Audio Engineering Society

Convention Paper 8882

Presented at the 134th Convention
2013 May 4–7 Rome, Italy

A modular framework for the analysis and synthesis of Head-Related Transfer Functions

Michele Geronazzo[1], Simone Spagnol[1], and Federico Avanzini[1]

[1] *Dept. of Information Engineering, University of Padova, Via Gradenigo 6/B, 35131 Padova, Italy*

Correspondence should be addressed to Michele Geronazzo (`geronazzo@dei.unipd.it`)

ABSTRACT

The paper gives an overview of a number of tools for the analysis and synthesis of head-related transfer functions (HRTFs) that we have developed in the past four years at the Department of Information Engineering, University of Padova, Italy. The main objective of our study in this context is the progressive development of a collection of algorithms for the construction of a totally synthetic personal HRTF set replacing both cumbersome and tedious individual HRTF measurements and the exploitation of inaccurate non-individual HRTF sets. Our research methodology is highlighted, along with the multiple possibilities of present and future research offered by such tools.

1. INTRODUCTION

In recent years spatial sound has become increasingly important in a plethora of application domains. Spatial rendering of sound is especially recognized to greatly enhance the effectiveness of auditory human-computer interfaces [1], particularly in those cases where the visual interface is limited in extension and/or resolution (as in mobile applications [2]). Furthermore, it aids improving the sense of presence in augmented/virtual reality systems and adds engagement to computer games. Winking to these ever-growing application domains, headphone-based reproduction systems driven by head tracking devices – if properly designed – allow tailoring immersive and realistic auditory scenes to any user without the need of expensive and cumbersome loudspeaker-based systems.

This paper gives a brief overview of a number of tools for the analysis and synthesis of head-related transfer functions (HRTFs, i.e. the frequency- and location-dependent acoustic transfer functions between the sound source and the eardrum of a listener [3]) we have developed, highlighting in particular our research methodology along with the diverse possibilities of present and future research offered by the mentioned tools. The main objective of our study in this context is the progressive development of a collection of algorithms for the construction of a totally synthetic HRTF set suitable for real-time rendering of custom spatial audio, taking the listener's anthropometric parameters as the sole input to the audio chain. Our modeling philosophy is the child of the structural approach by Brown and Duda [4]: the total contribution of the listener's body to the HRTF is split into smaller blocks or modules, and each module contains a measured, reconstructed or synthetic response, as will be made clearer throughout the paper. Our approach differentiates from recent trends in HRTF customization because no self-tuning of parameters or selection of responses from databases will be required in principle by the listener.

After the short picture of past and current research on HRTF-related issues given in Section 2, Section 3 will introduce the currently adopted framework for HRTF rendering and customization, including all of the hardware and software tools at our disposal. Section 4 will then discuss in more detail the structural modus operandi, with a focus on the multiple current and future research directions made possible by the available tools (further expanded in the conclusive Section 5).

2. BACKGROUND

In order to enable authentic auditory experiences, the correct sound pressure level (SPL) due to one or more acoustic sources positioned in a virtual space shall be reproduced at the eardrums of a listener by the pair of headphones. In the literature, sound transmission from the headphone to the eardrum is often represented through an analogue circuit model [5], reported in Fig. 1. With reference to such model, the current aim of our research is the correct reproduction of the sound pressure P_6 at the entrance of the blocked ear canal of the listener due to a sound source placed around him/her, even though nothing prevents a future extension of these studies to circuit points nearer to the eardrum impedence as soon as proper tools are available (e.g., HRTF measurements at the eardrum, ear canal models, etc.).

The classical solution that best approximates ideality involves the use of individual HRTFs measured on the listener himself/herself. By convolving a desired monophonic sound signal with a pair of personal head-related impulse responses (HRIRs),

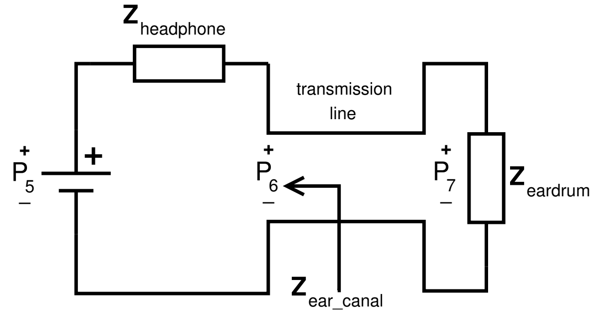


Fig. 1: Circuit model of sound transmission from the headphone to the eardrum (after [5]).

one per ear, adequately compensated for headphone-induced spectral coloration[1], one can reach almost the same localization accuracy as in free-field listening conditions [8], especially when head motion and/or artificial reverberation [9] are considered [10, 11].

