# LOW-ORDER MODELING OF HEAD RELATED TRANSFER FUNCTIONS BASED ON SPECTRAL SMOOTHING AND PRINCIPAL COMPONENT ANALYSIS

Michał Pec, Michał Bujacz, Paweł Strumiłło
Lodz University of Technology, Institute of Electronics
michal.pec@gmail.com, bujaczm@p.lodz.pl, pawel.strumillo@p.lodz.pl

## ABSTRACT

The presented paper discusses methods of reducing the order of HRTF filters. Two alternative methods of HRTF preprocessing were proposed: wavelet based approximation of HRTF magnitude spectrum and dimensionality reduction by means of Principal Component Analysis. Frequency components of HRTFs were weighted according to the non-linear frequency resolution of human hearing via transformation to the bark scale. The preprocessed HRTFs were modeled with either recursive or non-recursive filters in linear and non-linear frequency domains.

## 1. INTRODUCTION

Filtering of audio signals through Head Related Transfer Functions (HRTFs) is a method of obtaining spatialized sound through stereophonic headphones. In most applications utilizing this technique HRTFs are modeled with Finite Impulse Response Filters (FIR) of a sufficiently large order. Previous studies conducted by the authors implemented HRTFs using FIR filters of the 128th order and allowed virtual sound sources to be localized by listeners with average errors of $6.36°$ in azimuth and $9.47°$ in elevation [1]. Although, efficient implementation of a filter of even a higher order is possible on PC computer or a DSP processor, lots of portable devices such as mobile phones might have insufficient computing power for generating multiple virtual sound sources. Objective measures and listening tests show that the methods proposed in this paper allow for low order modeling of HRTFs without noticeable reduction in localization accuracy and perceived sound quality.

## 2. HEAD RELATED TRANSFER FUNCTIONS

The Head Related Transfer Function is defined as a frequency response of linear and time invariant acoustic system formed by the head, ear pinna and torso to a wideband acoustic stimulus placed in the far field. The position of sound source is defined in a vertical-polar coordinate system (shown in Fig. 1) given by two angles: azimuth ($\theta$) and elevation ($\varphi$).

The representation of an HRTF in the time domain is called the Head Related Impulse Response (HRIR).

$$HRIR(\theta,\varphi,t) = \mathcal{F}^{-1}(HRTF(\theta,\varphi,\omega)) \qquad (1)$$

The most reliable method of acquiring values of HRTFs is by measuring the response to a known acoustic wideband signal with microphones placed at the entrances to the ear channels. The measurement is conducted simultaneously for both ears for a large number of directions relative to the listener. The authors conducted a series of HRTF measurements using apparatus constructed in cooperation with the Wroclaw University of Technology [1]. The measurement set allows collection of HRTFs for a full azimuth range and elevation angles in range $(-45° : 90°)$.

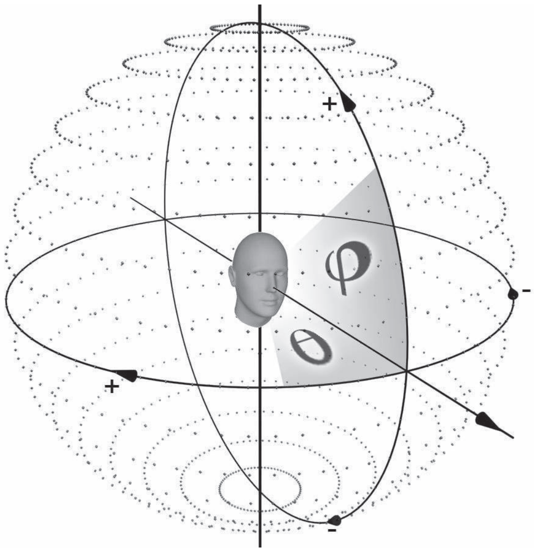



Figure 1: Vertical-polar coordinate system with azimuth ($\theta$) and elevation ($\varphi$) angles highlighted

## 3. TECHNIQUES OF LOW ORDER MODELING OF HRTFS

The most straightforward method to reduce the order of an FIR filter is the reduction of the number of its coefficients. This solution however causes oversimplification of the HRTF magnitude spectrum in the most critical frequency range. This is illustrated in Fig. 2.

