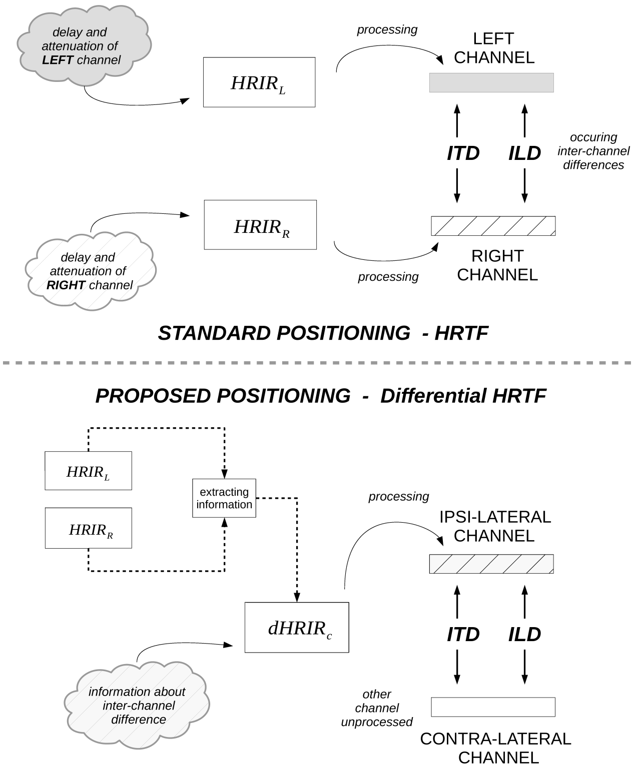

Fig. 3. HRTF *versus* DHRTF. HRTF (top) and DHRTF (bottom) positioning methods are compared here. In the HRTF method, both L and R channels are processed in parallel. In the DHRTF method, information about time and level differences is extracted from the HRTF pair and applied to only one channel of the stereo signal.

of sides. Application of the inverse Fourier transform on the DHRTF results in the *Differential Head Related Impulse Response* (DHRIR) that is used for the implementation. An example of an ordinary DHRTF is shown in Fig. 4, gray line. The concept of employing the ratio of the contra- and ipsi-lateral HRTFs being referred to as *Interaural Transfer Function* (ITF, IATF) has been previously used in several applications. The ITF was employed for cross-talk cancellation in (Gardner, 1998), for modelling of the contra-lateral HRTF from a measured ipsi-lateral (Avendano *et al.*, 1999), or for low-order approximation of the contra-lateral HRTF (Lorho *et al.*, 2000). However, it has not been used in a concept of direct virtual positioning. The authors use designation *Differential* HRTF to underline employment of the ITF as a one-channel positioning method (*differential* refers to difference of the two HRTFs in the logarithmic scale). In specific HRTF pairs, an unexpected phenomenon occurs. Magnitude of the HRTF corresponding to the ipsi-lateral channel may be lower on particular frequencies than that in the contra-lateral channel (against expectation that the signal in the contra-lateral channel is always weaker). This results in an artificial narrow-band notch (spike) exceeding the gain level of 0 dB. The perceptual effect of this spectral spike leads to highly noticeable disturbing artifacts perceived as unwanted pure tone character disturbance in the contra-lateral channel. However, the presence of the spike in the DHRTF is neither determined for specific spectral bands, nor for specific positions. Due to the principle of the method, the artifacts are generally likely to occur around $\vartheta = 0°$ and 180°. The artifact phenomenon is demonstrated in Fig. 4, black line. A more comprehensive analysis of the artifacts and their elimination (by employing spectral limitation and low-pass filtering for the DHRTF spectrum) can be found in (Storek *et al.*, 2016).

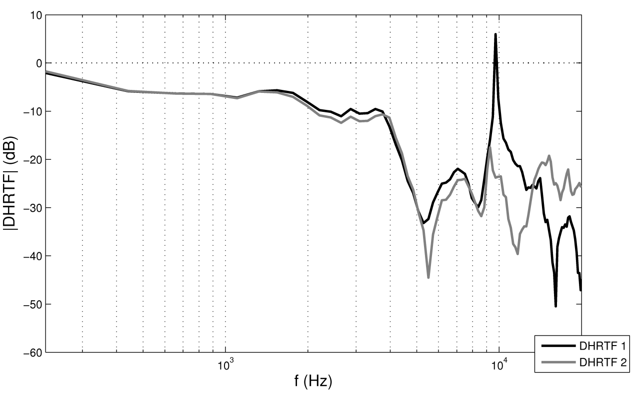

Fig. 4. DHRTF magnitude. The DHRTF for a source placed in the horizontal plane ($\vartheta = 70°$ and $\varphi = 0°$) is created from two different sets (different subjects) of HRTF. The presence of the spectral spike (black line) exceeding the level of 0 dB is undesirable due to the occurrence of positioning artifacts. The DHRTF represented by gray line is not subject to artifacts.

