Estimation of Spectral Notches from Pinna Meshes

Insights from a Simple Computational Model

Spagnol, Simone; Miccini, Riccardo; Onofrei, Marius George; Unnthorsson, Runar; Serafin, Stefania

Published in:
IEEE/ACM Transactions on Audio Speech and Language Processing

DOI (link to publication from Publisher):
10.1109/TASLP.2021.3101928

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

Estimation of Spectral Notches From Pinna Meshes: Insights From a Simple Computational Model

Simone Spagnol, *Senior Member, IEEE*, Riccardo Miccini, Marius George Onofrei, Runar Unnthorsson, and Stefania Serafin

***Abstract*—While previous research on spatial sound perception investigated the physical mechanisms producing the most relevant elevation cues, how spectral notches are generated and related to the individual morphology of the human pinna is still a topic of debate. Correctly modeling these important elevation cues, and in particular the lowest frequency notches, is an essential step for individualizing Head-Related Transfer Functions (HRTFs). In this paper we propose a simple computational model able to predict the center frequencies of pinna notches from ear meshes. We apply such a model to a highly controlled HRTF dataset built with the specific purpose of understanding the contribution of the pinna to the HRTF. Results show that the computational model is able to approximate the lowest frequency notch with improved accuracy with respect to other state-of-the-art methods. By contrast, the model fails to predict higher-order pinna notches correctly. The proposed approximation supplements understanding of the morphology involved in generating spectral notches in experimental HRTFs.**

***Index Terms*—Spatial audio, head-related transfer functions (HRTFs), audio signal processing, spatial hearing, HRTF individualization, pinna.**

I. Introduction

RESEARCH on Head-Related Transfer Functions (HRTFs) [1] has been rapidly progressing over the past decade. The availability of relatively low-cost hardware technologies for immersive visualization has shown the need for higher fidelity HRTF-based spatial sound simulation. Although recent advances in fast HRTF acquisition make it possible to capture HRTFs via acoustic measurement in minutes [2], [3], individual HRTFs are still hard to obtain for the general public. As a consequence, most applications have been relying on non-individual, or generic, HRTFs. Generic HRTFs are known to cause systematic localization errors such as front/back reversals, wrong elevation perception, and inside-the-head localization [4], even if the use of real-time head tracking and artificial reverberation is able to significantly reduce these issues [5].

HRTF individualization—or the process of providing the user with an HRTF that matches the temporal and spectral content of their own ears' responses—aims at solving the above issues. HRTF individualization approaches can be divided into three families: numerical simulation, indirect individualization based on anthropometry, and indirect individualization based on perceptual feedback [6]. The first class of methods, consisting in simulating the propagation of acoustic waves around a 3D scan of the subject's head and torso through computational techniques [7], has recently gained much attention from researchers. Although such methods are getting more and more accurate in predicting an individual HRTF from a 3D head mesh [8], a lack of large-scale perceptual evaluation studies prevents further discussion on their effectiveness.

On the other hand, one of the active but still unresolved research topics is to identify the physical mechanisms underlying the generation of the most important spectral HRTF cues. A thorough understanding of such generation mechanisms would allow the development of HRTF models that are easy to tune and computationally efficient [9]. Unfortunately, many available HRTF models focus on spatial rendering limited to the horizontal plane [10], [11], overlooking the relevant elevation cues.

While previous research shows some understanding of the relationship between pinna modes/reflections and HRTFs [12]–[14], more research is still needed to fully understand the main elevation cues and how they depend on the individual. This is partly due to generally low sample sizes in terms of number of measured individuals and the lack of commonly agreed protocols for measuring HRTFs and individual morphology, resulting in very different measurements for the same subject depending on the setup [15]. The lack of large and standardized HRTF datasets together with the relevant anthropometric data also complicates the use of machine learning and deep learning techniques, although some research in that direction is starting to appear [16]–[18].

In this paper we extend our previous research by using a new dataset of highly controlled HRTF measurements on a KEMAR

Manuscript received November 6, 2020; revised April 20, 2021 and June 1, 2021; accepted July 26, 2021. Date of publication August 4, 2021; date of current version August 19, 2021. This work was supported in part by the European Union's Horizon 2020 Research and Innovation programme under the Marie Skłodowska-Curie under Grant 797850 and in part by the NordForsk's Nordic University Hubs programme under Grant 86892. The associate editor coordinating the review of this manuscript, and approving it for publication was Prof. H. Hacihabiboglu. *(Corresponding author: Simone Spagnol.)*

Simone Spagnol was with the Department of Architecture, Design and Media Technology, Aalborg University, 2450 Copenhagen, Denmark. He is with the Faculty of Industrial Design Engineering, Delft University of Technology, 2628CE Delft, The Netherlands (e-mail: s.spagnol@tudelft.nl).

Riccardo Miccini, Marius George Onofrei, and Stefania Serafin are with the Department of Architecture, Design and Media Technology, Aalborg University, 2450 Copenhagen, Denmark (e-mail: rimiccini@yahoo.it; monofr11@student.aau.dk; sts@create.aau.dk).

Runar Unnthorsson is with the School of Engineering and Natural Sciences, University of Iceland, 107 Reykjavík, Iceland (e-mail: runson@hi.is).

Digital Object Identifier 10.1109/TASLP.2021.3101928

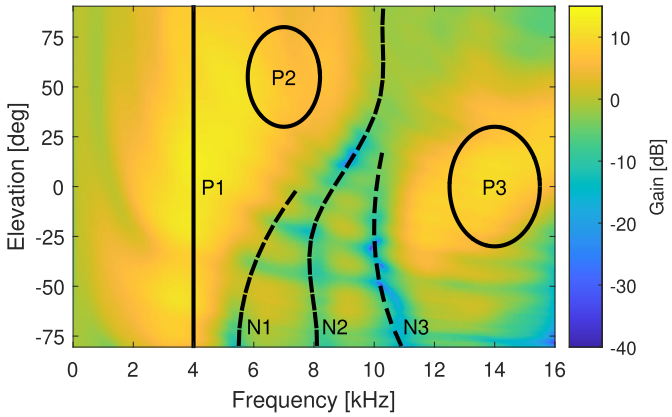


Fig. 1. Frontal median plane HRTF amplitude spectra for an example subject with the main spectral peaks (P1−P3) and notches (N1−N3) highlighted.

mannequin with several different artificial pinnae, captured in an anechoic room with increased vertical resolution. Furthermore, we propose a simple computational model to predict the most important elevation cues, i.e., pinna notches, from a 3D pinna mesh. The computational model applied to this dataset provides more insight on the generation mechanisms for the first pinna notch, known as N1.

The paper is organized as follows. Section II outlines the related work. Section III reports the data collection procedures, Section IV describes our custom procedure to extract the relevant features from HRTFs, and Section V introduces the computational model. Finally, Section VI presents our results, and Sections VII and VIII report the discussion and conclusions, respectively.

II. Related Work

Humans are capable of vertical spatial sound localization by parsing specific spectral cues. This is achieved with poorer resolution than for sound sources located in the horizontal plane, where interaural cues play a major role [19]. A number of seminal experiments advanced the understanding of the spectral cues responsible for vertical localization. Notably, Hebrank and Wright [20] established that spectral cues for vertical localization exist between 4 and 16 kHz, and that a sound must occupy this frequency range in order to be localized along the median plane. The pinna is known to be responsible for generating these cues, which come in the form of spectral peaks and notches generated through processes of resonance, reflection, and diffraction [21]. Fig. 1 highlights these spectral features in an example HRTF set collected across the frontal median plane.

Shaw [22] identified six resonant modes of the pinna excited by sounds from different directions. These modes were calculated and averaged among 10 different pinnae and include: one mode appearing for all directions at 4.2 kHz (omnidirectional mode), two modes appearing for directions above the head at 7.1 and 9.6 kHz (vertical modes), and three modes appearing around the horizontal plane at 12.2, 14.4, and 16.7 kHz (horizontal modes). Pinna modes are thought to cause the most prominent peaks in the HRTF [12]—see P1 to P3 in Fig. 1 which correspond to the omnidirectional, vertical, and horizontal modes, respectively. As Hebrank and Wright observed [20], perception above the head is associated with a 7 to 9 kHz peak and frontal perception with increased energy above 13 kHz. Taken together, these results highlight the relevance of pinna modes as elevation cues. The center frequencies of peaks are relatively insensitive to changes in elevation of the sound source [23], and models to estimate them from individual anthropometric parameters have been recently proposed [14], [24].

By contrast, the exact origin of spectral notches is more difficult to trace, although it has been long thought to refer to reflections on the concha wall causing the pinna to behave like a delay-and-add system in the time domain [25]. More recently, it has been hypothesized that notches are due to pressure nodes in the concha induced by interactions between propagating waves and the pressure anti-node with opposite phase forming in the upper pinna cavities [12]. Center frequencies of pinna notches, especially N1, are generally seen to increase with the elevation angle [20], [26], therefore representing salient elevation cues [27]. Furthermore, notches exhibit little variation with changes in azimuth [21] or distance [28], [29] and are deeper for sound images below the horizontal plane [30].

