The Selfear project: a mobile application for low-cost pinna-related transfer function acquisition

(Article begins on next page)

THE SELFEAR PROJECT: A MOBILE APPLICATION FOR LOW-COST PINNA-RELATED TRANSEFR FUNCTION ACQUISITION

Michele Geronazzo

Dept. of Neurological and Movement Sciences
University of Verona
michele.geronazzo@univr.it

Jacopo Fantin, Giacomo Sorato, Guido Baldovino, Federico Avanzini

Dept. of Information Engineering
University of Padova
Correspondence should be addressed to
avanzini@dei.unipd.it

ABSTRACT

Virtual and augmented reality are expected to become more and more influential even in everyday life in the next future; the role of spatial audio technologies over headphones will be pivotal for application scenarios which involve mobility. This paper faces the issue of head-related transfer function (HRTF) acquisition with low-cost mobile devices, affordable to anybody, anywhere and possibly in a faster way than the existing measurement methods. In particular, the proposed solution, called the SelfEar project, focuses on capturing individual spectral features included in the pinna-related transfer function (PRTF) guiding the user in collecting non-anechoic HRTFs through a self-adjustable procedure. Acoustic data are acquired by an audio augmented reality headset which embedded a pair of microphones at listener ear-canals. The proposed measurement session captures PRTF spectral features of KEMAR mannequin which are consistent to those of anechoic measurement procedures. In both cases, the results would be dependent on microphone placement, minimizing subject movements which would occur with human users. Considering quality and variability of the reported results as well as the resources needed, the SelfEar project proposes an attractive solution for low-cost HRTF personalization procedure.

1. INTRODUCTION

Binaural audio technologies have the aim of reproducing sounds in the most natural way, as if listeners were surrounded by realistic virtual sound-sources. This audio technology originated in late 19th century experiments [1], and it finds its roots in the recording of sounds through a "dummy head" that simulates the characteristics of the listener's head and incorporates two microphonic capsules inside the auditory ducts, emulating eardrums membranes [2]. Binaural audio could provide us with a 360 degrees listening experience, placing the virtual sound sources in defined points thanks to which our brain succeeds in perceiving the spatial qualities of source and envi-

ronment. It obtains its maximum efficiency through headphones reproduction, which keeps signal characteristics intact, without environmental reflections and reverberations. The rendering of virtual acoustic scenarios involves binaural room impulse responses (BRIR) that can be defined in two main components: the first one is connected to the environmental characteristics contained in the room impulse response (RIR), and the other one is related to the anthropometric characteristics of the listener, i.e. head-related impulse response (HRIR) [2]. All these impulse responses (IRs) have their counterparts in the frequency domain, formally their Fourier transforms: binaural room transfer function (BRTF), room transfer function (RTF), and head-related transfer function (HRTF). In particular, HRTFs describe a linear time-invariant filter where the acoustic filtering to which head, torso and ear of a subject concur is defined.

The ground-truth HRTF acoustic measurement offers an impulse response that has high-quality subject-related information and high-precision. However, professional HRTFs acquirement process requires time resources and expensive equipments that are rarely available for real applications. A more affordable procedure could discard some individual features to obtain a cheaper HRTF representation which still gives accurate psyco-acoustic information [3]. The HRTF acquirement process in a domestic environment is a challenging issue; recent trends are supported by low-cost devices for acquisition of 3D mesh images [4] and algorithms for HRTF modeling and customization [5]. These solutions unfortunately lack robust individual cues for external ear acoustics due to the fine anthropometric structure of the pinna. This information is collected in the so called pinna-related transfer function (PRTF) [6] which is also very difficult to model in numerical simulations [7, 8]. PRTFs contain salient localization cues for elevation perception (see [9] for a review), requiring an accurate representation in order to provide vertical dimension in binaural audio technologies.

