FREQUENCY ESTIMATION OF THE FIRST PINNA NOTCH IN HEAD-RELATED TRANSFER FUNCTIONS WITH A LINEAR ANTHROPOMETRIC MODEL

Simone Spagnol,
Faculty of Ind. Eng., Mech. Eng. and Computer Science,
University of Iceland
Reykjavík, Iceland
spagnols@hi.is

Federico Avanzini,
Department of Information Engineering,
University of Padova
Padova, Italy
avanzini@dei.unipd.it

ABSTRACT

The relation between anthropometric parameters and Head-Related Transfer Function (HRTF) features, especially those due to the pinna, are not fully understood yet. In this paper we apply signal processing techniques to extract the frequencies of the main pinna notches (known as N_1, N_2, and N_3) in the frontal part of the median plane and build a model relating them to 13 different anthropometric parameters of the pinna, some of which depend on the elevation angle of the sound source. Results show that while the considered anthropometric parameters are not able to approximate with sufficient accuracy neither the N_2 nor the N_3 frequency, eight of them are sufficient for modeling the frequency of N_1 within a psychoacoustically acceptable margin of error. In particular, distances between the ear canal and the outer helix border are the most important parameters for predicting N_1.

1. INTRODUCTION

Most of the current binaural sound rendering techniques rely on the use of Head-Related Transfer Functions (HRTFs), i.e. filters that capture the acoustic effects of the human head [1]. HRTFs allow loyal simulation of the signal that arrives at the entrance of the ear canal as a function of the sound source spatial position. The classic solution best approximating free-field listening conditions involves the use of individual HRTFs measured on the listener himself with the addition of head tracking and artificial reverberation [2]. However, obtaining personal HRTF data for a vast number of users is only possible with expensive equipment and invasive recording procedures [3]. This is the reason why non-individual HRTFs, acoustically measured on anthropomorphic mannequins, are often preferred in practice. The drawback with non-individual HRTFs is that these likely never match with the listener's unique anthropometry, and especially the outer ear, resulting in frequent localization errors such as front/back reversals, elevation angle misperception, and inside-the-head localization [4].

In order to efficiently face such issues, several techniques for synthetic HRTF design have been proposed during the last two decades. In the authors' opinion, the most attractive is represented by structural HRTF models [5]. According to this approach, the most important effects involved in spatial sound perception (acoustic delays and shadowing due to head diffraction, reflections on pinna contours and shoulders, and so on) are isolated and modeled separately with a corresponding filtering element. The advantages of such an approach over alternative binaural rendering techniques are twofold:

1. adaptability to a specific subject, based on anthropometric quantities (head radius, pinna shape, shoulder width, and so on);
2. computational efficiency, as models are structured in smaller blocks each simulating one physical effect, allowing low-cost implementation and low-latency reproduction on many devices.

However, previous studies on the relation between acoustic effects and anthropometry - including applications of anthropometric regression methods to measured HRTF data [6, 7, 8, 9] - have produced mixed results, highlighting that many of these relations are not fully understood yet.

We can identify two reasons why these studies fail at explaining such relations. First, they typically do not take into account prior knowledge of the structural components that are responsible for localization cues, blindly applying classical machine learning techniques to long anthropometric feature vectors including irrelevant parameters that only have the effect of increasing clutter. Second, the former studies consider the whole HRTF or a dimensionally reduced version of it (e.g. via Principal Component Analysis) as the set of target variables, without applying any substantial preprocessing step in order to extract local HRTF features. In particular, it is known that local minima (notches) and maxima (peaks) in the transfer function are salient for detecting the most "individual" dimension, i.e. the elevation of a sound source [10].

