DOI: 10.61782/fa.2023.0984

IMPACT OF INPUT PREPROCESSING FOR HRTF ELEVATION CLASSIFICATION OVER MULTIPLE DATASETS

Juan A. De Rus[1*] **Jesus Lopez-Ballester**[1] **Mario Montagud**[1,2]
Francesc J. Ferri[1] **Jose J. Lopez**[3] **Maximo Cobos**[1]
[1] Computer Science Department, Universitat de València, Spain
[2] i2CAT Foundation, Barcelona, Spain
[3] iTEAM, Universitat Politècnica de València, Spain

ABSTRACT

The localization of sound sources on the horizontal plane is mainly aided by perceived interaural level and time differences. However, identifying elevation cues in Head-Related Transfer Functions (HRTFs) remains challenging. Spectral cues play a key role in localizing sources in elevation and are highly individual, resulting from anatomic characteristics specific to each person, such as the shape of the pinnae, head, or torso. In a previous study, we proposed a simple 1D convolutional neural network (CNN) trained to classify HRTF signals into different elevation sectors to identify spectral elevation cues using explainability techniques. Although the model obtained promising results, it was only trained and validated on the CIPIC database. In this work, we focus on developing a model that can generalize across multiple HRTF datasets to achieve good classification performance across various subjects and measurements. Since each dataset is obtained in different conditions (e.g., source signal used, distance between emitters and receivers, spatial resolution, calibration), the preprocessing of the data may significantly impact the overall inter-dataset model performance. We explore different preprocessing techniques and evaluate their impact on the classification task to select meaningful standardization strategies for working with multiple HRTF datasets.

Keywords: *HRTF, elevation cues, dataset preprocessing, convolutional neural networks*

1. INTRODUCTION

HRTFs, or Head-Related Transfer Functions, describe the transmission of sound from a point in space to the human ear canal [1]. This concept finds application in various fields, such as personalized sound design for individuals, including noise cancellation [2] and the creation of immersive Virtual Reality (VR) environments [3]. Each person's HRTF is unique, influenced by factors such as head, torso, shoulder, and pinnae shape and other characteristics [4].

The main goal of our study is to develop a Convolutional Neural Network (CNN) classification model that effectively utilizes data from multiple HRTF datasets to identify the elevation location in the median plane. We aim to explore which localization cues are crucial in determining the elevation of a given HRTF response, and hypothesize that the model's ability to classify the elevation sector will inherently capture these features, allowing for generalization across all datasets. To achieve this, we employ data standardization techniques and compare the model's performance across the different datasets, highlighting the impact of dataset differences on the results.

1.1 Challenges in elevation sound source localization

To localize a sound source in the horizontal plane, various localization cues such as interaural time difference (ITD), interaural level difference (ILD), spectral cues (SC), and horizontal plane directivity (HPD) can be em-

[*] ***Corresponding author***: *juan.rus@uv.es.*

ployed. However, these cues are not equally effective for elevation location. In the vertical plane, we are limited to using spectral cues produced by anthropometric factors, including reflections and refractions from the pinna and torso [5].. Therefore, while ITD and ILD are usually sufficient to determine the horizontal localization of a sound source, spectral cues play a crucial role in determining the elevation location.

Previous research has demonstrated that the pinna's distortions manifest as spectral cues for elevation location beyond the 4 kHz frequency band [6], extending up to 10 kHz [7]. Additionally, at 12 kHz, a spectral cue appears as a peak, indicating sound coming from behind the listener [8]. Another effect to note is that the significant notch shifts to lower frequency bands as the sound moves from the zenith towards the lower frontal half of the median plane [9].