This paper highlights the issue of costs reduction in the HRTF acquirement process, with particular focus on PRTF extrapolation for the mobile audio augmented reality (mAAR) system. This system involve headphones, provided with embedded external microphones for binaural capture of multiple-channel audio stream from the environment, as well as algorithms for binaural audio reproduction. An attractive idea is to use embedded micro-
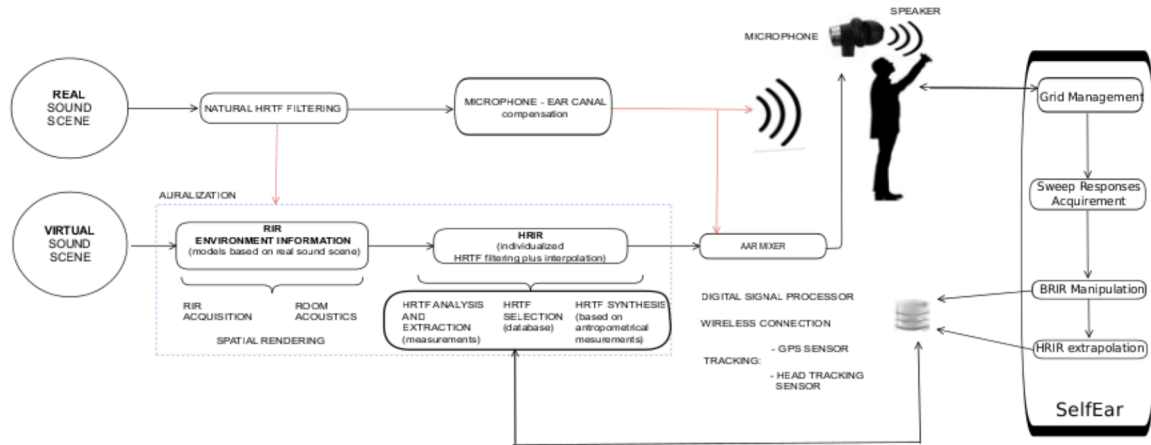
Figure 1: Schematic view of the SelfEar project in mAAR contexts.

phones in order to acquire HRTFs everywhere from sound stimuli played back by mobile device's speakers; the Self-Ear project has the purpose of developing the signal processing algorithms and interaction with the device in order to obtain a self-adjust procedure. Few studies have been conducted aiming to access the HRTF consistency in a non-anechoic environment for the acoustic contribution in mid-sagittal planes [10] which are relevant for individual spectral content introduced in PRTFs. The compromise on costs and portability unavoidably leads to mainly two different issues. Firstly, the mobile acquirement process implicates surrounding environment influences such as frequency coloration and phase shifts. Secondly, employing mobile device's speakers as sound source and consumer binaural microphones for the acquisition brings to less accurate recordings with respect to professional equipement.

In this paper we presented a series of measurements conducted in a silent booth on a KEMAR dummy head [11]. Our final goal was to compare responses obtained using the SelfEar system with those from professional equipment. In particular: Sec. 2 contains the description of a mobile audio augmented reality system and criteria for virtual sound externalization; in Sec. 3 the SelfEar project is presented. Section 4 describes acoustic measurements on a dummy head in non-anechoic environment. Finally, results are discussed in Sec. 5, and Sec. 6 concludes the proposed preliminary evaluation with promising research directions.

2. MOBILE AUDIO AUGMENTED REALITY

In a mAAR system (see fig.1), the listener might be able to enjoy a mix of real and virtual sound sources. The real sound sources are captured by headset microphones after natural acoustic filtering by the listener. A compensation filter considers errors introduced by different headphones and microphones positions compared to the unblocked entry point of the auditory channel resembling natural listener condition. The rendering of virtual sources needs a dynamic and parametric auralization process in order to create a perfect superposition with reality. Auralization employs BRIR, whose rendering must be coherently connected to the real surrounding environment in which the

subject is immersed. The cascade of RIRs and HRIRs should be personalized according to environment [12] and the listener [3]. Digital signal processing (DSP) algorithms implement corrective filters that compensate microphones, speakers and their interactions, taking into account psychoacoustic effects and artifacts that may be caused by wearing the earphone with respect to normal hearing conditions without headset.

Producing realistic virtual and augmented acoustic scenarios over headphones with particular attention to space properties and externalization issues remains one major challenge due to the interconnections of the above mentioned components of a mAAR system. Challenges and criteria for reality driven externalization can be summarized in four categories [13]:

- *ergonomic delivery system*: the ideal headphones should be acoustically transparent which means listeners are not aware of the sound emitted by transducers [14]. Low invasiveness of headphones cups are essential for such purpose [15].

- *tracking*: head movements in listening produces reliable dynamic interaural cues [16]; tracking listener position in the environment allows recognition of acoustic interaction and a common spatial representation between real and virtual scene;

- *room acoustics knowledge*: spatial impression and perception of the acoustic space involve the knowledge of real world early reflection and reverberation [17]; this information concurs to the availability of realistic spatial impression [18];

- *individual spectral cues*: head and pinna individually filter the incoming sound to listener ears; moreover individual correction must be considered for acoustic coupling between headphones and external ear [19].