This paper's main objective is to explore the relationship between the center frequencies of the three main notches in a set of frontal median-plane HRTFs and anthropometric parameters under the form of global pinna measurements (e.g. pinna height, concha width) as well as measurements that vary with the elevation angle of the sound source (i.e. distances between the ear canal and pinna edges). Indeed, the starting point of this paper (briefly reported in Section 2) is a previous work [11] that highlights how in median-plane HRTFs frequencies of the most prominent spectral minima are tightly related to the shape of the subject's pinna, and in particular to the above cited distances. Section 3 describes the methods for feature extraction and the anthropometric regression model, whose results are reported and discussed in Section 4. Section 5 concludes the paper.

2. A STRUCTURAL PINNA MODEL

The most relevant differences between the HRTFs of two subjects are due to different pinna features (shape, size, and orientation). The pinna has a fundamental role in shaping HRTFs thanks to two main acoustic phenomena, i.e., reflections and resonances. Consequently, the HRTF shows a sequence of peaks centered around the

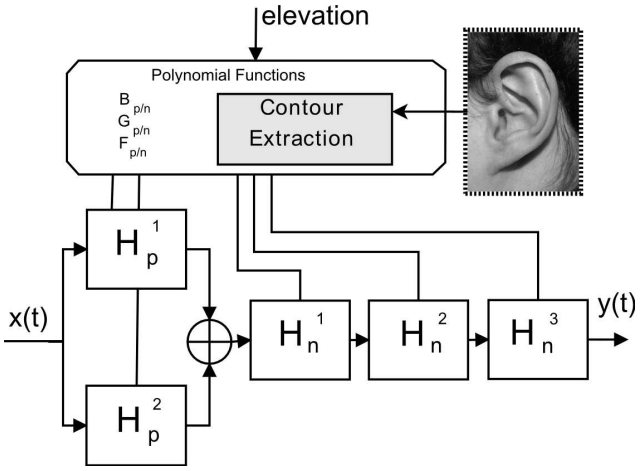

Figure 1: Schematic representation of the structural pinna model.

resonant frequencies and a sequence of notches located at all those frequencies where destructive interference between direct and reflected waves occurs. The spectral location of peaks and notches represents a pivotal cue to the characterization of the sound source's spatial position, in particular of its elevation [10].

As a matter of fact, Iida *et al.* [12] showed that a parametric HRTF recomposed using only the first peak in the HRTF spectrum coupled with the first two notches yields almost the same localization accuracy as the corresponding measured HRTF. Additional evidence in support of the lowest-frequency notches' relevance is given by Moore [13], who states that the threshold for perceiving a shift in the central frequency of a spectral notch is consistent with the localization blur (i.e., the angular threshold for detecting changes in the direction of a sound source) on the median plane. Also, Hebrank and Wright [14] judge increasing frontal elevation apparently cued by the increasing central frequency of a notch.

Previous literature suggests a number of solutions to synthetic modeling of the pinna-related component, known as Pinna-Related Transfer Function (PRTF). However, these models suffer from evident limits, e.g., the validity in an over restricted spatial region [15], and/or the absence of an explicit parametrization on the listener's anthropometry [16]. The authors propose a structural PRTF model composed of two filter blocks [11]. As Fig. 1 details, the first block (*resonant* block) includes two second-order peak filters placed in parallel, while the second block (*reflective* block) includes three second-order notch filters placed in series.

Similarly to previous works [17], the authors also studied the relation between notch frequencies in PRTFs and pinna geometry. A *ray-tracing* procedure on pinna images was exploited to map reflection points at a given distance from the reference ear-canal point, each of which is directly derived from a single notch frequency. The authors conclude that the use of negative reflection coefficients is crucial in determining notch frequencies. Therefore, the relation between notch frequency and reflection point-ear canal distance can be approximated by the following simple equation,

$$D_i(\phi) = \frac{c}{2F_i(\phi)}, \tag{1}$$

where c is the speed of sound, ϕ is the PRTF elevation angle, F_i is the center frequency of the i-th notch N_i, and D_i is the distance between the corresponding reflection point and the ear-canal point. Reflection points obtained from Eq. (1) were mapped on pinna images of a pool of experimental subjects, resulting in a close correspondence between reflection points and the three main contours, i.e. helix border, antihelix/concha inner wall, and concha border.

