3D EAR SHAPE AS AN ESTIMATOR OF HRTF NOTCH FREQUENCY

Marius George Onofrei, Riccardo Miccini
Dept. of Architecture, Design & Media Technology
Aalborg University
Copenhagen, Denmark
monofr11@student.aau.dk

Runar Unnthorsson
School of Engineering and Natural Sciences
University of Iceland
Reykjavík, Iceland
runson@hi.is

Stefania Serafin, Simone Spagnol
Dept. of Architecture, Design & Media Technology
Aalborg University
Copenhagen, Denmark
ssp@create.aau.dk

ABSTRACT

This paper makes use of a new dataset of Head-Related Transfer Functions (HRTFs) containing high resolution median-plane acoustical measurements of a KEMAR mannequin with 20 different left pinna models as well as 3D scans of the same pinna models. This allows for an investigation of the relationship between 3D ear features and the first pinna notch present in the HRTFs, with the final aim of developing an accurate and handy procedure for predicting the individual HRTF from non-acoustical measurements. We propose a method that takes the 3D pinna mesh and generates a dataset of depth maps of the pinna viewed from various median-plane elevation angles, each having an associated pinna notch frequency value as identified in the HRTF measurements. A multiple linear regression model is then fit to the depth maps, aiming to predict the corresponding first pinna notch. The results of the regression model show moderate improvement to similar previous work built on global and elevation-dependent anthropometric pinna features extracted from 2D images.

1. INTRODUCTION

Head-Related Transfer Functions (HRTFs) are essential when it comes to providing a spatial listening experience over headphones. They are the filters that describe the acoustic effects of the human head and are highly dependent on the direction of the incoming sound source relative to each ear. In particular, the pinna plays a major role in determining the spectral information contained in HRTFs which proves to be a critical cue for spatial localization, especially when it comes to identifying the elevation of a sound source [1]. While the major mechanisms for sound localization on the horizontal plane are the interaural differences [2], such cues have a low impact in the median plane, giving the HRTF and in particular the pinna component (known as PRTF – Pinna-Related Transfer Function) particular importance for spatial localization. The contribution of the pinna is also relevant off the median plane, as variations in azimuth up to 30 degrees from the median plane cause limited spectral changes in the PRTF [3].

Since the anthropometry of the pinna is unique to individuals, it implies that HRTFs are also unique to each individual. Indeed, the use of non-individual HRTFs in binaural sound rendering is known to produce sound localization errors such as wrong elevation perception, confusion between front and back sources, or even the feeling that the sound comes from inside the head [4]. Within such context, customizing a generic HRTF model onto a user based on individual morphological data is a particularly attractive alternative to time- and resource-consuming acoustical measurements.

A number of previous work suggest a relationship between pinna anthropometry and certain HRTF features [5, 6], particularly spectral notches, which are believed to be caused by reflections of the incoming sound inside the pinna [7]. The first (lowest-frequency) pinna notch, known in the literature as N_1, has long been known as the responsible cue for increasing frontal elevation [8]. Encouraging results were found when trying to predict the center frequency of N_1, starting from both global and elevation-dependent anthropometric pinna measurements [9, 10].

The present work expands on these results, aiming to make use of a new dataset of HRTFs which contains high-resolution acoustical measurements of a KEMAR mannequin equipped with 20 different artificial pinnae on its left channel in the frontal median plane. The dataset comes together with the 3D scans of the 20 pinna models, which allow prediction of the N_1 frequency estimated from the measured HRTF based on a regression model built on elevation-dependent depth maps of the pinnae. The remainder of the paper is organized as follows. Section 2 describes the methods for the acquisition of HRTFs and 3D models, post-processing and feature extraction, and the regression model, whose results are reported and discussed in Section 3. Section 4 concludes the paper.