3. THE SELFEAR PROJECT

3.1 Overview of the system

SelfEar is a mobile application designed to be executed on the Android platform in order to obtain user's personal HRIRs from a sound stimulus played by the mobile device. The phone/tablet must be held with the stretched arm and moved on the subject's median plane stopping at specific arm's elevation angle. The in-ear microphones capture the audio coming from the loudspeaker device, thus recording the position-, listener- and environment- specific BRIR, i.e. an acoustic self-portrait. The data collected through this application can be later employed in order to finally obtain an individualized HRIR. After post-processing procedures that compensate acoustic effect of acquiring conditions and playback device, individualized HRTFs can directly support spatial audio rendering and research framework [20]. Depending on the complexity of virtual scenarios, real-time HRTF synthesis is possible on mobile platform today. A promising technique involves HRTF selection through acoustic parameter extracted with SelfEar: the procedure selects the subject's best HRTF approximation based on existing HRTF databases (for instance CIPIC database [21]). [1]

3.2 Source manager

The spatial grid management system of SelfEar guides the user through the BRIR acquirement process defining a self-adjusted procedure depicted in Fig. 2. In the following, we describe each step, starting from the application launch to the session end, resulting in a set of individual BRIRs.

In the launching view of the SelfEar application, the user is asked to select the device's speakers position that may be on the top, front, bottom or back side of the device. This choice will have an effect on the device orientation during the sound stimulus playback in order to maximize speakers performance due to their directivity. The user can then press the "'Start'" button to begin the BRIR acquirement procedure; its steps follow this logical flow:

1. *Target reaching*: the current device elevation in the mid-sagittal plane appears on the scree above the target elevation (see the screenshot on the bottom right of Fig. 2). SelfEar computes data coming from the device's accelerometer on the three Cartesian axes, $a_{x,y,z}$, to calculate the current elevation on the horizon, ϕ_i, with the following formula:

$$\phi_i = \arctan\left(\frac{\pm a_y}{|a_z|}\right)$$

in case the speakers are located in the top or bottom side; whereas with the formula:

$$\phi_i = \arctan\left(\frac{\pm a_z}{|a_y|}\right)$$

in case in the front or back side. The numerator has the sign equals to:
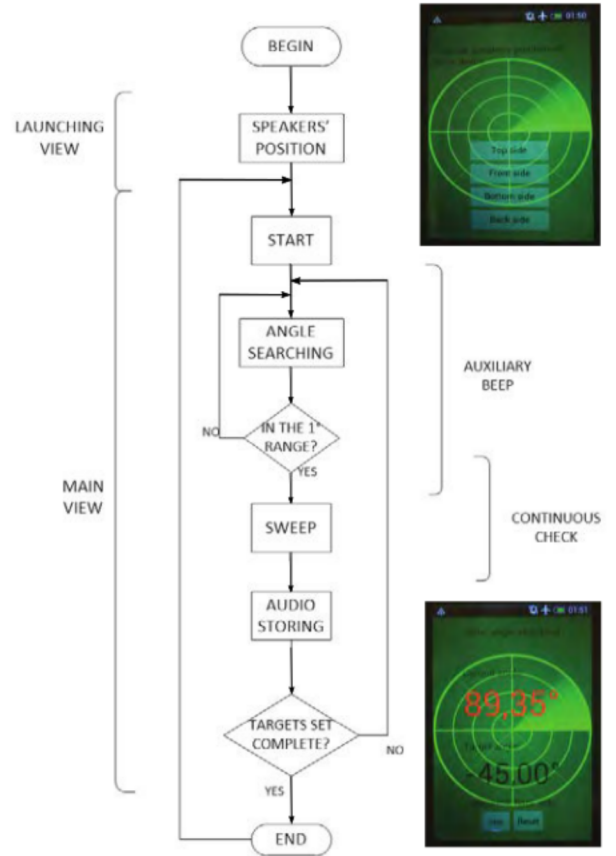

Figure 2: Block diagram of SelfEar procedure for BRIR acquisition in the median plane. Screenshots of the two application views are also reported.

+ for bottom- or back-sided speakers;

- for top- or front-sided speakers.

Target elevations sequence spans in ascending order among $[-40°, 40°]$ angles of the CIPIC HRTF database with equal spacing of $5.625°$. An auxiliary beep signal sonifies the distance between the actual and the target position supporting the elevation pointing procedure, which would be particularly useful in case the display is not visible due to the speaker's position (e.g. in the back side). The pause between one beep and another is directly proportional to the difference between the current measured angle, ϕ_i and the target, $\widehat{\phi_i}$, as shown in the following equation:

$$pause_i = \left|\phi_i - \widehat{\phi_i}\right| \cdot k$$

where i is an instant when a single beep terminates its playback and k is a constant value that makes perceptible the pause. [2] The goal for this step is to approach the target elevation within a precision of

---

[1] A collection of several acoustic measurements conducted on 50 different subjects (more than 1200 measurements each), also including anthropometric information.

[2] The formula returns a value in milliseconds, which would result in a too short pause to be heard without a constant multiplier. For the proposed implementation, we chose $k = 5$ with informal tests.
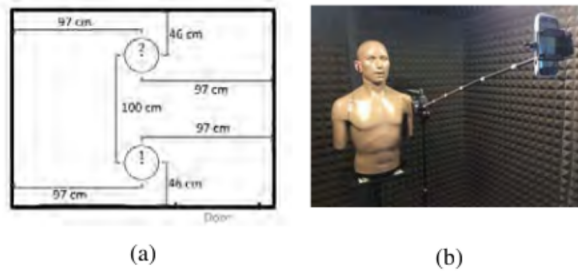
(a)   (b)

Figure 3: Measurement setup. (a) Source and receiver positions in the SSP. (b) SelfEar measument setup with selfie stick incorporated.

