Evaluating vertical localization performance of 3D sound rendering models with a perceptual metric

Michele Geronazzo*, Andrea Carraro†, Federico Avanzini‡

Department of Information Engineering
University of Padova

ABSTRACT

The *head-related transfer functions* (HRTFs) describe individual acoustic transformation that sound sources undergo due to human anatomy before arriving at the left and right tympanic membranes. The resulting spectral modifications are the main localization cues for elevation detection in space. In this paper, synthetic HRTF models able to render the vertical spatial dimension in virtual auditory displays, are evaluated via auditory models. Perceptually-motivated metrics describe the output of 4 virtual experiments that numerically simulate real listening experiments for 20 virtual subjects. The current implementation considers a limited set of parameters for a structural model of the pinna acting as a proof-of-concept of such approach. Accordingly, results confirm that the research framework is a flexible tool for systematic evaluation of different instances of structural model.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities; H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Signal analysis, synthesis, and processing; H.5.5 [Information Interfaces and Presentation]: Sound and Music Computing—Modeling;

1 INTRODUCTION

Sound localization has relevant implications in several everyday tasks and activities. The auditory modality continuously captures the acoustic scene, providing a listener with spatial information carried by temporal and spectral transformations of sound sources caused by both the environment and by the physicality of the listener himself. Knowledge of such a complex process is needed in order to develop accurate and realistic artificial sound spatialization in immersive virtual reality scenarios, including sensory substitution devices (e.g. for visual disabilities), in tele-operation remote system (e.g. robotic explorer), or entertainment and social applications (e.g. video-games). Moreover, emerging technologies in hear-through headsets and high resolution head-mounted displays call for the integration of binaural spatial audio in new portable applications. In particular, spatial audio technologies (see [7] for recent trends) through headphones usually involve *binaural room impulse responses* (BRIRs) to render a sound source in space. BRIR can be split in two separate components: *room impulse response* (RIR), which defines room acoustic properties, and *head-related impulse response* (HRIR), which acoustically describes individual contributions of listener's head, pinna, torso and shoulders.

According to the *spatial audio quality inventory* (SAQI) [17], localization accuracy is a relevant auditory quality in *Virtual Auditory Displays* (VADs) and it is usually quantified via psychoacustic experiments with human subjects. This paper deals with elevation localization cues, i.e. perception of sound source position in the vertical spatial dimension, which is mainly provided by monaural spectral features of the *head-related Transfer Function* (HRTF, the Laplace transform of HRIR). This information complements binaural cues such as *interaural time difference* (ITD) and *interaural level difference* (ILD) for localization in azimuth, i.e. the horizontal dimension.

HRTFs are usually measured over a discrete set of spatial locations with discrete frequency samples in anechoic chambers with special and expensive equipments. Finding a continuous functional HRTF representation (e.g., a parametric filter structure) allows dynamic rendering of arbitrary positions of sound sources in space and an accurate interpolation [15]. Such HRTF models need to be as similar as possible to the original measured responses in terms of auditory attributes. This similarity is typically assessed through time-consuming evaluation processes with human subjects. However, a complementary approach can be used, which consists in evaluating HRTF models with computational auditory models able to simulate the human auditory system. Model parameters can be tuned in order to describe virtual subjects and their individual characteristics, such as level of hearing impairments [9] or sensitivity to spectral shape [3]. If the auditory model is well calibrated to the reality, a perceptual metric can be developed to predict the perceptual performance of a VAD.

This paper intends to illustrate this approach by applying it to a specific case which is used here as a proof-of-concept. A filter model of the external ear previously presented in [13] is here evaluated by means of an auditory model for sound localization in the mid-sagittal plane [4] (i.e., the vertical plane dividing the listener's head in left and right halves) provided by the Auditory Modeling Toolbox[1]. For twenty virtual subjects in the CIPIC database [2], four virtual experiments are simulated: subjects listening with (i) their own ears, (ii) synthetic ears, and (iii)/(iv) ears of a mannequin.

