ON THE RELEVANCE OF THE DIFFERENCES BETWEEN HRTF MEASUREMENT SETUPS FOR MACHINE LEARNING

Johan Pauwels

Queen Mary University of London
j.pauwels@qmul.ac.uk

Lorenzo Picinali

Imperial College London
l.picinali@imperial.ac.uk

ABSTRACT

As spatial audio is enjoying a surge in popularity, data-driven machine learning techniques that have been proven successful in other domains are increasingly used to process head-related transfer function measurements. However, these techniques require much data, whereas the existing datasets are ranging from tens to the low hundreds of datapoints. It therefore becomes attractive to combine multiple of these datasets, although they are measured under different conditions. In this paper, we first establish the common ground between a number of datasets, then we investigate potential pitfalls of mixing datasets. We perform a simple experiment to test the relevance of the remaining differences between datasets when applying machine learning techniques. Finally, we pinpoint the most relevant differences.

Index Terms— HRTF, spatial audio, machine learning

1. INTRODUCTION

The subtle filtering effect caused by the absorption and scattering of sound on the ears, head and torso allow us to perceive the direction of audio sources. Accounting for these effects in the binaural rendering of virtual and augmented sonic environments is therefore necessary in order to achieve maximal realism and immersion. Due to an increasing number of applications in communication, healthcare and entertainment, the technology to create such environments has recently received growing interest from both academic and industry researchers.

Because of the differences in morphology, the filtering is specific to each individual. A personalised head-related transfer function (HRTF) encodes the direction-dependent filters in the frequency domain, whereas the equivalent representation in the time domain is called a head-related impulse response (HRIR). The acquisition of personalised HRTFs can be done in multiple ways [1], of which acoustic measurement and numerical simulation are the most common. These processes all require a dedicated and expensive setup, and are time-intensive, which makes them scale badly to a larger population.

In order to come come up with more scalable ways of obtaining individual HRTFs, several approaches have been designed and evaluated [2]. A promising approach is to use data-driven techniques, for example [3, 4, 5] use auto-encoders on HRTFs to discover a latent space that is subsequently used to enable the synthesis of HRTFs in various ways. Alternatively, [6, 7, 8, 9] synthesise HRTFs from anthropometric measurements, where the last uses pinna images as additional input. Another type of input that is used to generate HRTFs is a 3D model, as examined in [10]. Finally, spatial interpolation of HRTFs is explored in [11, 12].

Although one motivation behind much of this research is to create procedures that limit the need for the current complex HRTF acquisition methods, the development of these new techniques relies very much on the availability of measured HRTF data. More data typically results in better performance, whereas existing datasets contain only tens to low hundreds of examples. A potential breakthrough could be if we could simply mix data acquired from different measurement setups; whether that is possible or not, and to which extent, is the subject of this paper.

The various measurement setups result in multiple variable elements [13], so first we need to harmonise the data to remove a number of known differences. Then we test the relevance of the remaining differences for machine learning (ML) applications. We postulate that the subjects measured in different measurement setups are all independent samples of the same population. Therefore, in the absence of systematic differences in measurement setups, a classifier should not be able to distinguish different setups to a level higher than chance.

To test this hypothesis, a classifier will be trained to predict from which dataset an HRTF is taken, i.e. under which setup it has been measured. From a practical standpoint, the resulting system has no direct use. There are no HRTFs around for which this information is unknown, but this procedure can teach us whether ML algorithms can potentially exploit dataset-specific characteristics as a shortcut in more realistic ML tasks as well. For example, if an ML algorithm is able to pick up these differences, there is a chance for a machine learning system to internally create distinct sub-models for each setup, which voids the benefits of using multiple datasets to increase training data, and will eventually fail to generalise to unseen datasets. Whether the differences between HRTFs obtained from different setups are perceptually relevant or not is another question, which won't be addressed here.

Apart from the practical implications, also on the conceptual level an HRTF should only contain the filtering effects a specific person has on sound. Undesired characteristics of the setup would otherwise leak into binaural renderings made with that HRTF.