Figure 4: Magnitude comparison (in dB SPL) of BRTFs (thick lines) and relative PRTFs (thin lines) obtained using: receiver - the right headset microphone (H), source - the smartphone loudspeaker (S, dashed lines) and the Genelec loudspeaker (L, continuous lines).

$\pm 1°$. This step can be interrupted and resumed upon request by the user.

2. *Position check*: once ϕ_i enters the valid range, a stability timer of 2 seconds starts; should the number of times the user exits a range of $\pm 2°$ from the target reach three before the timer ends, the procedure goes back to the end of step 1.

3. *Sweep playback*: after the stability timer ends, the sound stimulus will be played from the device's speakers; should the user exit the $\pm 2°$ range just once during the sweep playback, the searching procedure for $\hat{\phi}_i$ is reset.

4. *BRIR storing*: once a sweep successfully terminates, the recorded audio is locally stored together with the elevation angle it refers to; the procedure then returns to step 1 with next target elevation in the sequence.

5. *End of session*: a session ends when elevations in the targets set are successfully reached.

4. ACOUSTIC MEASUREMENTS

Two measurement sessions were performed in a non-anechoic environment using a dummy head in order to minimize errors due to subject movement. We focused on the frontal direction $\phi = 0$ [6, 22] which is the spatial direction with highly significant PRTF spectral characteristics: the two main resonances (P1: omnidirectional mode, and P2: horizontal mode) and the three prominent notches (N1-3 corresponding to pinna reflections). Accordingly, we provided a detailed analysis of the acquired acoustic signals with different measurement setups, also reporting a qualitative evaluation of the SelfEar application for a set of HRIRs in the frontal mid-sagittal plane.

4.1 Setup

Facility and Equipment - All the measurement and experimental sessions were conducted inside a Sound Station Pro 45 (SSP), a 2×2 m silent booth with a maximum acoustic isolation of 45 dB.

Figure 3a shows the spatial setup of each experiment measurement in the SSP, identifying two positions: posi-
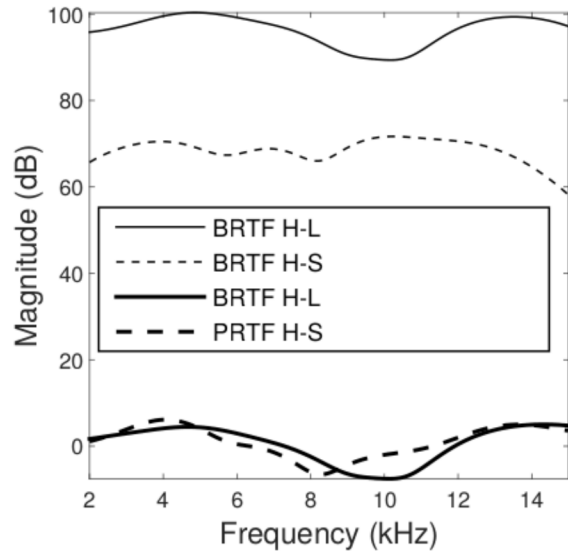
tion #1 relative to the source, while position #2 to the receiver.

Two types of playback device have been used in the experiments (acronyms also defined):

- *L* : a Genelec 8030A loudspeaker which has been calibrated to have an adequate SNR with a test tone at 500 Hz with 94 dB SPL;

- *S* : a HTC Desire C smartphone supported by a self-produced boom arm with a selfie stick incorporated;[3] in this case the maximum SPL reached is 51 dB at the reference frequency of 500 Hz.

Two type of receivers were also used in all the measurements (acronyms also defined):

- *H* : a pair of Roland CS-10EM in-ear headphones with embedded microphones;

- *K* : professional G.R.A.S microphones embedded in the head&torso simulator KEMAR; in the proposed setup, the right ear was equipped with ear canal simulator while the left ear not.

In all experiments, the center of sound source and receiver were placed at the same height. The source signal was a one second logarithmic sine sweep signal that comprises all the audible frequencies, from 20 Hz to 20 kHz, uniformly. The acoustic signals were recorded with the free software Audacity with a Motu 896 mk 3 audio interface and the processing was accomplished in Matlab (version 8.4).