As a consequence, if we have an image of the pinna we can extract the above three contours, transform them into a sequence of polar coordinate pairs $(D_i(\phi), \phi)$ with respect to the ear-canal point, and derive from Eq. (1) notch frequencies for every desired elevation ϕ and for each of the three contours. The only independent parameter used in the model is indeed sound source elevation, which drives the evaluation of three polynomial functions (F_n^i, $i = 1, 2, 3$) that interpolate the obtained notch frequencies for a certain sampling step $\Delta\phi$. For what concerns the bandwidth and gain of notches as parameters, no clear relation with the pinna shape was found. The authors previously approximated these parameters, as well as resonance parameters, using average values from a population of subjects [18].

3. METHODS

The raw dataset consists of measured Head-Related Impulse Responses (HRIRs) from the CIPIC database [19], a public-domain database of high spatial resolution HRIRs measured at 1250 directions for 45 different subjects. Since this work involves subject anthropometry in the form of both numeric data (anthropometric parameters included in the CIPIC database) and a picture of their left or right pinna, we consider the 33 of them for which both are available for our analysis.

Consistently with the CIPIC database spatial grid, we take the interaural polar coordinate system as reference. We restrict our analysis to the frontal half of the median plane (azimuth angle $\theta = 0^\circ$), with the elevation angle ϕ varying from $\phi = -45^\circ$ to $\phi = 45^\circ$ at 5.625-degree steps (17 HRIRs per subject). We choose to consider the median plane because relative azimuthal variations up to at least $\Delta\theta = 30^\circ$ at fixed elevation cause very slight spectral changes in the pinna-related component of the HRTF [20], hence the model can be generalized to a wider set of azimuth values. Elevations higher than 45° were discarded because of the general lack of spectral notches in the corresponding HRTFs [21].

3.1. Notch frequency extraction

In order to get the relevant notch frequencies, we apply the ad-hoc signal processing algorithm by Raykar *et al.* [20] to each HRIR. Briefly, the algorithm computes the autocorrelation function of the linear prediction residual and extracts notch frequencies as the local minima of its group-delay function falling beyond a fixed threshold (heuristically set to -0.5 samples). Then, for each available elevation ϕ, the extracted notches are grouped in frequency tracks along adjacent elevations through the McAulay-Quatieri partial tracking algorithm [22], originally used to group sinusoidal partials along consecutive temporal windows according to their spectral location. The matching interval for the tracking procedure is set to $\Delta = 1$ kHz. The very same procedure for notch extraction and grouping was successfully used in a previous work [23].

Only tracks with 3 notches at least are preserved. If more than three tracks satisfying such a requirement are available, only the three longest tracks are considered and each frequency point labeled F_1, F_2, and F_3 in increasing order of average frequency. In those cases where a subject lacks a notch track (9 cases out of

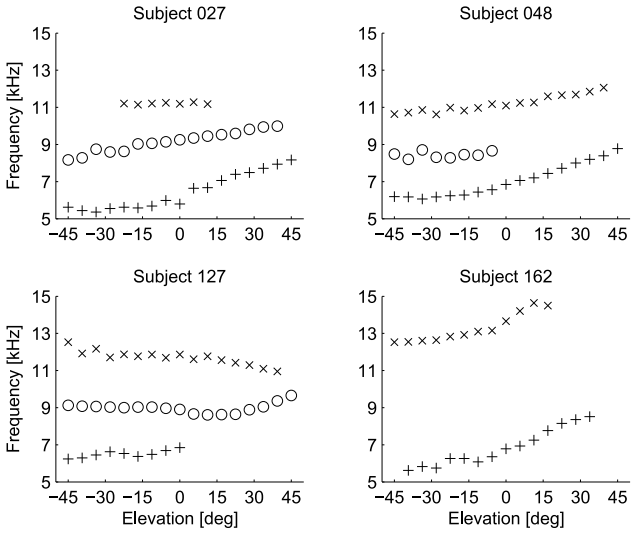

Figure 2: *Extracted notch tracks of four representative CIPIC subjects (N1:+, N2:○, N3:×).*