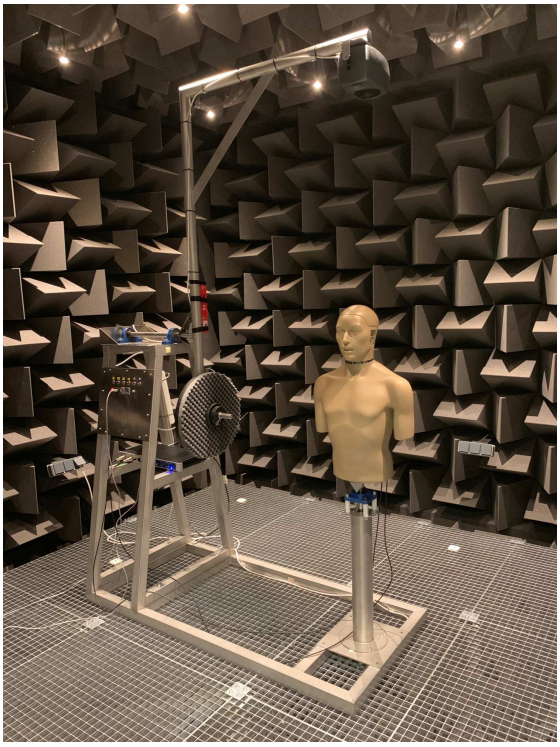

Figure 1. HRTF measurement setup and environment.

2. METHODS

2.1 Dataset

The dataset considered for analysis is an unreleased version of the Viking HRTF dataset [1] [11] where a focus on extra median-plane measurements has been set. Similarly to [11], an automatic, custom made HRTF measurement system was used. As can be seen in Fig. 1, the system consists of a KEMAR mannequin (45BB-4 configuration) with interchangeable pinnae and a Genelec 8020CPM-6 loudspeaker placed at a fixed distance of 1 m from the center of the mannequin's head, which adjusts for elevation via an L-shaped rotating arm.

As opposed to the Viking HRTF dataset which was measured inside a relatively silent yet standard room, the new measurements were carried out in the anechoic chamber recently installed at the University of Iceland, during the month of November 2019. 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 ISO 3745 and ISO 26101, the anechoic free field in the chamber has been certified (as of September 2019) compliant for measurements within an area 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.

A total of 20 different custom-made pinnae, shown in Fig. 2, were used in turn as the left ear of the mannequin, while the the right ear was always kept the same: a standard large KEMAR pinna (35 Shore-OO hardness). The custom-made pinnae are labeled in alphabetical order as subjects A to S, plus the KEMAR replica. For further information on the pinna sample and casting procedure, consult [11]. The dataset also considers additional cases, necessary either for facilitating post-processing or for quality checking the measurements. In particular, subject Z is a case where no pinna is mounted on the left side and the cavity is filled with a silicone baffle of the same material, used in order to isolate torso effects of the KEMAR mannequin. A number of measurements with standard KEMAR pinnae of different sizes and hardnesses were also carried out, which can be used to verify results of previous work [12].

[1] https://itsadive.create.aau.dk/index.php/viking-hrtf/

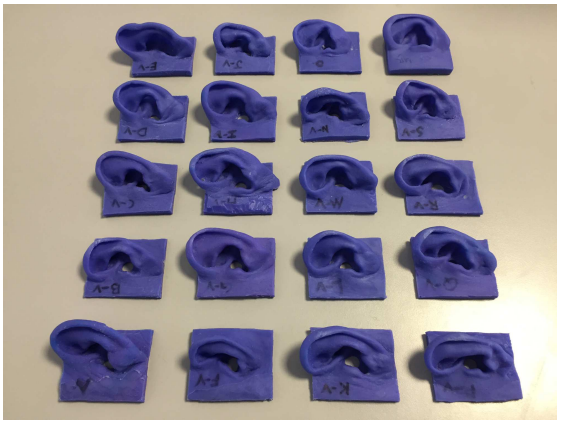

Figure 2. Custom-made pinnae used for the HRTF measurements.