The main contributions of this paper are twofold. First, it demonstrates that the proposed approach to systematic evaluation of HRTF representations through auditory models allows rapid prototyping of individual synthetic HRTFs for several listening conditions in VADs. Second, it shows that a perceptual metric for elevation perception can be applied to the estimation of filter structure and parameters in HRTF models (particularly, models of the pinna).

.

2 SPECTRAL FEATURES FOR VERTICAL LOCALIZATION

This work focuses on perception of sound source localization in the mid-sagittal plane though the analysis of spectral features in the *pinna-related transfer function* (PRTF) that describes the acoustic properties of the external ear, i.e. pinna, before arriving at the tympanic membrane.

The process is essentially monoaural, as binaural cues (ITD, ILD) undergo almost null variations with varying elevations due to irrelevant variation in ITD at both ears. For broadband sources,

*e-mail: geronazzo@dei.unipd.it
†e-mail: carraro3@dei.unipd.it
‡e-mail:avanzini@dei.unipd.it

2015 IEEE 2nd VR Workshop on
Sonic Interactions for Virtual Environments (SIVE)
Arles, France, 24 March 2015
978-1-4799-1969-7/15/$31.00 ©2015 IEEE

[1] http://amtoolbox.sourceforge.net/

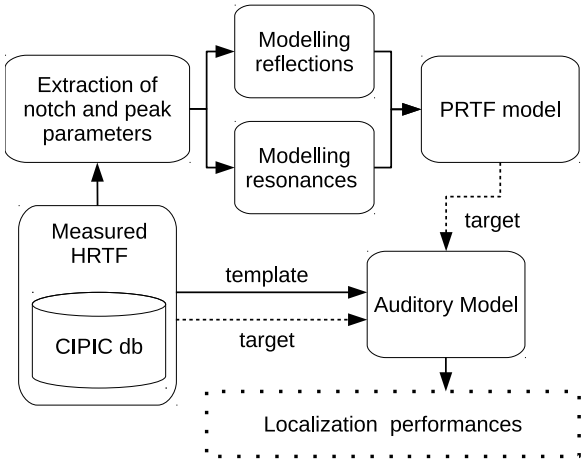

Figure 1: Schematic view of the proposed virtual experiments.

one can consider the contribution of torso reflections and acoustic shadow below 3 kHz in the PRTF spectrum, even if the perceptual relevance of such elevation cues is dominated by high-frequency content [1]. On the other hand, acoustic waves scattering in the proximity of the pinna creates a complex and individual topography of pressure nodes which is not clearly understood [23], thus resulting in elevation- and listener-dependent peaks and notches that characterize the PRTF amplitude spectrum from 3 to 16 kHz.

2.1 The pinna structural model

In the structural modeling approach, the acoustic contribution of head, torso and pinna are considered in separate filter realizations [8]. Among several pinna structural models (see [10] for an extensive review) in both time and frequency domain, the *"resonances-plus-reflections"* paradigm adopted in [12] simulates the physical phenomena underlying peaks (resonances) and notches (reflections) with a filter cascade of these two contributions.

With reference to filter realization in [13], source elevation ϕ as independent parameter, drives the evolution of resonances' center frequency $F_p^i(\phi)$, 3dB bandwidth $B_p^i(\phi)$, and gain $G_p^i(\phi)$, $i = 1,2$, and of the corresponding notch parameters ($F_n^j(\phi)$, $B_n^j(\phi)$, $G_n^j(\phi)$, $j = 1,2,3$).[2]

Resonances are represented as a parallel of two different second-order peak filters. The first peak ($i = 1$) has the form [24]

$$H_{\text{res}}^{(1)}(z) = \frac{1+(1+k)\frac{H_0}{2}+l(1-k)z^{-1}+(-k-(1+k)\frac{H_0}{2})z^{-2}}{1+l(1-k)z^{-1}-kz^{-2}}, \quad (1)$$

where

$$k = \frac{\tan\left(\pi\frac{B_p^1(\phi)}{f_s}\right)-1}{\tan\left(\pi\frac{B_p^1(\phi)}{f_s}\right)+1}, \qquad l = -\cos\left(2\pi\frac{F_p^1(\phi)}{f_s}\right), \quad (2)$$

$$V_0 = 10^{\frac{G_p^1(\phi)}{20}}, \qquad H_0 = V_0 - 1, \quad (3)$$

and f_s is the sampling frequency. The second peak ($i = 2$) is implemented as in [21],

$$H_{\text{res}}^{(2)}(z) = \frac{V_0(1-h)(1-z^{-2})}{1+2lhz^{-1}+(2h-1)z^{-2}}, \quad (4)$$

[2]For a given mid-sagittal PRTF, only available ϕ values allow the extraction of the filter parameters from the resonant or reflective component.