The differences between measurement setups have previously been studied by the "Club Fritz" project [14], where the same dummy head was acoustically measured in different laboratories and under different conditions. These measurements were compared also with HRTFs obtained by numerical simulation starting from a geometric model of the head and pinnas [15]. The main difference of our study compared to the Club Fritz project is that we look at multiple human subjects per measurement setup, with no subjects measured under multiple conditions, as opposed to a single dummy head. This introduces additional variance in the measurements, but we argue that this is counteracted by the use of multiple subjects. Only systematic differences are relevant, random variations will

be be smoothed out. Furthermore, we are not studying symmetry or inter-aural differences; only single ears are processed and the measurements of right ears are mirrored to reduce variance.

2. ESTABLISHING A COMMON GROUND FOR ALL DATASETS

In order to make the experiment as insightful and fair as possible, it is important to remove known differences between measurement setups as much as possible. Some of these differences will manifest themselves just in the numerical values of the HRTFs, whereas others directly impact the format of the HRTF. Notably, sample rate, HRIR duration and the measurement positions all affect the storage format of the data.

We start by selecting ten datasets of HRTFs, based on their public availability and the presence of shared measurement positions. These datasets are listed in Table 1. Eight out of ten are acoustically measured, the remaining two are created with Boundary Element Method simulations from 3D meshes [23]. In order to avoid any potential spectral distortion associated with interpolating spatial positions [24], we only use the intersection of the positions in all datasets, without spatial resampling. In this sense, our harmonisation procedure is more conservative than the one proposed in [25], but otherwise it is similar. For simplicity, we restrict ourselves to the horizontal plane[1]. The resulting twelve positions are $0°, \pm30°, \pm60°, \pm90°, \pm120°, \pm150°, 180°$.

To minimise the differences in loudspeaker and microphone characteristics between recording setups, free-field compensated versions of all HRTFs are used. Not all datasets provide raw recordings and a free-field reference separately, so the versions compensated by the original authors are used. The exact signal processing steps to obtain these is itself a source of variation [14]. Therefore minimum-phase versions are calculated to compensate for the differences in onset due to varying windowing parameters.

A common samplerate and HRIR duration is determined by taking the lowest value out of all datasets, respectively 44.1 kHz and 5.333 ms. All HRIRs are resampled and truncated to these values, giving 235 samples long HRIRs.

Because the amplitude of the stimuli and the distance to the loudspeakers differ according to setup, the HRIRs need to be scaled by a dataset-specific factor. We take the measurement with the highest RMS (at the ipsilateral position) for both ears of each subject, then calculate the median over an entire dataset. All datasets are then scaled by their respective factor such that the median of all loudest measurement positions is unity.

The resulting HRIRs are then converted to HRTFs using the Fourier transform, and only their magnitude is retained (future developments might be looking also at the phase component). The DC offset is removed and all frequency bins above 18 kHz are discarded, following [14].

The final data representation are matrices of 12 angles by 95 frequency bins for each ear of every subject in all datasets, which we believe to be harmonised as well as possible given the available HRTF formats. This will serve as input to the classifier, and its target class will be the name of the respective dataset.

---

[1]Strictly speaking, the HUTUBS dataset does not contain measurements in the horizontal plane, but at an elevation of -0.72°. Since all angles are measured up to a considerably larger tolerance, it was deemed close enough.

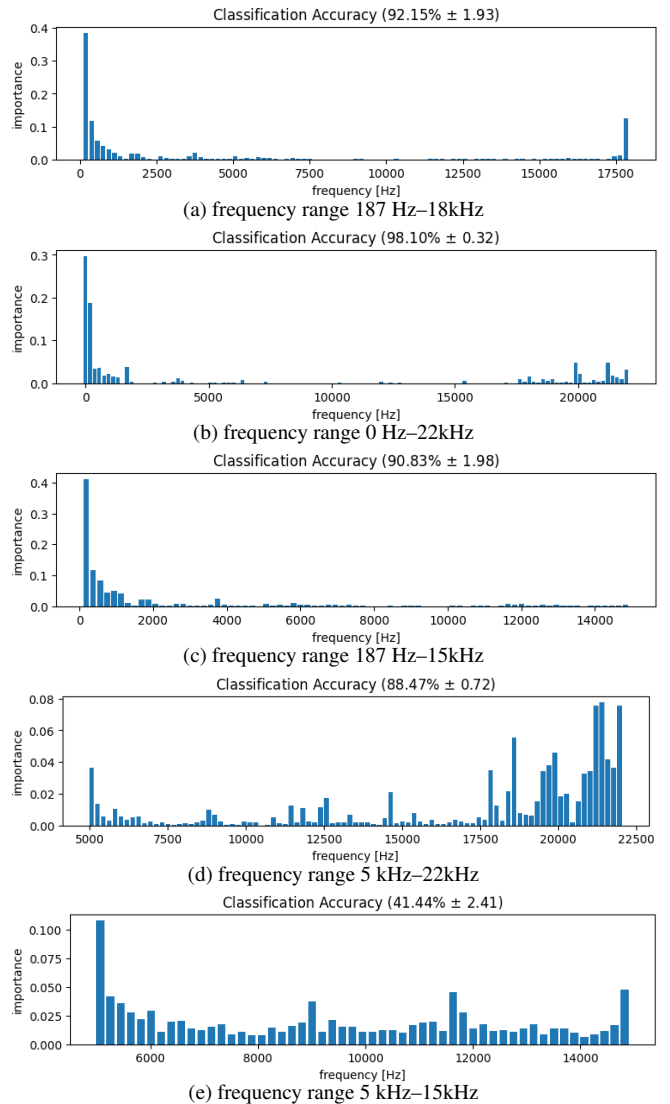## 3. CLASSIFICATION OF HRTFS BY MEASUREMENT SETUP