The logarithmic sweep method was adopted for recording the acoustical responses [13]. The duration of the sweep was 0.9 s, spanning a frequency interval between 20 Hz and 20 kHz at a sampling rate of 48 kHz. The average SPL level at the left ear with the source directly above the mannequin was 90 dB SPL at 1 kHz. For each left pinna, sweep response measurements were recorded at 1° steps from -80° to $+90^\circ$ elevation in the frontal half of the median plane (azimuth = 0°), totaling 3420 sweep response measurements. Positive/negative elevations correspond to directions above/below the horizontal plane, respectively.

2.2 HRIR determination

In order to retrieve the Head-Related Impulse Response (HRIR) and subsequently the HRTF from the measured sweep responses a post-processing script was developed. The onset of the sweep response was determined by means of evaluating the cross-correlation function, $\Psi\,[n]$, of the recorded sweep response and the input sweep signal. The lag n_i giving the maximum value of the cross-correlation function is in fact the optimal lag for the raw sweep response that results in maximum correlation with the input sweep signal. In [11], an additional step was performed to ensure that the onset point was not driven by possible wall reflections. Since the new measurements were carried out in an anechoic chamber, this step can be disregarded.

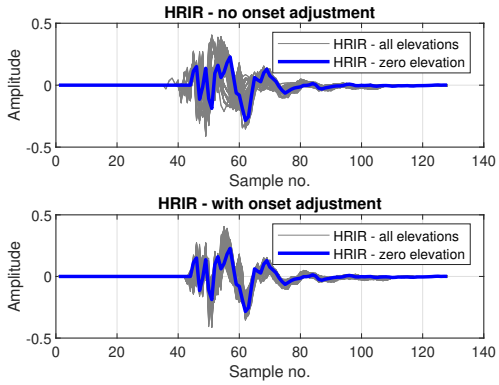

Figure 3. Resulting HRIRs in the frontal median plane for subject A (top) and adjusted HRIRs to give maximum correlation to the zero-elevation HRIR (bottom).

Furthermore, that additional step was particularly aimed to address the case of far-ear HRTFs while the focus of the new measurements is the median plane.

The impulse response is then calculated by performing an inverse filtering of the measured sweeps via the following equation:

$$h\left[n\right] = \Re\left(\mathcal{F}^{-1}\left(\mathcal{F}\left(y\left[n\right]\right) \cdot S^{-1}\left[n\right]\right)\right), \qquad (1)$$

where $\mathcal{F}$ is the DFT function (with the FFT length equal to the length of the sweep signal), $\mathcal{F}^{-1}$ is the inverse DFT function, and $S^{-1}\left[n\right]$ is the inverse reference spectrum of the input sweep signal. A 128-sample half-Hann window is then applied to each impulse response $h[n]$ with the aim of removing unwanted early and late reflections occurring later than approximately 2.5 ms.

It was found that the method is not completely robust for all elevations as there are some slight onset differences in the HRIRs for different elevations, which causes some peaks in the signal to be offset. Since this results in jumps between adjacent HRTF magnitudes, an additional correction is introduced. A similar cross-correlation procedure as described above is performed to adjust the resulting HRIRs such that they give maximum correlation to the HRIR resulting from elevation $\phi = 0°$. Figure 3 shows this adjustment, where it can be seen that the signals are better aligned when this supplementary correction is used.

2.3 HRTF feature extraction

As a first step prior to extracting the HRTF features of interest, it is desirable to remove any effects that are not related to the pinna. A noticeable presence of shoulder reflections can be observed in the HRTFs, and therefore a compensation with the responses of subject Z has been designed. For all available elevations, a recursive IIR digital filter of the 64th order is fitted to the magnitude response of subject Z's HRTFs – displayed in Fig. 4 – using the Yule-Walker method [14], such that the filtered response would be approximately flat. Filters are not fitted to the smallest

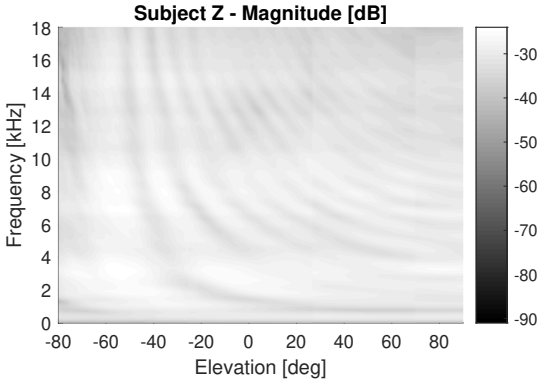

Figure 4. HRTF magnitude response of the KEMAR mannequin with no left pinna in the frontal median plane.