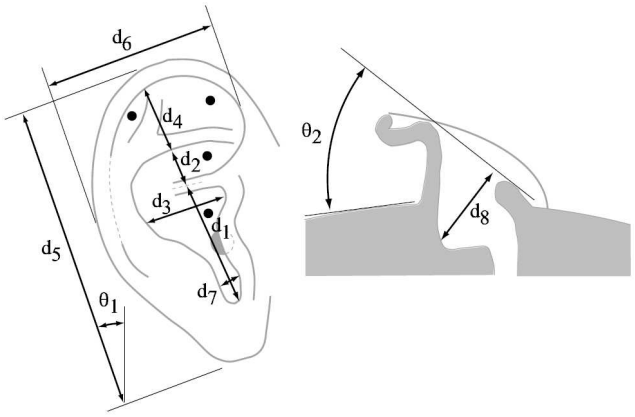

Figure 3: *The* 10 *anthropometric parameters of the pinna included in the CIPIC database (figure reproduced from [19]).*

33), labels are assigned according to the closest notch track frequency median among all subjects with three tracks. Figure 2 reports notch tracks of four representative subjects. Overall, the notch frequency extraction step yields 367 different observations for F_1, 401 for F_2, and 303 for F_3. Given the limited amount of data and the exploratory nature of this work, all observations will be used to train the following regression model, and 10-fold cross-validation will be performed to give an estimate of model fit.

3.2. Anthropometric feature extraction

Thirty-seven global anthropometric measurements for each one of the 33 considered subjects are available in the CIPIC database, 17 for the head and torso and 10 for each pinna. In this work, because of the focus on pinna notches, we consider the single pinna parameters only, which are reported in Fig. 3 and described in Table 1 for convenience. Table 1 also reports pairwise correlation values calculated on all available CIPIC subjects. Notice the over-

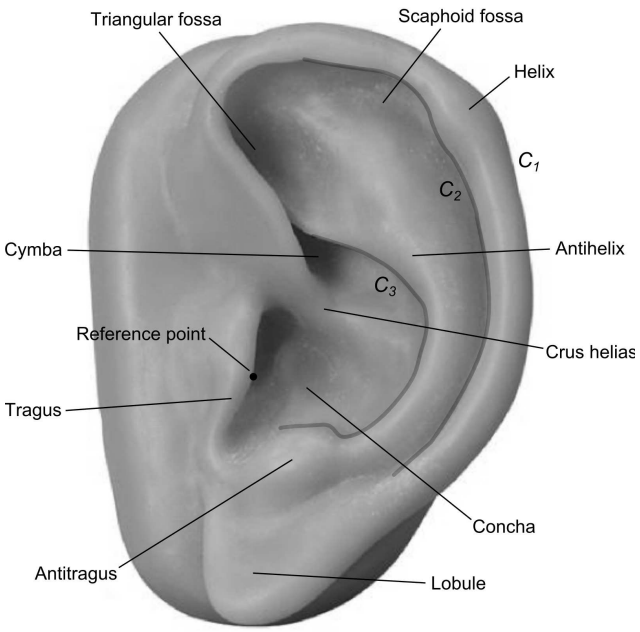

Figure 4: *Anatomy of the pinna and the three extracted pinna contours C_1, C_2, and C_3.*

all weak correlation values, apart from some cases where measurements overlap (e.g. $d_4 - d_5$, $d_3 - d_7$), which denotes an acceptable degree of orthogonality among the considered parameters.