4. Objective comparison

The three positioning algorithms described above were examined for specific features. Energy of the channel response in dependence on azimuth (gain curves, obtained as sum of squares of the response) of the three methods is shown in Fig. 5. A considerable increase of gain in the DHRTF around the front and back positions ($\vartheta = 0°$, 180°) results from the occurrence of the *negative ILD* (see Fig. 4). Unlike the amplitude panning method, HRTF shows a different course for some particular positions, even though the same trend of rising gain for ipsi-lateral channel of the gain curve is preserved. The most significant is the variation of the total gain. Notice also the different total gain corresponding to the front and back source positions. This phenomenon results from the shadowing effects of the pinna structure for the back source position in higher frequencies. Another significant feature is a non-zero gain for the side position of the contra-lateral ear. This feature has an important role in natural sounding of the processed stimuli. In open space listening the contra-lateral total gain is reduced approximately by only 18 dB to the ipsi-

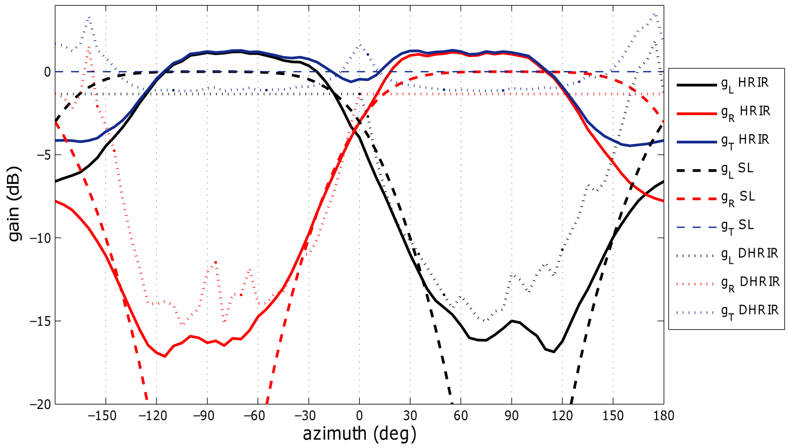

Fig. 5. Azimuth related gains of AP, HRTF, and DHRTF. Logarithmic expressions of channel gains are shown here for the three compared methods. Gains of the left, g_{L}, and right, g_{R}, channels are shown together with the total gain $g_{\mathrm{T}} = g_{\mathrm{L}} + g_{\mathrm{R}}$. Solid lines show HRIR gains of the HRTF, dashed lines show sine law gains of the AP, and dotted lines show DHRIR gain of the DHRTF method.

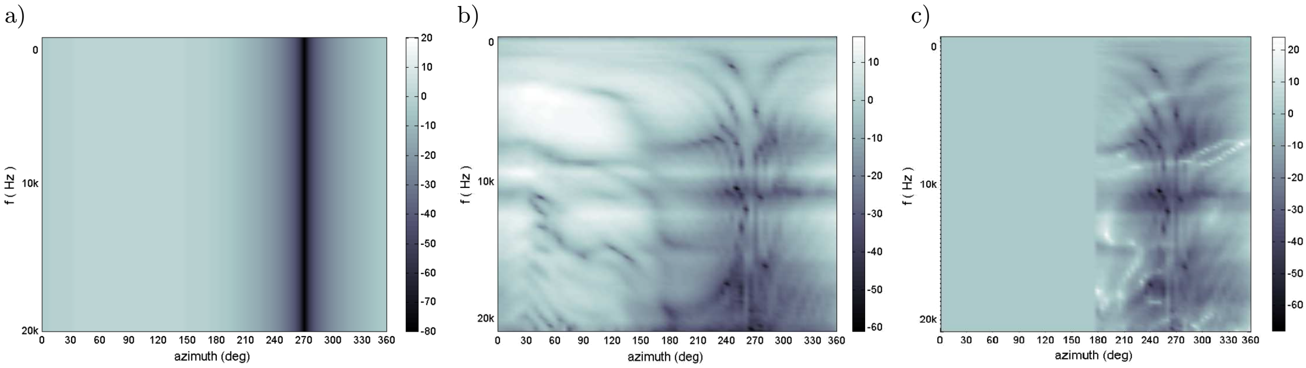

Fig. 6. Azimuth transfer functions. The behavior of channel transfer functions of the right channel (artificial head) corresponding to particular methods are shown. The gray map is for 360 degrees of the horizontal plane (ϑ, on x-axis, position is sampled by 5° step, y-axis shows sound frequencies). The shades of grey correspond to the attenuation magnitudes in dB. Range $\vartheta \in (0°,\ 180°)$ denotes the right half (ipsi-lateral) of the auditory space, whilst range $\vartheta \in (180°,\ 360°)$ covers the left (contra-lateral) half-plane. a) AP: the magnitude of the transfer function remains constant under entire frequency range, b) HRTF: unique spectral features are commonly observed within the HRTF, c) DHRTF: since this demonstration corresponds to the *right* ear, the transfer function remains constant for the ipsi-lateral position of the source.

lateral total gain, as the HRTF method shows. This behaviour is quite well followed also by the DHRTF method.