1.2 Past work

In a previous study [10], we trained a CNN to classify HRTFs in various elevation sectors ranging from "Forward-Down" to "Back-Down" and the laterals as shown in Tab. 1 using data from the CIPIC dataset [11]. We use spherical coordinates with "side" convention which uses lateral angle ranging $[-90, 90]$ in horizontal plane and polar angle ranging $[-90, 270]$ in the median plane. We chose CNNs because of their pattern recognition capabilities [12], which have been successfully utilized in capturing spatial audio features in HRTFs [13], as well as in other sound-oriented tasks like acoustic scene classification [14], music tagging [15], speech recognition [16], or automatic discrimination between front and back locations in binaural recordings [17].

Our fully convolutional model consisted of three 1D convolutional blocks with ReLU activation and max-pooling between blocks, followed by a last convolutional layer and global average pooling to summarize filter responses before the last dense layer with softmax activation (Fig. 1). This simple model achieved significant accuracy. More details of the model and its training can be found in [10]

Next, we explored the use of common explainable artificial intelligence (XAI) techniques [18] to determine what the model was examining to make its predictions, including which parts of the data were most important or salient for the prediction. We compared the results with those of the literature. The two XAI techniques we employed were Class Activation Mapping (CAM) [19] and Gradient CAM (GradCam) [20].

Table 1. Differences in recording setup: ear and source distances (meters), source signal, and anechoic chamber used.

Class	Polar Angle	Lateral Angle
Front Down	$[-90, -20]$	$[-60, 60]$
Front Level	$(-20, 20]$	$[-60, 60]$
Front Up	$(20, 70]$	$[-60, 60]$
Up	$(70, 110]$	$[-60, 60]$
Back Up	$(110, 160]$	$[-60, 60]$
Back Level	$(160, 200]$	$[-60, 60]$
Back Down	$(200, 270]$	$[-60, 60]$
Lateral Up	$[0, \infty)$	$\|lateral\| > 60$
Lateral Down	$(-\infty, 0)$	$\|lateral\| > 60$

1.2.1 Findings

We extracted significant frequency bands that presented more saliency when being classified corresponding to different elevation sectors in the data We also found high saliency (which may indicate possible elevation cues) in the low frequency band below 500 Hz for the backward regions from "back-down" to "up" not present in the forward regions (Fig. 2).

Furthermore, we found a possible effect of complementarity between opposite regions like "front-level" and "back-level" with opposite saliencies on the 2-10 KHz frequency band being almost identical out of this range (Fig. 3).

2. INTERDATASET MODEL ACCURACY

As previously explained, the focus of this work is on the development of models for the elevation classification of HRTFs, with a particular emphasis on interdataset results. In other words, we are interested in examining how well a model performs when trained on one HRTF dataset but tested on data from a different dataset.

2.1 Data format used

To address the challenge of dealing with different datasets, each with its unique characteristics and idiosyncrasies, various efforts have been made to standardize data formats, such as the Marl-Nyu format for Matlab [21]. In our case, to facilitate working with different datasets, we opted to use the Spatially Oriented Format for Acoustics (SOFA) Conventions [22]. The SOFA format is characterized by the inclusion of self-contained information regard-

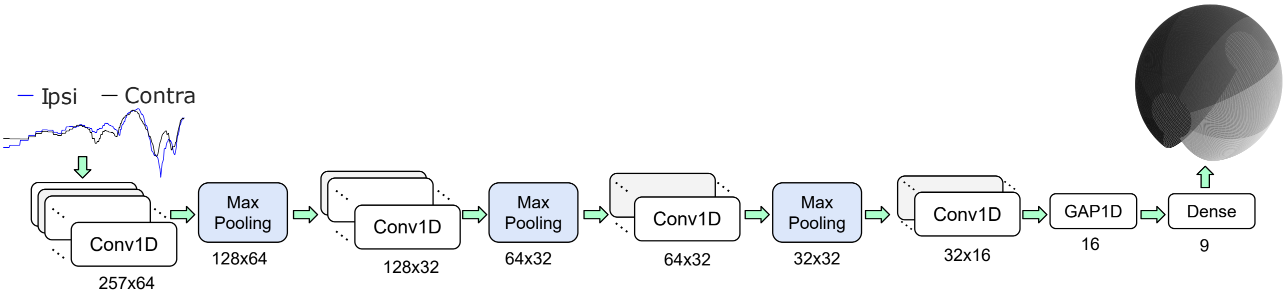


Figure 1. Topology of the convolutional architecture developed on previous study to classify HRTFs into nine elevation sectors.