In addition, elevation-dependent parameters strictly related to the ray-tracing rationale outlined in the previous Section are extracted according to the following procedure. The three contours corresponding to the outer helix border, the inner helix border, and the concha border/antitragus (see Fig. 4 for visual evidence of these contours) are traced by hand with the help of a pen tablet and stored as sequences of pixels.[1] Then, the point of maximum protuberance of the tragus is chosen as the reference ear-canal point for the computation of distance parameters. For each considered elevation angle $\phi \in [-45, 45]$, distances in pixels between the reference point and the point intersecting each traced pinna contour along the ray originating from the reference point with slope $-\phi$ are finally converted to centimeters based on the measuring tape reported in each picture close to the pinna and stored as $r_i(\phi)$, where $i \in \{1, 2, 3\}$ refers to the associated contour C_i.

The three parameters r_1, r_2, and r_3 are, as one may expect due to the roughly elliptic shape of the pinna, highly correlated. In particular, the correlation between r_1 and r_2 is 0.95 and these both correlate 0.75 with r_3. The very high correlation value between r_1 and r_2 can also be explained by the fact they both refer to the helix.

[1] Notice that these three contours do not all correspond to the three hypothesized reflection contours described in the previous Section. The reason we chose them is that in practical applications the same three contours can be robustly extracted through depth edge detection techniques based on multi-flash imaging [24]. However, in this work we chose manual tracing over automatic edge detection because single pictures were available only and because intensity-based edge detection methods (e.g. Canny, Sobel) fail in low-contrast areas such as those in the available pictures.

Table 1: *Pearson correlation coefficients between each pair of anthropometric parameters of the pinna.*

parameter	description	d_2	d_3	d_4	d_5	d_6	d_7	d_8	θ_1	θ_2
d_1	cavum concha height	-0.06	0.10	0.19	0.51	0.21	0.23	0.24	-0.04	0.20
d_2	cymba concha height		-0.02	0.33	0.47	0.13	0.11	0.30	0.02	-0.11
d_3	cavum concha width			0.03	0.19	0.47	0.59	0.27	0.23	0.02
d_4	fossa height				0.67	0.53	0.03	0.30	-0.06	-0.28
d_5	pinna height					0.52	0.22	0.44	-0.11	0.00
d_6	pinna width						0.16	0.45	0.09	-0.24
d_7	intertragal incisure width							0.15	-0.09	0.03
d_8	cavum concha depth								0.01	0.14
θ_1	pinna rotation angle									-0.12
θ_2	pinna flare angle									

3.3. Anthropometric regression

In order to investigate the dependence of F_1, F_2 and F_3 frequencies on anthropometry, multiple linear regression is performed on the thirteen measured parameters (d_i, $i = 1 \dots 8$, θ_j, $j = 1 \dots 2$, $r_k(\phi)$, $k = 1 \dots 3$, $\phi \in [-45, 45]$) for all subjects and elevations where the notch is available. Since our focus is on anthropometric parameters, the elevation angle ϕ is not considered as a regressor. We compute first a model for each notch including all variables; then, if the model is capable of significantly accounting for a large portion of variance in the corresponding notch frequency data, we design a more conservative and meaningful model in accordance with the following greedy algorithm.

1. Calculate pairwise correlations between notch frequencies and each anthropometric parameter.
2. Compute a univariate linear regression model with the parameter showing the highest unsigned correlation value in Step 1 as regressor and notch frequency as outcome.
3. Calculate pairwise correlations between residuals of the previous model and each of the remaining anthropometric parameters.
4. Compute a multivariate linear regression model by adding the parameter showing the highest unsigned correlation value in Step 3 to the list of regressors used in the previous model.
5. If the model significantly improves the previous instance at the $p = 0.05$ level (according to an analysis of variance), label the model as "good" and return to Step 3. Otherwise, output the last "good" model.

This model allows to check which variables significantly account for variation in the corresponding notch frequency, as well as the number of variables needed in order to have an acceptable approximation of the outcome.