While a number of characteristic peak and notch patterns have previously been identified, their contribution to vertical sound localization is still a topic of inquiry. Iida *et al.* [31] achieved localization performances similar to using the subjects' own HRTF for the front and rear portions of the median plane by synthesizing a parametric HRTF composed of only the first peak (P1) and the first two notches (N1, N2). They therefore concluded that N1 and N2 are the most relevant elevation cues, while P1 could be used by the human hearing system as a reference for analyzing notches. In a more recent experiment [32], they further improved localization performances for a larger subset of elevations by including a second peak (P2). Models that mimic the characteristic peak/notch patterns seen in HRTFs have also been proposed throughout the past decades, including Watkins' double-delay-and-add time-domain model [33], Shaw's physical flange-and-cavity model [34], and the diffraction-reflection model by Lopez-Poveda and Meddis [21]. The main drawback of these models is that it is unclear how they can be customized to fit a particular listener.

As a matter of fact, relating the individual variations of spectral features in HRTFs to pinna anthropometry is a key aspect of HRTF individualization, and has been previously investigated to some extent. In a recent work, Iida *et al.* [35] synthesized the HRTF of listeners from their individual anthropometric parameters using multiple regression, obtaining similar spectral features as in the respective measured HRTFs. Another promising approach to HRTF individualization involves embedding the HRTF into a compressed representation, which can then be estimated using anthropometric parameters. Common choices for such compressed representation include Principal Component Analysis (PCA) [36], [37], Surface Spherical Harmonics [38], and, more recently, deep autoencoders [11]. The anthropometric parameters can be related to the compressed representation using multiple regression analysis or artificial neural networks [39].

Furthermore, advances in the fields of computer learning and computer vision allow for automatic feature extraction from images of the pinnae [40], [41]. Pinna images are featured in

recent work such as that of Lee and Kim [42], where they are used together with anthropometric measurements as input data for an artificial neural network trained to synthesize HRTF spectra. The authors of the present paper previously proposed a structural model of the pinna [13] whose parameters are given by a simple ray-tracing algorithm that converts 2D reflection paths on three distinct pinna edges into notch frequencies. Using the same algorithm together with a subset of standard anthropometric parameters, a linear regression model to estimate N1 frequencies from individual anthropometry was also introduced [43], as well as a marginally improved model based on PCA [44].

With the objective of gaining additional understanding of the mechanisms that generate spectral notches in experimental HRTFs, this paper offers the following key contributions:

- a new dataset of highly controlled HRTF measurements of a KEMAR mannequin with different artificial pinnae and corresponding 3D pinna meshes, specifically designed to investigate the effect of a given pinna on the HRTF;
- a simple computational model for predicting individual notches from a pinna mesh, which builds on and extends to the 3D case the ray-tracing algorithm previously described in [13];
- a performance evaluation of the computational model on the new HRTF dataset and a discussion of its strengths and shortcomings.

III. Data Collection

In this work we use a new dataset of HRTFs measured on a KEMAR mannequin [45] equipped with custom silicone pinnae modeled on 20 different artificial heads. This set of acoustic measurements is an improved version of the previously released Viking HRTF dataset [46], [47], with a focus on extra median-plane measurements. Moreover, the new measurements were carried out in an anechoic environment. In order to minimize the impact of issues related to HRTF measurements on human subjects and of anatomical components other than the single pinna, we chose to create our own set of acoustic measurements rather than relying on available HRTF datasets.

A. Artificial Pinnae and 3D Meshes

In order to provide a diverse sample of pinna shapes for the KEMAR mannequin, we set up and applied a custom procedure to cast silicone replicas of pinnae from artificial heads. In short,[1] the main steps of the procedure consisted in sequentially creating

1) a first negative silicone mold out of the artificial head's pinna;
2) a positive replica with a hard material (Jesmonite). This step was in order to shape the standard rectangular base of the KEMAR, and to allow drilling a hole to accommodate the KEMAR microphone inside the concha;
3) a second negative silicone mold with rectangular base included;
4) the final silicone (25 Shore-A hardness) replica.

[1]For more details on the pinna casting procedure, please consult [46].

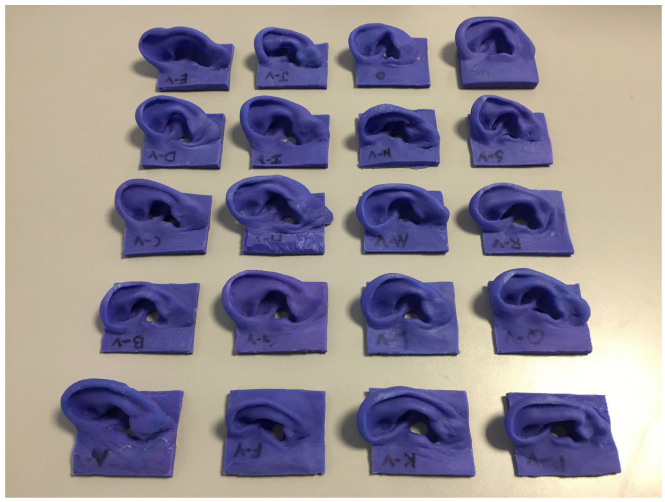


Fig. 2. The sample of silicone pinna replicas.

We applied the above casting procedure to a series of left ears from 20 different artificial heads. These included the KEMAR with standard large anthropometric pinna (GRAS KB5001) and 19 dummy heads made out of plaster, borrowed from the Saga Museum in Reykjavík. The dummy heads, labeled A to S in alphabetical order, were manufactured between 2001 and 2003[2] by casting the heads of 19 Icelandic humans (7 female), aged between 7 and 77 at the time of manufacturing. The final result is shown in Fig. 2.

3D scans of all the artificial pinnae were then acquired with a Creaform Go!SCAN 20 white-light handheld scanner at 1 mm resolution. Every pinna was scanned on both the front and back (rectangular base) sides and the two scans were then merged using the VXelements software. Each single scan took approximately two minutes. The same software was also used to automatically close occasional holes in the mesh based on the neighboring vertices, and to manually align to a global coordinate system where the x- (width) and y- (height) axes are parallel to the shorter and longer sides of the rectangular base, respectively, and the z- (depth) axis is normal to the back side of the base. An example pinna mesh is shown in Section V. The origin of each mesh was manually selected as the point lying approximately in the center of the microphone hole at a depth that roughly corresponds to the location of the microphone diaphragm when the corresponding artificial pinna is inserted in the KEMAR's left pinna slot. This was made possible thanks to the availability of detailed pictures of the HRTF measurement sessions described in the following Subsection.

B. Acoustic HRTF Measurements

The HRTF measurement system, pictured in Fig. 3, consisted of an aluminum scaffolding hosting a spinning platform for a KEMAR mannequin (45BB-4 configuration), as well as a pivoting arm equipped with a Genelec 8020 C loudspeaker at its farthest extremity. The two degrees of freedom offered by the platform and arm allowed the speaker to travel around a

[2][Online]. Available: https://sagamuseum.is/sagadesign/index.html

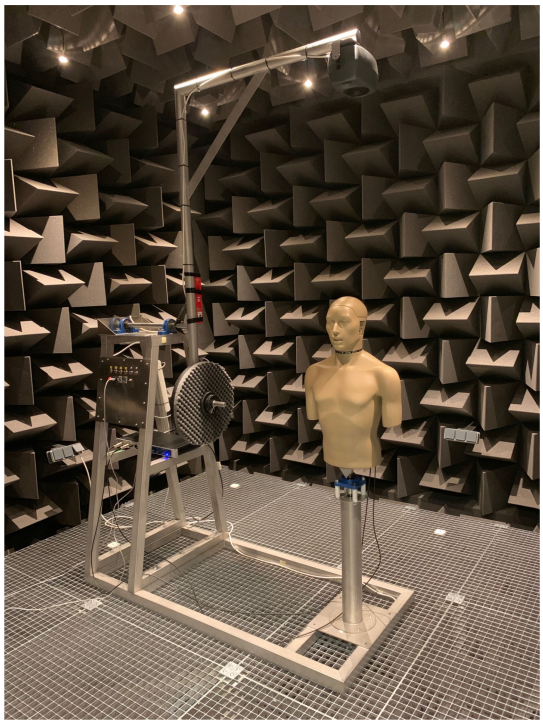


Fig. 3. The HRTF measurement system: mechanical apparatus, loudspeaker, and KEMAR mannequin.

1-m radius spherical surface centered on the middle point of the mannequin's interaural axis. This distance guarantees the collection of far-field spectral cues with reasonable accuracy [29]. The two rotation axes were driven by independent high-torque stepper motors (JVL MST001 A, 1.2 Nm) with integrated gearboxes, controlled using an Arduino with serial connection. A dedicated RME Fireface 802 audio interface connected both the loudspeaker and the KEMAR half-inch pressure microphones (GRAS 40AO) to the host workstation. Prior to the measurement sessions, a rotary laser level was used to ensure proper alignment of the various components.

The HRTF measurements were carried out during the month of November 2019 inside the anechoic chamber recently installed at the University of Iceland. The chamber has a size of $5.2 \times 4.3 \times 3.9$ m (LWH), and the mannequin was placed roughly in the center of it. Based on the ISO 3745 and ISO 26 101 standards, the anechoic free field in the chamber has been certified (as of September 2019) compliant for measurements within a volume equivalent to a distance of approximately 0.2 m from wedge tips on the walls, ceiling and floor in the frequency range from 200 Hz to 10 kHz. In order to remotely monitor the data acquisition process and to detect possible mechanical or electrical failures, video footage was streamed to the host workstation in real time via a webcam mounted inside the anechoic chamber and pointing towards the measurement system.