Unfortunately, obtaining personal HRTF data for a vast number of users is simply unpracticable because specific hardware, anechoic spaces, and long collection times are strictly required [4]. This is the main reason why non-individual HRTFs, usually measured on anthropomorphic mannequins [12], are often preferred in practice. The drawback with non-individual HRTFs is that such peculiar transfer functions likely never match with the listener's unique anthropometry, and especially his/her outer ear [13], resulting in frequent localization errors such as front/back reversals [14], elevation angle misperception [15], and inside-the-head localization [16, 17].

Consequently, computational models of HRTFs parameterized on the anthropometry of the listener or tunable by the listener himself/herself have surfaced

---

[1] Headphones, when used for the reproduction of binaural signals, have to be equalized if high localization accuracy is needed. Unfortunately, the transfer function between headphone and eardrum heavily varies from person to person and with small displacements of the headphone itself. Such variation is particularly marked in the high-frequency range where important elevation cues generally lie. Thus, an inaccurate compensation likely leads to spectral colorations that affect both source elevation perception and sound externalization [6]. Although various techniques have been proposed in order to face such a delicate issue (see [7] for a review), modeling the correct equalization filter is still a hot open research theme.

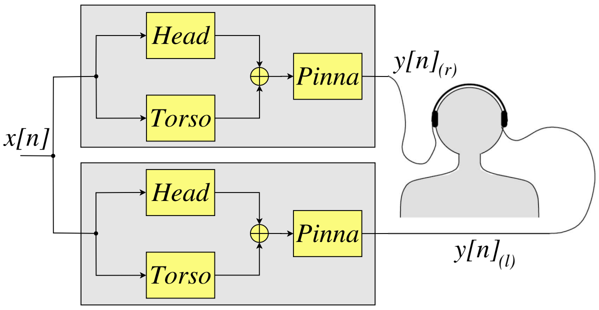


Fig. 2: A generic structural HRTF model.

in the past two decades. According to Brown and Duda [4], these models can be classified in three groups:

1. *pole/zero models*: filter design, system identification, and neural network techniques are applied in order to fit multiparameter models to experimental data (e.g. [18]);
2. *series expansions* based e.g. on principal component analysis (PCA) [19] or surface spherical harmonics (SSH) [20] applied to collections of HRIRs or HRTFs;
3. *structural models*: the contributions of the listener's head, pinnae, shoulders and torso to the HRTF are isolated and arranged in different filter structures each accounting for some well-defined physical phenomenon, as Fig. 2 roughly sketches. The linearity of these contributions allows reconstruction of the global HRTF from a proper combination of all the considered effects [21].

Although recent trends in HRTF customization mainly have focused on series expansions with self-tuning of weights [22, 23] or simply non-individualized HRTF selection [24, 25, 26], structural HRTF modeling remains the most attractive alternative from both the viewpoints of computational efficiency and physical meaning: parameters of the rendering blocks sketched in Fig. 2 can be estimated from real data, fitted to low-order filter structures, and finally related to meaningful anthropometric measurements. Our methodology follows the structural modeling approach, subtly generalized such as to allow for the inclusion of measured or extrapolated data in one or more of its components. For instance, one would desire to combine a filter model of the pinna with the measured HRTF of a generic pinnaless mannequin, or to feed a post-processed HRTF including the response of the pinna alone (*pinna-related transfer function*, PRTF [27]) into a filter snowman model [28]. We will refer to this approach as *mixed structural modeling* throughout the remainder of this paper.

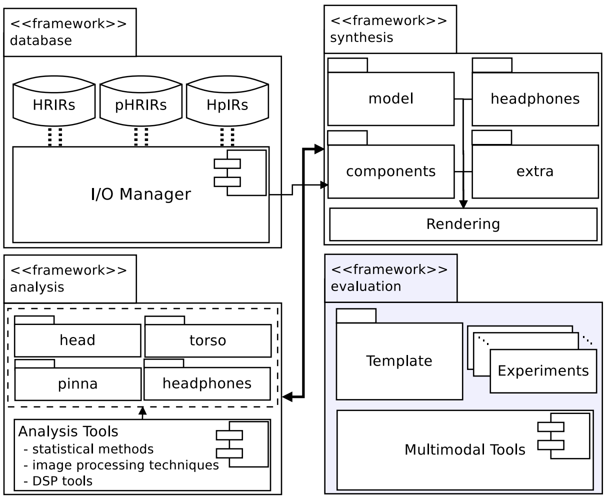


Fig. 3: UML-like representation of the research framework structure.