4. RESULTS

Table 2 reports the results of the three multiple linear regression models (each called M_i in association with F_i, $i = 1, 2, 3$) with all 13 anthropometric parameters as regressors and each notch frequency as outcome. M_1 outperforms the other two models in terms of R^2 value, denoting a much stronger fit for notch N_1 than for N_2 or N_3. The very low R^2 values for both M_2 and M_3 let us conclude that the 13 parameters are not sufficient to explain tight relations between anthropometry and notches N_2 and N_3.

Let us now concentrate on model M_1. Slopes of most of the thirteen parameters, as well as the intercept, are highly significantly different from zero, with only the two rotation angles θ_1 and θ_2 resulting weak parameters as well as r_2, which is however tightly correlated with r_1 as previously discussed. Focusing on the most significant parameters ($p < 0.001$), coefficient signs reveal that both r_1 and r_3 have negative slope, indicating that as the distance between reference point and pinna contour increases F_1 decreases, as hypothesized in our previous ray-tracing model. Elevation-independent parameters have, by contrast, different signs; we will return to this point later on in this discussion.

Model M_1 suffers from considerable variance inflation related to some of its parameters. As an example, the Variance Inflation Factor (VIF) for parameters d_5, r_1 and r_2 is 6.8, 14.6, and 14.9 respectively, denoting the presence of unnecessary variables in the model. As a consequence, we build a more conservative model according to the algorithm outlined in Section 3.3. The algorithm chooses eight variables in the following order: r_1, d_6, d_1, r_3, d_2, d_3, d_8, d_7, with the model built at each step with the inclusion of a new variable significantly improving the previous model at the $p = 0.001$ level according to analysis of variance. The first chosen variable, r_1, accounts alone for a large percentage of variation, equal to $R^2 = 0.51$. This result can be related to the pinna model in Section 2, which assumed notch N_1 to be caused by reflections on the outermost pinna contour.

Again, results of the final model, which we call $\hat{M}_1$, are reported in Table 2. Notice how close the summary results are to those of M_1, confirming that the 8 chosen variables are sufficient in explaining the variance accounted for in M_1. Also, coefficient signs do not reverse passing from one model to the other, hence the previous observations on r_1 and r_3 still hold. For what concerns elevation-independent parameters, notice that the height parameters d_1 and d_2 carry a negative sign, indicating that an increase in height implies lower notch frequencies, in accordance with results by Middlebrooks [25]. By contrast, the width and depth parameters d_3, d_6, d_7 and d_8 have positive signs in 3 out of 4 cases.

Let us now turn to some considerations of psychoacoustic nature. The residual standard error of $\hat{M}_1$ is 0.59 kHz, meaning that our approximation introduces significant errors if used for predicting F_1, and the displacement of pivotal elevation cues such as pinna notches is known to have an impact on localization accuracy. For instance, it is known and verifiable in our data that 1-kHz shifts of N_1 can correspond to an increase/decrease of the elevation angle of 20° or more [14]. However, from previous literature [13]

Table 2: *Summary of the resulting linear regression models. For each model, the coefficient of each regressor and its level of significance (coefficient different from zero at level $p = 0.05$ (*), $p = 0.01$ (**), and $p = 0.001$ (***)) are reported, along with the RMS error and the R^2 value arising from cross-validation. Units of measurement are: cm ($d_1 - d_8$, $r_1 - r_3$), rad (θ_1, θ_2), kHz (outcomes $F_1 - F_3$).*