The order of the filter needed to model the transfer function depends on how complex is its shape. Since human hearing has limited and non uniform frequency resolution, the notches and peaks present in HRTF magnitude spectrum may be modeled with different accuracy. Moreover, only the range between approx. 3 - 10kHz is crucial for spatial hearing, so HRTF magnitude parts outside this range may be approximated with less precision. Methods utilizing this fact base on smoothing of the magnitude spectrum of HRTF [5] or dividing the frequency to a number of subranges and mod-

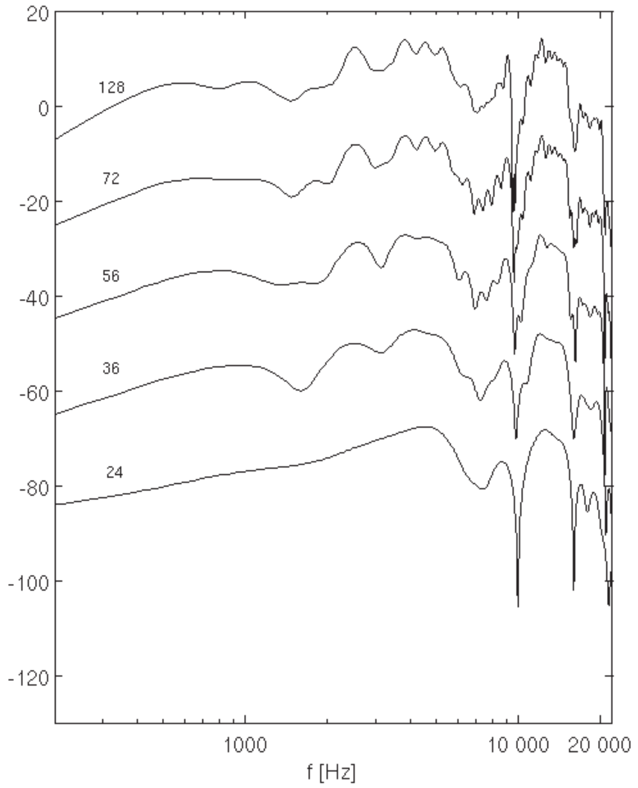



Figure 2: HRTF magnitude spectra for different impulse response lengths (24, 36, 56, 72, 128). For better readability consecutive plots are shifted by 20dB

eling each separately with different filters [6].

## 4. PREPROCESSING OF HRTF SPECTRUM

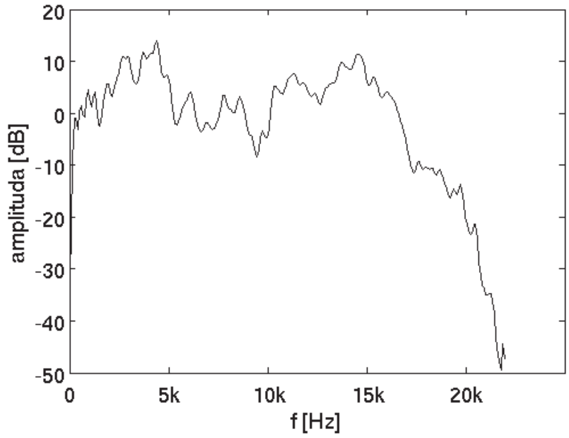
Since HRTF is a system of linear phase, its minimum phase decomposition is possible [2]. The all-pass component contains all information about the phase and the minimum-phase component models the magnitude response. Minimum phase decomposition is crucial since only the magnitude spectrum undergoes preprocessing.