Fig. 1: Frequency importance and five-fold cross validation test accuracy using gradient boosted trees

As mentioned before, the measurements taken at the right ear are mirrored around the median plane such that they can be considered as additional examples of left ear measurements, again to simplify the classification problem. For each dataset, we therefore have twice the number of datapoints as the number of subjects listed in Table 1. As can be seen in the table, the number of points per dataset differ by orders of magnitude, which is challenging when training a classifier. Therefore we create a balanced dataset by taking the first 36 datapoints of each class, which is the minimum as determined by the SADIE II dataset.

Due to the spatial symmetry between ears of the same subject [26], the left and the mirrored-right datapoints can obviously not be considered as independent. Consequently, they are treated as indivisible groups when splitting the dataset into training and test

name	human subjects	samplerate	HRIR duration	radius	# directions	method
SADIE II [16]	18	96* kHz	5.333* ms	1.2m	2114*	measured
ITA Aachen[17]	46	44.1 kHz	5.805 ms	1.2m	2304	measured
ARI	221	48 kHz	5.333 ms	1.2m	1550	measured
Ircam Listen	50	44.1 kHz	11.610 ms	1.95m	187	measured
Ircam BiLi [18]	55	96 kHz	21.333 ms	2.06m	1680	measured
SONICOM	120	96* kHz	5.333* ms	1.7m	864	measured
Princeton 3D3A [19]	38	96 kHz	21.333 ms	0.76m	648	measured*
HUTUBS [20]	87	44.1 kHz	5.805 ms	1.47m	440	measured*
CHEDAR [21]	1253	48 kHz	10 ms	1m*	2522*	simulated
Widespread [22]	1005	48 kHz	10 ms	1m*	2522	simulated
Common ground	18	44.1 kHz	5.333 ms	n/a	12	n/a

Table 1: The used HRTF datasets and relevant key properties. An asterisk * signifies that variants with a different value for that property exist in the dataset, but that this particular one was used.

sets.

The effectiveness of the classifier is measured using its accuracy on the test set, i.e. the percentage of unseen examples that are recognised correctly. While every effort is made to achieve an accuracy that is as high as possible, in the ideal scenario this would be limited to the level of random guessing (10%). That would mean that the preprocessing, as detailed in section 2, is enough to make measurement setups indistinguishable from each other, and therefore that datasets can be combined without drawbacks.

Perhaps surprisingly, our results show that the average testing accuracy over 5 cross-validation folds can achieve more than 95% with relatively simple CART decision tree [27] and linear support vector machine (SVM) [28] classifiers. The average training accuracy reaches 100%, indicating some overfitting. The few incorrectly recognised examples are distributed over multiple datasets, without strong trends. We can conclude that each dataset has a pattern of frequencies observable in one or a combination of spatial positions that sets it apart from other datasets regardless of the specific subject, i.e. it acts as a fingerprint that can be used to identify the measurement setup.

Originally, these experiments were performed using deep learning based classifiers, but these were substituted for simpler classical machine learning algorithms once it was confirmed that these findings could be replicated with classical ML classifiers. Not only are the latter faster to train and easier to interpret, they also allow to draw stronger conclusions from these results. These standard classifiers operate directly on the features, in our case the HRTF spectra, meaning that the identifying patterns are formed by combining frequencies directly. In contrast, the premise of deep learning is that first optimal features get derived from the inputs before those features are combined to predict the output.