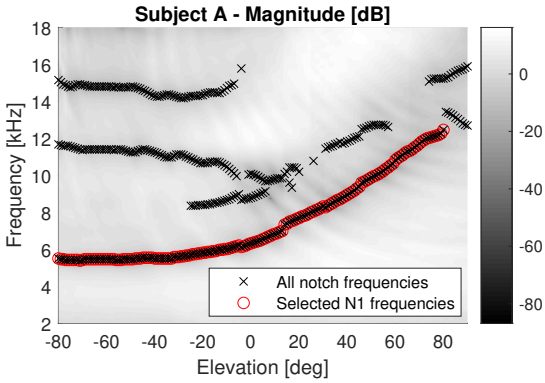

Figure 5. HRTF magnitudes of subject A and resulting notch frequencies.

and largest 5% of frequencies in order to achieve a better stability and fit in the frequency range of interest. The filter is applied to the post-processed HRIRs of all other subjects. Although shoulder reflection effects can still be noticed in the majority of the filtered HRIRs due to slight frequency offsets, their magnitude is significantly reduced.

Subsequently, notch frequencies are extracted using two alternative methods. The first one is the signal processing algorithm by Raykar et al. [15], which computes the auto-correlation function of the linear prediction residual and extracts notch frequencies as the local minima of its group-delay function falling beyond a fixed threshold. In the current implementation of the algorithm the linear predictor (LP) analysis step is skipped as it was found that it introduces additional notches not visible in the original signal spectrum. The second one is a more direct algorithm, where notches are extracted directly from the HRTF spectrum. A threshold for notch selection is computed for each elevation as the mean of the spectrum magnitude between 3 kHz and 16 kHz.

The two methods generally align with regards to the identified notches. However, there exist outlier cases where one method is better at finding notches than the other. Hence, the decision taken is to merge the two notch sets. Finally,

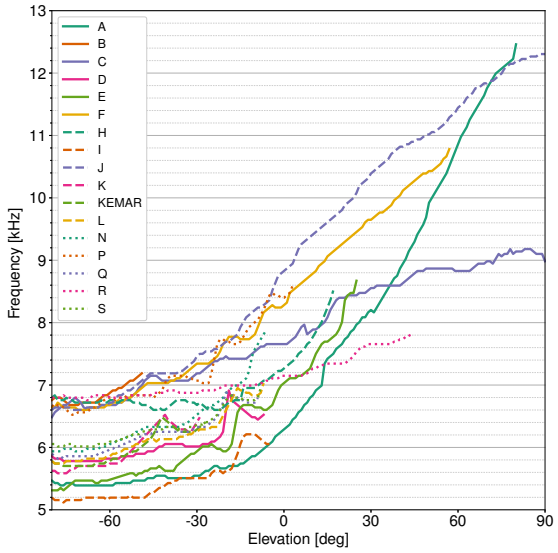

Figure 6. First notch (N_1) track for all available subjects.

the first notch N_1 is manually selected as the lowest frequency track starting between 4 and 7 kHz at lower elevations. Since three subjects, M, G and O, did not show a clear N_1 track in this frequency region, they have been disregarded from the following analysis. This results in a total of $1,599$ data points. Figure 5 shows the selected N_1 frequencies for subject A superposed to the HRTF plot, while Fig. 6 shows the identified N_1 tracks for all 17 subjects.