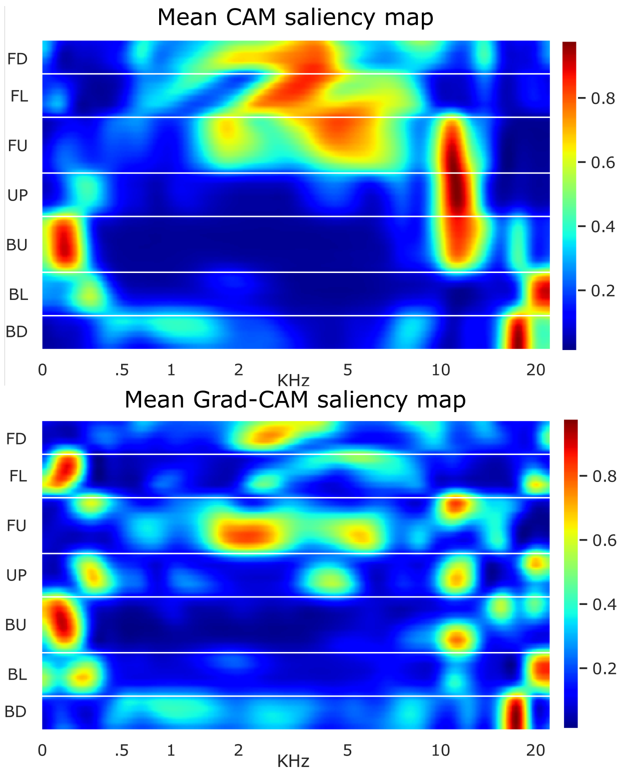


Figure 2. CAM (top) and Grad-CAM (down) saliency maps averaged across subjects and azimuth, showed per elevation class.

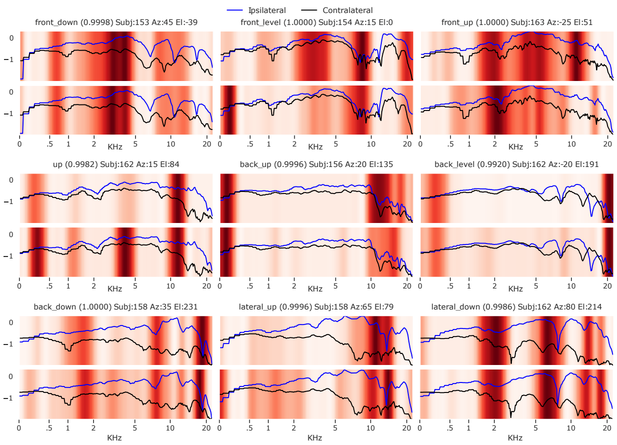


Figure 3. CAM (top) and Grad-CAM (bottom) saliency maps over the most representative sample of each class. The predicted class probability is given in parentheses. The color bar is red-scaled, then white shades indicate low relevance and red ones high relevance.

ing the measurement setup description and all the relevant elements, such as the listener, the source, and the room, in each file.
Moreover, the SOFA format allows for the use of the data as HRTFs in the frequency domain or as Head-Related Impulse Responses (HRIRs) in the time domain. In this work, we use HRIRs and apply our own preprocessing. All the data used in this work has been acquired in .SOFA format and is currently available at `https://www.sofaconventions.org/mediawiki/index.php/Files`.