Other important characteristics of the methods is their directional-dependent transfer function. See frequency relations for one (right) ear with the use of grey maps in Fig. 6 for the three methods. Panel (a) illustrates the transfer function (frequency independent) for AP that shows a pronounced attenuation at $\vartheta = 270°$ (dark stripe), which is related to the direction of the ear opposed to the sound source. The HRTF function does not have such marked attenuation as can be observed in panel (b). Specific frequency dependent features are apparent over the azimuth range. An unique character can be observed within the DHRTF in panel (c). Since all the gray maps refer to the transfer function of the right channel, all the salient features in the right half-space of panel (c) corresponding to contra-lateral position of the source are preserved, while the left half-space corresponds to constant 0 dB level (as when the ipsi-lateral signal remains original).

By analysing the transfer function of both channels for the three described methods a position dependent ILD is obtained for a full 360 degree range in the horizontal plane, as shown in Fig. 7. The figure introduces the resulting ILD for the AP (a), HRTF positioning (b), and DHRTF positioning methods (c). It is apparent that the features of panels (b) and (c) are completely identical.

Several hypotheses on the DHRTF performance resulted from the objective analysis. Due to the one-channel processing, the DHRTF method was expected to perceptibly change the timbre of the positioned sound and specific loudness fluctuations along the horizontal plane were predicted to occur. The hypotheses were to be confirmed or disproved by the listening tests.

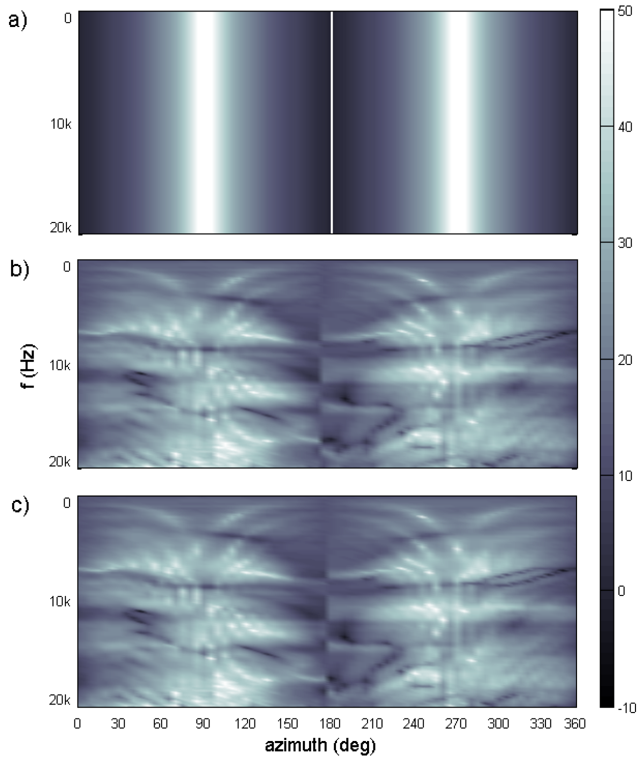

Fig. 7. ILD azimuth functions. The gray map describes frequency dependence of ILD across positions of AP (a), HRTF (b), and Differential HRTF (c) positioning methods for the range of 360° in the horizontal plane. The bar corresponds to the magnitude in dB.

5. Subjective comparison

In order to investigate how the stimuli positioned by the DHRTF method are perceived by the listeners and what the difference in perception compared to the other two positioning methods is, subjective listening tests were performed. The outputs of the objective comparison of the methods resulted in the selection of three parameters to be assessed in the listening test. The parameters were not primarily focused on the investigation of localisation precision or the JND (Just Noticeable Difference), since the JND has been investigated in previous work (Storek, 2013). The factors to be assessed and rated by the subjects were as follows:

- **Spatial impression** represents the effect of spatial fidelity and credibility of the sound source located at particular positions; i.e. natural sounding.
- **Coloration** regards affection of the sound timbre. The main goal was to verify whether the DHRTF would incline to disturbing coloration of the final positioned sound due to the one-channel filtering.
- **Loudness** was expected to vary along particular positions according to Fig. 5. Varying loudness might be perceptible specifically when the positioned sound source moves.

All HRTFs used in this article are from the freely available CIPIC HRTF database (Algazi *et al.*, 2001).