### 4.1. Non-linear frequency mapping

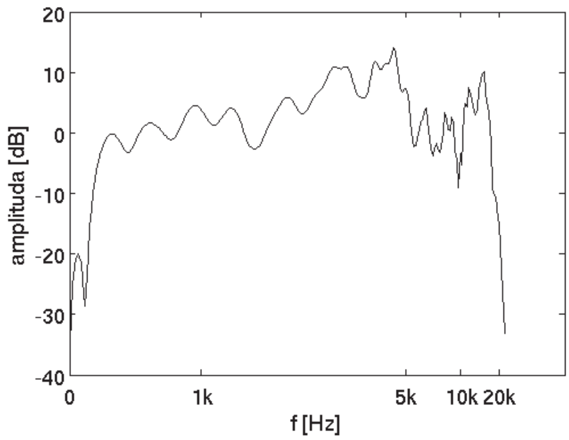
As mentioned in section 3, human hearing has a limited frequency resolution. Moreover, the resolution is not constant in the whole audible acoustic frequency range, but declines with increasing frequency. For this reason an assumption was made that modeling of components of HRTF that are imperceptible is unjustified. For further processing, the HRTF magnitude spectrum was converted into the Bark psychoacoustic scale [3]. Conversion from one frequency domain to another is done by means of bilinear conformal mapping [4]:

$$\zeta = \frac{z + \lambda}{1 + z\lambda} \tag{2}$$

where $\zeta$ is a frequency in the new frequency domain, $z$ is a frequency in the old domain. Parameter $\lambda$ is called the "all-pass coefficient". Using a sampling range $fs = 44100\frac{1}{s}$ the Bark transformation occurs when $\lambda = 0,7561$. The Bark transformation effects in "stretching" of the mid-low frequency range and "coarctation" of the high frequency range.




(a) linear frequency




(a) frequency transformed to Bark scale

Figure 3: Comparison of the HRTF magnitude spectra in linear and Bark scale frequency domains

Figure 3 shows an example comparison of the magnitude spectra in linear and bark scales. It is noticeable that the frequency range from 0 to 10kHz covers approximately 80% of the scale. All subsequent stages of the HRTF magnitude processing are performed in the Bark scale.

### 4.2. Wavelet based smoothing of magnitude spectrum

Although the conversion to Bark scale caused the frequency components whose share in spatial hearing is minor to have lower weights, the imperceptible peaks and notches are still present in the magnitude spectrum. A procedure of smoothing of the HRTF magnitude spectrum was proposed. An efficient method of smoothing of signals is wavelet filtering. In the proposed solution Stationary

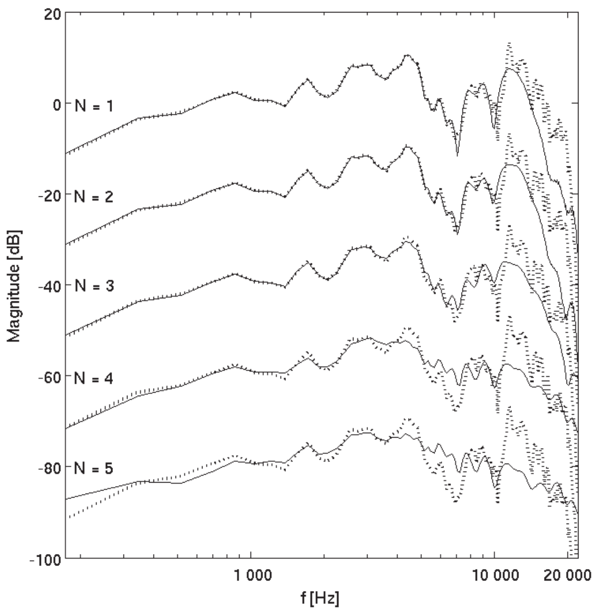



Figure 4: The HRTF magnitude spectra approximated using SWT coefficients for different levels of decomposition. The dotted line is the original magnitude spectrum. For better readability the plots are shifted by 20dB.

Wavelet Transform (SWT) [7] was used. SWT in contrast to Discrete Wavelet Transform does not contain a decimation step after each level of decomposition. For this reason, a smoothed version of HRTF magnitude spectrum is obtained using only the approximation coefficients. In the decomposition process a *Coiflet 1* filter bank was used. Fig. 4 shows the HRTF magnitude spectra reconstructed using approximation coefficients on different levels of decomposition.

It is noticeable that for level 1 - 3 approximation has a minor influence on the magnitude spectrum's shape in the frequency range up to 10kHz. Serious distortion is visible for higher frequencies. For further analysis HRTFs smoothed on 3rd level of approximation were used.