2.2 Used datasets, characteristics and differences

In this work, we utilized data from 11 distinct HRTF datasets: RIEC [23], FABIAN [24], CIPIC [11], HUTUBS [25], AACHEN [26], LISTEN [27], ARI [28], Crossmod [29], SADIE [30], BiLi [31], and 3D3A [32]. In order to highlight the main differences between these datasets that could negatively impact the compatibility of models trained on different data, we extracted information from various sources including the datasets' websites, associated papers, and SOFA files.

2.2.1 *Recording conditions: chambers, source Signals, distances*

Table 2. Differences in recording setup: ear and source distances (meters), source signal, and anechoic chamber used.

Data	**Ear**	**Source**	**Signal** [1]	**Anech.**
RIEC	0.09	1.5	OATSP	Yes
Fabian	0.0662	1.7	Swept S.	Yes
CIPIC	0.09	1.5	R. Noise	No [2]
Hutubs	0.75	1.47	M.E.S.S.	Yes
AACHEN	0.07	1.2	Swept S.	Semi
Listen	0.09	2.06	Exp. S.S.	Yes
ARI	0.09	1.2	Exp. S.S.	Semi
Crossmod	0.09	2.06	Exp. S.S.	Yes
SADIE	0.09	1.2	O. Swept S.	Yes
BiLi	0.09	2.06	Exp. S.S.	Yes
3D3A	0.085	0.76	M.E.S.S.	Yes

As we can see on Tab. 2 there is no standard for the average distance between ears and the distance to the source, which can impact the resulting HRTFs' amplitude levels. Moreover, in the datasets we have observed, Swept Sines derived techniques are predominantly used for the source signal, although other methods such as pseudo-aleatory noise and the Optimized Aoshima's Time-Stretched Pulse (OATSP) [33] have also been employed.

However, the quality of the resulting HRTFs can be affected by the chambers used for the recordings, which are not always anechoic. This can introduce unwanted reflections and distortions that affect the accuracy and reliability of any subsequent analysis or modeling based on these HRTFs.

2.2.2 *Sample data: duration, sampling Rate, and number of subjects*

Table 3. Differences in Sample Data Across Datasets

Data	**Samples**	**SR (kHz)**	**Nº Subj.**
RIEC	512	48	105
Fabian	256	44.1	22
CIPIC	200	44.1	45
Hutubs	256	44.1	96
AACHEN	256	44.1	48
Listen	8192	44.1	50
ARI	256	48	200
Crossmod	8192	44.1	24
SADIE	256	48	20
BiLi	512	96	56
3D3A	2048	96	38

The variations in sampling rate (SR) and total number of samples for each HRTF are significant factors to consider when combining data from different datasets for model training. During preprocessing, it is important to take into account the above factors, ensuring that the final data representations used for model training are consistent across the different datasets.

2.2.3 *Spatial and class distribution*

The spatial distribution of measurements in terms of the angles taken can vary significantly across datasets. Differences can arise in the density of samples, the range of angles in elevation, and the evenness of sample distribution. In some datasets, there are imbalances with a greater density of samples along either the elevation or the azimuth angles. Additionally, there may be differences in the class distribution of the datasets, which could impact the performance of the HRTF models trained on them (See Tab. 4).

2.2.4 *Post-processed samples*

In addition, some datasets have undergone post-processing before their release, as summarized in Tab. 5. This post-processing includes equalization, frequency cut-offs, numerical simulations, temporal windowing, gain calibration, low-frequency compensation, low-frequency extension, and diffuse field equalization. Combining raw and processed data may result in some effects.