	M_1		$\hat{M}_1$		M_2		M_3	
RMS error [kHz]	0.59		0.59		0.75		1.27	
R^2	0.76		0.76		0.34		0.15	
parameter	coefficient	level	coefficient	level	coefficient	level	coefficient	level
d_1	-1.85	***	-1.48	***	-1.22	***	-1.61	**
d_2	-2.21	***	-1.70	***	-1.74	***	-3.44	**
d_3	-1.01	***	-1.03	***	0.74	**	0.16	
d_4	-0.51	**			-0.27		2.54	***
d_5	0.47	**			0.70	**	0.13	
d_6	1.76	***	1.79	***	0.11		-1.44	***
d_7	1.15	**	1.45	***	-1.59	***	2.55	**
d_8	0.74	***	0.79	***	-0.25		-1.44	**
θ_1	-0.07				-0.84	*	2.26	**
θ_2	-0.27				-1.06	*	-1.18	
r_1	-1.54	***	-1.16	***	-0.23		1.03	
r_2	0.42	*			-0.47	*	-1.43	*
r_3	-0.81	***	-0.78	***	-0.24		-0.12	
intercept [kHz]	9.84	***	10.44	***	11.19	***	17.21	***

we also know that two steady notches in the high-frequency range (around 8 kHz) differing in center frequency are distinguishable on average if such difference is around 10% of the lowest center frequency at least, independent of notch bandwidth.

If we extend the above assumption to the frequency range of N_1, we can detect all those notch frequencies predicted by $\hat{M}_1$ that exceed 10% of the corresponding extracted notch frequency F_1. This results in approximately 80% of predicted notch frequencies lying below such a threshold. If we take into account that different sources of error may exist in our data, for instance

- a completely automatic extraction of notch frequencies that may introduce artifacts during both the notch picking and tracking procedures, and
- the arbitrary placement of the ear-canal reference point in the computation of elevation-dependent parameters, which does not necessarily correspond to the microphone position in CIPIC measurements [11],

we can assess the overall fitness of our model. Still, having in regard that notch detectability heavily depends on stimulus intensity and intersubject variation [26], individual psychoacoustic tests are needed in order to ascertain whether the proposed approximation correlates to elevation performance in the median plane.

5. CONCLUSIONS

In this paper we studied linear models for approximating frequencies of pinna notches in frontal median-plane HRTFs from anthropometric parameters. Results for notch N_1 show an encouraging correspondence between anthropometry and HRTF features, confirming the possibility of predicting pinna notches from pictures of the ear. Even though the proposed model represents a step forward towards a full understanding of the physical mechanisms lying behind the generation of pinna notches, it bears some limitations.

First of all, the model is built from data included in the CIPIC HRTF database. Even though it is the most diffuse among the scientific community, the CIPIC database suffers from measurement error (such as left/right asymmetries [27]) and lack of documentation (e.g. no reference about microphone position) which likely contributed to increase the clutter in our data. More recent and documented databases such as the Aalto PRTF database [21] or the SYMARE database [28] will be used in future works in order to carry a more controlled analysis on the available data.

Secondly, the analysis was only conducted in the frontal half of the median plane. In order for results to be general, a wider and denser grid of HRTFs will need to be considered and the proposed notch extraction/tracking procedures to be accordingly tuned and extended to account for a 2-D (or possibly even 3-D, if distance dependence is included) representation of the notch data.

Last but not least, our model precludes the inclusion of possible nonlinear effects, which will be investigated in the future through the use of state-of-the-art machine learning techniques for nonlinear regression. The contingent availability of a large amount of data will allow the construction of proper training and test sets in order to have a technically sound data analysis.

6. ACKNOWLEDGMENTS

The authors wish to thank Mr. Silvio Galesso for his help with anthropometric data extraction. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 643636.[2] This work was supported by the research project Personal Auditory Displays for Virtual Acoustics, University of Padova, under grant No CPDA135702.

[2] Sound of Vision, `www.soundofvision.net`

7. 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, April 2001.

[2] 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, October 2001.

[3] B. Xie, *Head-Related Transfer Function and Virtual Auditory Display*, J.Ross Publishing, Plantation, FL, USA, 2nd edition, June 2013.

[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, June 1996.

[5] 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, September 1998.

[6] L. Li and Q. Huang, "HRTF personalization modeling based on RBF neural network," in *Proc. 38th IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP 2013)*, Vancouver, BC, Canada, May 2013, pp. 3707–3710.