---

[3] Since the 1-m selfie-stick is longer than an average arm of the user, we assume that PRTF spectral details for elevation perception are invariant with distance [23].
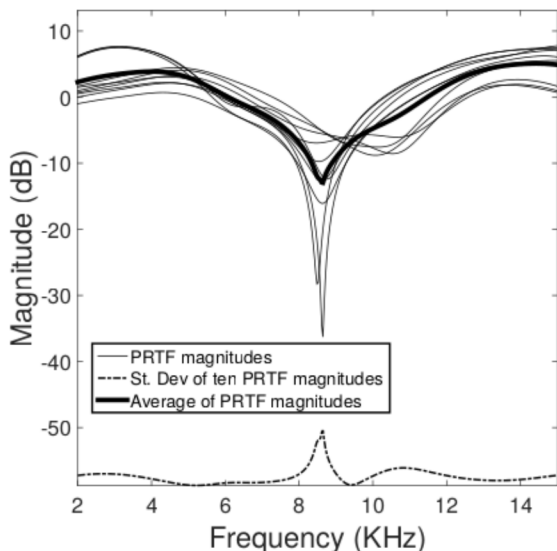
Figure 5: PRTF magnitudes in ten repositioning of the headset in the right ear canal of the KEMAR mannequin. Thick line represents the average magnitude. The standard deviation is shifted by -60 dB for convenience.

Calibration: diffuse-field measurement - A self-produced structure was used for diffuse-field measurements in order to acquire environmental- and setup- specific acoustical features. It consists of two pieces of iron wire that fall from the booth ceiling at a distance of 17.4 cm apart, corresponding to the same distance of KEMAR microphones.

We acquired diffuse-field measurements for all pairs of source and receiver, leading to a total of four measurements.

4.2 Acoustic data

Measurement session one - In this session, the Genelec loudspeaker and the KEMAR were placed inside the SSP , respectively in positions #1 and #2 of Fig. 3a. In the first step, right and left ear response of KEMAR were measured thus obtaining an *at the eardrum* measurement for the right ear and a *blocked ear canal* measurement for the left ear. The second step involved the headset inserted in the right ear canal; we conducted ten measurements related to different earphones placements in order to analyze measurement variability introduced by microphone position.

Measurement session two - In this session, the selfie-stick structure held the smartphones which was placed inside the SSP in position #1 of Fig. 3a; on the other hand, the KEMAR wearing the right headphone was placed in position #2 of Fig. 3a. The self-stick structure kept the smartphone at the distance of one meter from the KEMAR and allowed a fine angular adjustment. Measurements spanned 15 angles between $-40°$ and $+40°$ on the median plane. Finally, we obtained two sets of 15 measurements for the left KEMAR ear (without headphones) and the right headphone microphone.
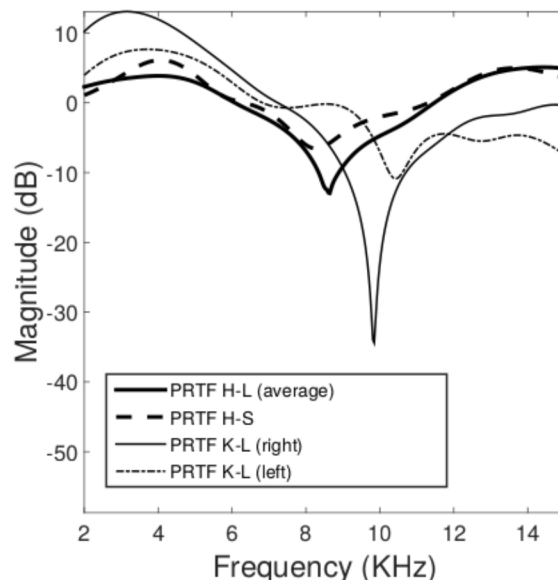

Figure 6: PRTF magnitude comparison: a) average PRTF from Fig.5; b) source: smartphone - receiver: headphone microphones; c) source: Genelec loudspeaker - receiver: KEMAR microphone in the right ear with ear canal; d) source: Genelec loudspeaker - receiver: KEMAR microphone in the left ear without ear canal simulator.

4.3 Analysis

For each measurement, the onset detection was computed applying a cross-correlation function with the original sweep signal and the BRIR was then extracted deconvolving sweep responses with the same sweep. Late reflections caused by the SSP and the presence of the equipment in the SSP were removed subtracting the corresponding diffuse-field responses from BRIRs. This processing ensured the acquirement of HRTFs. Accordingly, PRTFs were obtained by windowing each impulse response with a 1 -ms hanning window (48 samples) temporally-centered on the maximum peak and normalized on the maximum value in amplitude [6]. All of normalized PRTF were then band-pass filtered between 2 kHz and 15 kHz, ensuring the extraction of salient peaks and notches caused by pinna acoustics.