5.1. Listening test

A graphical user interface was designed and used for presenting stimuli to the subject and gathering the subjects' responses. Each trial of the test consisted of presenting four stimuli to the subject; three positioned stimuli to be assessed and one monaural reference stimulus (the original sound to be positioned). The reference was always presented first and the order of the following samples positioned by particular methods was randomised. After the initial presentation of all the stimuli the subject had unlimited option to listen to the presented sounds again by clicking on buttons corresponding to particular sounds. The subject was asked to adjust the value of *sliders* representing particular parameters (spatial impression, coloration, loudness) of each *unknown* positioning method. The slide scale consisted of 0.5 interval steps from 1 to 5 and were identical for all the three parameters. Verbal equivalents of slider value ratings are summarised in Table 1 in exact wording, as they were presented

Table 1. Rating keys. Word expressions corresponding to numeric values were assigned to the assessed segments.

Score	Spatial impression		Coloration		Loudness	
	Rating	Description	Rating	Description	Rating	Description
1	very poor	dull sound inside the head	much worse	timbre is much different much worse than the reference	well perceptible	DECREASE of loudness
2	poor	–	slightly worse	–	barely perceptible	DECREASE of loudness
3	average	credible source position but no natural character	inaudible	the same timbre perception of both stimuli	same impression	NO CHANGE of loudness
4	good	–	slightly better	–	barely perceptible	INCREASE of loudness
5	excellent	sound outside the head in specific position	much better	timbre is different – much better than the reference	well perceptible	INCREASE of loudness

in subjects' instructions. This approach was chosen according to recommendations in (Otcenasek, 2008). Finally, the subject was asked to select the most preferred stimulus intuitively according to the quality of spatial impression and natural character of the sound.

5.2. Stimuli description

Three different stimuli, with lengths ranging from 1.6 s to 3.4 s, were chosen for the test; snare drum phrase, speech segment, guitar chord. Each of the stimuli was positioned using the particular methods: for AP the samples of each channel were multiplied by corresponding gains, and direct convolution of the stimuli and 200-samples long filter response (FIR of order 199) was implemented for HRTF and DHRTF. The convolution within the DHRTF was performed only for one channel, as results from its definition. Spatial division for the front half-plane was chosen simply in the range from $\vartheta = -90°$ to $\vartheta = +90°$ with a step of $\Delta\vartheta = 30°$. Therefore, the 3 stimuli and 7 positions result in 21 trials of the test. Each trial was expected to last no more than one minute, thus the session length did not exceed 25 minutes in order to maintain the subjects' motivation to fulfill the task correctly (Otcenasek, 2008).

The length of the original HRTF data set as well as the resulting DHRTF data set consisted of 200 samples of standard sampling frequency 44.1 kHz. Therefore, the maximal time length of HRIR corresponds to 4.54 ms. The DHRTFs were selected from two available HRTF sets for acoustic manikin (Algazi *et al.*, 2001) in order to avoid the spectral spike occurrence. It is important to notice that in terms of assessing loudness, mutual gain of the particular methods was normalised to the same mean gain. The gains are shown in Fig. 5. It is also assumed that the differences within the methods are much more significant than differences resulting from occasional deviations of the subjects' anthropometric parameters from the manikin's (Fels, Vorländer, 2009). Therefore, subjective dependences of the individual HRTFs were not taken into account.

6. Results

The test was performed on 26 subjects, aged from 19 to 43. Both musically skilled subjects and people with no musical background were included in this set. The results were statistically analysed by the software GraphPad Prism. The following graphs present the results of each assessed parameter by *boxes* representing 25% to 75% percentiles and *whiskers* showing the sample standard deviation. The mean value of each data set is represented by a horizontal line in the box. The results were subjected to multiple factor analysis of variance, RM-ANOVA (Repeated Measures Analysis of Variance), with two factors: *positioning method* (AP, HRTF, DHRTF) and *position* ($\vartheta \in \{-90, -60, -30, -0, +30, +60, +90\}$).

Figure 8 shows the results for the ratings of *spatial impression*. The most weak spatial effect was provided by the *amplitude panning* method, while the best results of spatial depth were produced by the HRTF method. The results of the DHRTF method appear in the middle range, inclining more to the character of the HRTF. The results of the spatial impression parameter formed a *V-shape*, with the tip of the "V" letter pointing to $\vartheta = 0°$ as the cues for perception of the space depth are connected to the synchronous ITD and ILD. This becomes more robust towards to the side positions. For the central position of $\vartheta = 0°$ the average values are almost identical. The analysis of variance revealed the following statistical outcomes: for variance within the *method* $F(2, 24) = 109.1$, $p < 0.0001$, and for variance within *position* $F(6, 20) = 11.08$, $p < 0.0001$. This means that both factors *positioning method* and *position* are statistically significant in rating of spatial impression. Results for channel *coloration* are shown in Fig. 9. The line at the value of 3 denoted *Imp.* on the y-axis refers to the level of imperceptibility. Despite the fact that the AP is the only method, which does not include channel filtering its rating is inferior to the other two methods, specifically from the side positions. This effect is probably connected with the unnatural character of the sound resulting from a close-to-zero gain in the contra-lateral channel in these positions. The HRTF and DHRTF have comparable values of their means along the az-