The sweep method was adopted for recording all acoustical responses [48]. Specifically, the excitation signal $x[n]$ emitted by the loudspeaker was a 0.9 s logarithmic sweep spanning frequencies from 20 Hz to 20 kHz at a sampling rate of 48 kHz. The average SPL level at the left KEMAR pinna with the source

Fig. 4. KEMAR large anthropometric pinna replica (left), pinna replica of artificial head S (center), and silicone baffle (right) mounted on the KEMAR's left channel.

directly above the mannequin was 90 dB SPL at 1 kHz. A standard large anthropometric pinna (GRAS KB5000) was installed on the right channel of the KEMAR throughout the whole measurement schedule. Conversely, the left pinna changed at every HRTF measurement session as follows,

- sets A to S: the 19 left pinna replicas of the corresponding artificial heads;
- set T: the KEMAR anthropometric left pinna replica;
- sets X and Y: the original KEMAR anthropometric left pinna in its soft (35 Shore-OO) and stiffer (55 Shore-OO) variants, respectively;
- set Z: a flat 25 Shore-A silicone baffle filling the pinna slot flush with the head (so as to simulate a "pinna-less" condition).

Fig. 4 shows the KEMAR replica, the pinna replica of one artificial head, and the silicone baffle mounted on the mannequin's left channel.

For each left pinna, sweep measurements were recorded in the frontal half of the median plane in the range of elevations from $-80°$ to $+90°$,[3] with a 1-degree resolution. This means that during these measurements the mannequin was kept at a fixed $0°$ azimuth and only the arm motor was operated, in order to collect data for different elevations. The entire acquisition process for a single measurement session was fully automated by a MATLAB script that required about 20 minutes to run. Upon launch, the script would simultaneously reproduce the input stimulus on the loudspeaker and collect audio data from the two KEMAR channels; it would then rotate the pivoting arm to the next angle and repeat the process. In order to minimize unwanted arm vibrations, a two-second pause took place between all motor commands and audio acquisitions. After the entire set of measurements for the given left pinna was acquired, the arm was brought back to its starting position with the help of a bubble level. Subsequently, a new left pinna was mounted on the mannequin's head and another set of measurements could be instantiated. We recorded a total of 23 (left pinna samples) $\times$ 171 (elevations) $=$ 3933 sweep responses.

On top of the above measurements, free-field reference measurements for the system were collected, allowing for removal of the influence of any element other than the mannequin. This was

[3]Positive/negative elevations correspond to directions above/below the horizontal plane, respectively.

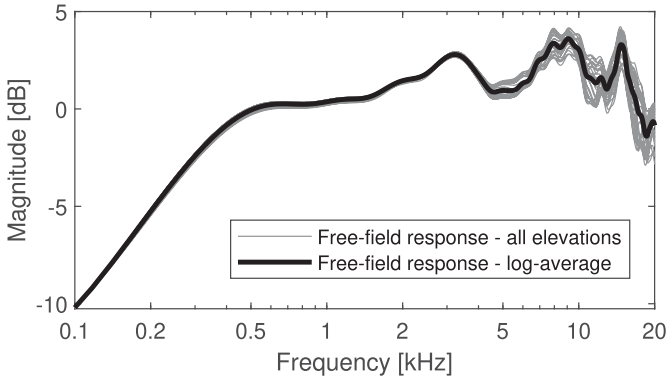


Fig. 5. Amplitude spectra of the free-field reference measurements. Thin gray lines show the responses for all elevations; the thicker black line shows the log-average.

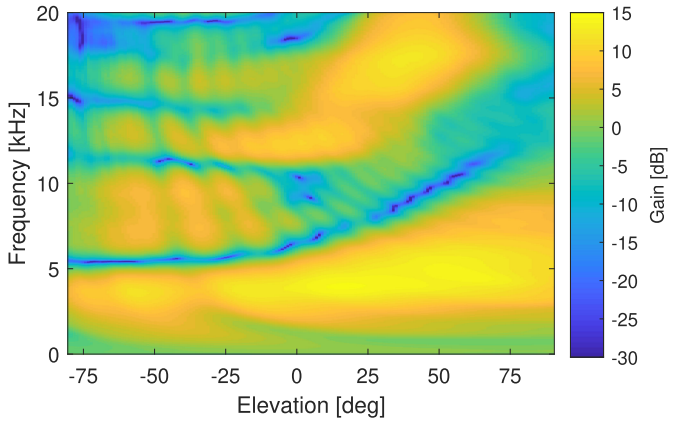


Fig. 6. Amplitude spectra of HRTF set A. The influence of the pinna (as elevation-dependent peaks and notches) and torso (as striations) can be observed.

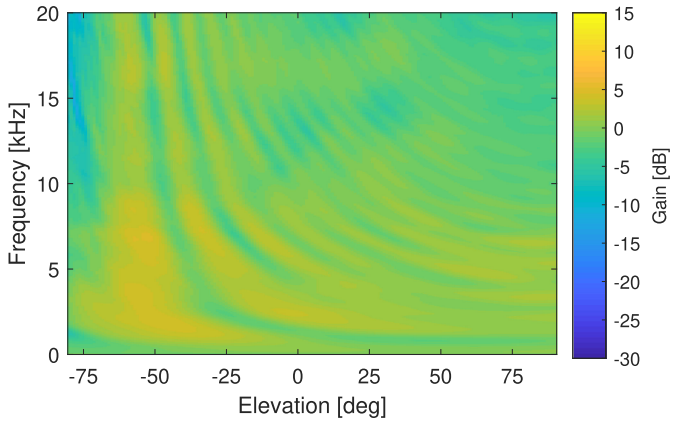


Fig. 7. Amplitude spectra of HRTF set Z. Only the influence of the torso can be observed.

done by removing the KEMAR, replacing it with a thin wooden pole, and mounting the left KEMAR microphone onto it, so that its position would be roughly at the center of the interaural axis with the head absent, and its orientation would closely match that of the mannequin's left channel. Reference measurements were taken with the same protocol as for the HRTF measurements except that a 5° elevation step was used. Their amplitude spectra are displayed in Fig. 5. Notice that fluctuations above 1 kHz closely follow the high-frequency response of the loudspeaker at 0° incidence,[4] meaning that effects from other components are negligible. However, the log-average curve should be preferred over the single measurements because of small high-frequency deviations possibly introduced by the wooden support.

IV. HRTF Feature Extraction

In order to recover Head-Related Impulse Responses (HRIRs) from sweep responses in accordance with the sweep method [48] and the free-field compensation method [49], a post-processing script was written in MATLAB. This means that each recorded signal $y[n]$ was filtered with (1) the inverse reference spectrum of the excitation signal $x[n]$, low-passed and high-passed with second-order digital Butterworth filters to compensate for the original zero sound pressure level below 20 Hz and above 20 kHz, and (2) the inverse complex log-average free-field response, where the phase spectrum of the free-field response was the minimum phase corresponding to the amplitude spectrum. Before step (2), a 128-sample falling half-Hann window was applied to each HRIR starting from its onset with the aim of removing the possible impact of reflections occurring far away from the mannequin.

As an example, Fig. 6 displays the HRTFs collected in set A calculated with a DFT size of 2048 points. Well-known spectral effects can be recognized here, such as the shoulder reflection ridge between 1 and 2 kHz followed by striations at higher frequencies [50], the omnidirectional peak around $3-4$ kHz [34], and the elevation-dependent pattern of other peaks and notches [1]. Predictably, a higher number of notches appear at lower elevations [51]. These features can be clearly identified in the other HRTF sets too, except for the pinna-less HRTFs in set Z where the main visible effect is obviously related to torso reflections acting as comb filters, see Fig. 7.

In order to extract the relevant spectral cues relative to the pinna, it is desirable to remove from the HRTF any other effect that is not caused by the interaction between the sound and the pinna itself. Isolating the response of the pinna alone (PRTF) can be done by using the pinna-less responses. Working under the assumption that the effects due to different anatomical components (head, pinna, and torso) on the HRTF are additive [9], the amplitude response of the PRTF for a given HRTF set i and elevation ϕ_k can be obtained by spectral division of the HRTF by the corresponding pinna-less response [29]:

$$|PRTF_i(f,\phi_k)| = \frac{|HRTF_i(f,\phi_k)|}{|HRTF_Z(f,\phi_k)|}. \quad (1)$$

Although the above assumption does not completely hold true, we found that good results are obtained when the spectral division is preceded by a further windowing of both the HRIR and pinna-less HRIR using a 1-ms falling half-Hann window [52]. This operation again does not fully suppress torso effects, especially at lower elevations where torso reflections occur within a few tenths of ms from the HRIR onset [53]

[4]See [Online]. Available: https://www.genelec.com/previous-models/8020c.

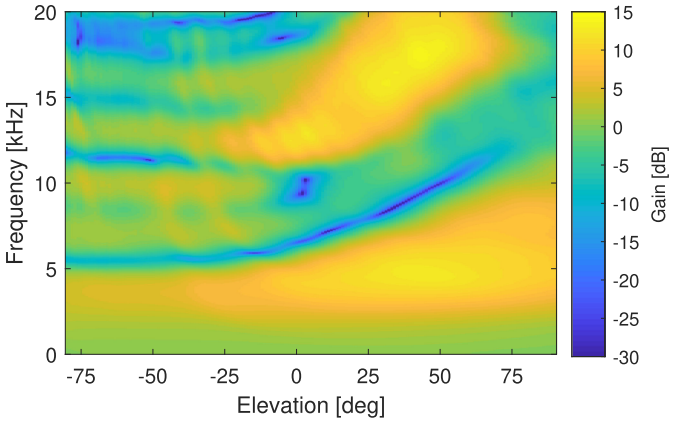


Fig. 8. Amplitude spectra of PRTF set A, obtained by spectral division of the windowed set-A HRTFs by the windowed pinna-less HRTFs.