Figure 4 shown the comparison between the magnitudes in dB SPL of the BRIR extracted from the measurements using as source (i) the Genelec loudspeaker, (ii) the smartphone loudspeaker, and the headset on the right KEMAR ear as receiver. It has to be noted that the sound pressure levels of the two loudspeakers differed from 30 dB SPL on average denoting a low signal-to-noise ratio while using smartphone loudspeaker. The same figure also depicts the two corresponding normalized PRTFs in order to assess the diffuse-field effects on the results. For smartphone measurements, the contribution of the diffuse-field compensation is clearly visible due to non-negligible acoustic contribution of the low-cost loudspeaker.

In Fig. 5, the dB magnitude of PRTFs of ten repositionings and their average are reported. The standard deviation is also reported in order to analyze variability in the mea-
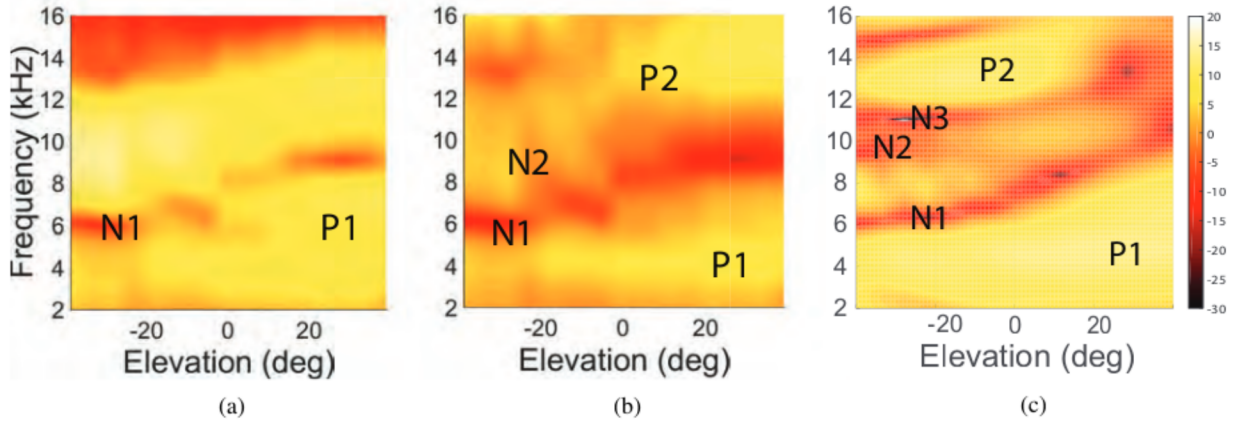
Figure 7: PRTFs in the median plane. (a) SelfEar acquisition - no compensation; (b) SelfEar acquisition - with diffuse-field compensation; (c) CIPIC KEMAR, Subject 165 - with free-field compensation. Plots also contain labels for the main peaks (P1-2) and notches (N1-3), where present.

surements introduced by headphone/microphone position. The maximum variability occurred in proximity of salient PRTF notches at 9 and 11 kHz which exhibited high sensitivities to topological changes between headphones and ear structure [8].

The main quantitative evaluation was performed in the frontal source position, $\phi = 0$, comparing the normalized PRTFs in different conditions. Figure 6 showed comparisons among PRTF magnitudes of measurements acquired with and without headset involving both Genelec and smartphone loudspeaker.

For this four PRTFs, the average spectral distortion (SD) error has been calculated [9] pairwise in the frequencies of interest $2\,kHz \leq f \leq 15\,kHz$ (value are showed in Table 1). These comparisons lead to several considerations:

- Pinna acoustics, K-L$_{right}$ vs. K-L$_{left}$: different ear shapes (right vs. left), and the ear canal acoustics (right with ear canal simulator and left with the blocked ear canal) differed remarkably; all comparisons between the 3rd and 4th column reflected these differences;

- Loudspeakers, H-S$_{right}$ vs. H-L$_{right}$: different loudspeakers introduced negligible spectral distortion in the proposed setup ($< 2\,dB$);

- SelfEar procedure, H-S$_{right}$ vs. K-L$_{left}$: difference between SelfEar acquisition of PRTFs and traditional measurement setup introduced the lower SD error in the available set (removing the control comparison on loudspeakers);

PRTF	H-L	H-S	K-L(right)	K-L(left)
H-L	0	1.79	6.92	5.25
H-S		0	7.35	4.64
K-L(right)			0	5.47
K-L(left)				0

Table 1: Spectral distortion among PRTFs of Figure 6. All values are in dB.

Figure 7 allows a visual comparison from the results obtained using SelfEar acquirement procedure on the considered elevation angles (with and without diffuse-field compensation), and the CIPIC measurements on the same angles range for Subject 165. The data were interpolated in order to have a smooth spatial transition.