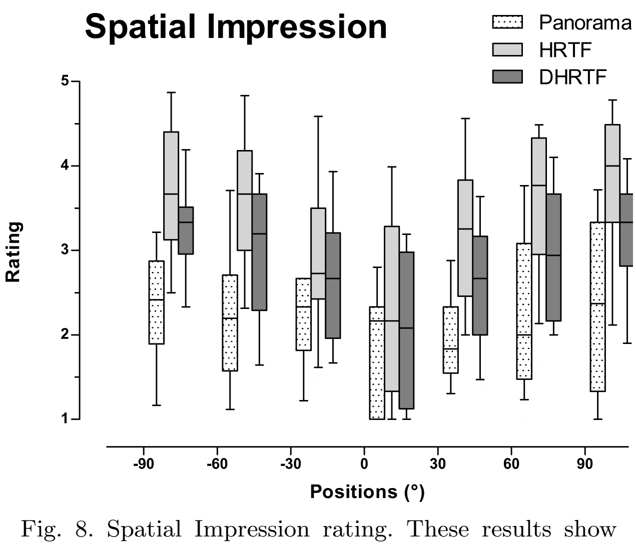

Fig. 8. Spatial Impression rating. These results show ratings of the parameter *spatial impression*. Amplitude panning shows the lowest spatial effect beside HRTF which provides the best results. DHRTF inclines more to the attributes of HRTF, however, with a slightly worse impression. The *boxes* represent 25% to 75% percentiles and *whiskers* show the standard deviation. The mean value of each data set is represented by a horizontal line in the box.

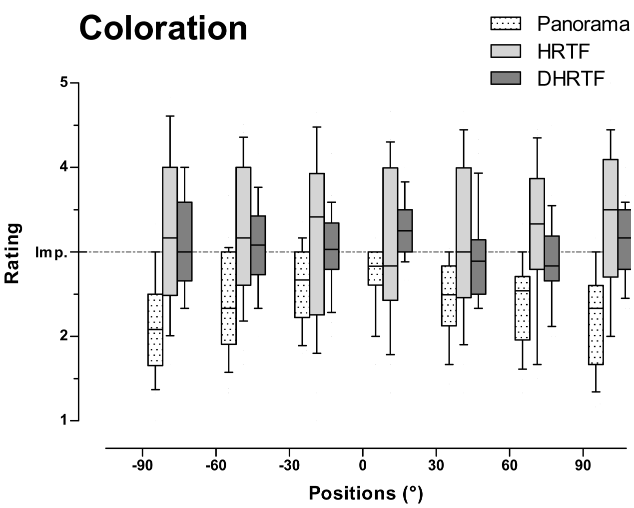

Fig. 9. Coloration rating. Results are shown for the *coloration* parameter. Though the average values of DHRTF and HRTF are similar, coloration is more affected by the HRTF according to the larger variation of the rating. This timbre change is either preferred or rated as worse. The unnatural character also probably contributes to the low rating of the AP method. The *boxes* and *whiskers* are the same as in Fig. 8.

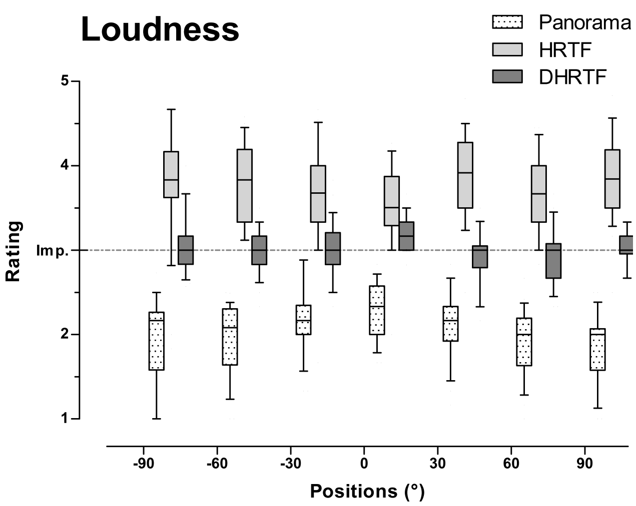

Fig. 10. Loudness rating. Results are shown for the *loudness* parameter. A boost of middle frequencies causes primarily higher perception of loudness in the HRTF method. In the DHRTF method, the middle frequencies differed minimally. The *boxes* and *whiskers* are similar to Fig. 8.

imuth. However, the deviation of the HRTF is remarkably higher. This phenomenon results from easily perceived stimuli timbre changes within the HRTF method that is caused by a boost in mid frequencies of the positioned sound (see the gain of the ipsi-lateral HRTF, right, in Fig. 2). This effect was assessed by both *better* and *worse* options, specifically, when musically skilled subjects preferred the mid-boost character. The DHRTF method preserves the original timbre of the stimuli the most against the previous hypothesis. This is most likely caused by maintaining the unprocessed channel as dominant resulting in the perception of the sound timbre close to the origin even in side positions, where the difference is maximal. RM-ANOVA revealed the following outcomes for variance within *method* $F(2,24) = 60.67$, $p < 0.0001$, and for variance within *position* $F(6,20) = 1.18$, $p = 0.32$. The values show that only the factor of *positioning method* is statistically significant for *coloration* parameter. A slight trend of dependence on *position* is observable for AP, when the outer positions are assessed worse, probably due to the unnatural character.