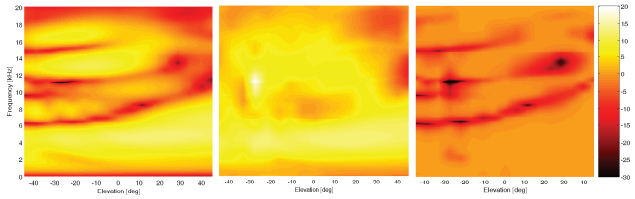

Figure 2: Structural decomposition algorithm on CIPIC 165. Full PRTF (left), resonant component (center), and reflective component (right).

$$h = \frac{1}{1+\tan\left(\pi\frac{B_p^2(\phi)}{f_s}\right)}, \quad (5)$$

while l and V_0 are defined as in Eqs. (2) and (3) with polynomial index $i = 2$. The former implementation has unitary gain at low frequencies, while the latter has a negative dB magnitude in the same frequency range. The parallel structure of these two filters ensures a neutral low-frequency contribution that can be assigned to a head model.

Reflections are represented as the cascade of three notch filters, i.e. $H_{\text{refl}}^{(j)}$ with $j = 1,2,3$, with a parametric equalizer implementation through Butterworth design method, ensuring transfer function and filter order compatibilities with peak realization, and a more versatile bandwidth specification with equal filter order.

3 The virtual experiments

The proposed evaluation methodology simulates virtual experiments where subjects are asked to give an absolute localization judgment about an auditory stimulus (see Fig. 1 for a schematic view). The adopted auditory model was introduced by Baumgartner et al. [4] following a *"template-based"* paradigm [16] that implements a comparison between the internal representation of an incoming sound at the eardrum and a reference template. Spectral features of sound events filtered with different HRTFs correlate with the direction of arrival, leading to a spectro-to-spatial mapping and a perceptual metric for elevation performances.

3.1 Subjects

Twenty virtual subjects took part in the virtual experiments. Two of them are KEMAR[3] mannequins with small (subject 021) and large pinnae (subject 165), respectively. For each virtual subject, individual HRTF measurements are available in the CIPIC database [2] and have been studied in [22]: 2500 measured HRIRs at $f_s = 44.1$ kHz (200 samples), 25 azimuths $\times$ 50 elevations $\times$ 2 ears. The interaural-polar coordinate system defines the spatial grid. Vertical dimension, i.e. elevation ϕ, is uniformly sampled on the range $-45°$ to $+230.625°$ with a $5.625°$ step.

3.2 Materials

3.2.1 Parameters of the PRTF model

An analysis-by-synthesis approach is used to estimate the parameters of peaks and notches in individual responses. In particular, a *structural decomposition algorithm* [12] is employed, in which the mid-sagittal PRTF magnitude spectrum of each subject is iteratively compensated, through a sequence of synthetic multi-notch filters until no local notches larger than -5 dB are left. When convergence is reached, the remaining residual describes the resonant component, while the combination of all the estimated multi-notch

[3]Knowles Electronic Manikin for Acoustic Research, one of the most commonly used mannequins for non-individual HRTF measures.

filters characterizes the reflective component. Figure 2 depicts an example of outcome from the decomposition phase. Filter parameters for peaks and notches with $F_p^i(\phi)$ and $F_n^i(\phi)$ in the range 4-15 kHz are extracted for ϕ from -45° to 45°, 17 available elevations. For $\phi > 45^\circ$ notch tracks are not deep enough to introduce perceptually relevant localization cues [6] .