2.4 Pinna feature extraction

The Viking dataset comes with 3D scans of all the custom pinna samples used in the tests, available as STL files. The scans were captured at 1 mm resolution with a Creaform Go!SCAN 20 handheld scanner. It is hypothesized that the use of 3D data can be an advantage to the traditionally used anthropometric measurements and/or 2D images of the pinna when trying to predict notch frequencies. An improvement to the previously used features can be to add a depth dimension to the images. Based on this hypothesis, we generate a dataset of depth maps of the pinnae "viewed" from the various elevation angles considered in the median plane.

Depth is calculated using a ray-tracing algorithm with rays propagating from the source of the sweep signal, which is considered to be a point at a distance of exactly 1 m from the ear canal entrance along the direction of each considered elevation angle. The center point of the ear canal entrance is assumed to be the location of the microphone diaphragm. The vertices of the pinna mesh are projected onto the plane passing through the center point, $(x, y, z) = (0, 0, 0)$, whose normal vector originating from the center point connects it to the signal source. A rectangular grid of 100×100 units is considered on this projection plane with a fixed size taken as the largest grid size needed to fully project any pinna mesh at any considered angle. In order to exclude effects of the rectangular pinna supports, only values above the microphone location ($z > 0$) are included. Projections of points closer than 1 m are also disregarded in order to consider only potential reflection locations. Figure 7 shows a resulting depth map for subject A at elevation $\phi = -30°$. The depth map is illustrated as a colored surface plot on the projection plane, with the colorbar indicating the depth of intersection points.

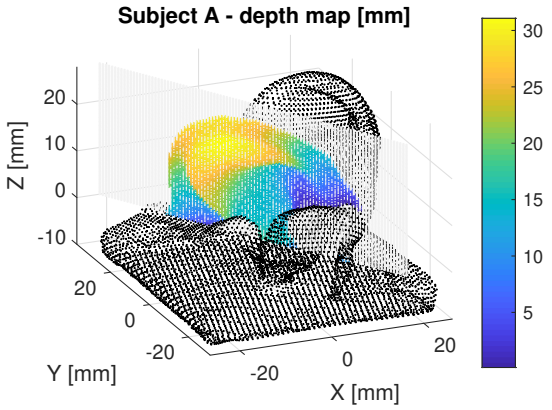

Figure 7. Example of depth map calculation for subject A, elevation $\phi = -30°$. Black points are the vertices of the pinna mesh, vertical gray lines represent the grid onto which pinna vertices are projected (or equivalently, the plane passing through the center point of the mesh whose normal vector is oriented towards the signal source), and colored points represent the resulting depth map.

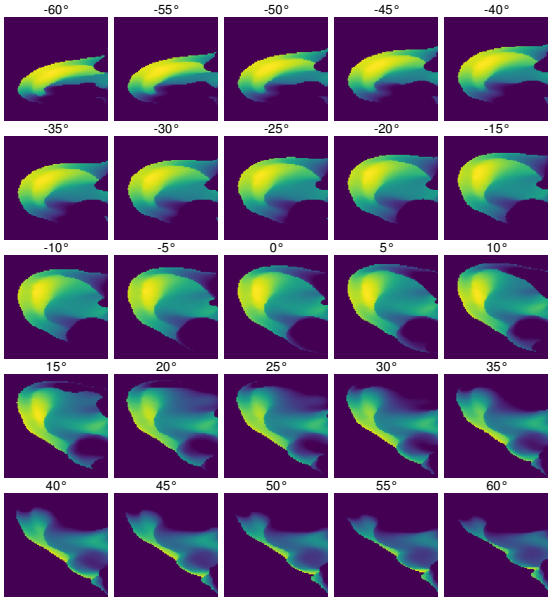

Figure 8. Example of depth maps at various elevations for subject A.

Furthermore, Fig. 8 shows various depth maps for subject A. Depth values are normalized with respect to the maxi-

mum depth recorded for all subjects. Values corresponding to no intersection or where the intersection is smaller than the distance to the source (1 m) are set to zero. This results in non-zero values of the depth map only in correspondence with surfaces which can produce reflections contributing to possible destructive interference and therefore the presence of notches in HRTFs.