Regarding the *loudness* assessment, see the graph in Fig. 10. The perception of loudness did not vary significantly across the positions. In accordance with the total gain curves (see Fig. 5) a slight rise for the DHRTF and small decay for the HRTF at central position is noticeable. The gains for all the methods were aligned using the same mean value; however, the results for loudness variation show a difference. It is important to note that the total gains for HRTF and DHRTF methods were derived based on the energy of their frequency-dependent impulse response. Under normal conditions the loudness perception depends also on the spectral character of the processed sound. A typical shape of the HRTF contains a resonance peak between 4 and 8 kHz (see Fig. 2), which corresponds to the most sensitive area of the human ear (Algazi, Duda, 2011). This results in the previously discussed mid-boost effect which may be finally reflected as an increased perception of loudness. The decreased AP rating is probably also a result of spectral independence of the changes. The perception of loudness variance for the DHRTF method is minimal, except for the small increase in the central position. The analysis of variance revealed the following outputs: for variance within *method* $F(2,24) = 706.1$, $p < 0.0001$, and for variance within *position* $F(6,20) = 1.23$, value $p = 0.30$. This means that only the factor of *positioning method* is statistically significant. The results disprove the previous hypothesis for the DHRTF that the loudness will fluctuate significantly along the positions due to the non-uniform (one-channel) filtering. Despite the examined positions were roughly distributed in the frontal plane, a follow-up experiments performed in (Storek *et al.*, 2016) confirmed this statement by employing moving virtual sound objects.

The last task of each trial of the test was to select the most preferred stimuli. The results presented in Fig. 11 show that the method preferences were not consistent within stimuli and this is possibly related to their spectral content. For sharp stimuli with strong high-frequency content such as the snare drum phrase (a) the HRTF positioning resulted in a strong boost and an even more sharpened sound. Therefore, the *milder* DHRTF was mostly chosen in this case. How-

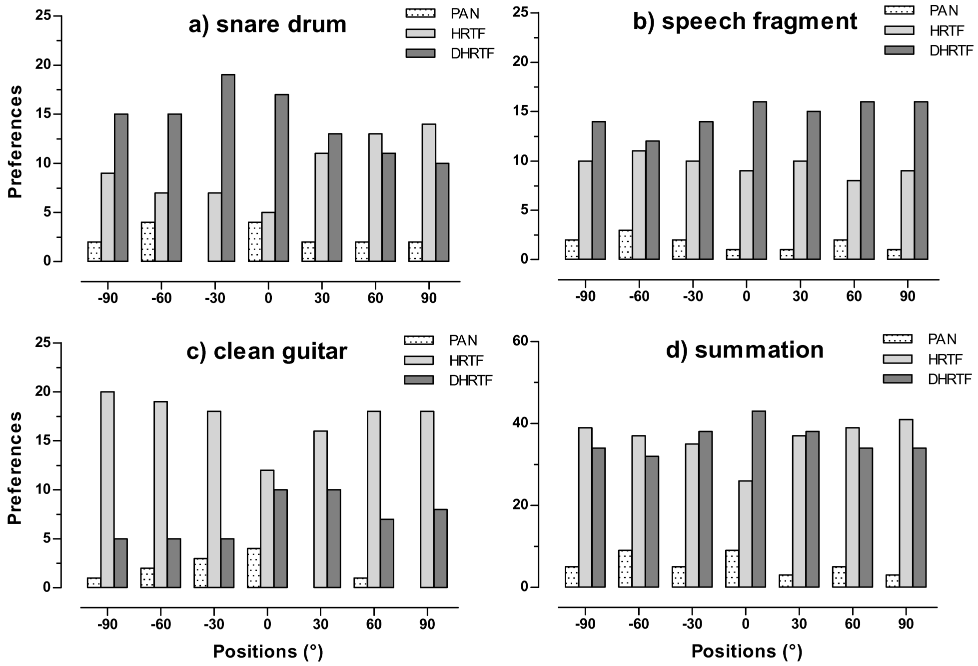

Fig. 11. Preferences based on content. Histograms of the method preference that depended on the character of the stimuli are shown. Sharp sounds with high-frequency content are sensitive to significant boosting by HRTF. This effect might be desirable for tonal instrument characters (guitar).

Table 2. Methods summary. Attributes of particular positioning methods.