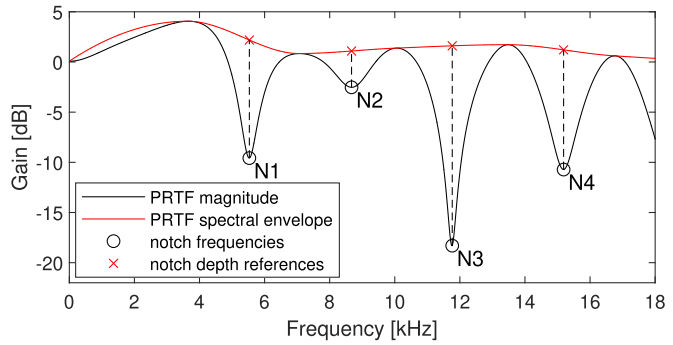


Fig. 9. PRTF for set A (elevation $= -80°$) and its estimated spectral envelope. Notch frequencies are selected as the PRTF local minima; notch depths are calculated as the differences between the magnitudes of the two curves at the notch frequencies (dashed lines).

and tend to overlap with pinna reflections. The result for set A is reported in Fig. 8. It can be noticed that nearly all torso reflections below 10 kHz have been removed and notch tracks appear considerably smoother than in the corresponding Fig. 6 HRTFs.

Center frequencies of pinna notches can now be directly extracted from each PRTF as the locations of the local minima, i.e., the samples that are smaller than their two neighboring samples in the frequency range between 4 and 16 kHz where the spectral cues generated by the directional filtering of the pinna lie [20]. Selection of the local minima was performed through an inverted basic peak picking algorithm. The depth value of the extracted notches was calculated as the difference between the values (in dB) of the local minima and of the PRTF spectral envelope at the same frequency. The spectral envelope was obtained by piecewise cubic interpolation of the spectral peaks [54], extracted through the same simple peak picking algorithm. Fig. 9 shows an example of notch extraction on a representative PRTF.

By plotting the extracted notch frequencies for each PRTF set in the elevation-frequency plane, one can easily visualize the evolution of the main notches along the elevation angle, and label them starting from lower elevations by increasing order of center frequency. Fig. 10 again shows the case of set A, where

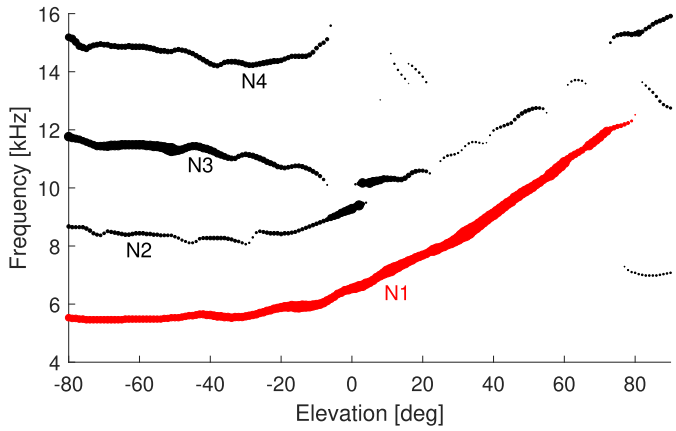


Fig. 10. Scatterplot of pinna notch frequencies for set A. The size of each point is proportional to the depth of the corresponding notch. The first notch, N1, is highlighted.

four main notches starting at the lowest elevation angle can be identified (and labeled N1 to N4). Notice that in this case N2 and N3 meet at about 10 kHz around $0°$ elevation, forming a single notch track from that elevation onwards. In general, and in accordance with previous literature, for each PRTF set two to four main notches are found at lower elevations and these become progressively shallower towards higher elevations.

V. The Computational Model

Spectral pinna notches have long been assumed to be caused by the destructive interference resulting from the sum of an incident wave and its reflected, time-delayed versions reaching the ear canal [25], [52], [55]. A single reflected component combines with the incident component after a time delay τ, such that

$$\tau = \frac{d}{c} \tag{2}$$

where d is the path difference traveled by the reflected wave with respect to the incident wave and c is the speed of sound.

In previous work [13] we assumed that a single reflection point corresponding to each pinna notch relates to one of five main pinna surfaces approximated as 2D contours traced on the helix, concha, and crus helias. No previous assumption was made on the sign of the reflection coefficient. The results of that study, calculated on the CIPIC HRTF database [56], showed that N1 is likely related to a negative reflection from the helix, N2 to a negative reflection from the concha inner wall, and N3 to a negative reflection from the concha border. The study also highlighted how the results for N1 are in agreement with previous studies on its generation mechanisms and on the pinna structures related to it [57]–[59], as well as offering a few speculations on how a negative reflection coefficient could be produced [13], [60].

If we assume a negative reflection coefficient, then destructive interference occurs whenever the reflected wave is delayed by a multiple of a full wavelength, or

$$d = n\lambda, \qquad n = 1, 2, \ldots, \tag{3}$$

and therefore spectral notches are produced at frequencies

$$f_n = \frac{c}{\lambda} = \frac{cn}{d}, \qquad n = 1, 2, \ldots \tag{4}$$

Similarly to torso reflections, each pinna reflection should translate into a comb filter-like effect in the measured signal spectrum, i.e., a series of periodic notches. However, experimental PRTFs do not typically show periodically related notches—see, for instance, again Fig. 8. Therefore, as in previous works [13], [52], it is assumed that a single reflection path gives rise to a single notch at frequency

$$f_1 = \frac{c}{d}. \tag{5}$$

Note that the analysis presented in [13] does not prove that negative reflections are actual physical phenomena that occur in the pinna. In order to reflect such a fact and to allow for some degree of approximation in the geometrical analysis that follows, we refer to the model we will now develop from Eq. (5) through basic ray tracing as a simple computational model, rather than a reflection model based on physical principles.

Given a 3D pinna mesh, we can find all its points which directly reflect a ray towards the ear canal entrance. Under the assumption that the sound propagates from the source as a planar wave, this is done through an algorithm that sequentially calculates, for every considered elevation angle ϕ,

1) the vectors from a 1-m far source at elevation ϕ to each vertex with positive z-coordinate (incoming rays);
2) the vectors from each vertex with positive z-coordinate to the mesh origin (reflected rays);
3) the vertex normals of the mesh, each defined as the normalized unweighted average of the surface normals of the faces that contain that vertex;
4) the angles between vertex normals and incoming rays;
5) the angles between vertex normals and reflected rays.

Our criterion for selecting the vertices of interest can now be formulated in terms of the above vectors and angles. Specifically, we select the vertices

1) whose normals subtend an angle smaller than a threshold $\theta_{\max}$ with both the incoming and reflected rays; and
2) whose reflected rays do not cross any mesh face before joining the mesh origin.

In simpler words, we are considering only those vertices that are "facing" both the sound source and the ear canal entrance and that are "visible" from the ear canal entrance itself. As an example, Fig. 11 shows the 828 selected vertices for head A's left pinna mesh when $\phi = -45°$.

Using Eq. (5) we can predict the frequencies at which destructive interference occurs. We set $c = 343$ m/s and, for each selected vertex v, the path difference is calculated as

$$d = \|\overrightarrow{sv}\| + \|\overrightarrow{vo}\| - 1, \tag{6}$$

where $\overrightarrow{sv}$ is the incoming ray, $\overrightarrow{vo}$ is the reflected ray, and the subtraction accounts for the path traveled by the incident wave. Notice that, because of the small protuberances around the ear canal entrance (tragus and antitragus), it is not guaranteed that a direct path exists from the source to the origin which does not cross any mesh face. However, for the sake of simplicity, the relative path length is kept constant at 1 m.

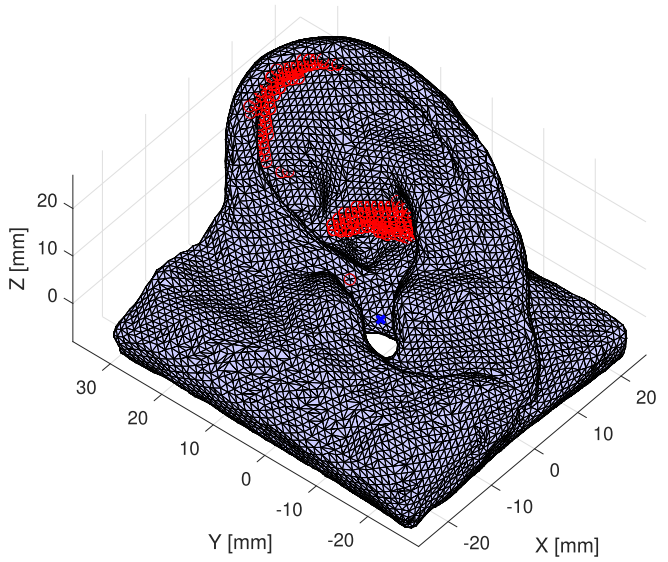


Fig. 11. 3D mesh of head A's left pinna and selected vertices for $\phi = -45°$ (highlighted in red). The origin point (0,0,0) is represented by the blue cross. Note that the origin appears off-center because of the angle of view; the z-axis approximately passes through the center of the microphone hole.