3. THE FRAMEWORK

The first logical distinction among the fundamental components of the presented framework regards the file system level. The *database* folder acts as the main data container, while all of the algorithms that extract relevant features in the available data are stored in the *analysis* folder. The *synthesis* folder contains the tools for spatial audio rendering designed and developed with an eye to real-time constraints. Finally, the experimental tools for subjective evaluation of models and the related data are organized in the *evaluation* folder. The UML-like diagram in Fig. 3 depicts all the components building the framework.

3.1. HRIR and HpIR Databases

The included data is under the form of several sets of head-related impulse responses (HRIRs) recorded for a high number of subjects in different spatial locations, and sets of headphone impulse responses

(HpIRs) for the characterization of different headphone models used for compensation in the reproduction process. Beside the full-body HRIR repository, a similar container includes the partial head-related impulse responses (pHRIRs) recorded by isolating specific body parts (e.g. pinna-related impulse responses (PRIRs) measured on the isolated pinna, or impulse responses measured on a pinna-less mannequin) or resulting from the decomposition carried on by the algorithms we will mention in the next subsection.

There exists a number of publicly available HRIR databases, the most notables of which are the CIPIC HRTF database [29] [2] and the LISTEN HRIR database [30] [3]. The main differences among these and other databases concern the type of stimulus used, the spatial grid of the measured HRIRs, and the microphone configuration (blocked- or open-ear-canal, distance from the eardrum, etc.).

An attempt to unify such variability resulted in the MARL-NYU data format [31], into which the CIPIC, LISTEN, FIU [32] [4] and KEMAR MIT [33] [5] databases were stored and organized. Our contribution to the standardization process begun in [31] lies in the introduction of some missing relevant data:

1. the raw HRIR data beside the already available compensated version;
2. the labeling of each HRIR's onset sample;
3. the management of heterogeneous spatial grids.

The general database organization was also applied to the Aalto [34] [6], ARI [7], and PKU&IOA [35] [8] HRTF databases, and to the Aalto PRTF database [36] [9] which collects pinna-related impulse responses in the mid-sagittal plane as pHRIRs. More details about the organization of the HRIR, pHRIR, and HpIR repositories can be found in [37].

3.2. Signal Analysis Tools

The *analysis* folder contains all the Matlab scripts and data structures exploitable for HRTF analysis. Our typical work flow follows an analysis-by-synthesis paradigm where the step-by-step modeling of salient features plays a significant role in the analysis of the acoustic signal. A notable instance of such paradigm is represented by the PRTF separation algorithm [38], which iteratively extrapolates the reflective component of a PRTF while keeping its resonant structure intact by direct subtraction of multi-notch filter structures. A similar algorithm, used in [39], separates the near- and far-field contributions of a rigid sphere approximating the head, allowing to model the two contributions independently through different filter structures.

Other kinds of algorithms represent the "boundaries" of such methodology. For instance, an image processing algorithm that extracts the relevant anthropometric parameters from a picture of the pinna [40] is currently being developed. A script for PCA modeling of HRTF data that helps understanding the degree of variability of the transfer functions with respect to specific features is available [41]. Last but not least, headphone equalization algorithms implementing various inverse filtering techniques are included.

3.3. Synthesized Audio Rendering

The audio engine, stored in the *synthesis* folder, includes four modules organized in separate subfolders:

- *model*: real-time realizations of the synthetic structural components;
- *components*: collection of tools that perform real-time convolutions between audio files and HRIRs/pHRIRs;
- *headphones*: management tool for headphone compensation filters;
- *extra*: utility bundle for I/O operations, sensors and basic binaural processing tools.