	Method		
	AP	HRTF	DHRTF
Processing requirements	multiplication 2 channels	convolution 2 channels	convolution 1 channel
Spatial impression	poor	excellent	excellent
Various elevation	no	yes	no
Front/back positioning	no	yes	partly
Channel coloration	none	high	tiny

ever, the *high-band* and *mid-band* enhancement may have even improved the entire stimuli sounding, as in the case of the guitar chord sound (c) due to its tonal character. This effect contributed to a good spatial quality, thus the HRTF was selected by the majority in this case. The subjects preferred mostly the DHRTF also for the male speech fragment (b). In the final summary (d), the DHRTF and HRTF were most preferred and basically equal, as compared to the AP method (DHRTF 46.3%, HRTF 46.5%, AP 7.2%), which was preferred only by a minority.

The attributes of each method are summarised in Table 2. While amplitude panning offers simple implementation at the cost of poor spatial impression, the HRTF demonstrates a complex approach with good spatial results. The DHRTF enables the reduction of processing to only one channel, while preserving remarkable spatial outputs and negligible channel coloration.

7. Discussion

The DHRTF based method can have useful applications in headphone listening. Any listener may expect sound reproduction to have the following qualities: it is pleasant, it feels natural, and it achieves the desired sound location perception. To test how the DHRTF method satisfies these requirements, parameters related to the qualities described above were chosen: spatial impression relates to location effects, coloration captures both how pleasant and natural the sound is, albeit mostly for a trained ear, and loudness should change smoothly and in a sense that it is related to all the qualities mentioned above.

The objective analysis highlights points where artifacts and noises can distort listening. The DHRTF method performed remarkably well in the subjective evaluation.

The DHRTF might prove advantageous in comparison with the HRTF. Two-channel processing may increase the requirements for computational resources when multiple sources are rendered simultaneously. This situation might arise in computer games or in training assistive programs for the visually impaired (Huang, Benesty, 2004; Seki, Sato, 2011). The DHRTF can be also effectively used in music post-processing (mixing), since the method provides very low timbre affection along with solid spatialisation. Some other sound examples to test with the three methods can be found in the collection made available by R.O. Duda (1996).

8. Conclusions

The performance of the DHRTF positioning method was investigated in this article. Subjective tests have shown that the proposed DHRTF positioning method shows promising and statistically significant results in comparison with the other widely used methods of virtual sound source positioning: amplitude panning and HRTF positioning. Due to its one-channel filtering, the DHRTF can be applied in devices and setups with limited access to computational resources. The results discovered that an important advantage of the DHRTF method is the preservation of the original sound timbre, which may be utilised in musical applications requiring separation of the sources in the stereo base (e.g. common mixing procedure in song production) while preserving the original timbre for aesthetic purposes. Such mixing procedure would deliver more natural spatial separation of the sources (instruments) than the commonly used amplitude panning, not affecting timbre of particular tracks as when the HRTF method is employed. The future research will be aimed at investigation of the artifacts and possibilities to remove them.

Acknowledgment

This work was supported by the Internal Grant Agency of the Czech Technical University in Prague, grant No. SGS 14/204/OHK3/3T/13 to D.S. and F.R., by International Visegrad Fund Agency, Visegrad-Taiwan Scholarship, No. 51390012 to D.S. Thanks to Elisa Brann, Eduard Kuriscak and Tomas Mika.

References

1. Adams N.H., Wakefield G.H. (2008), *State-space synthesis of virtual auditory space*, IEEE T. Audio Speech, **16**, 5, 881–890.
2. Algazi V.R., Duda R.O. (2011), *Headphone-based spatial sound*, IEEE Signal Proc. Mag., **28**, 1, 33–42.
3. Algazi V.R., Duda R.O., Thompson D.M., Avendano C. (2001), *The CIPIC HRTF database*, [in:] IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics, pp. 99–102.
4. Avendano C., Duda R.O., Algazi V.R. (1999), *Modeling the contralateral HRTF*, [in:] Audio Engineering Society Conference: 16th International Conference: Spatial Sound Reproduction, Audio Engineering Society.
5. Baumgartner R., Majdak P., Laback B. (2014), *Modeling sound-source localization in sagittal planes for human listeners*, The Journal of the Acoustical Society of America, **136**, 2, 791–802.
6. Blanco-Martin E., Casajús-Quirós F.J., Gómez-Alfageme J.J., Ortiz-Berenguer L.I. (2011), *Objective measurement of sound event localization in horizontal and median planes*, J. Audio Eng. Soc., **59**, 3, 124–136.
7. Blauert J. (1997), *Spatial hearing: the psychophysics of human sound localization*, MIT press.
8. Blauert J. (2013), *The technology of binaural listening*, Springer Verlag, Berlin, eBook.
9. Duda R.O. (1996), *Auditory localization demonstrations*, Acta Acust. United Ac., **82**, 2, 346–355.
10. Fels J., Vorländer M. (2009), *Anthropometric parameters influencing head-related transfer functions*, Acta Acust. United Ac., **95**, 2, 331–342.
11. Gardner W.G. (1998), *3-D audio using loudspeakers*, Springer Science & Business Media.
12. Hartmann W.M., Rakerd B. (1989), *On the minimum audible angle – a decision theory approach*, J. Acoust. Soc. Am., **85**, 5, 2031–2041.
13. Huang Y., Benesty J. (2004), *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, Springer, Boston, MA, USA.
14. Kostal L., Marsalek P. (2010), *Neuronal jitter: can we measure the spike timing dispersion differently*, Chin. J. Physiol., **53**, 454–464.
15. Langendijk E.H., Bronkhorst A.W. (2002), *Contribution of spectral cues to human sound localization*, The Journal of the Acoustical Society of America, **112**, 4, 1583–1596.
16. Lorho G., Huopaniemi J., Zacharov N., Isherwood D. (2000), *Efficient HRTF synthesis using an interaural transfer function model*, [in:] Signal Processing Conference, 2000 10th European, pp. 1–4, IEEE.
17. Majdak P., Goupell M.J., Laback B. (2010), *3-D localization of virtual sound sources: effects of visual environment, pointing method, and training*, Atten. Percept. Psycho., **72**, 2, 454–469.
18. Malinina E.S., Andreeva I.G. (2010), *The role of spectral components of the head-related transfer functions in evaluation of the virtual sound source motion in the vertical plane*, Acoust. Phys., **56**, 4, 576–583.
19. Marsalek P. (2001), *Neural code for sound localization at low frequencies*, Neurocomputing, **38**, 1443–1452.