[1] OATSP: Optimized Aoshima's Time-Stretched Pulse
Swept S: Swept Sine
R. Noise: Pseudo-Aleatory Noise
M.E.S.S.: Multiple Exponential Sine Sweep

[2] Room with absorbers

Table 4. Spatial Coordinates and Class Balance

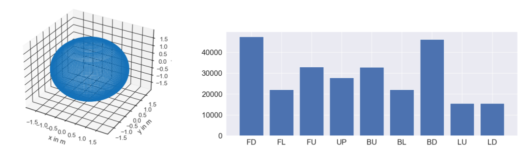
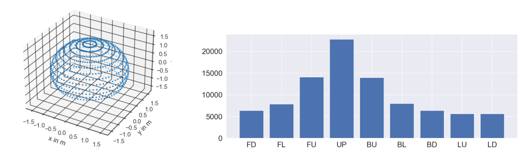
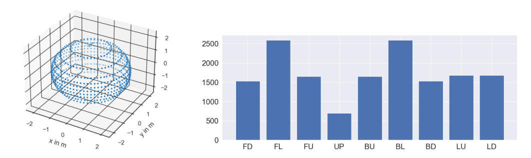
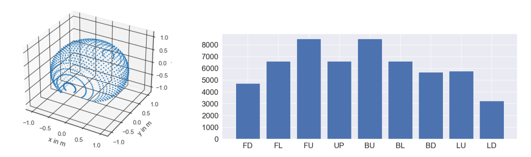
Even FABIAN	
Azimuth RIEC	
Azimuth CROSSMOD	
Elevation CIPIC)	

Table 5. Processing summary of datasets in SOFA files: Raw data, Equalization, Frequency Cut Off, Temporal Windowing, Low Frequency Compensation, Gain Calibration, Diffuse Field Equalization, Numerical Simulation.

Dataset	**Raw**	**Eq**	**FCO**	**TW**	**LFC**	**GC**	**DF Eq**	**NS**
RIEC				✓	✓	✓		
Fabian								✓
CIPIC	✓							
Hutubs			✓					
AACHEN			✓					
Listen	✓							
ARI			✓					
Crossmod	✓							
SADIE				✓	✓		✓	
BiLi		✓		✓				
3D3A		✓						

3. PRE-PROCESSING TECHNIQUES TESTED

Our objective was to standardize data from different datasets recorded under varying conditions in order to improve the performance of models trained with one type of data in predicting the location of HRTFs from different datasets. To accomplish this task, we experimented with various preprocessing techniques applied to the HRTFs in the frequency domain and used them to train different models.

3.1 Normalization

We tested three normalization techniques: no normalization, min-max normalization, and Average Equator Energy normalization [34]. In min-max normalization, each HRTF sample has a peak at 1 and a notch at -1. In Average Equator Energy normalization, we divide each HRTF sample by the average HRTF energy at the equator (zero elevation angle).

3.2 Mel warping

Although this technique was not used for standardizing the data, we employed it to evaluate the model's response by converting the HRTFs to the Mel Scale, which is a frequency scale that is more aligned with human perception. To achieve this, we divided the frequency range into equally spaced points on the Mel Scale and selected the frequency bins that were closest to their respective frequencies in Hz.

3.3 Frequency Cut off

We conducted tests with different effective ranges of frequencies, including the full range [0-22050 Hz] as well as several restricted ranges such as [20-22050 Hz], [20-16000 Hz], [20-22000 Hz], [50-22050 Hz], [50-16000 Hz], [50-22000 Hz], [500-22050 Hz], [500-16000 Hz], and [500-22000 Hz]. These tests were performed with the understanding that some datasets have a restricted functional range of frequencies and some may have lower frequencies simulated or processed. In addition, since in our previous study we found useful elevation cues below 5 kHz, we wanted to evaluate results when frequencies below are suppressed.

3.4 Scale amplitude

We experimented with two different HRTF scales, linear and log10. The linear scale maintains a constant amplitude ratio between the input and output signals, while the log10 scale compresses the amplitude of the input signals to produce a more perceptually uniform output.