4. CLASSIFICATION OF INDIVIDUAL HRTF POSITIONS BY MEASUREMENT SETUP

The previous experiment demonstrated that, when all source positions are observed at once, at least one reveals the measurement setup. This excellent classification accuracy could be due to a single or a combination/contrasting of multiple positions, which might also differ between specific datasets. In order to study the direction-independence of the differences between measurement setups, another classification experiment is performed.

The multi-directional HRTFs are broken down into individual datapoints and the measurement angle information is discarded, while the target class remains the name of the dataset. The number of datapoints per class therefore increases by twelve, from 36 to 432. This classification is expected to be harder, since it requires finding identifying frequency patterns for each source position.

Nevertheless, a decision tree trained with 5-fold cross-validation still recognises the dataset with around 81% accuracy on average and a linear SVM achieves around 83%. The accuracy rises further when slightly more complicated classifiers [29] are used, over 86% for a non-linear SVM with a radial basis function kernel and more than 92% for gradient-boosted trees. In addition to being the highest performing classifier, gradient boosted trees have the advantage that they are easy to interpret and therefore can provide the relative importance of each frequency in the decision process, visualised in Figure 1a. Because we removed the DC offset as described in section 2, the lowest remaining frequency is $\frac{44100}{235} \approx 187$ Hz. The highest remaining frequency after harmonising the datasets is 18 kHz.

It is immediately clear that the lowest and highest frequencies are contributing the most to the classification process. Note that this importance is not correlated to signal strength. The low frequencies contain most of the signal's energy, and it is likely that the room reflections are being identified as a fingerprint of the measurement setup. In contrast, the high frequencies contain little energy, so the high-frequency "noise" is not masked and is consistent enough to contribute to the fingerprint.

To further investigate the importance of those particular frequency ranges, we change the harmonisation procedure slightly and repeat the experiments with different frequency ranges for the HRTFs. Using the full frequency range up to 22 kHz, including DC offset, makes it even easier to identify the dataset, as seen in Figure 1b. Excluding more high frequencies, which are inaudible for a large part of the adult population anyway, has relatively little effect, as seen for example in Figure 1c where all frequencies above 15 kHz are dropped. The classifier seems to partially compensate by using the low frequency info more effectively. The reverse is also true, to a lesser extent. Removing the lower frequencies (below 5 kHz in Figure 1d) makes the classifier turn to the highest frequencies to regain most of the classification accuracy. This action is far more destructive for an HRTF though. Only when both the high and low frequencies are removed does the accuracy drop significantly, though it remains significantly above chance level (about 42% for a frequency range of 5 to 15 kHz, visualised in Figure 1e).
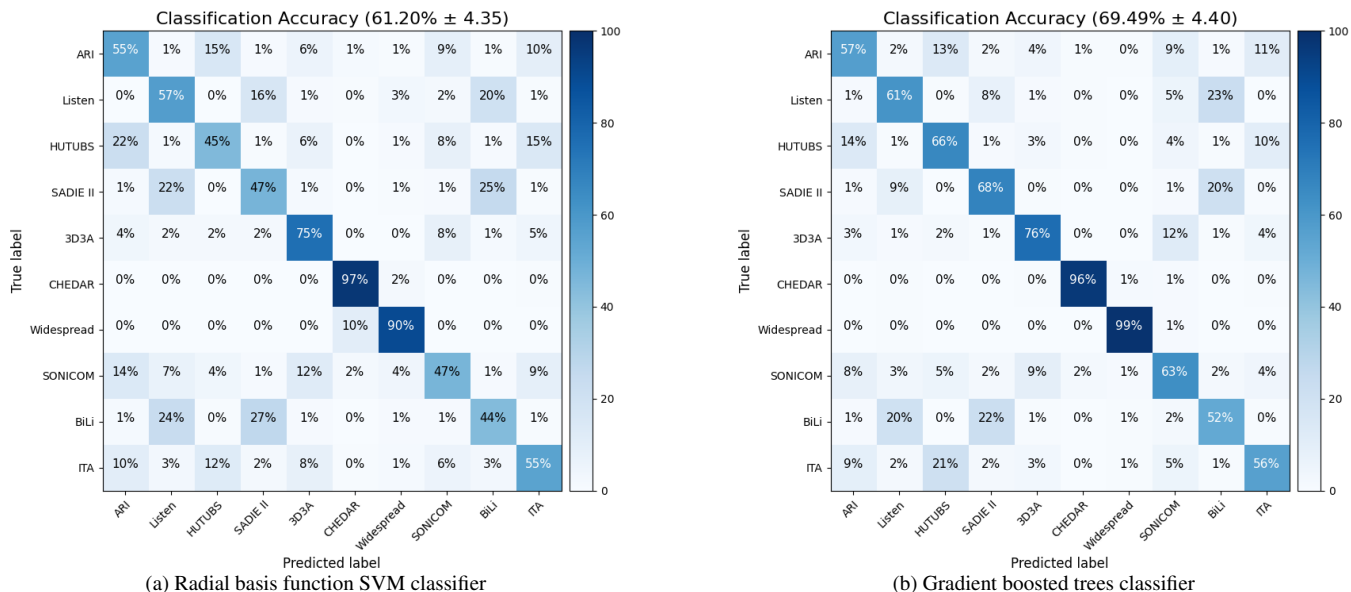
	ARI	Listen	HUTUBS	SADIE II	3D3A	CHEDAR	Widespread	SONICOM	BiLi	ITA
ARI	55%	1%	15%	1%	6%	1%	1%	9%	1%	10%
Listen	0%	57%	0%	16%	1%	0%	3%	2%	20%	1%
HUTUBS	22%	1%	45%	1%	6%	0%	1%	8%	1%	15%
SADIE II	1%	22%	0%	47%	1%	0%	1%	1%	25%	1%
3D3A	4%	2%	2%	2%	75%	0%	0%	8%	1%	5%
CHEDAR	0%	0%	0%	0%	0%	97%	2%	0%	0%	0%
Widespread	0%	0%	0%	0%	0%	10%	90%	0%	0%	0%
SONICOM	14%	7%	4%	1%	12%	2%	4%	47%	1%	9%
BiLi	1%	24%	0%	27%	1%	0%	1%	1%	44%	1%
ITA	10%	3%	12%	2%	8%	0%	1%	6%	3%	55%