Various combinations of one or more instances for each module are possible in order to realize a candidate version of the 3D audio engine. All instances are implemented in Pure Data [10], a graphical pro-

---

[2] http://interface.cipic.ucdavis.edu/

[3] http://recherche.ircam.fr/equipes/salles/listen/

[4] http://dsp.eng.fiu.edu/HRTFDB/main.htm

[5] http://sound.media.mit.edu/resources/KEMAR.html

[6] http://www.acoustics.hut.fi/go/aes133-hrtf/

[7] http://www.kfs.oeaw.ac.at

[8] http://www.cis.pku.edu.cn/auditory/Staff/Dr.Qu.files/Qu-HRTF-Database.html

[9] http://www.dei.unipd.it/∼spagnols/PRTF_db.zip

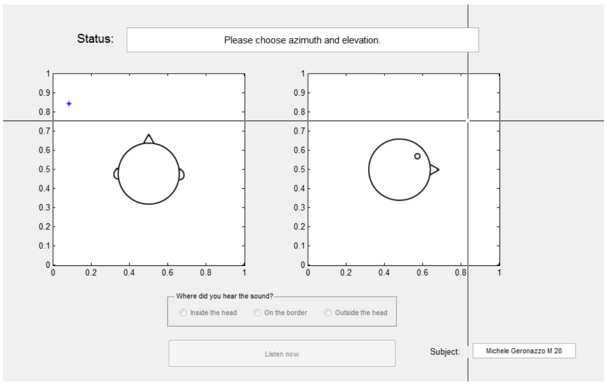
[10] http://puredata.info/

gramming environment for audio processing, in the form of C/C++ externals. All the tentative prototypes are catalogued in a further folder (the *rendering* folder), each accompanied by a descriptor file including the list of the modules used in that instance.

The intrinsic modularity of our approach leads us to the implementation of one structural filter block for each relevant body part:

- a pinna filter realization that acts as a synthetic PRTF, consisting of a peak and notch filter structure [42, 43] where each filter is fed by three parameters (peak/notch central frequency, bandwidth, and gain) each stored in a configuration file indexed by the current elevation angle;
- a spherical model of the head taking into account the far-field and near-field scattering effects around its rigid body. The parametrization is made onto the sphere radius selected as a weighed combination of the listener's head dimensions [44] and the near-field contribution is reduced down to a first-order shelving filter [39];
- a spherical torso approximation (as in the snowman model [28]) that models elevation cues at low frequencies.

An expanded mention goes to the contents of the *extra* folder, into which a collection of third-party utility tools is kept updated and integrated in our prototyping environment mainly developed in Matlab, Pure Data and C/C++ libraries. A basic external for the rendering process is CW_binaural~ [45] that implements real-time convolution of sound inputs with selected HRIRs. It has the peculiar feature of being able to load an arbitrary discrete set of HRIRs in .wav format and realize different kinds of interpolations between adjacent spatial positions. This tool is at the basis of our dynamic 3D audio rendering system, where the successful transposition of dynamic sources into a virtual world not only depends on the accuracy of the interpolation scheme but is also heavily conditioned by the quality of the motion tracking system. In decreasing order of degree of immersion, a PhaseSpace Impulse MoCap system, a head-pose estimation system via webcam with the faceAPI software library, and a Trivisio Colibri wireless inertial motion tracker mounted on top of a pair of headphones are already integrated in our environment.


Status: Please choose azimuth and elevation.

Where did you hear the sound?
○ Inside the head ○ On the border ○ Outside the head

Subject: Michele Geronazzo M 28

Fig. 4: A GUI for localization experiments.

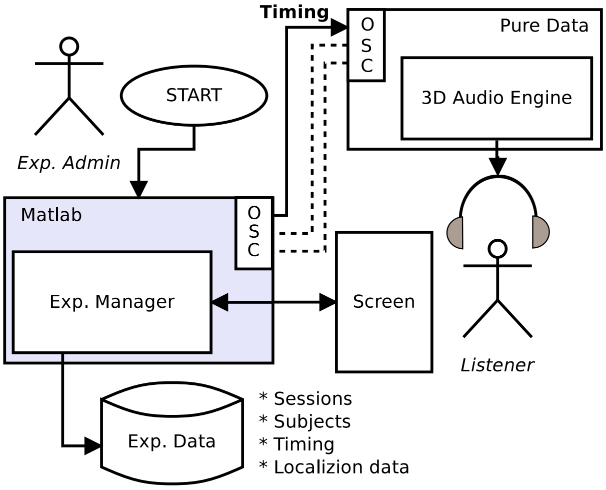


Fig. 5: A high-level technical description of a typical experimental scenario.

3.4. Experimental Environment

An environment for subjective localization tests is stored in the *evaluation* folder. A GUI in Matlab offers the main environment for the playback and judgment of the HRTF data and models to be evaluated, reported in the screenshot of Fig. 4. The subject listens to the sound stimulus, interactively selects the perceived sound location and/or other properties, and switches to the next sound stimulus.