2.5 Regression model

Depth maps, together with the recorded N_1 frequencies for all subjects can now be used as a basis for a regression model aiming to predict the N_1 frequencies in HRTFs. For subjects where N_1 frequencies are observed (17 total subjects) the 100×100 depth map at each elevation ϕ acts as the set of input variables for calculating $N_1(\phi)$, for a total of $10,000$ features.

For the sake of comparison against the results of previous studies, in this work we chose to use a multiple linear regression model. The assumptions for applying linear regression using an ordinary least squares estimator were directly checked on the available data, revealing that the relation between depth features and notch frequency is not linear, that features are highly multicollinear, and that residuals are heteroscedastic. As a consequence, we apply a principal component analysis (PCA) to our data and use the first 256 components (explaining approximately 98% of the total variance in the data) as the set of predictor variables. The new set of variables fulfils all the assumptions for linear regression except for homoscedasticity of residuals; this has no big impact on the results as long as no statistical inferences are conducted [16].

In order to maximize the available data for training, a leave-one-out cross-validation at subject level is carried out. This means, for example, that when evaluating the regression model for test subject A the training is performed on the 16 other subjects. This implies that no test subject appears in the training set. However, as subjects have a different number of recorded N_1 frequencies, the split ratio varies in the cross-validation. The used metrics for evaluating the multiple regression model are the classical R^2 of the fit and the mean absolute error (MAE), calculated as

$$MAE = \frac{1}{n}\sum_{j=1}^{n} |y_j - \hat{y_j}| \tag{2}$$

where n is the number of observations for the test subject, y_j is the ground-truth frequency value, and $\hat{y_j}$ is the predicted value. The latter metric was preferred over root-mean-square error because it provides an intuitive representation of the average residual, and because it is robust to outliers and large errors.

The considered HRTF dataset has a high median-plane resolution, which is of course an advantage when it comes to identifying the progression of notch frequencies along elevation. However, due to the fact that N_1 tends to only be present at lower elevations [17], training the regression model on depth maps associated with the entire elevation range is inappropriate. From Fig. 6 it can be seen that only 4 subjects have values above $45°$ elevation. Another aspect to be noticed is that at the lowest elevation angles, up to about $-50°$, notch tracks for all subjects appear quite flat. Including this data in the learning model may be problematic, as the depth maps in that range change considerably but their target values do not. As this may pollute the model, which we name M_1, we decided to concentrate on learning N_1 frequencies within the interval $\phi \in [-45, 45]°$ and disregard all other elevation values. The chosen interval also allows comparison with the results of a previous work which considered prediction of N_1 frequencies from anthropometric measurements in the same elevation range [10]. Furthermore, a model M_0 is trained on the entire available dataset, i.e. elevation interval $\phi \in [-80, 90]°$, to highlight the aforementioned shortcomings.

	M_0		M_1	
Subject	MAE	R^2	MAE	R^2
A	519.19	0.94	228.94	0.96
B	365.53	0.24	–	–
C	877.29	0.76	727.90	0.62
D	600.85	0.33	666.68	0.58
E	941.45	0.87	461.87	0.89
F	594.41	0.84	532.25	0.89
H	285.78	0.54	305.89	0.82
I	1352.90	0.16	779.15	0.75
J	1263.95	0.93	877.25	0.95
K	274.64	0.91	135.64	0.00
L	566.15	0.74	134.66	0.81
N	996.41	0.27	437.83	0.62
P	421.69	0.42	414.49	0.09
Q	1136.40	0.72	791.34	0.70
R	593.30	0.59	687.54	0.72
S	206.11	0.74	382.34	0.61
KEMAR	660.91	0.75	68.13	0.61
mean	685.70	0.63	476.99	0.66
std	351.70	0.26	258.58	0.27
weighted mean	733.97	0.68	531.89	0.74

Table 1. Mean absolute error (in Hz) and R^2 of predicted versus measured N_1 frequencies for the two considered multiple linear regression models M_0 and M_1, divided per test subject.