(a) Radial basis function SVM classifier

	ARI	Listen	HUTUBS	SADIE II	3D3A	CHEDAR	Widespread	SONICOM	BiLi	ITA
ARI	57%	2%	13%	2%	4%	1%	0%	9%	1%	11%
Listen	1%	61%	0%	8%	1%	0%	0%	5%	23%	0%
HUTUBS	14%	1%	66%	1%	3%	0%	0%	4%	1%	10%
SADIE II	1%	9%	0%	68%	0%	0%	0%	1%	20%	0%
3D3A	3%	1%	2%	1%	76%	0%	0%	12%	1%	4%
CHEDAR	0%	0%	0%	0%	0%	96%	1%	1%	0%	0%
Widespread	0%	0%	0%	0%	0%	0%	99%	1%	0%	0%
SONICOM	8%	3%	5%	2%	9%	2%	1%	63%	2%	4%
BiLi	1%	20%	0%	22%	1%	0%	1%	2%	52%	0%
ITA	9%	2%	21%	2%	3%	0%	1%	5%	1%	56%

(b) Gradient boosted trees classifier

Fig. 2: Confusion matrices for the classification of individual source position in HRTFs limited to frequency range 1–18 kHz

5. DISCOVERING SIMILAR SETUPS THROUGH CONFUSION PAIRS

Imperfect classifiers reveal which classes are the hardest to separate through their confusion matrix. Although such analysis is inherently linked to the type of classifier (different types might struggle with different pairs) without guarantee of generalisation, it is nonetheless insightful to have a look at some confusion matrices from the previous experiment. In Figure 2, the confusion for a non-linear SVM classifier and a gradient boosted trees classifier is shown. The range of the HRTF input was limited to 1–18 kHz, such that a fair amount of errors was made, and all source positions are processed individually.

Some interesting observations are common to both classifiers. The CHEDAR and Widespread datasets are particularly easy to distinguish, likely because they both are generated as numerical simulations from morphable models of pinna, head and torso for the former and only pinna for the latter. Furthermore, the Listen, BiLi and SADIE II form a group of datasets that is harder to distinguish. The first two were measured at the same institution, so even though the measurement procedure changed somewhat in between them, such similarity can be expected. The confusion with the SADIE II dataset is harder to explain, but could be due to similar postprocessing of the measurements or sheer luck. ARI, ITA and HUTUBS form a similar group without obvious explanation.