### 4.3. Dimensionality reduction of HRTF

Head Related Transfer Function is a frequency response of a complex acoustic system whose each component has a different directional characteristic. Modifications of the signal spectrum introduced by the ear pinna have highest influence on the sensitivity to change in sound source position for small changes in angles. On the other hand, frequency components interacting with the head and torso have a higher influence on the perception of large angle differences. A set of HRTFs for both ears is a multidimensional space of parameters of different importance in regards to the sensitivity to sound source position changes. Omitting the parameters with the smallest variations with changes in the angular position of a sound source can be used to further simplify the shape of the magnitude spectrum, which translates into a reduction in the order of the filter required to model it.

Dimensionality reduction was performed by means of Principal Component Analysis. Dimensionality reduction in PCA is obtained by transformation of a dataset into a new set of so called principal components which are not correlated to each other and ordered according to decreasing covariance. Only a few of principal components can cover almost all the variance in a data set [8]. When the original data set is reconstructed with a limited number of principal components, the ones that have the lowest influence on the variance are discarded. In case of HRTFs it results in the smoothing of the magnitude spectrum. It was assumed that the HRTF magnitude spectra for each sound source position are observations of a set of independent variables which are frequency components. HRTFs for each ear were processed separately. As it was mentioned particular components of the human acoustic system (e.g. the head, shoudlers, torso and ear pinna) have different influence on the variance of shape of the HRTF spectrum in different ranges of angular sound source positions. Changes introduced by the overall head shape when a sound source moves from one side of head to another are clearly noticeable, but they are negligible when the sound source moves by a small angle. On the other hand, influence of the ear pinna is most noticeable for small angles, but negligible for large angle changes. Influence of changes caused by head shape in full angular range might dominate and dwarf very important changes made by pinna shape in narrower ranges. For this reason the analysis of the whole HRTF set at one time is not a reasonable approach. Instead, a decision was made to divide the HRTF set into small subsets according to the azimuth angle. PCA decomposition and reconstruction is then performed for every subset separately. Since in every sector different principal components may be used for reconstruction, there is a risk of discontinuities on the boundaries of adjacent sectors. To avoid this drawback, sectors overlap. Size of a sector in which data is reconstructed covers the azimuth range of $\Delta\theta = 60^{\circ}$ but azimuth range for analysis is $40^{\circ}$ wider. In reconstruction only 6 principal components were used, which ensures that over 90% of variance of original HRTF data set is preserved.

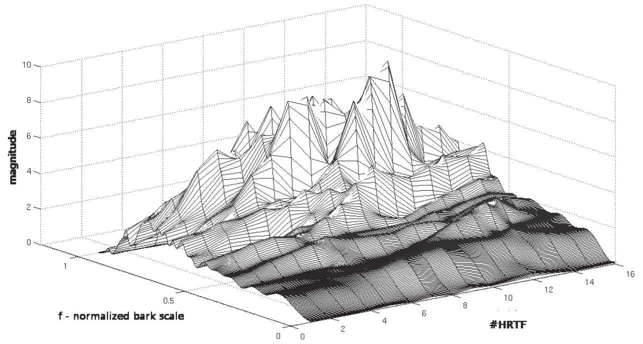
Figure 5 shows original and reconstructed magnitude spectra for one azimuth and sixteen elevation positions. The effect of smoothing is very clear not only in the frequency domain but also between corresponding coefficients in adjacent HRTFs.

## 5. DESIGNING OF DIGITAL FILTERS

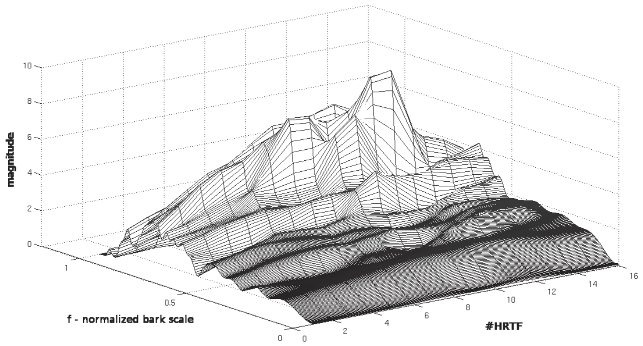
All operations in the preprocessing stage touched only the magnitude spectrum, so prior to the filter designing a complex form of the HRTFs must be obtained by restoring the phase information. After that, impulse responses of reduced length can be calculated by means of minimization of the mean square error between the original and simplified spectra.