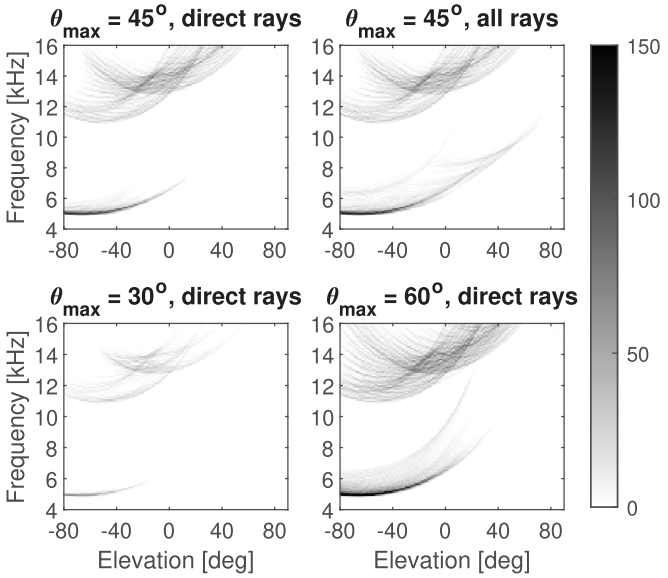


Fig. 12. Distribution of predicted notch frequencies according to the computational model applied to head A's left pinna. The first peak, starting at about $f_1 = 5.1$ kHz, is hypothesized to represent the center frequency of N1. Four different vertex selection conditions are compared, depending on the choice of $\theta_{\max}$ and the inclusion or exclusion of shadowed rays.

For each elevation ϕ we construct a histogram by distributing the predicted frequencies into linearly spaced frequency bins. It is now possible to visualize the predicted notch frequencies as a sequence of histograms, one per elevation. Fig. 12 reports again the case of head A, with bin spacing $\Delta f = 100$ Hz. In general, we identify two distinct clusters of notch frequencies evolving along the elevation angle: a concentrated cluster due to selected

vertices on the helix border and cavity, and a more dispersed one due to selected vertices on the concha wall. From the plots we can notice that the choice of $\theta_{\max} = 45°$ generally results in a good trade-off between the concentration and length along the elevation angle of the predicted notch patterns, especially for the first cluster. Fig. 12 also shows the case where selection condition (2) above is relaxed to consider all reflected rays, including those crossing a mesh face before joining the origin. Although this case results in longer tracks, additional data points hardly corresponding to actual pinna notches are introduced around the first cluster.

We decided, therefore, to set $\theta_{\max} = 45°$ and to keep considering condition (2) as originally intended in the remainder of this work. We hypothesize that the peak value of the first cluster, representing a large fraction of rays sharing the same path difference, corresponds to the center frequency of N1. Therefore, after selecting the lowest-frequency cluster we calculate the maximum bin count per histogram (i.e., elevation) within this selection, where available. The corresponding bin center represents the predicted N1 frequency.

VI. Results

A. Robustness and Accuracy of HRTF Measurements

Thanks to the availability of reference HRTF sets X and Y (see Section III-B) measured on the original KEMAR left pinnae in two different variants, and to the constant presence of the same right pinna in all measurements, it is possible to evaluate the robustness of our median-plane measurements as well as their fidelity to measurements taken from a previous dataset.

In particular, we calculated the mean spectral distortion between each pair of HRTF sets a and b as [61]

$$SD(a^{l|r}, b^{l|r}) = \frac{1}{N_\phi} \sum_i \sqrt{\frac{1}{N_f} \sum_j \left(20\log_{10} \frac{\left| H_a^{l|r}(\phi_i, f_j) \right|}{\left| H_b^{l|r}(\phi_i, f_j) \right|} \right)^2}, \tag{7}$$

where $l|r$ represents either the left or right channel, ϕ_i is an available elevation angle, f_j is an available frequency bin, N_ϕ is the total number of different elevation angles (in our case $N_\phi = 171$), and N_f is the number of frequency bins in a given frequency range. In order to best capture differences in spectral pinna cues, we consider the frequency range between 4 and 16 kHz as before.

The results of this analysis show that while the right channel is affected by measurement noise only (mean $SD(a^r, b^r) = 0.18$ dB, std $= 0.09$ dB), the mean spectral distortion between each pair of our 20 custom-made artificial pinnae installed on the left channel, shown in Fig. 13, is generally high (mean $SD(a^l, b^l) = 6.13$ dB, std $= 1.21$ dB), reflecting anthropometric differences. By contrast, the KEMAR pinna replica (set T) scores a lower mean distortion with respect to the reference sets X ($SD(\mathrm{T}^l, \mathrm{X}^l) = 1.93$ dB) and Y ($SD(\mathrm{T}^l, \mathrm{Y}^l) = 2.49$ dB). Interestingly, these values are both lower than the inter-channel mean spectral distortion of set X, which featured two original (albeit slightly different) KEMAR pinnae of the same material:

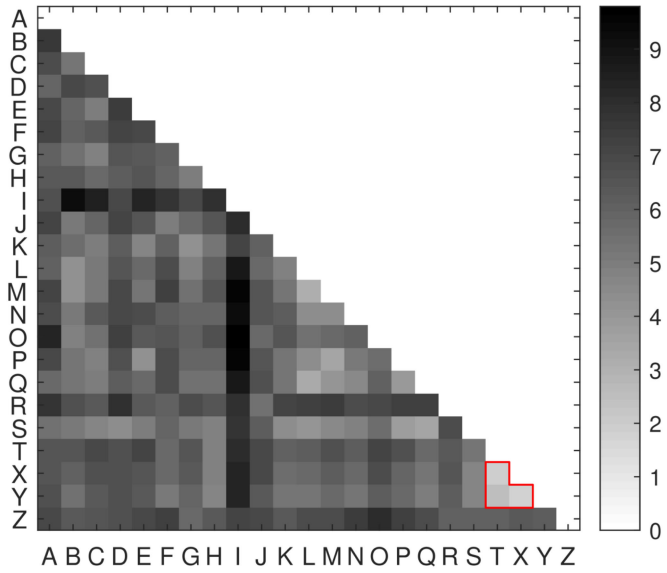


Fig. 13. Mean spectral distortion [dB] between each pair of HRTF sets, left channels only. Due to symmetry, values above the diagonal line are not repeated. The low spectral distortion between HRTF sets T, X and Y is highlighted.

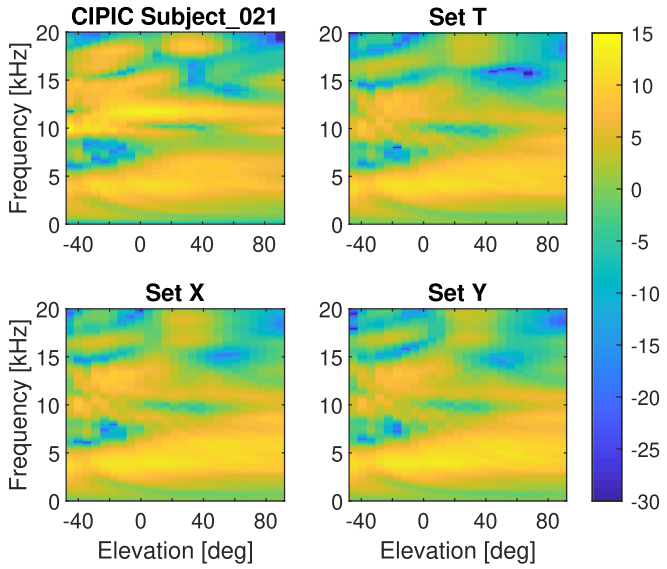


Fig. 14. Amplitude spectra [dB] of HRTF sets T, X, Y, and CIPIC subject 021, left channel. Because of the different resolution in elevation (5.625° in the CIPIC database, 1° in our measurements), only the HRTFs for the closest elevation values to each available CIPIC measurement in the frontal median plane are shown.

$SD(\mathrm{X}^l, \mathrm{X}^r) = 3.91$ dB. This result suggests that although differences in hardness may influence the spectral similarity of HRTFs, even the slightest difference in pinna shape appears more influential.

Finally, in Fig. 14 we compare the three HRTF sets measured with the different variants of the KEMAR left pinna (sets T, X, and Y) to previous measurements from the CIPIC HRTF database [56], where Subject 021 is the KEMAR with large standardized pinnae. We can qualitatively notice a close agreement among the four sets especially up to about 10 kHz, above

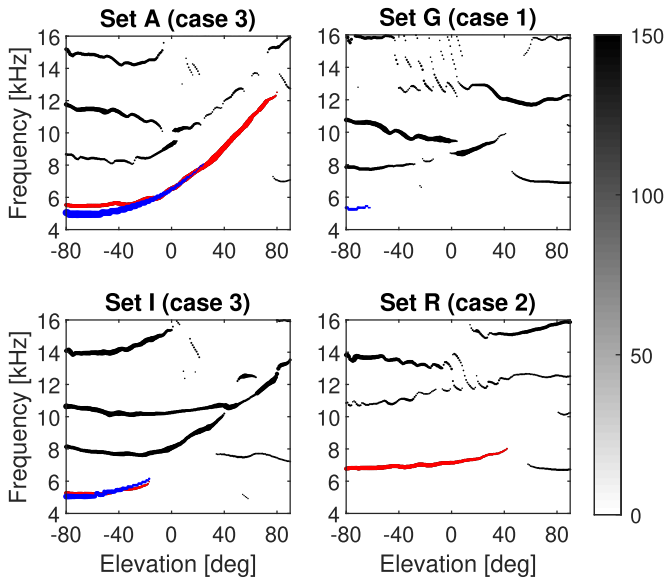


Fig. 15. Distribution of predicted notch frequencies according to the simple computational model and extracted notch frequencies for four different sets. The extracted and predicted N1 are marked with red and blue points, respectively. The size of each point is proportional to notch depth and maximum bin count, respectively.

which some common pinna cues can still be recognized (e.g., the 15-kHz notch). If we take into account that the external shape of the GRAS KB5001 pinna is identical to that of the standardized KEMAR pinna except for the concha and canal that have been recently modified to closely mimic the properties of a real human ear,[5] overall this result highlights the accuracy of our reference KEMAR measurements.