For most of the subjects, the algorithm extracts three main notch tracks for a total of 3 triplets of notch parameters, i.e. $F_n^i(\phi)$, $G_n^i(\phi)$, and $B_n^i(\phi)$ (3 dB bandwidth relative to notch depth). All parameters are directly extracted from reflective responses and then assigned to a filter specification object of Matlab *DSP System Toolbox* [4] which models notch contribution. Only notches with gain lower than -15 dB, interpret $B_n^i(\phi)$ as bandwidth relative to -3-db gain (absolute notch amplitude). This ensures a good approximation of narrow notches with a 2^{nd} order filter without the extraction of different $B_n^i(\phi)$.

Finally, extraction of the parameters of peaks follows the procedure in [13], considering the mean magnitude spectrum of all resonant components in the med-sagittal plane belonging to all virtual subjects. For every available ϕ, the procedure extracts two maxima of the mean magnitude spectrum, which outputs the gain $G_p^i(\phi)$ and central frequency $F_p^i(\phi)$ of each resonance peak, $i = 1,2$, and the corresponding 3 dB bandwidth $B_p^i(\phi)$ in order to compute filter responses from Eqs. 1 and 4.

3.2.2 Elevation prediction with the auditory model

The adopted auditory model is based on two different processing phases before predicting the absolute elevation. During peripheral processing, an internal representation of the incoming sound is created. The *target* sound, which is the HRTF model in our virtual experiments, is converted into a *directional transfer function* (DTF) [16] and processed. In the second phase, the new representation is compared with a *template*, i.e. individual DTFs computed from individual HRTFs, thus simulating the localization process of our brain.

In practice, the target and template DTF sets are filtered with gammatone filterbank in order to simulate the auditory processing of the inner ear. In support of the comparison phase, the algorithm defines:

- a common set of target and template elevation angles;
- a set of channels corresponding to ERB frequency bands.

Given a specific target/template angle and channel, the algorithm computes the gain at the central frequency of each band and target/template internal representation. The *inter-spectral difference* (ISD) for each band is extracted from the differences in dB between the two signals; then, the *spectral standard deviation* (STD) is computed from the ISD values of each band.

For each target angle, the probability that the virtual listener points to a specific angle defines the *similarity index* (SI). The index value results from the distance (in degrees) between the target angle and the response angle which is the argument of a Gaussian distribution with zero-mean and standard deviation called *uncertainty*, U. The lower U, the higher is assumed the sensitivity of the virtual listener in discriminating different spectral profiles resulting in a measure of probability and not a deterministic value.[5]

Accordingly, the similarity indexes are bounded from 0 to 1, and the sum of the SI for a certain target angle is equal to 1. Simulation data are stored in probability mass vector, where each target angle has the probability that the virtual listener points at each available local angle.

[4] www.mathworks.com/products/dsp-system/

[5] This is in agreement with reality, where spatial perception is influenced by psychological an temporary factors [3].

Table 1: ΔPEs of the simulations w.r.t. individual precision errors.

	PE	ΔPE		
Subject	individual	struct. model	CIPIC 021	CIPIC 165
CIPIC 003	17.7921	+8.0534	+5.6201	+6.0908
CIPIC 008	20.6832	+15.6319	+6.4706	+11.2464
CIPIC 009	23.4100	+6.2619	+4.9159	+4.4283
CIPIC 010	19.8699	+12.9163	+7.4151	+4.9875
CIPIC 011	22.7646	+4.0959	+6.6252	+6.8473
CIPIC 012	16.7770	+11.7753	+12.1164	+7.7046
CIPIC 015	19.4375	+13.2956	+5.5895	+5.3325
CIPIC 017	19.2560	+8.6293	+5.5960	+4.8373
CIPIC 019	20.5713	+7.5697	+7.4433	+6.2953
CIPIC 020	22.5487	+6.6617	+7.4831	+2.4943
CIPIC 027	19.5813	+10.5434	+13.4948	+10.2176
CIPIC 028	16.9980	+11.9318	+8.2182	+6.3205
CIPIC 033	19.5783	+9.3286	+11.6488	+11.8247
CIPIC 040	18.8795	+13.8228	+7.0053	+4.4024
CIPIC 044	18.4665	+11.1379	+10.0326	+12.1177
CIPIC 048	23.3474	+7.5680	+2.6215	+2.7329
CIPIC 050	19.4822	+9.5782	+13.7177	+16.9702
CIPIC 134	26.1789	+3.2941	+5.3976	+4.4625