4. EXPERIMENTS

We conducted extensive experiments on each dataset, using the various methods and parameters described earlier. We trained individual models for each dataset, as well as a combined model with a small percentage of samples from each dataset (10% of the training data from the other datasets) and some combined models with data from only a few selected datasets (the percentage varied depending on the number of datasets selected). We then tested each model against its own test data as well as test data from different datasets. We were careful to use different subjects for training and testing, and to maintain consistency in the preprocessing of data; for instance, if a model was trained with data using a Mel Warping transformation, we tested that model against different datasets with the same preprocessing. However, due to time constraints and the need to train a large number of models for each experiment, we limited the training of each model to 100 epochs, with a patience of 20 (i.e., the training was stopped if there was no improvement in performance for 20 epochs).

4.1 Results on preprocessing influence

Although some methods showed better accuracy results for specific datasets and conditions, overall, we did not find any parameter that was statistically significant than the others for every case (see Tab. 6). We conducted experiments by changing only one variable at a time while keeping the rest fixed, and after analyzing the results, we selected the final tests with the following parameters: Average Equator Energy Normalization, Full Range of frequencies (without cut off), no Mel Warping, and linear amplitude. These parameters were chosen because they yielded slightly better results.

Table 6. P-Values after testing different parameters of: Dataset, Amplitude, Normalization, Mel Warp, Frequecy Cut Off

Data	Ampl	Norm	Mel W.	Freq Cut
2.12e-9	.64	.9	.35	.0503

4.2 Inter-dataset results

As expected, each model achieved good results when tested against the dataset to which its train samples belong. However, we obtained poor accuracy results against datasets other than the one used for training. The combined dataset achieved the best result (Fig. 4), as it achieved acceptable accuracy against all datasets, even though it was trained with a reduced number of samples from each. It is worth noting that the limited number of training epochs may have affected the results.

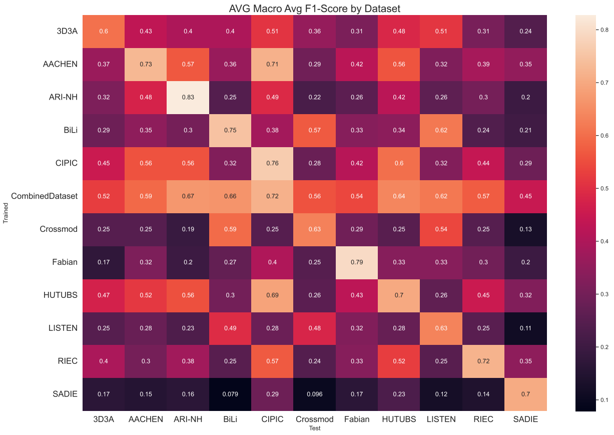


Figure 4. Inter-Dataset Accuracy Results

5. DISCUSSION

We observed an interesting phenomenon where certain sets or cliques of datasets demonstrated a higher accuracy within themselves than against other datasets. Two distinct cliques were identified: one comprising the Crossmod, BiLi, and Listen datasets, and the other consisting of CIPIC, Riec, Hutubs, and AACHEN.

The datasets in the first clique, Crossmod, BiLi, and Listen, share several similarities, such as the same distance between ears (0.09), same distance to the source (2.06), use of an anechoic chamber, and same source signal (Exponential Sine Sweep). Conversely, the datasets in the second clique, CIPIC, Riec, Hutubs, and AACHEN, do not appear to share any discernible characteristic that would explain the observed inter-dataset accuracy results.

6. FUTURE WORK

Our future work involves applying Explainable Artificial Intelligence (XAI) techniques to analyze the results obtained from working with various datasets. Our goal is to identify the factors responsible for the poor performance of classification models trained using different data and determine the most significant factors that contribute to this behavior. We will investigate different conditions that may have an impact on HRTF recordings, such as the distance between ears, distance to the emitter, source signal used, and data processing techniques. Additionally, we will analyze how these conditions affect the saliency of HRTF frequency bands and their importance to the classification model. Once the causes are identified, we aim to propose standardization techniques suitable for working with heterogeneous HRTF datasets in related problems such as HRTF personalization.