[7] Q. Huang and L. Li, "Modeling individual HRTF tensor using high-order partial least squares," *EURASIP J. Adv. Signal Process.*, vol. 2014, no. 58, pp. 1–14, May 2014.

[8] P. Bilinski, J. Ahrens, M. R. P. Thomas, I. J. Tashev, and J. C. Platt, "HRTF magnitude synthesis via sparse representation of anthropometric features," in *Proc. 39th IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP 2014)*, Firenze, Italy, May 2014, pp. 4501–4505.

[9] F. Grijalva, L. Martini, S. Goldenstein, and D. Florencio, "Anthropometric-based customization of head-related transfer functions using Isomap in the horizontal plane," in *Proc. 39th IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP 2014)*, Firenze, Italy, May 2014, pp. 4506–4510.

[10] S. K. Roffler and R. A. Butler, "Factors that influence the localization of sound in the vertical plane," *J. Acoust. Soc. Am.*, vol. 43, no. 6, pp. 1255–1259, June 1968.

[11] 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, March 2013.

[12] 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, August 2007.

[13] 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. Am.*, vol. 85, no. 2, pp. 820–836, February 1989.

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

[15] P. Satarzadeh, R. V. Algazi, and R. O. Duda, "Physical and filter pinna models based on anthropometry," in *Proc. 122nd Conv. Audio Eng. Soc.*, Vienna, Austria, May 2007, pp. 718–737.

[16] K. J. Faller II, A. Barreto, and M. Adjouadi, "Augmented Hankel total least-squares decomposition of head-related transfer functions," *J. Audio Eng. Soc.*, vol. 58, no. 1/2, pp. 3–21, January/February 2010.

[17] 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. (ICASSP 2011)*, Prague, Czech Republic, May 2011, pp. 2408–2411.

[18] M. Geronazzo, S. Spagnol, and F. Avanzini, "A head-related transfer function model for real-time customized 3-D sound rendering," in *Proc. INTERPRET Work., SITIS 2011 Conf.*, Dijon, France, November-December 2011, pp. 174–179.

[19] 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, October 2001, pp. 1–4.

[20] V. C. Raykar, R. Duraiswami, and B. Yegnanarayana, "Extracting the frequencies of the pinna spectral notches in measured head related impulse responses," *J. Acoust. Soc. Am.*, vol. 118, no. 1, pp. 364–374, July 2005.

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

[22] R. J. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 4, pp. 744–754, August 1986.

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

[24] S. Spagnol, D. Rocchesso, M. Geronazzo, and F. Avanzini, "Automatic extraction of pinna edges for binaural audio customization," in *Proc. IEEE Int. Work. Multi. Signal Process. (MMSP 2013)*, Pula, Italy, September-October 2013, pp. 301–306.

[25] J. C. Middlebrooks, "Individual differences in external-ear transfer functions reduced by scaling in frequency," *J. Acoust. Soc. Am.*, vol. 106, no. 3, pp. 1480–1492, September 1999.

[26] A. Alves-Pinto and E. A. Lopez-Poveda, "Detection of high-frequency spectral notches as a function of level," *J. Acoust. Soc. Am.*, vol. 118, no. 4, pp. 2458–2469, October 2005.

[27] S. Spagnol and F. Avanzini, "Anthropometric tuning of a spherical head model for binaural virtual acoustics based on interaural level differences," in *Proc. 21st Int. Conf. Auditory Display (ICAD)*, Graz, Austria, July 2015, pp. 204–209.

[28] C. Jin, P. Guillon, N. Epain, R. Zolfaghari, A. van Schaik, A. I. Tew, C. Hetherington, and J. Thorpe, "Creating the Sydney York morphological and acoustic recordings of ears database," *IEEE Trans. Multimedia*, vol. 16, no. 1, pp. 37–46, January 2014.