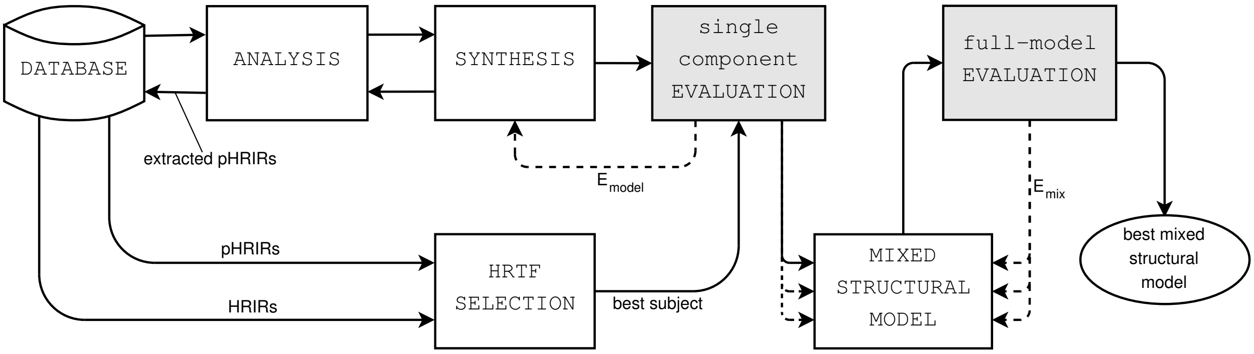


Fig. 6: Typical research workflow towards a complete structural model.

The data exchange between the Matlab environment and the audio engine is granted by the OSC (Open Sound Control) protocol, running on top of UDP. Fig. 5 reports a technical description of a typical experiment, where a *pnet* library and the *dumpOSC* external are responsible for communication on the Matlab side and on the Pure Data side respectively.

The basic features for the experimenter such as subject and session management are also available. Subjects' records and their personal information can be manipulated on demand. Each experimental session is stored in an independent file labeled by a session number id. The *task_info* struct contains the descriptive information of the task and the timestamps for each trial conducted within the session. The latter field operates as a primary key to read and understand the experimental data stored in a table specifically designed for the purpose of the experiment. Common statistical analysis softwares can directly import the already organized data.

4. MODUS OPERANDI

The research process aims to build a completely customizable structural model through subsequent refinements, starting from a selection of recorded HRIRs to a totally synthetic filter model. The intermediate steps are balanced mixtures of selected pHRIRs and synthetic structural components.

A candidate mixed structural model is described by a set of parameters and components; the evaluation step guides the exclusion of certain combinations of such components. The obtained 3D audio engine shall maximize the *efficiency-to-parametrization rate*: this means that synthetic submodels with a high customization degree but poor localization performance compel us to conceive a different or modified submodel. In such a case, a previous version of that submodel, whether it be synthetic, directly recorded or extracted by analysis, remains the best option for the component's acoustic contribution up to that point. As soon as the partial selection process reaches its maximum, we will have the optimal solution for our complete structural model. We now describe in more detail our typical research workflow, referring to Fig. 6 throughout the whole section.

4.1. HRTF Selection

The simplest HRTF selection possible is realized by choosing the same set of HRTFs for all of the listeners. Whether it be the best available localizer or a mannequin with mean anatomical features, no prediction can be made on the localization performance of a specific listener. An insight knowledge of the relation between localization cues and anthropometric features can guide the selection process as in [46]. If measurements of pinna dimensions for each subject in the considered HRTF database are available, a simple "best match" with directly measured pinna dimensions of a common listener can be exploited. Eq. 1 shows a possible distance function on the left pinna applied to the CIPIC HRTF database:

$$s = \min_{i=1\ldots N}\left(d_1^i + d_2^i + d_4^i - p^l\right), \tag{1}$$

where N is the number of subjects in the database, d_k is the vector associated to the k-th anthropometric feature of the pinna, i.e. cavum concha height

(d_1), cymba concha height (d_3) and fossa height (d_4) respectively, and p^l is the distance from the superior internal helix border to the intertragic incisure of the l-th listener. At the end of this procedure the s-th subject is declared the best-matching elevation localizer for listener l and this correspondence is saved in the experimental environment.