7. ACKNOWLEDGMENTS

This work has been supported by Grants FPU20/05384 and RYC2020-030679-I from MCIN/AEI/10.13039/501100011033, as well as by the "European Social Fund (ESF) Investing in Your Future". Grant TED2021-131003B-C21 funded by MCIN/AEI/10.13039/501100011033 and by the "EU Union NextGenerationEU/PRTR". The authors acknowledge also the Artemisa computer resources funded by the EU ERDF and Comunitat Valenciana, and the technical support of IFIC (CSIC-UV).

8. REFERENCES

[1] H. Møller, M. F. Sørensen, D. Hammershøi, and C. B. Jensen, "Head-related transfer functions of human subjects," *Journal of the Audio Engineering Society*, vol. 43, no. 5, pp. 300–321, 1995.

[2] H. Nakashima, R. Kouyama, N. Hiruma, and Y.-i. Fujisaka, "Binaural wind noise detection, cancellation and its evaluation for hearing aids based on HRTF cues," pp. 004896–004899, 2015.

[3] M. Geronazzo, E. Sikström, J. Kleimola, F. Avanzini, A. De Götzen, and S. Serafin, "The impact of an accurate vertical localization with HRTFs on short explorations of immersive virtual reality scenarios," in *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 90–97, IEEE, 2018.

[4] M. Zhu, M. Shahnawaz, S. Tubaro, and A. Sarti, "HRTF personalization based on weighted sparse representation of anthropometric features," in *2017 International Conference on 3D Immersion (IC3D)*, pp. 1–7, IEEE, 2017.

[5] G.-T. Lee, S.-M. Choi, B.-Y. Ko, and Y.-H. Park, "HRTF measurement for accurate identification of binaural sound localization cues," *arXiv preprint arXiv:2203.03166*, 2022.

[6] K. Iida and Y. Ishii, "Individualization of the head-related transfer functions on the basis of the spectral cues for sound localization," in *Principles and applications of spatial hearing*, pp. 159–178, World Scientific, 2011.

[7] A. Alves-Pinto, A. R. Palmer, and E. A. Lopez-Poveda, "Perception and coding of high-frequency spectral notches: potential implications for sound localization," *Frontiers in neuroscience*, vol. 8, p. 112, 2014.

[8] J. Hebrank and D. Wright, "Spectral cues used in the localization of sound sources on the median plane," *The Journal of the Acoustical Society of America*, vol. 56, no. 6, pp. 1829–1834, 1974.

[9] R. A. Butler and K. Belendiuk, "Spectral cues utilized in the localization of sound in the median sagittal plane," *The Journal of the Acoustical Society of America*, vol. 61, no. 5, pp. 1264–1269, 1977.

[10] J. A. De Rus, A. Lopez-García, J. Lopez-Ballester, J. J. Lopez, A. M. Torres, F. J. Ferri, M. Montagud, and M. Cobos, "On the Application of Explainable Artificial Intelligence Techniques on HRTF Data," in *24th International Congress on Accoustics Proceedings*, (Gyeongju, Korea), Oct. 2022.

[11] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No. 01TH8575)*, pp. 99–102, IEEE, 2001.

[12] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," *arXiv preprint arXiv:1511.08458*, 2015.

[13] E. Thuillier, H. Gamper, and I. J. Tashev, "Spatial audio feature discovery with convolutional neural networks," in *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp. 6797–6801, IEEE, 2018.

[14] J. Abeßer, "A review of deep learning based methods for acoustic scene classification," *Applied Sciences*, vol. 10, no. 6, p. 2020, 2020.

[15] T. Kim, J. Lee, and J. Nam, "Sample-level cnn architectures for music auto-tagging using raw waveforms," in *2018 IEEE international conference on*

acoustics, speech and signal processing (ICASSP), pp. 366–370, IEEE, 2018.