B. Performance of the Computational Model

The simple computational model presented in Section V outputs one sequence of histograms for each of the 20 considered left pinnae, allowing us to compare the predicted notch frequencies against those extracted from the corresponding HRTFs with the algorithm outlined in Section IV. Fig. 15 reports such a comparison for four representative pinnae/HRTF sets. In general, while the majority of our sets do not present any clear relationship between higher-order notches (N2, N3, ...) and modeled concha contributions, there appears to be a substantial overlap between N1 and the modeled helix contributions.

However, the latter observation does not hold true for all our pinnae. We can identify the following three cases:

Case 1: Sets G, H, M, O. In these HRTF sets, the first available notch falls above 7 kHz even in the lower elevation range, suggesting that the extracted notch might not be N1. As a matter of fact, the predicted N1 frequency is much lower and is generally associated to a faint cluster of points.

Case 2: Sets C, P, R. While the extracted N1 falls within a plausible frequency range (6 to 7 kHz at lower elevations), there is no corresponding helix cluster. This is because there is no straight path from the helix points to the mesh origin.

[5][Online]. Available: https://www.gras.dk/products/product/730-kb5001

TABLE I
NOTCH FREQUENCY MEAN ABSOLUTE/SIGNED ERROR (IN HZ), NOTCH FREQUENCY MISMATCH, AND PEARSON'S CORRELATION COEFFICIENT OF NOTCH FREQUENCIES COMPUTED BY COMPARING THE MEASURED N1 FREQUENCIES FOR CASE 3 WITH THE CORRESPONDING FREQUENCIES PREDICTED USING THE COMPUTATIONAL MODEL

	Notch frequency			
x	$MAE(x)$	$MSE(x)$	$m(x)$	$r_f(x)$
A	264.52	-242.02	4.71%	0.99
B	125.21	36.35	1.82%	0.86
D	122.99	-21.02	2.06%	0.94
E	532.71	-454.59	9.60%	0.98
F	264.03	156.40	3.30%	0.99
I	136.67	-0.98	2.53%	0.96
J	378.05	31.08	5.20%	0.97
K	291.90	-291.90	5.28%	0.85
L	91.54	-0.86	1.52%	0.89
N	352.77	302.34	5.18%	0.98
Q	119.76	56.36	1.93%	0.85
S	95.10	-47.01	1.50%	0.93
T	297.76	-290.98	4.78%	0.43
median	264.03	-0.98	3.30%	0.94

Case 3: Sets A, B, D, E, F, I, J, K, L, N, Q, S, T. For all these pinnae/HRTF sets, the extracted and predicted N1 frequencies overlap. In the remainder of this Section, we focus on this case.

For a given pinna/HRTF set x, having defined $\Phi(x)$ as the set of elevations where both an extracted and predicted N1 are present, the error metrics we present are the following:

1) the *mean absolute error* between predicted ($\hat{f}_x$) and extracted (f_x) N1 frequencies,

$$MAE(x) = \frac{1}{|\Phi|}\sum_{i\in\Phi}\left|\hat{f}_x(\phi_i) - f_x(\phi_i)\right|; \qquad (8)$$

2) the *mean signed error* between predicted and extracted N1 frequencies,

$$MSE(x) = \frac{1}{|\Phi|}\sum_{i\in\Phi}\hat{f}_x(\phi_i) - f_x(\phi_i); \qquad (9)$$

3) the *notch frequency mismatch* [13] between predicted and extracted N1 frequencies, i.e., the average percentual ratio between the absolute error and the extracted frequency value,

$$m(x) = \frac{1}{|\Phi|}\sum_{i\in\Phi}\frac{\left|\hat{f}_x(\phi_i) - f_x(\phi_i)\right|}{f_x(\phi_i)}\cdot 100\%; \qquad (10)$$

4) the sample Pearson correlation coefficient $r_f(x)$ between pairs of predicted and extracted N1 frequencies.

These four metrics are used for testing both the accuracy and elevation trend of the predicted N1 frequencies. In particular, the notch frequency mismatch metric enables comparison with the findings by Moore *et al.* [27] where two steady notches in the high-frequency range (in that case around 8 kHz) differing just in central frequency are not distinguishable on average if the mismatch is less than approximately 9%, regardless of their bandwidth.

Table I reports complete error metrics for the 13 sets. Generally, there is close agreement between extracted and predicted N1 for the common elevations as shown by the MAE metric as well as the sample Pearson correlation coefficient. In particular,

not only the absolute frequency values are similar, but also the elevation trend of extracted and predicted N1 frequencies, with an initial plateau at lower elevations followed by a frequency rise. Remarkably, in all cases except one (set E), the notch frequency mismatch is well below the above mentioned 9% threshold. In addition, we observe no particular estimation bias from the MSE metric, with a comparable number of cases either overestimating or underestimating notch frequency on average.

VII. Discussion

The results reported in Section VI-A suggest that the collected acoustical measurements are replicable, robust, and faithful to reference KEMAR HRTFs, providing a level of accuracy for investigating the relation between HRTFs and pinna anthropometry previously not possible. In turn, this gives substance to the results of Section VI-B, which report a general agreement between the ground-truth N1 data and the predictions of our simple computational model. Conversely, the same model does not support any clear relationship between anthropometry and higher-order pinna notches.

It must be acknowledged that there exist at least two sources of error that could have slightly undermined the accuracy of our input data. The first one is related to the 3D pinna meshes. As reported in Section III-A, the origin of each pinna mesh has been selected manually prior to calculation of the reflected rays and, in particular, its placement along the z-axis has been solely based on 2D data from pictures taken during the HRTF measurement sessions. Furthermore, one variable not taken into account here is the possible deformation of the pinna replica once inserted into the KEMAR slot due to slightly off-center microphone holes. Even so, the origin placement procedure has been based on observations of the pinna replica alone without using the HRTF in any way. This represents a more unbiased estimate with respect to [13], where the assumed microphone location had been optimized by minimization of the error between extracted and predicted HRTF notches for every subject.

The second possible error source comes from the entirely automatic notch extraction procedure described in Section IV. This was preferred over the classic signal processing algorithm by Raykar *et al.* [52] as the latter outputs a number of notches that heavily depends on a user-defined threshold. When such a threshold is relaxed, additional notches appear, which were not visible in the original HRTF amplitude spectra. On the other hand, our algorithm extracts a certain number of PRTF notches that do not depend on any user-defined parameter, and this often results in shorter N1 tracks than observed in previous literature. However, we believe that a conservative yet objective estimate of notch frequencies represents the best possible scenario for a solid data analysis.

For 13 HRTF sets out of 20, we found a clear correspondence between our model predictions and ground-truth N1 frequencies. The notch frequency mismatch metric scores a median value of 3.3% across the 13 sets, which is less than half the median mismatch value previously found in [13], reported at 7.4% out of 17 sets. While the two studies refer to two different HRTF datasets with different representations of individual anthropometric data, this result endorses the extension of our model to the 3D case. The improved representation of pinna structures through 3D data and the robustness of our HRTF measurements are to be seen as additional factors for the said improvement. Furthermore, the average MAE in this study is considerably lower than that found in preliminary works that used linear regression models to predict N1 from anthropometric parameters [44] or depth maps of pinnae [62]. This suggests that the computational model is able to extract strong features from 3D data.

On the other hand, the computational model fails for 7 other HRTF sets. In 4 of them (case 1 in Section VI-B) there is strong evidence that N1 cannot be identified in the corresponding HRTFs because it is not generated at all. The issue of missing notches was also seen in a previous study on the CIPIC and ARI databases [63], where it was hypothesized as the cause of poor vertical localization performances. For 3 of these 4 HRTF sets the computational model actually predicts a notch track. This is, however, generally lower in frequency and overlaps with the omnidirectional 4 kHz resonance, which might explain the lack of prominent notches.

For the remaining 3 HRTF sets (case 2 in Section VI-B), N1 appears in the HRTF but not in the model prediction because there is no straight path from the helix to the mesh origin. For instance, in the specific case of set R, a considerably deep N1 track seen in the HRTF (see Fig. 15) cannot be associated to any ray-traced path in the corresponding pinna. It is therefore plausible to hypothesize that in these few cases the generation mechanism for N1 does not correspond to a contribution from the helix. As previously pointed out by Mokhtari *et al.* [14], there exists no established method for correctly labeling transfer function peaks (or notches, in our case) depending on frequency and spatial location, and since different physical mechanisms might appear in different pinnae (such as the presence or absence of a cavum-fossa vertical mode) these notches labeled as N1 might share the same generation mechanism of a higher-order notch in a different HRTF.

As for higher-order notches, it has been confirmed that a simple model such as ours is not able to replicate their trend along the elevation angle. This is in line with our previous results [43], [44] that pointed out the inaccuracy of linear regression models in predicting higher-order notches from both global and elevation-dependent anthropometric parameters. Conversely, this seems to contradict the findings in [13] where a median N2 frequency mismatch value as low as 5.3% was found. Nevertheless, that finding might have been biased by the optimization of the assumed microphone location, as previously mentioned. In general, our computational model calculates a discrete number of paths with very different path lengths associated to the concha wall for the same direction, while in practice, as pointed out by Lopez-Poveda and Meddis [21], significant reflections occur on an infinite number of points along the posterior concha wall for all source locations. In this regard, a more detailed physical model of the concha might be necessary to explain how these notches are produced and how they can be adequately approximated.