6. CONCLUSION AND FUTURE WORK

In this paper, we demonstrated that machine learning classifiers can easily identify the measurement setup in HRTFs, for all employed source positions, and despite preprocessing aimed at harmonising datasets as well as possible. The implication is that any machine learning workflow that involves HRTFs is at risk of generalising badly to other measurement setups, and therefore that cross-dataset testing is of paramount importance. It also means that simply mixing (harmonised) datasets to increase the number of training examples will not automatically lead to an increase in robustness.

It would be interesting to pinpoint the unique identifying characteristics for each dataset in the future, possibly looking also at phase-related differences. This could be done by training binary classifiers between a single dataset and all others. The effect of different HRTF representations should be examined as well, for instance the spherical harmonics decomposition used by [30] to combine multiple datasets. Furthermore, the perceptual relevance of all differences should be studied.

Ideally, the differences between measurement setups are not just diagnosed, but also reduced. It might be possible to do so manually through specific compensation filters once the exact differences are pinpointed. Alternatively, since the experimental procedure proposed in this paper is effectively a discriminator that should be fooled, it could be possible to use adversarial training to create a system that learns how to perform such equalisation automatically. In any case, perceptual evaluation of such compensation and/or other modifications would be required to ensure that the HRTFs keep their relevant characteristics. After all, flat spectra are impossible to distinguish, but also useless.

7. ACKNOWLEDGEMENTS

8. REFERENCES

[1] Corentin Guezenoc and Renaud Seguier, "HRTF Individualization: A Survey," in *Audio Engineering Society Convention 145*. Oct. 2018, Audio Engineering Society.

[2] Lorenzo Picinali and Brian FG Katz, "System-to-user and user-to-system adaptations in binaural audio," in *Sonic Interactions in Virtual Environments*, pp. 115–143. Springer, 2023.

[3] Kazuhiko Yamamoto and Takeo Igarashi, "Fully perceptual-based 3D spatial sound individualization with an adaptive vari-

ational autoencoder," *ACM Transactions on Graphics*, vol. 36, no. 6, pp. 212:1–212:13, Nov. 2017.

[4] Tzu-Yu Chen, Tzu-Hsuan Kuo, and Tai-Shih Chi, "Autoencoding HRTFS for DNN Based HRTF Personalization Using Anthropometric Features," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 271–275.

[5] Riccardo Miccini and Simone Spagnol, "HRTF Individualization using Deep Learning," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, Mar. 2020, pp. 390–395.

[6] Chan Jun Chun, Jung Min Moon, Geon Woo Lee, Nam Kyun Kim, and Hong Kook Kim, "Deep Neural Network Based HRTF Personalization Using Anthropometric Measurements," in *Audio Engineering Society Convention 143*. Oct. 2017, Audio Engineering Society.

[7] Haytham Fayek, Laurens van der Maaten, Griffin Romigh, and Ravish Mehra, "On Data-Driven Approaches to Head-Related-Transfer Function Personalization," in *Audio Engineering Society Convention 143*. Oct. 2017, Audio Engineering Society.

[8] Yuxiang Wang, You Zhang, Zhiyao Duan, and Mark Bocko, "Global HRTF Personalization Using Anthropometric Measures," in *Audio Engineering Society Convention 150*. May 2021, Audio Engineering Society.

[9] Geon Woo Lee and Hong Kook Kim, "Personalized HRTF Modeling Based on Deep Neural Network Using Anthropometric Measurements and Images of the Ear," *Applied Sciences*, vol. 8, no. 11, pp. 2180, Nov. 2018.

[10] Yaxuan Zhou, Hao Jiang, and Vamsi Krishna Ithapu, "On the Predictability of HRTFs from Ear Shapes Using Deep Networks," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, June 2021, pp. 441–445.

[11] Grady Kestler, Shahrokh Yadegari, and David Nahamoo, "Head Related Impulse Response Interpolation and Extrapolation Using Deep Belief Networks," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 266–270.

[12] Pongsakorn Siripornpitak, Isaac Engel, Isaac Squires, Samuel J. Cooper, and Lorenzo Picinali, "Spatial up-sampling of HRTF sets using generative adversarial networks: A pilot study," *Frontiers in Signal Processing*, vol. 2, 2022.