20. Marsalek P., Kofranek J. (2004), *Sound localization at high frequencies and across the frequency range*, Neurocomputing, **58**, 999–1006.
21. Oreinos C., Buchholz J.M. (2013), *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., **99**, 5, 836–844.
22. Ortega-González V., Garbaya S., Merienne F. (2010), *Reducing reversal errors in localizing the source of sound in virtual environment without head tracking*, [in;] Haptic and Audio Interaction Design, pp. 85–96.
23. Otcenasek Z. (2008), *On Subjective Evaluation of Sound* [in Czech], Akademie muzickych umeni, Prague, Czech Republic.
24. Pec M., Bujacz M., Strumiłło P. (2007), *Personalized head related transfer function measurement and verification through sound localization resolution*, [in:] Proceedings of the 15th European Signal Processing Conference, pp. 2326–2330.
25. Pulkki V. (2001), *Localization of amplitude-panned virtual sources II: Two-and three-dimensional panning*, J. Audio Eng. Soc., **49**, 9, 753–767.
26. Rumsey F. (2011), *Whose head is it anyway? Optimizing binaural audio*, J. Audio Eng. Soc., **59**, 9, 672–675.
27. Sanda P., Marsalek P. (2012), *Stochastic interpolation model of the medial superior olive neural circuit*, Brain Res., **1434**, 257–265.
28. Seki Y., Sato T. (2011), *A training system of orientation and mobility for blind people using acoustic virtual reality*, IEEE T. Neur. Sys. Reh., **19**, 1, 95–104.
29. Shinn-Cunningham B.G., Santarelli S., Kopco N. (2000), *Tori of confusion: Binaural localization cues for sources within reach of a listener*, J. Acoust. Soc. Am., **107**, 3, 1627–1636.
30. Sodnik J., Susnik R., Tomazic S. (2004), *Acoustic signal localization through the use of head related transfer functions*, Systemics, Cybernetics and Informatics, **2**, 6, 56–59.
31. Sodnik J., Susnik R., Tomazic S. (2006), *Principal components of non-individualized head related transfer functions significant for azimuth perception*, Acta Acust. United Ac., **92**, 2, 312–319.
32. Storek D. (2013), *Virtual sound source positioning by differential head related transfer function*, [in:] Audio Engineering Society Conference: 49th International Conference: Audio for Games, Audio Engineering Society.
33. Storek D., Bouse J., Rund F., Marsalek P. (2016), *Artifact reduction in positioning algorithm using differential HRTF*, Journal of Audio Engineering Society, **64**, 208–217.
34. Suzuki S., Murase M., Wakunami K., Takashi T. (2008), *The effect of head motion and HRTF on human auditory localization by headphone presented sound*, [in:] The 3rd International Symposium on Biomedical Engineering, pp. 1–4.
35. Wersenyi G. (2009), *Effect of emulated head-tracking for reducing localization errors in virtual audio simulation*, IEEE T. Audio Speech, **17**, 2, 247–252.
36. Xie B., Zhang T. (2010), *The audibility of spectral detail of head-related transfer functions at high frequency*, Acta Acust. United Ac., **96**, 2, 328–339.
37. Yao S.-N., Chen L.J. (2013), *HRTF adjustments with audio quality assessments*, Archives of Acoustics, **38**, 1, 55–62.
38. Zhang P.X., Hartmann W.M. (2010), *On the ability of human listeners to distinguish between front and back*, Hearing Research, **260**, 1, 30–46.
39. Zölzer U. (2011), *DAFX: digital audio effects*, Wiley Online Library, Hoboken, NJ, USA.