Since the HRTFs were converted to Bark scale, the resulting impulse responses cannot be directly used as coefficients of an FIR filter. Special digital fitler structures which perform filtering using coefficients derived in non-linear frequency domain were proposed in [9]. These so called warped filters (WFIR and WIIR) are proven to be very effective in acoustic systems.

Coefficients of recursive filters were calculated by means of Prony's method basing on impulse responses of 128 coefficients. The designed filters have an equal number of zeroes and poles. Similar as in the case of non-recursive filters, the obtained coefficients cannot be used in typical a IIR filter structure. To obtain coefficients of digital filters that may be realized in normal FIR and IIR filter structures, the calculated warped impulse responses were transformed back to a linear frequency scale by the use of bilinear




(a)




(b)

Figure 5: HRTF magnitude spectra for one azimuth angle and sixteen elevation angles - (a) - original data, (b) - reconstructed with four principal components

transform:

$$z = \frac{\zeta + \lambda'}{1 + \zeta\lambda'} \tag{3}$$

where $\lambda' = -\lambda$. Coefficients of the IIR filter were obtained as previously using Prony's method.

## 6. EVALUATION OF THE PROPOSED METHODS

The influence of the preprocessing methods and filter order reduction on the quality of modeling of HRTFs was evaluated by means of computational tests and listening trials.

### 6.1. Computational evaluation of the HRTF modeling methods

An objective method to verify the newly designed low order filter was to compare its output with a reference hi-order filter. A short pink noise burst of 1024 samples was used as the test signal. The noise was bandpass filtered in range 1.5-10kHz to highlight the frequencies most significant for spatial hearing. As a reference filter a 256th order FIR filter was used with coefficients taken directly from the measured HRIR. The test signal was filtered through the reference filter and all the proposed simplified filters and the responses were converted to the Bark scale. The responses of the tested filters were compared to the response of the reference filter according to the following measures:

- Pearson's correlation coefficient
- RMS error

$$RMSE(X_1, X_2) = \frac{\sqrt{\sum_{i=1}^{N}(x_{1i} - x_{2i})^2}}{N} \tag{4}$$

where $x_1$ and $x_2$ are the elements of sets $X_1$ and $X_2$ and $N$ is the element count.

- Chebyshev norm error

$$\|E\|_\infty = \|X_1 - X_2\|_\infty = max|X_1 - X_2| \tag{5}$$

The following filters were analyzed:

- FIR of number of coefficients: 96, 72, 48, 24
- WFIR of number of coefficients: 96, 72, 48, 24
- WIIR of orders: 36, 24, 16, 8

The coefficients of the filters were calculated after the following preprocessing operations:

- SWT smoothing on 3rd level of decomposition
- dimensionality reduction by means of PCA using 6 principal components

A normal FIR filter without any preprocessing whose coefficients were taken directly from the HRIR and windowed to lengths: 96, 72, 48, 24 was also analyzed.

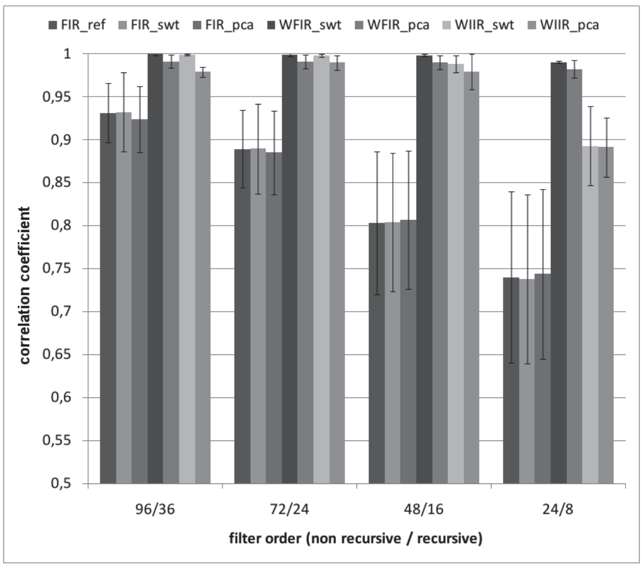



Figure 6: Averaged values of the correlation coefficient for all tested filters. Value of standard deviation is marked on every bar.