VIII. Conclusion

The simple computational model proposed in this paper helped us gain better understanding of the mechanisms that

generate spectral notches in experimental HRTFs. The role of the helix in producing N1, previously hypothesized in [13], is hereby confirmed from notch frequency predictions that are extremely close to experimental data. Conversely, higher-order notches—most likely associated to interactions within the concha—would require the development of an alternative model of the concha including diffraction and reflection effects to be fully interpreted.

In its current form, the computational model is able to predict the evolution of N1 along the median plane from individual 3D pinna morphology. While it is not sufficient to infer the individual HRTF, the model can help selecting the closest non-individual HRTF in a dataset based on minimization of the error between non-individual and predicted N1. As a matter of fact, it has been previously found that N1-based HRTF selection improves localization performance with respect to generic HRTFs [63], [64]. Assessing whether this also applies to our computational model through individual localization tests is currently planned as future work.

We believe that the HRTF dataset presented in this paper can be useful for studying the effect of the pinna on spatial sound perception. Although the amount of time and resources available only allowed us to produce and test a limited number of artificial pinna samples, the casting procedure presented herein can be applied to any artificial head, and preliminary investigations suggest the possibility of extending it to human heads. It is also worth remarking that artificial pinnae can be easily replicated from the available negative molds, allowing modifications to pinna structures (e.g., removing the helix or filling cavities) for the sake of further understanding the physical mechanisms generating spectral cues. Ultimately, a broader sample size would grant the application of more advanced machine learning techniques for extracting the relevant anthropometric parameters from 3D pinna models and relating them to HRTF features.

REFERENCES

[1] 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.*, vol. 49, no. 4, pp. 231–249, Apr. 2001.

[2] S. Li and J. Peissig, "Measurement of head-related transfer functions: A review," *Appl. Sci.*, vol. 2020, no. 10, Jul. 2020.

[3] J. He, R. Ranjan, and W. S. Gan, "Fast continuous HRTF acquisition with unconstrained movements of human subjects," in *Proc. 41st IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2016, pp. 321–325.

[4] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hammershøi, "Binaural technique: Do we need individual recordings?," *J. Audio Eng. Soc.*, vol. 44, no. 6, pp. 451–469, Jun. 1996.

[5] 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.*, vol. 49, no. 10, pp. 904–916, Oct. 2001.

[6] C. Guezenoc and R. Seguier, "HRTF individualization: A survey," in *Proc. 145th Conv. Audio Eng. Soc.*, New York, NY, USA, 2018, Art. no. 10129.

[7] B. F. G. Katz, "Boundary element method calculation of individual head-related transfer function. I. rigid model calculation," *J. Acoust. Soc. Amer.*, vol. 110, no. 5, pp. 2440–2448, Nov. 2001.

[8] F. Brinkmann, M. Dinakaran, R. Pelzer, P. Grosche, D. Voss, and S. Weinzierl, "A cross-evaluated database of measured and simulated HRTFs including 3D head meshes, anthropometric features, and headphone impulse responses," *J. Audio Eng. Soc.*, vol. 67, no. 9, pp. 705–718, Sep. 2019.

[9] C. P. Brown and R. O. Duda, "A structural model for binaural sound synthesis," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 5, pp. 476–488, Sep. 1998.

[10] X.-Y. Zeng, S.-G. Wang, and L.-P. Gao, "A hybrid algorithm for selecting head-related transfer function based on similarity of anthropometric structures," *J. Sound Vib.*, vol. 329, no. 19, pp. 4093–4106, Sep. 2010.

[11] T.-Y. Chen, T.-H. Kuo, and T.-S. Chi, "Autoencoding HRTFs for DNN based HRTF personalization using anthropometric features," in *Proc. 44th IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2019, pp. 271–275.

[12] 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," *J. Acoust. Soc. Amer.*, vol. 132, no. 6, pp. 3832–3841, Dec. 2012.

[13] 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.

[14] P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Vertical normal modes of human ears: Individual variation and frequency estimation from pinna anthropometry," *J. Acoust. Soc. Amer.*, vol. 140, no. 2, pp. 814–831, Aug. 2016.

[15] A. Andreopoulou, D. R. Begault, and B. F. G. Katz, "Inter-laboratory round robin HRTF measurement comparison," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 5, pp. 895–906, Aug. 2015.

[16] C. J. Chun, J. M. Moon, G. W. Lee, N. K. Kim, and H. K. Kim, "Deep neural network based HRTF personalization using anthropometric measurements," in *Proc. 143rd Conv. Audio Eng. Soc.*, New York, NY, USA, 2017, Art. no. 9860.

[17] R. Miccini and S. Spagnol, "HRTF individualization using deep learning," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces Workshops*, 2020, pp. 390–395.

[18] R. Miccini and S. Spagnol, "A hybrid approach to structural modeling of individualized HRTFs," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces Workshops*, 2021, pp. 80–85.

[19] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, 2nd ed., Cambridge, MA, USA: MIT Press, Oct. 1996.

[20] J. Hebrank and D. Wright, "Spectral cues used in the localization of sound sources on the median plane," *J. Acoust. Soc. Amer.*, vol. 56, no. 6, pp. 1829–1834, Dec. 1974.

[21] E. A. Lopez-Poveda and R. Meddis, "A physical model of sound diffraction and reflections in the human concha," *J. Acoust. Soc. Amer.*, vol. 100, no. 5, pp. 3248–3259, Nov. 1996.

[22] E. A. G. Shaw, "Acoustical features of human ear," in *Binaural Spatial Hearing Real Virtual Environments*. Mahwah, NJ, USA: R. H. Gilkey and T. R. Anderson, Lawrence Erlbaum Associates, 1997, pp. 25–47.

[23] Y. Kahana and P. A. Nelson, "Boundary element simulations of the transfer function of human heads and baffled pinnae using accurate geometric models," *J. Sound Vib.*, vol. 300, no. 3-5, pp. 552–579, 2007.

[24] P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Frequency and amplitude estimation of the first peak of head-related transfer functions from individual pinna anthropometry," *J. Acoust. Soc. Amer.*, vol. 137, no. 2, pp. 690–701, Feb. 2015.

[25] D. W. Batteau, "The role of the pinna in human localization," *Proc. R. Soc. London. Ser. B., Biol. Sci.*, vol. 168, no. 11, pp. 158–180, Aug. 1967.

[26] P. J. Bloom, "Creating source elevation illusions by spectral manipulation," *J. Audio Eng. Soc.*, vol. 25, no. 9, pp. 560–565, Sep. 1977.

[27] B. C. J. Moore, S. R. Oldfield, and G. J. Dooley, "Detection and discrimination of spectral peaks and notches at 1 and 8 kHz," *J. Acoust. Soc. Amer.*, vol. 85, no. 2, pp. 820–836, Feb. 1989.

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

[29] S. Spagnol, "On distance dependence of pinna spectral patterns in head-related transfer functions," *J. Acoust. Soc. Amer.*, vol. 137, no. 1, pp. EL58–EL64, Jan. 2015.

[30] S. Carlile and D. Pralong, "The location-dependent nature of perceptually salient features of the human head-related transfer functions," *J. Acoust. Soc. Amer.*, vol. 95, no. 6, pp. 3445–3459, Jun. 1994.

[31] K. Iida, M. Itoh, A. Itagaki, and M. Morimoto, "Median plane localization using a parametric model of the head-related transfer function based on spectral cues," *Appl. Acoust.*, vol. 68, no. 8, pp. 835–850, Aug. 2007.

[32] K. Iida and Y. Ishii, "Effects of adding a spectral peak generated by the second pinna resonance to a parametric model of head-related transfer functions on upper median plane sound localization," *Appl. Acoust.*, vol. 129, pp. 239–247, Jan. 2018.

[33] A. J. Watkins, "Psychoacoustical aspects of synthesized vertical locale cues," *J. Acoust. Soc. Amer.*, vol. 63, no. 4, pp. 1152–1165, Apr. 1978.