[13] Song Li and Jürgen Peissig, "Measurement of Head-Related Transfer Functions: A Review," *Applied Sciences*, vol. 10, no. 14, pp. 5014, Jan. 2020.

[14] Areti Andreopoulou, Durand R. Begault, and Brian F. G. Katz, "Inter-Laboratory Round Robin HRTF Measurement Comparison," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 5, pp. 895–906, Aug. 2015.

[15] Raphaël Greff and Brian F. G. Katz, "Round Robin Comparison of HRTF Simulation Systems: Preliminary Results," in *Audio Engineering Society Convention 123*. Oct. 2007, Audio Engineering Society.

[16] Cal Armstrong, Lewis Thresh, Damian Murphy, and Gavin Kearney, "A Perceptual Evaluation of Individual and Non-Individual HRTFs: A Case Study of the SADIE II Database," *Applied Sciences*, vol. 8, no. 11, pp. 2029, Nov. 2018.

[17] Ramona Bomhardt, Matias de la Fuente Klein, and Janina Fels, "A high-resolution head-related transfer function and three-dimensional ear model database," *Proceedings of Meetings on Acoustics*, vol. 29, no. 1, pp. 050002, Nov. 2016.

[18] Thibaut Carpentier, Hélène Bahu, Markus Noisternig, and Olivier Warusfel, "Measurement of a head-related transfer function database with high spatial resolution," in *7th Forum Acusticum(EAA)*, Krakow, Poland, Sept. 2014.

[19] Rahulram Sridhar, Joseph G. Tylka, and Edgar Choueiri, "A Database of Head-Related Transfer Functions and Morphological Measurements," in *Audio Engineering Society Convention 143*. Oct. 2017, Audio Engineering Society.

[20] Fabian Brinkmann, Manoj Dinakaran, Robert Pelzer, Peter Grosche, Daniel Voss, and Stefan Weinzierl, "A Cross-Evaluated Database of Measured and Simulated HRTFs Including 3D Head Meshes, Anthropometric Features, and Headphone Impulse Responses," *Journal of the Audio Engineering Society*, vol. 67, no. 9, pp. 705–718, Sept. 2019.

[21] Slim Ghorbal, Xavier Bonjour, and Renaud Séguier, "Computed HRIRs and Ears Database for Acoustic Research," in *Audio Engineering Society Convention 148*. May 2020, Audio Engineering Society.

[22] Corentin Guezenoc and Renaud Séguier, "A wide dataset of ear shapes and pinna-related transfer functions generated by random ear drawings," *The Journal of the Acoustical Society of America*, vol. 147, no. 6, pp. 4087–4096, June 2020.

[23] Harald Ziegelwanger, Wolfgang Kreuzer, and Piotr Majdak, "Mesh2HRTF: An open-source software package for the numerical calculation of head-related transfer functions," in *Proceedings of the 22nd International Congress on Sound and Vibration*, Florence, IT, 2015.

[24] Xiao-Li Zhong and Bo-Sun Xie, "Maximal azimuthal resolution needed in measurements of head-related transfer functions," *The Journal of the Acoustical Society of America*, vol. 125, no. 4, pp. 2209–2220, Apr. 2009.

[25] Areti Andreopoulou and Agnieszka Roginska, "Towards the Creation of a Standardized HRTF Repository," in *7th Forum Acusticum(EAA)*. 2011, Audio Engineering Society.

[26] Xiao-Li Zhong, Feng-chun Zhang, and Bo-Sun Xie, "On the spatial symmetry of head-related transfer functions," *Applied Acoustics*, vol. 74, no. 6, pp. 856–864, June 2013.

[27] Leo Breiman, *Classification and Regression Trees*, Routledge, New York, Oct. 2017.

[28] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin, "LIBLINEAR: A Library for Large Linear Classification," *Journal of Machine Learning Research*, vol. 9, no. 61, pp. 1871–1874, 2008.

[29] Trevor Hastie, Robert Tibshirani, and Jerome Friedman, *The Elements of Statistical Learning*, Springer Series in Statistics. Springer, New York, NY, 2009.

[30] Jingwei Xi, Wen Zhang, and Thushara D. Abhayapala, "Magnitude Modelling of Individualized HRTFs Using DNN Based Spherical Harmonic Analysis," in *2021 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct. 2021, pp. 266–270.