In figures 6, 7, 8 the respective values of the correlation coefficient, RMS error and Chebyshev error are shown for all tested filters. The presented values were averaged for 10 HRTF datasets. Subscripts in the symbols in the pictures have the following meaning:

- *ref* - reference filter with a limited number of coefficients (w/o any preprocessing)
- *swt* - filter was designed basing on HRTF magnitude smoothed with an SWT transform

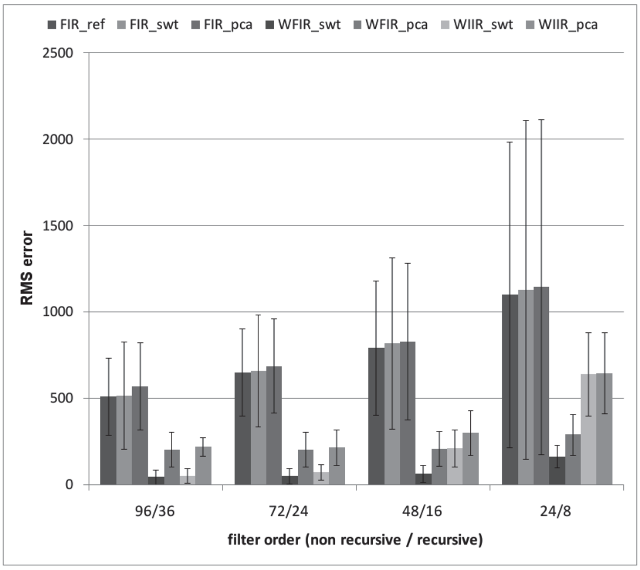



Figure 7: Averaged values of the rms error for all tested filters. Value of standard deviation is marked on every bar.

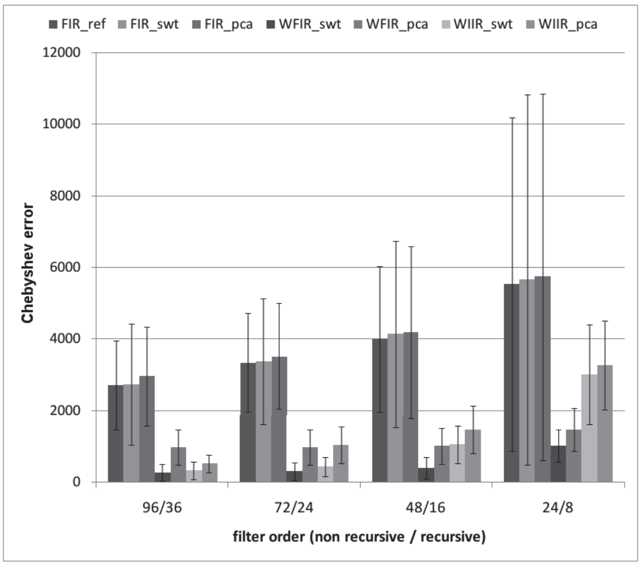



Figure 8: Averaged values of Chebyshev error for all tested filters. Value of standard deviation is marked on every bar.

- *pca* - filter was designed basing on HRTFs after dimensionality reduction

Values of the correlation coefficient indicate excellent fit of WFIR and WIIR characteristics. It is noticeable that the WFIR filter of 24 coefficients gives a result closer to the reference than an FIR filter of 96 coefficients obtained by windowing the original HRIR. Values of the correlation coefficient (much higher) and values of errors (much lower) show much better performance of warped filters against their linear counterparts. Moreover the spread of the results (standard deviation shown on plots) is much lower for warped filters. There is also one more observation: an 8th order of WIIR filter seems to be insufficient for modeling HRTFs. The comparison of the results for both preprocessing methods shows that their effect is very similar. Slightly better results were obtained when wavelet smoothing was used.