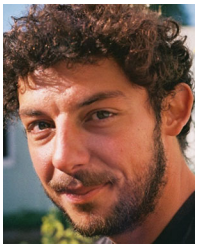
[34] E. A. G. Shaw, "The acoustics of the external ear," in *Acoustical Factors Affecting Hearing Aid Performance*. Baltimore, MD, USA: G. A. Studebaker and I. Hochberg, Univ. Park Press, 1980, pp. 109–125.
[35] K. Iida, H. Shimazaki, and M. Oota, "Generation of the amplitude spectra of the individual head-related transfer functions in the upper median plane based on the anthropometry of the listener's pinnae," *Appl. Acoust.*, vol. 155, pp. 280–285, Dec. 2019.
[36] 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. Amer.*, vol. 91, no. 3, pp. 1637–1647, Mar. 1992.
[37] J. Chen, B. D. Van Veen, and K. E. Hecox, "A spatial feature extraction and regularization model for the head-related transfer function," *J. Acoust. Soc. Amer.*, vol. 97, no. 1, pp. 439–452, Jan. 1995.
[38] M. J. Evans, J. A. S. Angus, and A. I. Tew, "Analyzing head-related transfer function measurements using surface spherical harmonics," *J. Acoust. Soc. Amer.*, vol. 104, no. 4, pp. 2400–2411, Oct. 1998.
[39] H. Hu, L. Zhou, H. Ma, and Z. Wu, "HRTF personalization based on artificial neural network in individual virtual auditory space," *Appl. Acoust.*, vol. 69, no. 2, pp. 163–172, Feb. 2008.
[40] K. J. Faller II and K. P. Hoang, "Estimation of parameters of a head-related transfer function (HRTF) customization model," in *Proc. 164th Meetings Acoust.* Kansas City, MO, USA: Acoustical Society of America, 2012, Art. no. 030007.
[41] S. Spagnol, M. Geronazzo, D. Rocchesso, and F. Avanzini, "Synthetic individual binaural audio delivery by pinna image processing," *Int. J. Pervasive Comput. Commun.*, vol. 10, no. 3, pp. 239–254, Jul. 2014.
[42] G. W. Lee and H. K. Kim, "Personalized HRTF modeling based on deep neural network using anthropometric measurements and images of the ear," *Appl. Sci.*, vol. 8, no. 11, Nov. 2018, Art. no. 2180.
[43] S. Spagnol and F. Avanzini, "Frequency estimation of the first pinna notch in head-related transfer functions with a linear anthropometric model," in *Proc. 18th Int. Conf. Digit. Audio Effects*, 2015, pp. 231–236.
[44] R. Miccini and S. Spagnol, "Estimation of pinna notch frequency from anthropometry: An improved linear model based on principal component analysis and feature selection," in *Proc. 1st Nordic Sound Music Comput. Conf.*, 2019, pp. 5–8.
[45] M. D. Burkhard and R. M. Sachs, "Anthropometric manikin for acoustic research," *J. Acoust. Soc. Amer.*, vol. 58, no. 1, pp. 214–222, Jul. 1975.
[46] S. Spagnol, K. B. Purkhús, S. K. Björnsson, and R. Unnthórsson, "The viking HRTF dataset," in *Proc. 16th Int. Conf. Sound Music Comput.*, 2019, pp. 55–60.
[47] S. Spagnol, R. Miccini, and R. Unnthórsson, "The viking HRTF dataset v2," *Zenodo*, Oct. 2020, doi: 10.5281/zenodo.4160401.
[48] S. Müller and P. Massarani, "Transfer-function measurement with sweeps," *J. Audio Eng. Soc.*, vol. 49, no. 6, pp. 443–471, Jun. 2001.
[49] H. Møller, "Fundamentals of binaural technology," *Appl. Acoust.*, vol. 36, no. 3/4, pp. 171–218, 1992.
[50] V. R. Algazi, R. O. Duda, R. Duraiswami, N. A. Gumerov, and Z. Tang, "Approximating the head-related transfer function using simple geometric models of the head and torso," *J. Acoust. Soc. Amer.*, vol. 112, no. 5, pp. 2053–2064, Nov. 2002.
[51] S. Spagnol, M. Hiipakka, and V. Pulkki, "A single-azimuth pinna-related transfer function database," in *Proc. 14th Int. Conf. Digit. Audio Effects*, 2011, pp. 209–212.
[52] 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. Amer.*, vol. 118, no. 1, pp. 364–374, Jul. 2005.
[53] V. R. Algazi, C. Avendano, and R. O. Duda, "Elevation localization and head-related transfer function analysis at low frequencies," *J. Acoust. Soc. Amer.*, vol. 109, no. 3, pp. 1110–1122, Mar. 2001.
[54] D. Schwarz and X. Rodet, "Spectral envelope estimation, representation, and morphing for sound analysis, transformation, and synthesis," in *Proc. Int. Comput. Music Conf.*, 1999, p. 1.
[55] J. Chen, B. D. Van Veen, and K. E. Hecox, "External ear transfer function modeling: A beamforming approach," *J. Acoust. Soc. Amer.*, vol. 92, no. 4, pp. 1933–1944, Oct. 1992.
[56] 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.*, 2001, pp. 1–4.
[57] H. Takemoto, P. Mokhtari, H. Kato, R. Nishimura, and K. Iida, "Pressure distribution patterns on the pinna at spectral peak and notch frequencies of head-related transfer functions in the median plane," in *Proc. Int. Work. Princ. Appl. Spatial Hearing*, 2009, pp. 179–194.
[58] P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Acoustic sensitivity to micro-perturbations of KEMAR's pinna surface geometry," in *Proc. 20th Int. Congr. Acoust.*, 2010, pp. 1–8.
[59] P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Pinna sensitivity patterns reveal reflecting and diffracting surfaces that generate the first spectral notch in the front median plane," in *Proc. 36th IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2011, pp. 2408–2411.
[60] P. Satarzadeh, R. V. Algazi, and R. O. Duda, "Physical and filter pinna models based on anthropometry," in *Proc. 122nd Conv. Audio Eng. Soc.*, 2007, pp. 718–737.
[61] S. Spagnol, E. Tavazzi, and F. Avanzini, "Distance rendering and perception of nearby virtual sound sources with a near-field filter model," *Appl. Acoust.*, vol. 115, pp. 61–73, Jan. 2017.
[62] M. G. Onofrei, R. Miccini, R. Unnthórsson, S. Serafin, and S. Spagnol, "3D ear shape as an estimator of HRTF notch frequency," in *Proc. 17th Int. Conf. Sound Music Comput.*, 2020, pp. 131–137.
[63] M. Geronazzo, S. Spagnol, and F. Avanzini, "Do we need individual head-related transfer functions for vertical localization? The case study of a spectral notch distance metric," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 7, pp. 1243–1256, Jul. 2018.
[64] M. Geronazzo, S. Spagnol, A. Bedin, and F. Avanzini, "Enhancing vertical localization with image-guided selection of non-individual head-related transfer functions," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2014, pp. 4496–4500.

Simone Spagnol (Senior Member, IEEE) received the Ph.D. degree in information engineering from the University of Padova, Italy, in 2012. He then was a Postdoctoral Researcher with the Iuav University of Venice, Venice, Italy, University of Padova, and University of Iceland, Reykjavik, Iceland, before being awarded a Marie Skłodowska-Curie fellowship at Aalborg University, Denmark. He is currently a Postdoctoral Researcher with Delft University of Technology, Delft, The Netherlands.

His scientific activity is mainly devoted to the management and representation of auditory information, and to the use of audio solutions in the health and care domain. He has authored more than 60 peer-reviewed publications, and he was the recipient of four best paper awards as first author. He has chaired the Scientific Committee of the 2020 and 2021 editions of the International Sound and Music Computing (SMC) Conference, and was a Guest Editor of *Wireless Communications* and *Mobile Computing* during 2017–2018. He has been a key Researcher and a Principal Investigator in several international and national research projects, including two Horizon 2020 EU projects.

Riccardo Miccini was born in Recanati, Italy, in 1993. He received the B.E. degree in electronics and computer engineering from the Technical University of Denmark, Kongens Lyngby, Denmark, in 2017 and the M.S. degree in sound and music computing from Aalborg University Copenhagen, Copenhagen, Denmark, in 2021. As part of his master's, he has been researching applications of deep neural networks for sound synthesis, speech enhancement, and HRTF individualization.

From 2016 to 2018, he was a Software Developer, while from 2019 to 2020, he was a Student Assistant for the Horizon 2020 IT'S A DIVE research project. During that time, he has authored or coauthored several articles on HRTF individualization for conferences and workshops. He is currently an Audio Engineer in Copenhagen, Denmark. His research interests include machine learning, sound synthesis, and data visualization.



Marius George Onofrei was born in Constanta, Romania, in 1988. He is currently a student in the sound and music computing M.S. programme with Aalborg University, Aalborg, Denmark. Here, he was involved in many projects in the fields of sound synthesis, machine learning for media technology or music signal analysis while finally settling on specializing in the field of physical modeling for sound synthesis.

He previously received the M.S. degree in structural engineering from Aalborg University, Aalborg, Denmark in 2013 and has had a successful career afterwards, working with the analysis of metocean data, with a focus on the likelihood of rogue waves, and working in the wind energy field in structural design of offshore wind-turbine foundations.

Runar Unnthorsson received the Ph.D. degree in mechanical engineering from the University of Iceland, Reykjavik, Iceland, in 2008. His work was on using acoustic emissions for monitoring carbon fiber reinforced polymers subjected to multi-axial fatigue loading. Since 2011, he has been with the faculty of industrial engineering, mechanical engineering, and computer science with the University of Iceland. He is currently a Professor with the faculty. His main research interests include performance engineering and the applications of acoustics or vibrations for sensory substitution, non-destructive evaluations, tactile or acoustic displays, and product design.

He coordinated the Horizon 2020 RIA project Sound of Vision (no. 643636) which was carried out in the years 2015–2017. The project was the recipient of the European Commission's 2018 Innovation Radar Prize in the category Tech for Society for the development of an assistive device for the visually impaired. In 2017, his ACUTE lab was awarded the second prize for its tactile display at the University of Iceland's Science and Innovation Awards. The ACUTE lab is currently working on the development of the tactile display with support from the Technology Development Fund (tths.is).

Stefania Serafin received the Ph.D. degree from Stanford University, Stanford, CA, USA, in 2004. She is currently a Professor of sonic interaction design with Aalborg University Copenhagen, Copenhagen, Denmark. She has previously an Assistant and Associate Professor with Aalborg University Copenhagen, Copenhagen, Denmark. She is the President of the Sound and Music Computing Association and the Project Leader of the Nordic Sound and Music Computing Network.