3.3 Perceptual metric

The proposed auditory model simulations use $U = 2$, a value that reasonably approximates the uncertainty of a real listener in determining the position of sound source in space [18]. Additionally, assigning the same U value to every virtual subject ensures a uniform set of listeners which is impracticable in real experiments. The following virtual experiments are performed:

1. template and target equal to **individual HRTFs** of the same subject. This experiment is the *ground truth* condition of the virtual listening experience, as it represents the condition in which a virtual subject is listening with his own ears, and accordingly it exhibits localization performances that are comparable with those reported in the literature for real scenarios;
2. template: individual HRTFs, target: **pinna structural model** defined in Sec. 2.1, whose parameters have been fitted to individual HRTFs through the parameter estimation procedure described in Sec. 3.2.1.
3. template: individual HRTFs, target: HRTFs of CIPIC subject 021 - **KEMAR with large pinnae**. This simulation is a control condition that resembles applicative scenarios where only generic HRTFs are available for sound spatialization;
4. template: individual HRTFs, target: HRTFs of CIPIC subject 165 - **KEMAR with small pinnae**. This simulation is a second control condition (similar to the previous one) that resembles applicative scenarios where only generic HRTFs are available for sound spatialization;

The three latter experiments represent conditions in which a virtual subject is listening with either the individual pinna structural model, the large pinnae KEMAR, and the small pinnae KEMAR.

The perceptual metric is constructed by combining the outputs of all the virtual experiments. In particular, the precision for every j-th elevation response close to the target position is defined in the *polar error* (PE) [18]:

$$PE_j = \sqrt{\frac{\sum_{i \in A} (\phi_i - \varphi_j)^2 p_j[\phi_i]}{\sum_{i \in A} p_j[\phi_i]}}$$

where $A = \left\{ i \in N : 1 \leq i \leq N_\phi, \left| \phi_i - \varphi_j \right| mod 180^\circ < 90^\circ \right\}$ defines local polar-angle responses within $\pm 90^\circ$ w.r.t. the local response