### 6.2. Listening trials

The aim of the listening tests was to determine the influence of the proposed methods of low order modeling of HRTF on the perception of spatial sounds. Ten volunteers took a part in this experiment. Three test signals were used:

- one-second sequence of pink noise band filtered in range 50Hz÷10kHz
- one-second sequence of a synthetized organ sound
- two-second sample of female voice

The following set of filters was used during the listening trials (symbols used in the plots are placed in the brackets) :

- WFIR, 24 coeffs after SWT smoothing (WFIR-24, sm3)
- WFIR, 24 coeffs after dimensionality reduction (WFIR-24, pca6)
- WIIR, 16th ofder after SWT smoothing (WIIR-16, sm3)
- WIIR, 8th order after dimensionality reduction (WIIR-8, pca6)
- FIR, 72 coeffs without any preprocessing (FIR-72, ref)

HRTFs used for filter design were personalized, i.e. measured individually for every participant. Virtual sound source presented to the listener could be located in one of 5 fixed spatial positions:

- $\theta = 30°$, $\varphi = 45°$
- $\theta = -60°$, $\varphi = -27°$
- $\theta = 70°$, $\varphi = -36°$
- $\theta = -80°$, $\varphi = 36°$
- $\theta = 0°$, $\varphi = 18°$

During the test, for each position and filter type two sounds were presented to the listener:

- test sound filtered by reference filter (FIR filter of 256 coefficients)
- test sound filtered by low order filter

The second sound was played after a half-second pause. The experiment was conducted in a double-blind way, i.e. neither the listener nor the researcher knew which filter was used at a given time and what was the sound source's position.

Each participant was asked to tell the similarity of the presented sounds by means of two measures:

- similarity of sound timbre
- difference in perceived virtual sound source placement

as numbers in range 1 to 4. The highest note means, that the sounds were indistinguishable (the same localization, the same timbre). The lower the assigned note, the higher perceived difference (sound perceived in different position, different timbre). Both notes were fully subjective.

In figures 9, 10 and 11 averaged results of the trials with noise, synthesized sound and human voice are shown. The perceived