4.2. Structural HRTF Selection & Modeling

The keen observer shall criticize the restriction of the latter selection procedure to the sole contribution of the pinna. Of course, there exists no guarantee for the accuracy in azimuth localization. However, the former procedure can be applied to the Aalto PRTF database or alternatively to a collection of mid-sagittal PRTFs extracted in the analysis step. On the other side, the selection of a best matching contribution from the head by means of recorded impulse responses (such as the pinnaless KEMAR HRIRs) or of extracted ITD (interaural time difference) / ILD (interaural level difference) information may adopt the same principle. Such an alternative leads to a finer selection of each structural component and is at the base of the progression of our mixed structural modeling approach.

HRTF sets resulting from further refinements of selection criteria should now be compared to our candidate synthetic filter models, the parameters of which are strictly related to the anthropometric quantities used for HRTF selection. The simplest example is the spherical model optimization in [44] where the sphere radius is customized through an empirical equation derived from ITD measurements of a population of 25 subjects. Eq. 2 reports the weighted sum of head dimensions used to calculate the optimal sphere radius while Eq. 3 represents a trivial ITD selection.

$$a_{opt} = w_1 X_1 + w_2 X_2 + w_3 X_3, \tag{2}$$

$$s = \min_{i=1 \dots N} a^i_{opt} - a^l_{opt}, \tag{3}$$

where X_1 is head half-width, X_2 head half-height and X_3 head half-length, w_j is the j-th weight, a^i_{opt} is the optimal sphere radius for the i-th database subject and a^l_{opt} is the optimal sphere radius of the l-th listener. At the end of this procedure we obtain two alternatives to test and compare: a parameterized spherical filter model and a selected set of real ITDs.

Restricting our attention to the cited structural components (head and pinna), 3×3 instances of mixed structural models already arise from the combination of the following alternatives:

- pinna structural block: measured KEMAR PRTFs, Aalto PRTF database selection, selection of extracted PRTFs from HRTF databases;
- head structural block: measured pinnaless KEMAR HRTFs, selection of extracted ITDs, parameterized spherical filter model.

4.3. Evaluation

The candidate models are subjected to three complementary evaluations:

- **objective evaluation**: signal-related error metrics such as spectral distortion and spectral cross-correlation;
- **auditory model evaluation**: auditory filterbanks and statistical prediction models [47];
- **subjective evaluation**: listening tests of localization and realism.

The space of possible structural model instances is reduced by a two-stage evaluation procedure, made by a single-component and a full-model evaluation. The single-component evaluation focuses on the minimization of the error E_{model} in Fig. 6, defined as the mean localization error of the best synthetic version available. A dimensionally reduced localization space (e.g. mid-sagittal plane data only) supports this early stage. The full-model evaluation takes the best representative solutions for each structural component in order to test the combined effects and the orthogonality of the models within full-space 3D virtual scenes. The minimization of E_{mix}, defined as the mean localization error of the mixed structural model, leads the mixing process.

5. DISCUSSION AND PERSPECTIVES

The research framework presented in this paper answers both the requirements of structural modularity and systematic HRTF model evaluation. A well-defined modus operandi with the aim of designing new synthetic filter models and HRIR/pHRIR selection processes is expected to progressively set the bar closer and closer to a complete structural HRTF model. The modular approach can be also extended to a multimodal domain where the integration of a 3D audio rendering engine with other sensory modalities such as video and haptics (e.g. Phantom devices) requires an evaluation in terms of integration, cross-augmentation and/or sensory substitution. A possible technical solution is the X3D ISO standard XML-based file format for representing 3D virtual environments and H3DAPI [11] as the handler for unified graphic and haptic scene graphs.

Some possible future directions that well represent further improvements and developments of new tools for such work methodology are now listed:

- the definition of a standardized format for anthropometric features (possibly borrowed from biometric researches) so as to integrate this information in the HRTF database;
- a study of the acoustic contribution of the pinna outside the mid-sagittal plane, that requires a more complex analysis than that performed in [38];
- the introduction of an ear canal model to approximate the correct P_7 (see Fig. 1) at the eardrum along with a possible formalization of a structural HpIR model;
- the manipulation of reliable and complex auditory models so as to facilitate a systematic exclusion procedure of HRIR/pHRIR distance functions and parameters of the synthetic models' components;
- the extension of the mixed structural model by the inclusion of computer-simulated HRIRs/pHRIRs calculated from mesh models of human heads [48], spatially discretized so as to be included in our unified database.

6. REFERENCES

[1] D. R. Begault. *3-D Sound for Virtual Reality and Multimedia*. Academic Press Professional, Inc., Cambridge, MA, USA, 1994.