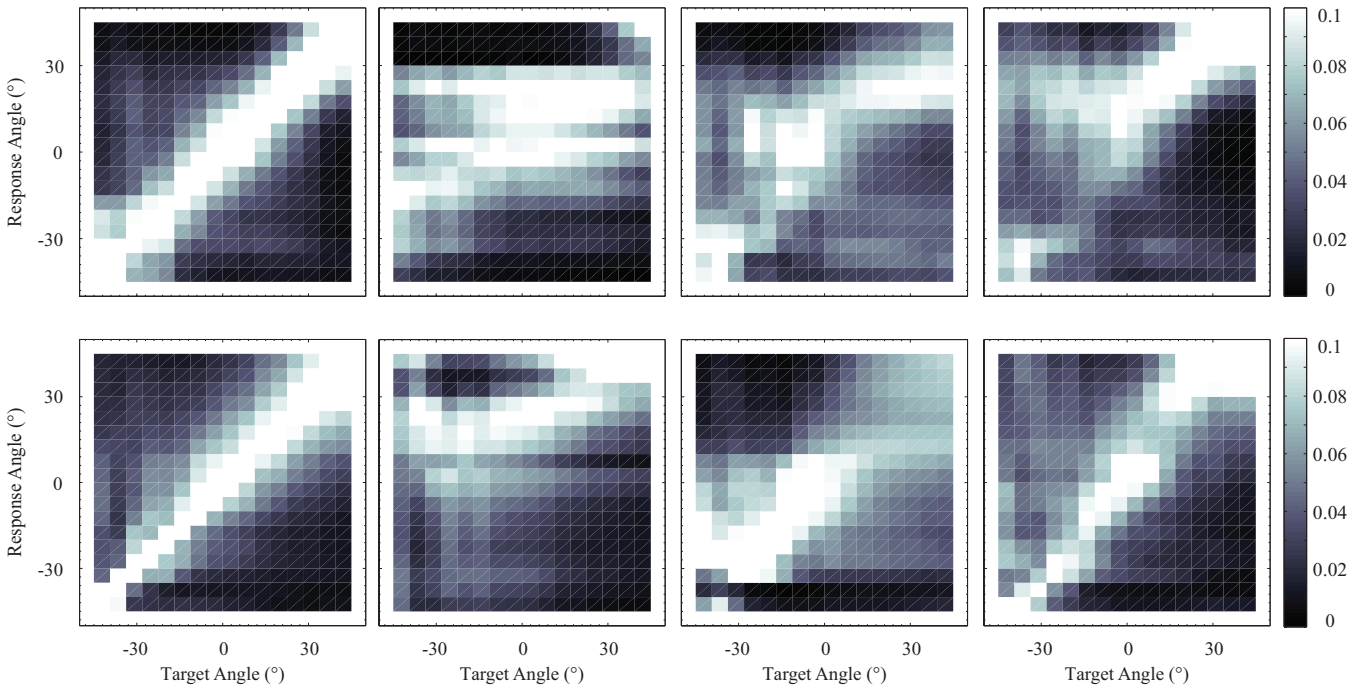

Figure 3: Examples of localization prediction. Response predictions for two subjects, CIPIC 011 (upper row) and CIPIC 048 (lower row), while listening to sound sources located at target elevation angles in the mid-sagittal plane, for the simulations 1. to 4. (from left to right). Probabilistic response predictions are encoded by brightness according to the color bar to the right.

ϕ_i and the target position φ_j, and $p_j[\phi_i]$ denotes the probability mass vector. Accordingly, localization performances for a single virtual experiment are quantified by the mean *PE* across elevation responses.

4 Results & Discussion

Table 1 shows simulation results for 18 CIPIC subjects in the four experiments. Individual precision errors, i.e. mean *PEs* for simulation #1, are taken as reference values for the remaining simulations, resulting in ΔPEs calculation for structural model, CIPIC 021 and CIPIC 165 simulations (see the last three columns of Table 1). Global statistics on ΔPEs reveals better performances for simulations with KEMARs (CIPIC 021, mean: 7.86, SD:3.12; CIPIC 165, mean: 7.18, SD:3.83) than the simulation with pinna structural models (mean: 9.56, SD:3.38). Moreover, 7 out of 18 subjects exhibit better *PEs* with pinna structural models.

The global evaluation phase reveals that the current set of model parameters and filters is not suitable for being tested with real subjects because it does not provide global improvements in localization performances with respect to listening with generic HRTFs such as dummy heads. It has to be noticed that if one does not make use of perceptual metrics, the number of model instances to be tested with real subjects require time- and resource- consuming procedures. This often implies listening tests with experts, which are unrepresentative of the entire population of listeners.

Individual localization performances reveal strengths and weaknesses in the model. Figure 3 shows probabilistic response predictions for two exemplary subjects, CIPIC 011 and CIPIC 048: the first is more accurate with structural models while the latter performs better with KEMAR HRTFs.

Focusing on CIPIC 011, the structural model clearly resolves up-down confusions (dark areas in the top-left and bottom-right of the 1^{st}-row, 2^{nd}-column plot) while KEMAR simulations produce high uncertainty, especially for lower elevations. On the other hand, elevation perception with the model is more compressed between $-30° < \phi < 30°$. The pinna structural model captures spectral differences between lower, central and upper target angles, without well differentiating adjacent angles.

As shown by the 2^{nd}- and 4^{th}-column plots in the second row, subject CIPIC 048 exhibits patterns between response-target angles which are similar to those of CIPIC 165 , i.e. KEMAR with small pinnae ($\Delta PE < 3°$). This suggest that spectral features of the two subjects are similar. The simulation with structural models shows a marked *elevation bias* towards the upper hemisphere, resulting in an impracticable audio rendering solution. This effect happens when spectral features in a set of PRTFs (the pinna structural model) are located systematically higher in frequency than individual PRTFs [20]. As a matter of fact, synthetic notches in the mid frequency range do not describe properly individual elevation changes.