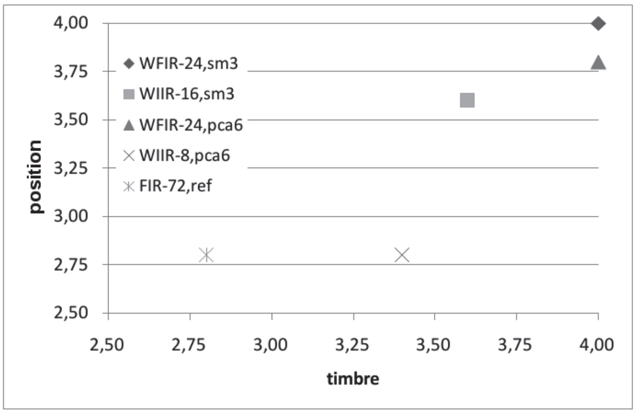



Figure 9: Averaged values of listeners notes for noise sound

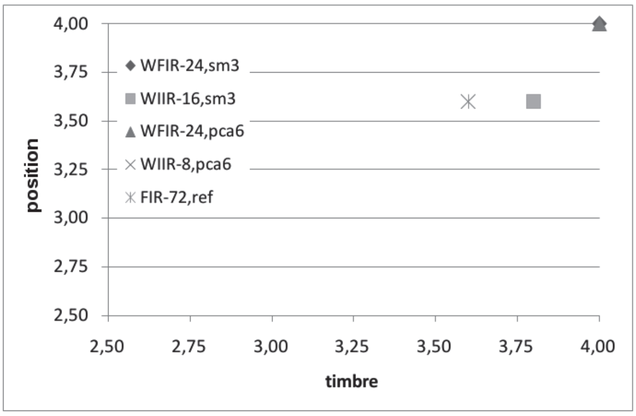



Figure 10: Averaged values of listeners notes for synth sound

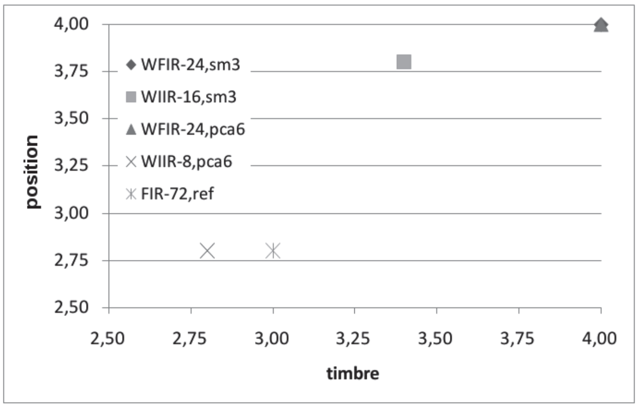



Figure 11: Averaged values of listeners notes for voice sound

quality of position was plotted against the perceived quality of timbre. The presented results are in line with the results of the objective tests. The WFIR filters of 24th order outperform all other ones in both domains: position and timbre. The experiment also confirmed, that an 8th order of WIIR filter is insufficient for modeling of HRTFs.

## 7. CONCLUSIONS

In the paper we proposed a number of low-order modeling methods for head related transfer functions (HRTFs). The methods are based on two approximation approaches that allow to smooth the HRTF spectra, namely: a wavelet approximation using the Stationary Wavelet Transform (SWT) and dimensionality reduction by means of the Principal Components Analysis (PCA). For the smoothed spectra the corresponding FIR and IIR filters were derived. These filters were defined both in a linear frequency scale and a nonlinear Bark scale. For the latter scale, special warped FIR and IIR filters (correspondingly WFIR and WIIR filters) were designed. The proposed filters of different orders that modeled the HRTFs were verified in both objective tests and in listening trials with 10 volunteers. It was shown that the proposed methods allow for a considerable reduction of filter orders, without sacrificing both the precision of perceived virtual sounds locations and the quality of the reproduced sounds. In particular it was concluded that the WFIR filter of the 24-th order can be used to safely substitute the FIR filter of order 72 modeling the HRTFs. These results are of primary importance for real time systems employing HRTF technology.

## 8. REFERENCES

[1] A. Dobrucki,P. Plaskota,P. Pruchnicki,M. Pec,M. Bujacz,P. Strumillo, “Measurement System for Personalized Head-Related Transfer Functions and Its Verification by Virtual Source Localization Trials with Visually Impaired and Sighted Individuals”, *J. Audio Eng. Soc*, vol. 58, no. 9, pp. 724-738 (2010 Sep)

[2] A. Kulkarni, S.K. Isabelle, H. S. Colburn, ”On the minimum-phase approximation of head-related transfer functions.”*IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 84–87, 1995.

[3] H. Fastl, E. Zwicker. ”*Psychoacoustics: facts and models*”, Springer, 2006.

[4] J.O. Smith, J.S. Abel.”Bark and erb bilinear transforms.” *IEEE Transactions on Speech and Audio Processing*, 7 pp. 697–708, 1999.

[5] J-M. Jot,V. Larcher,O. Warusfel,”Digital signal processing issues in the context of binaural and transaural stereophony.”*Proceedings of Convention of the Audio Engineering Society*, 1995.

[6] T. Matsumura,N. Iwanaga,W. Kobayashi,T. Onoye,I. Shirakawa, ”Embedded 3-D sound movement system based on feature extraction of head-related transfer function.” *IEEE Transactions on Consumer Electronics*, 51 pp.262–269, 2005.

[7] J.E. Fowler, ”The redundant discrete wavelet transform and additive noise.”*IEEE Signal Processing Letters*, 12 pp. 629 – 632, 2005.

[8] I.T. Jolliffe,*”Principal Component Analysis”*.Springer, 2002.

[9] M. Karjalainen,A. Harma,U.K. Laine.“Realizable warped IIR filters and their properties.”*Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1997.