[2] A. Härmä, J. Jakka, M. Tikander, M. Karjalainen, T. Lokki, J. Hiipakka, and G. Lorho. Augmented reality audio for mobile and wearable appliances. *J. Audio Eng. Soc.*, 52(6):618–639, June 2004.

[3] C. I. Cheng and G. H. Wakefield. Introduction to head-related transfer functions (HRTFs): Representations of HRTFs in time, frequency, and space. *J. Audio Eng. Soc.*, 49(4):231–249, April 2001.

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

[5] H. Møller. Fundamentals of binaural technology. *Appl. Acoust.*, 36:171–218, 1992.

[6] B. Masiero and J. Fels. Perceptually robust headphone equalization for binaural reproduction. In *Proc. 130th Conv. Audio Eng. Soc.*, London, UK, May 2011.

[7] Z. Schärer and A. Lindau. Evaluation of equalization methods for binaural signals. In *Proc. 126th Conv. Audio Eng. Soc.*, Munich, Germany, May 2009.

[8] A. W. Bronkhorst. Localization of real and virtual sound sources. *J. Acoust. Soc. Am.*, 98(5): 2542–2553, November 1995.

[9] V. Välimäki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel. Fifty years of artificial reverberation. *IEEE Trans. Audio, Speech, Lang. Process.*, 20(5):1421–1448, July 2012.

[10] D. R. Begault, E. M. Wenzel, and M. R. Anderson. Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *J. Audio Eng. Soc.*, 49(10):904–916, October 2001.

[11] F. L. Wightman and D. J. Kistler. Resolution of front-back ambiguity in spatial hearing by listener and source movement. *J. Acoust. Soc. Am.*, 105(5):2841–2853, May 1999.

---

[11] http://www.h3dapi.org/

[12] M. D. Burkhard and R. M. Sachs. Anthropometric manikin for acoustic research. *J. Acoust. Soc. Am.*, 58(1):214–222, July 1975.

[13] A. Abaza, A. Ross, C. Hebert, M. A. F. Harrison, and M. S. Nixon. A survey on ear biometrics. *ACM Trans. Embedded Computing Systems*, 9(4):39:1–39:33, March 2010.

[14] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman. Localization using nonindividualized head-related transfer functions. *J. Acoust. Soc. Am.*, 94(1):111–123, July 1993.

[15] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hammershøi. Binaural technique: Do we need individual recordings? *J. Audio Eng. Soc.*, 44(6):451–469, June 1996.

[16] G. Plenge. On the differences between localization and lateralization. *J. Acoust. Soc. Am.*, 56 (3):944–951, September 1974.

[17] D. S. Brungart. Near-field virtual audio displays. *Presence*, 11(1):93–106, February 2002.

[18] E. C. Durant and G. H. Wakefield. Efficient model fitting using a genetic algorithm: Pole-zero approximations of HRTFs. *IEEE Trans. Speech Audio Process.*, 10(1):18–27, January 2002.

[19] D. J. Kistler and F. L. Wightman. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *J. Acoust. Soc. Am.*, 91(3): 1637–1647, March 1992.

[20] M. J. Evans, J. A. S. Angus, and A. I. Tew. Analyzing head-related transfer function measurements using surface spherical harmonics. *J. Acoust. Soc. Am.*, 104(4):2400–2411, October 1998.

[21] V. R. Algazi, R. O. Duda, R. P. Morrison, and D. M. Thompson. Structural composition and decomposition of HRTFs. In *Proc. IEEE Work. Appl. Signal Process., Audio, Acoust.*, pages 103–106, New Paltz, New York, USA, October 2001.

[22] S. Hwang, Y. Park, and Y. Park. Modeling and customization of head-related impulse responses based on general basis functions in time domain. *Acta Acustica united with Acustica*, 94 (6):965–980, November 2008.

[23] K. H. Shin and Y. Park. Enhanced vertical perception through head-related impulse response customization based on pinna response tuning in the median plane. *IEICE Trans. Fundamentals*, E91-A(1):345–356, January 2008.

[24] B. U. Seeber and H. Fastl. Subjective selection of non-individual head-related transfer functions. In *Proc. 2003 Int. Conf. Auditory Display (ICAD03)*, pages 259–262, Boston, MA, USA, July 2003.

[25] R. H. Y. So, B. Ngan, A. Horner, J. Braasch, J. Blauert, and K. L. Leung. Toward orthogonal non-individualised head-related transfer functions for forward and backward directional sound: Cluster analysis and an experimental study. *Ergonomics*, 53(6):767–781, June 2010.