From the proposed evaluation phase, several improvements can be discussed in order to solve the above criticalities:

- capturing spectral peculiarities: wrong spectral references like different peak contributions and the suppression of weak notches lead to an equal distributed probability among responses;
- avoiding elevation bias: weak spectral features like limited notch bandwidth lead to a suppression of spectral information due to auditory filter-bank selectivity;

Recall that the structural model uses average rather than individ-

ual parameter values for the resonant component (peak filters): individual values should be fed to the model in order to assess the impact of this average resonant component on performance degradation [13]. Moreover, using a different notch filter design, which exaggerates bandwidth and/or gain specification, would grant a clear identification of notches, to the detriment of resonance accuracy. Moreover, one can design higher order filters for notches, to the detriment of the computational costs of such model.

Once these issues will be handled, new virtual experiments will be performed following the proposed methodology and perceptual metric in order to report correlations among spectral features, structural filter parameters and localization predictions.

5 Conclusion & Future Works

In this paper, the proposed evaluation method employs auditory models for a systematic performance analysis of vertical localization in VADs, with particular attention to personalized HRTF rendering through headphones. The current implementation considers the structural modeling approach for the pinna as a proof-of-concept of such research methodology.

The limited set of parameters being evaluated in this work reveals several criticalities in the current filter complexity and parameters. Nevertheless, the adopted methodology has proved to be robust and flexible, and allows to switch among different instances of structural models and parameters. New simulations are left as future works for an extensive study on different configurations.

Thanks to the recent introduction of the spatially oriented format for acoustics (SOFA) [19], the increase in number of publicly available HRTFs sharing the same data format makes a large number of virtual subjects available for extensive simulations. This is in remarkable contrast with psychoacoustic experiments where very limited numbers (as low as four or five) of subjects are often employed. Furthermore, several virtual listeners can be modeled starting from the same HRTF set and employing more sophisticated auditory models [5]. Changing the uncertainty value U, simulating different parameters for frequency selectivity, or simulating sensorimotor uncertainty in a virtual pointing task, allows to prototype and test several features of structural models, e.g. with a single notch filter or with an artificially exaggerated notch depth.

Finally, the proposed methodology can be adopted to evaluate synthetic HRTFs constructed using a *mixed structural model* (MSM) approach [14, 10]. In this approach, individual anthropometric quantities guide structural models and nonindividual HRTF selection techniques in creating new synthetic HRTFs. In particular, the MSM defines a structural combination where each component can be a synthetic or measured transfer function. Finding the most effective MSM is even more crucial for an extensive usage of personal auditory displays in the diffusion of spatial audio contents on the web [11].

Acknowledgements

This work was supported by the research project Personal Auditory Displays for Virtual Acoustics, University of Padova, under grant no. CPDA135702.

References

[1] V. R. Algazi, C. Avendano, and R. O. Duda. Elevation localization and head-related transfer function analysis at low frequencies. *The Journal of the Acoustical Society of America*, 109(3):1110–1122, 2001.

[2] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano. The CIPIC HRTF database. In *Proc. IEEE Work. Appl. Signal Process., Audio, Acoust.*, page 1–4, New Paltz, New York, USA, Oct. 2001.

[3] G. Andéol, E. A. Macpherson, and A. T. Sabin. Sound localization in noise and sensitivity to spectral shape. *Hearing Research*, 304:20–27, Oct. 2013.

[4] R. Baumgartner, P. Majdak, and B. Laback. Assessment of sagittal-plane sound localization performance in spatial-audio applications. In J. Blauert, editor, *The Technology of Binaural Listening*, Modern Acoustics and Signal Processing, pages 93–119. Springer Berlin Heidelberg, Jan. 2013.

[5] R. Baumgartner, P. Majdak, and B. Laback. Modeling sound-source localization in sagittal planes for human listeners. *The Journal of the Acoustical Society of America*, 136(2):791–802, 2014.