3. RESULTS

Table 1 shows the results of the two considered multiple linear regression models. The MAE and R^2 for each subject are given as a consequence of the leave-one-out cross-validation method, as well as the mean and standard deviation of the individual metrics. Since some subjects record N_1 frequencies at more elevations than others, we also include a weighted mean of the two metrics, where the value for each subject is weighted with the number of predicted data points.

This effectively calculates the overall performance of the regression model by equally considering all individual data points and is a more comparable measure to the results of [10] where an overall MAE of 607 Hz for the test data was found, as opposed to the 532 Hz MAE found for mod-

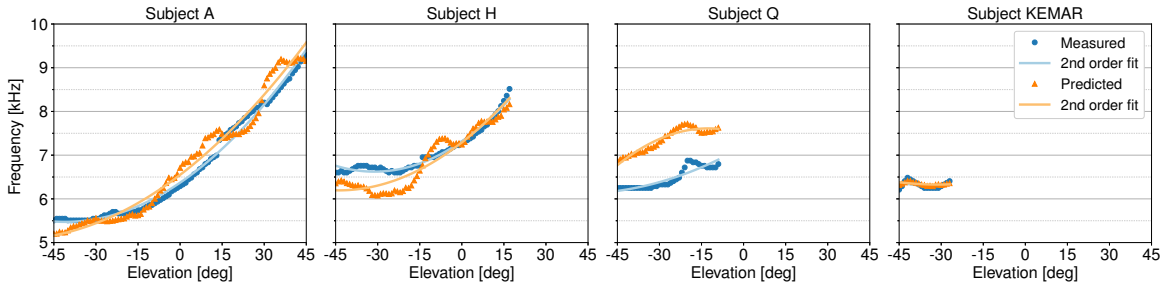

Figure 9. Measured and predicted (with model M_1) N_1 tracks for four test subjects.

el M_1. This moderate improvement can be attributed to the inclusion of 3D information in the form of depth maps used as input in the current regression model, as opposed to the features used in the previous work [10] which were based on both generic and elevation-dependent anthropometric measurements from the CIPIC database. This result takes on further significance considering the low number of available training subjects in this study (16, because of no N_1 points for subject B in the $[-45, 45]^\circ$ elevation range), and also corroborates the choice of using an HRTF dataset based on highly controlled and repeatable measurements [11].

When comparing the results for models M_0 and M_1, we observe worse performances in M_0. The same trend is seen when looking at the MAE and R^2 metrics for the majority of individual subjects. As expected, the reason why M_1 outperforms M_0 is related both to the low variation at subject level in notch frequency values for low elevations, and to the general lack of notch points for high elevation values.

Another approach to evaluating the models is to consider the perceptual threshold of discrimination of two spectral notches in the high-frequency range. Previous literature indicates that differences of around 10% of the lower notch frequency are needed to distinguish between the two, independent of notch bandwidth [18]. Therefore it is possible to determine the percentage of predictions in the validation set that fall within this range. The same trend seen in the previous evaluation is again present: model M_0 scores a percentage of 54.78%, while there is considerable improvement for M_1 which scores 71.27%, also improving the 61.2% score reported in [10].

It must be noted that although the evaluations presented here show a certain degree of accuracy, the progression of N_1 frequencies from the regression models does not look as smooth as in the measurements. Figure 9 shows that even for cases where prediction is accurate, like for subjects A and H, it seems that a second-order polynomial fit to the predicted data points would be more suitable. In other cases (e.g. subject Q), while the N_1 trend along elevation looks correctly reproduced, a considerable offset of almost 1 kHz can be noticed. Finally, despite the low number of data points available for the KEMAR subject, the reconstruction of its N_1 track with model M_1 is nearly perfect.