5. DISCUSSION

From Christensen *et al.* [24] it is already known that the receiver position and its displacement from the ideal HRTF measurement point, i.e. at the entrance of the ear canal, highly influence HRTF directivity patterns for frequencies higher than $3 - 4\,kHz$. Our work is in agreement with their measurements showing a shift of notch central frequencies up to $2\,kHz$ with very high variability in magnitude among different microphone placements (see standard deviation of Fig. 5) and a maximum difference of 10 dB.

Shifts in peak/notch central frequencies are also visible in Fig. 6 due to topological differences between observation point, depending on microphone position, and acoustic scattering object, i.e. presence/absence of ear canal and differences between left and right ears. Spanning a wider range of frontal elevation positions allowed any measurement system to acquire relevant PRTF spectral features: in PRTFs from the CIPIC KEMAR (see labels in Fig. 7(c)), P1 has central frequency at $4\,kHz$ and P2 at $13\,kHz$, moreover N1 moves from 6 to 9 kHz, N3 from 11.5 to 14 kHz with increase in elevation; finally, N2 stars from 10 kHz and progressively disappears once reaching the frontal direction.

SelfEar application is capable of acquire P1 and N1 effectively considering both diffuse-field compensated PRTFs or not compensated BRIRs. Since the environment had not negligible contribution, the visual comparison between Fig. 7(a) and (b) stresses the importance of being able to accurately extract PRTFs from BRIRs. In particular from Fig. 7(b), one can identify also P2 and a little presence of N2. However, N3 was completely absent suggesting an acoustic interference introduced by headphones in pinna

concha. Following the resonances-plus-reflections model for PRTFs [6, 9], we can speculate about the absence of concha reflections due to headphone presence; moreover, the volume of the concha was dramatically reduced in this condition, thus producing changes in resonant modes of the pinna structure [8]. Furthermore, SD value of comparison $H - S$ vs. $K - L_{left}$ is 4.64 dB which suggests a good reliability in performances comparable to the personalization method in [9] (SD values between 4 and 8 dB) and to the state-of-the art numerical HRTF simulations in [8] (SD values between 2.5 and 5.5 dB).

It is worthwhile to notice that notch and peak parameters, i.e. central frequency, gain, and bandwidth, can be directly computed from available PRTFs. These spectral features can be exploited in synthetic PRTF models and/or HRTF selection procedure following a mixed structural modeling approach [3]. Finally, there is nothing to prevent a direct usage of PRTFs extracted by SelfEar in binuaral audio rendering.

6. CONCLUSION AND FUTURE WORK

The SelfEar application allows low-cost HRTF acquisition in the frontal median plane capturing peculiar spectral cues of the listener's pinna, i.e. PRTF. The application take advantage of a AAR technological framework for mobile devices. Once properly compensated, extracted PRTFs are comparable in terms of salient acoustical features to those measured in anechoic chamber.

The proposed system was tested following a robust measurements setup without a human subject in a silent booth which is an acoustically treated environment. Thus, a robust procedure is require for PRTF capturing in domestic environments, statistically assessing the influence of noisy and random acoustic events, as well as subject movements during the acquisition. For such purpose, signal processing algorithms for event detection, noise cancellation and movement tracking are crucial in signal compensation and in pre- and post- processing stages.

A natural evolution of this application will take into account also sagittal planes, i.e. plane around listeners with azimuth $\neq 0$, with particular attention to frontal directions which are easily accessible with arm movements and are crucial for auditory displays such as sonified screens [25]. Optimized procedures will be studied in order to reduce the number of required source positions and to control mobile position and orientation with respect to user movements; the SelfEar application will implement computer vision algorithms able to track listener's head-pose in real-time with embedded camera and depth sensors.

In addition to HRTF acquisition functionality, we will include capabilities of full BRIR acquisition in SelfEar, storing RIR and HRIR responses separately in order to directly render mAAR scenarios coherently in real-time. Extrapolated RIR will parametrize computational room acoustic models for the purpose of dynamic auralization, such as image-source and raybeam-tracing modeling for the first reflections and statistical handling of late reverberation [12].

Finally, it is indisputable that psycho-acoustic evaluation with human subjects is necessary in order to confirm the reliability of the SelfEar application providing effective individualized HRIRs in rendering virtual sound sources.

7. REFERENCES

[1] S. Paul, "Binaural Recording Technology: A Historical Review and Possible Future Developments," *Acta Acustica united with Acustica*, vol. 95, no. 5, pp. 767–788, Sep. 2009.

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

[3] 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)*, Santorini, Greece, Jul. 2013, pp. 1–8.

[4] H. Gamper, M. R. P. Thomas, and I. J. Tashev, "Anthropometric Parameterisation of a Spherical Scatterer ITD Model with Arbitrary Ear Angles," in *2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct. 2015, pp. 1–5.