[6] J. Blauert. *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT Press, Cambridge, MA, USA, 1983.

[7] J. Blauert. *The Technology of Binaural Listening*. Modern Acoustics and Signal Processing. Springer Berlin Heidelberg, 2013.

[8] C. P. Brown and R. O. Duda. A structural model for binaural sound synthesis. *IEEE Trans. Audio, Speech, Lang. Process.*, 6(5):476–488, 1998.

[9] M. Florentine, S. Buus, B. Scharf, and E. Zwicker. Frequency selectivity in normally-hearing and hearing-impaired observers. *J Speech Hear Res*, 23(3):646–669, Sept. 1980. PMID: 7421164.

[10] M. Geronazzo. *Mixed structural models for 3D audio in virtual environments*. Ph.D. thesis, University of Padova, Padova, Italy, Apr. 2014.

[11] M. Geronazzo, J. Kleimola, and P. Majdak. Personalization support for binaural headphone reproduction in web browsers. In *Proc. 1st Web Audio Conference*, Paris, France, Jan. 2015.

[12] M. Geronazzo, S. Spagnol, and F. Avanzini. Estimation and modeling of pinna-related transfer functions. In *Proc. of the 13th Int. Conference on Digital Audio Effects (DAFx-10)*, pages 431–438, Graz, Austria, Sept. 2010.

[13] M. Geronazzo, S. Spagnol, and F. Avanzini. A head-related transfer function model for real-time customized 3-d sound rendering. In *Proc. INTERPRET Work., SITIS 2011 Conf.*, pages 174–179, Dijon, France, Dec. 2011.

[14] M. Geronazzo, S. Spagnol, and F. Avanzini. Mixed structural modeling of head-related transfer functions for customized binaural audio delivery. In *Proc. 18th Int. Conf. Digital Signal Process. (DSP 2013)*, pages 1–8, Santorini, Greece, July 2013.

[15] N. A. Gumerov, A. E. O'Donovan, R. Duraiswami, and D. N. Zotkin. Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation. *The Journal of the Acoustical Society of America*, 127(1):370–386, 2010.

[16] E. H. A. Langendijk and A. W. Bronkhorst. Contribution of spectral cues to human sound localization. *The Journal of the Acoustical Society of America*, 112(4):1583–1596, 2002.

[17] A. Lindau, V. Erbes, S. Lepa, H.-J. Maempel, F. Brinkman, and S. Weinzierl. A spatial audio quality inventory (SAQI). *Acta Acustica united with Acustica*, 100(5):984–994, Sept. 2014.

[18] P. Majdak, R. Baumgartner, and B. Laback. Acoustic and non-acoustic factors in modeling listener-specific performance of sagittal-plane sound localization. *Frontiers in Psychology*, 5, Apr. 2014.

[19] P. Majdak, Y. Iwaya, T. Carpentier, R. Nicol, M. Parmentier, A. Roginska, Y. Suzuki, K. Watanabe, H. Wierstorf, and H. Ziegelwanger. Spatially oriented format for acoustics: A data exchange format representing head-related transfer functions. In *Audio Engineering Society Convention 134*. Audio Engineering Society, 2013.

[20] J. C. Middlebrooks. Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. *The Journal of the Acoustical Society of America*, 106(3):1493–1510, 1999.

[21] S. J. Orfanidis, editor. *Introduction To Signal Processing*. Prentice Hall, 1996.

[22] S. Spagnol, M. Geronazzo, and F. Avanzini. On the relation between pinna reflection patterns and head-related transfer function features. *IEEE Trans. Audio, Speech, Lang. Process.*, 21(3):508–519, Mar. 2013.

[23] H. Takemoto, P. Mokhtari, H. Kato, R. Nishimura, and K. Iida. Mechanism for generating peaks and notches of head-related transfer functions in the median plane. *The Journal of the Acoustical Society of America*, 132(6):3832–3841, 2012.

[24] U. Zölzer. *DAFX: digital audio effects, 2nd Edition*. Wiley, Chichester, West Sussex, U.K., 2 edition, 2011. edited by Udo Zölzer.