4. CONCLUSIONS

The work described in this paper is based on a recently collected dataset of HRTF measurements carried out on a KEMAR mannequin with interchangeable pinnae, for which 3D scans of the used pinna models are available. An approach for using 3D data to predict the frequency progression of the first pinna notch N_1 in frontal median-plane HRTFs is proposed. The method uses information learned from the data-set in order to build a multiple linear regression model that could hopefully generalize outside of the dataset. While results clearly show that 3D features and HRTF features are related, the accuracy of prediction could be further improved. This might be done by including additional training data, by using more complex (possibly non-linear) regression models, or by performing feature selection. For instance, a machine learning approach such as random forest regression would be an appropriate solution, since it has no strict prerequisite as opposed to linear regression. There is, however, progress in terms of including 3D morphological information as opposed to the classical anthropometric features, as shown by comparison of the current results with the regression model previously used in [10].

While this paper focuses on N_1 estimation, more HRTF features need to be estimated from morphological user data in order to build a perceptually convincing individualized HRTF set. Future work will address the estimation of higher frequency notches in HRTFs, whose identification is complicated by higher noise and ambiguity. Models for a wider spatial range will also be considered. Ultimately, the broader goal of the project is an accurate tuning of low-order structural HRTF models [19,20] and HRTF selection methods [21,22] based on morphological data [23]. Applications of these methods are expected to range from travel aids for the visually impaired to entertainment and rehabilitation systems [24, 25].

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 797850, and from NordForsk's Nordic University Hubs programme under grant agreement No. 86892.

5. REFERENCES

[1] 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.

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

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

[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] K. Iida, Y. Ishii, and S. Nishioka, "Personalization of head-related transfer functions in the median plane based on the anthropometry of the listener's pinnae," *J. Acoust. Soc. Am.*, vol. 136, no. 1, pp. 317–333, July 2014.

[6] 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. Am.*, vol. 137, no. 2, pp. 690–701, February 2015.

[7] 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.

[8] 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.

[9] 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. Digital Audio Effects (DAFx-15)*, Trondheim, Norway, December 2015, pp. 231–236.

[10] 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 and Music Computing Conf. (Nordic SMC 2019)*, Stockholm, Sweden, Nov. 2019, pp. 5–8.

[11] S. Spagnol, K. B. Purkhús, S. K. Björnsson, and R. Unnthórsson, "The Viking HRTF dataset," in *Proc. 16th Int. Conf. Sound and Music Computing (SMC 2019)*, Malaga, Spain, May 2019, pp. 55–60.

[12] 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.

[13] S. Müller and P. Massarani, "Transfer-function measurement with sweeps," *J. Audio Eng. Soc.*, vol. 49, no. 6, pp. 443–471, June 2001.

[14] B. Friedlander and B. Porat, "The modified Yule-Walker method of ARMA spectral estimation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-20, no. 2, pp. 158–173, March 1984.

[15] 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.

[16] A. F. Hayes and L. Cai, "Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation," *Behav. Res. Methods*, vol. 39, no. 4, pp. 709–722, Nov. 2007.

[17] 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.

[18] 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.

[19] 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.

[20] 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, January 2017.

[21] S. Spagnol, "HRTF selection by anthropometric regression for improving horizontal localization accuracy," *IEEE Signal Process. Lett.*, vol. 27, pp. 590–594, Apr. 2020.

[22] S. Spagnol, "Auditory model based subsetting of head-related transfer function datasets," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP 2020)*, Barcelona, Spain, May 2020, pp. 391–395.

[23] R. Miccini and S. Spagnol, "HRTF individualization using deep learning," in *Proc. 2020 IEEE Conf. Virtual Reality and 3D User Interfaces Work. (VRW 2020)*, Athens, GA, USA, Mar. 2020, pp. 390–395.

[24] F. Avanzini, S. Spagnol, A. Rodá, and A. De Götzen, "Designing interactive sound for motor rehabilitation tasks," in *Sonic Interaction Design*, K. Franinovic and S. Serafin, Eds. Cambridge, MA, USA: MIT Press, March 2013, ch. 12, pp. 273–283.

[25] S. Spagnol, G. Wersényi, M. Bujacz, O. Balan, M. Herrera Martínez, A. Moldoveanu, and R. Unnthórsson, "Current use and future perspectives of spatial audio technologies in electronic travel aids," *Wireless Comm. Mob. Comput.*, vol. 2018, p. 17 pp., March 2018.