[5] S. Spagnol, M. Geronazzo, D. Rocchesso, and F. Avanzini, "Extraction of Pinna Features for Customized Binaural Audio Delivery on Mobile Devices," in *Proc. 11th Int. Conf. on Advances in Mobile Computing & Multimedia (MoMM13)*, Vienna, Austria, Dec. 2013, pp. 514–517.

[6] M. Geronazzo, S. Spagnol, and F. Avanzini, "Estimation and Modeling of Pinna-Related Transfer Functions," in *Proc. of the 13th Int. Conf. on Digital Audio Effects (DAFx-10)*, Graz, Austria, Sep. 2010, pp. 431–438.

[7] H. Ziegelwanger, P. Majdak, and W. Kreuzer, "Numerical Calculation of Listener-specific Head-related Transfer Functions and Sound Localization: Microphone Model and Mesh Discretization," *J. Acoust. Soc. Am.*, vol. 138, no. 1, pp. 208–222, Jul. 2015.

[8] S. Prepeliță, M. Geronazzo, F. Avanzini, and L. Savioja, "Influence of Voxelization on Finite Difference Time Domain Simulations of Head-Related Transfer Functions," *J. Acoust. Soc. Am.*, vol. 139, no. 5, pp. 2489–2504, May 2016.

[9] 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.*, vol. 21, no. 3, pp. 508–519, Mar. 2013.

[10] A. Ihlefeld and B. Shinn-Cunningham, "Disentangling the Effects of Spatial Cues on Selection and Formation of Auditory Objects," *J. Acoust. Soc. Am.*, vol. 124, no. 4, pp. 2224–2235, 2008.

[11] W. G. Gardner and K. D. Martin, "HRTF Measurements of a KEMAR," *J. Acoust. Soc. Am.*, vol. 97, no. 6, pp. 3907–3908, Jun. 1995.

[12] L. Savioja and U. P. Svensson, "Overview of Geometrical Room Acoustic Modeling Techniques," *J. Acoust. Soc. Am.*, vol. 138, no. 2, pp. 708–730, Aug. 2015.

[13] J. Loomis, R. Klatzky, and R. Golledge, "Auditory Distance Perception in Real, Virtual and Mixed Environments," in *Mixed Reality: Merging Real and Virtual Worlds*, Y. Ohta and H. Tamura, Eds.   Springer, 1999.

[14] J. Ramo and V. Valimaki, "Digital Augmented Reality Audio Headset," *J. of Electrical and Computer Engineering*, vol. 2012, p. e457374, Oct. 2012.

[15] R. W. Lindeman, H. Noma, and P. G. d. Barros, "An Empirical Study of Hear-Through Augmented Reality: Using Bone Conduction to Deliver Spatialized Audio," in *2008 IEEE Virtual Reality Conference*, Mar. 2008, pp. 35–42, 00013.

[16] W. O. Brimijoin, A. W. Boyd, and M. A. Akeroyd, "The Contribution of Head Movement to the Externalization and Internalization of Sounds," *PLoS ONE*, vol. 8, no. 12, p. e83068, Dec. 2013.

[17] N. Sakamoto, T. Gotoh, and Y. Kimura, "On -Out-of-Head Localization- in Headphone Listening," *J. of the Audio Eng. Soc.*, vol. 24, no. 9, pp. 710–716, Nov. 1976.

[18] J. S. Bradley and G. A. Soulodre, "Objective Measures of Listener Envelopment," *J. Acoust. Soc. Am.*, vol. 98, no. 5, pp. 2590–2597, Nov. 1995.

[19] F. L. Wightman and D. J. Kistler, "Headphone Simulation of Free-Field Listening. II: Psychophysical validation," *J. Acoust. Soc. Am.*, vol. 85, no. 2, pp. 868–878, 1989.

[20] M. Geronazzo, S. Spagnol, and F. Avanzini, "A Modular Framework for the Analysis and Synthesis of Head-Related Transfer Functions," in *Proc. 134th Conv. Audio Eng. Society*, Rome, Italy, May 2013.

[21] 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.*, New Paltz, New York, USA, Oct. 2001, pp. 1–4.

[22] F. Asano, Y. Suzuki, and T. Sone, "Role of Spectral Cues in Median Plane Localization," *J. Acoust. Soc. Am.*, vol. 88, no. 1, pp. 159–168, 1990.

[23] D. S. Brungart and W. M. Rabinowitz, "Auditory localization of nearby sources. Head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 106, no. 3, pp. 1465–1479, 1999.

[24] F. Christensen, P. F. Hoffmann, and D. Hammershøi, "Measuring Directional Characteristics of In-Ear Recording Devices," in *In Proc. Audio Eng. Soc. Con. 134*.   Audio Engineering Society, May 2013.

[25] A. Walker and S. Brewster, "Spatial Audio in Small Screen Device Displays," *Pers. Technol.*, vol. 4, no. 2, pp. 144–154, Jun. 2000.