[16] S. Kwon, "A cnn-assisted enhanced audio signal processing for speech emotion recognition," *Sensors*, vol. 20, no. 1, p. 183, 2019.

[17] S. K. Zieliński, P. Antoniuk, H. Lee, and D. Johnson, "Automatic discrimination between front and back ensemble locations in hrtf-convolved binaural recordings of music," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2022, no. 1, p. 3, 2022.

[18] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, *et al.*, "Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai," *Information fusion*, vol. 58, pp. 82–115, 2020.

[19] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2921–2929, 2016.

[20] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, pp. 618–626, 2017.

[21] A. Andreopoulou and A. Roginska, "Towards the creation of a standardized HRTF repository," in *Audio Engineering Society Convention 131*, Audio Engineering Society, 2011.

[22] P. Majdak, Y. Iwaya, T. Carpentier, R. Nicol, M. Parmentier, A. Roginska, Y. Suzuki, K. Watanabe, H. Wierstorf, H. Ziegelwanger, *et al.*, "Spatially oriented format for acoustics: A data exchange format representing head-related transfer functions," in *Audio Engineering Society Convention 134*, Audio Engineering Society, 2013.

[23] K. Watanabe, Y. Iwaya, Y. Suzuki, S. Takane, and S. Sato, "Dataset of head-related transfer functions measured with a circular loudspeaker array," *Acoustical science and technology*, vol. 35, no. 3, pp. 159–165, 2014.

[24] F. Brinkmann, A. Lindau, S. Weinzierl, M. Müller-Trapet, R. Opdam, M. Vorländer, *et al.*, "A high resolution and full-spherical head-related transfer function database for different head-above-torso orientations," *Journal of the Audio Engineering Society*, vol. 65, no. 10, pp. 841–848, 2017.

[25] F. Brinkmann, M. Dinakaran, R. Pelzer, P. Grosche, D. Voss, and S. Weinzierl, "A cross-evaluated database of measured and simulated hrtfs including 3d head meshes, anthropometric features, and headphone impulse responses," *Journal of the Audio Engineering Society*, vol. 67, no. 9, pp. 705–718, 2019.

[26] R. Bomhardt, M. de la Fuente Klein, and J. Fels, "A high-resolution head-related transfer function and three-dimensional ear model database," in *Proceedings of Meetings on Acoustics 172ASA*, vol. 29, p. 050002, Acoustical Society of America, 2016.

[27] O. Warusfel, "Listen hrtf database." `http://recherche.ircam.fr/equipes/salles/listen/`, 2023.

[28] I. für Schallforschung, "Hrtf-database." `https://www.oeaw.ac.at/en/isf/das-institut/software/hrtf-database`.

[29] "Crossmod hrtfs." `https://sofacoustics.org/data/database/crossmod(hrtf)/`.

[30] C. Armstrong, L. Thresh, D. Murphy, and G. Kearney, "A perceptual evaluation of individual and non-individual hrtfs: A case study of the sadie ii database," *Applied Sciences*, vol. 8, no. 11, p. 2029, 2018.

[31] F. Rugeles Ospina, M. Emerit, and B. F. Katz, "The three-dimensional morphological database for spatial hearing research of the bili project," in *Proceedings of Meetings on Acoustics 169ASA*, vol. 23, p. 050001, Acoustical Society of America, 2015.

[32] R. Sridhar, J. G. Tylka, and E. Choueiri, "A database of head-related transfer functions and morphological measurements," in *Audio Engineering Society Convention 143*, Audio Engineering Society, 2017.

[33] Y. Suzuki, F. Asano, H.-Y. Kim, and T. Sone, "An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses," *The Journal of the Acoustical Society of America*, vol. 97, no. 2, pp. 1119–1123, 1995.

[34] Y. Zhang, Y. Wang, and Z. Duan, "HRTF Field: Unifying Measured HRTF Magnitude Representation with Neural Fields," *arXiv preprint arXiv:2210.15196*, 2022.