[26] B. F. G. Katz and G. Parseihian. Perceptually based head-related transfer function database optimization. *J. Acoust. Soc. Am.*, 131(2): EL99–EL105, February 2012.

[27] V. C. Raykar, R. Duraiswami, and B. Yegnanarayana. Extracting the frequencies of the pinna spectral notches in measured head related impulse responses. *J. Acoust. Soc. Am.*, 118(1): 364–374, July 2005.

[28] V. R. Algazi, R. O. Duda, and D. M. Thompson. The use of head-and-torso models for improved spatial sound synthesis. In *Proc. 113th Conv. Audio Eng. Soc.*, pages 1–18, Los Angeles, CA, USA, October 2002.

[29] 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.*, pages 1–4, New Paltz, New York, USA, October 2001.

[30] G. Eckel. Immersive audio-augmented environments - the LISTEN project. In *Proc. 5th IEEE*

Int. Conf. Info. Visualization (IV'01), pages 571–573, Los Alamitos, CA, USA, July 2001.

[31] A. Andreopoulou and A. Roginska. Towards the creation of a standardized HRTF repository. In *Proc. 131st Conv. Audio Eng. Soc.*, New York, NY, USA, October 2011.

[32] N. Gupta, A. Barreto, M. Joshi, and J. C. Agudelo. HRTF database at FIU DSP lab. In *Proc. 35th IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP 2010)*, pages 169–172, Dallas, TX, USA, March 2010.

[33] W. G. Gardner and K. D. Martin. HRTF measurements of a KEMAR. *J. Acoust. Soc. Am.*, 97(6):3907–3908, June 1995.

[34] J. Gómez Bolaños and V. Pulkki. HRIR database with measured actual source direction data. In *Proc. 133rd Conv. Audio Eng. Soc.*, San Francisco, CA, USA, October 2012.

[35] T. Qu, Z. Xiao, M. Gong, Y. Huang, X. Li, and X. Wu. Distance-dependent head-related transfer functions measured with high spatial resolution using a spark gap. *IEEE Trans. Audio, Speech, Lang. Process.*, 17(6):1124–1132, August 2009.

[36] S. Spagnol, M. Hiipakka, and V. Pulkki. A single-azimuth pinna-related transfer function database. In *Proc. 14th Int. Conf. Digital Audio Effects (DAFx-11)*, pages 209–212, Paris, France, September 2011.

[37] M. Geronazzo, F. Granza, S. Spagnol, and F. Avanzini. A standardized repository of head-related and headphone transfer function data. In *Proc. 134th Conv. Audio Eng. Soc.*, Rome, Italy, May 2013.

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

[39] S. Spagnol, M. Geronazzo, and F. Avanzini. Hearing distance: A low-cost model for near-field binaural effects. In *Proc. EUSIPCO 2012 Conf.*, pages 2005–2009, Bucharest, Romania, September 2012.

[40] S. Spagnol, M. Geronazzo, and F. Avanzini. Fitting pinna-related transfer functions to anthropometry for binaural sound rendering. In *Proc. IEEE Int. Work. Multi. Signal Process. (MMSP'10)*, pages 194–199, Saint-Malo, France, October 2010.

[41] S. Spagnol and F. Avanzini. Real-time binaural audio rendering in the near field. In *Proc. 6th Int. Conf. Sound and Music Computing (SMC09)*, pages 201–206, Porto, Portugal, July 2009.

[42] 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):508–520, March 2013.

[43] 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, November-December 2011.

[44] V. R. Algazi, C. Avendano, and R. O. Duda. Estimation of a spherical-head model from anthropometry. *J. Audio Eng. Soc.*, 49(6):472–479, June 2001.

[45] D. Doukhan and A. Sédès. CW_binaural~: A binaural synthesis external for Pure Data. In *Proc. 3rd Puredata Int. Conv. (PdCon09)*, São Paulo, Brazil, July 2009.

[46] J. C. Middlebrooks. Individual differences in external-ear transfer functions reduced by scaling in frequency. *J. Acoust. Soc. Am.*, 106(3): 1480–1492, September 1999.

[47] E. H. A. Langendijk and A. W. Bronkhorst. Contribution of spectral cues to human sound localization. *J. Acoust. Soc. Am.*, 112(4):1583–1596, October 2002.

[48] B. F. G. Katz. Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation. *J. Acoust. Soc. Am.*, 110(5):2440–2448, November 2001.