DOI: 10.61782/fa.2023.1051

forum acusticum 2023

HEAD-RELATED INTENSITY-BASED TRANSFER FUNCTION DATASET

Kelly Evans[1] **Jiarui Wang**[1*] **Thushara Abhayapala**[1]
Jihui Aimee Zhang[2,1]
[1] The Australian National University, Australia
[2] University of Southampton, UK

ABSTRACT

Recently, there has been growth in 3D audio applications in fields such as virtual reality (VR) and augmented reality (AR). These fields require highly realistic sound to be delivered binaurally to human users. This has generally been achieved by obtaining scalar measurements of sound pressure in or near the human ear, to create a head-related transfer function (HRTF) dataset that characterises the interference of the human anatomy. This paper presents a novel method of improving HRTF datasets by obtaining a head-related intensity-based transfer function (HRIBTF) that characterises the sound at the entrance to the ear canal in terms of intensity, rather than pressure. The vector properties of sound intensity have been proven to inherently provide more useful information in the subjective perception of the sound, which is highly applicable to experiences such as VR and AR. For each ear, the intensity vectors are measured using three micro-electro-mechanical systems (MEMS) microphone pairs. An analysis of the HRIBTF dataset presented in this paper provides promising conclusions of the applications of sound intensity measurements, in terms of the relationship between the intensity vector and source locations.

Keywords: *Intensity, head-related transfer function, MEMS microphone.*

**Corresponding author*: *u5879960@anu.edu.au.*

1. INTRODUCTION

The ability of human to accurately perceive or localise a sound event is mostly attributed to the head-related transfer functions (HRTFs). HRTF is a scalar quantity often measured at the blocked ear canal entrance. HRTFs are widely used in binaural reproduction [1, 2]. HRTFs are highly individualised and difficult to measure [3, 4]. Generic HRTFs can lead to reduced perception quality [5], while personalised HRTF estimates require personal anthropometric data [6–8].

Sound intensity is a vector quantity that is more relevant to human's perception of sound. Sound intensity vector has been used in sound field reproduction algorithms [9–12]. Compared with pressure based methods, intensity based methods have less stringent requirements on the geometry of the loudspeaker array as well as numerical accuracy. Similar ideas can be applied to binaural reproduction. It can be assumed that binarual reproduction using intensity vectors can achieve adequate perception without using personalised HRTF. The velocity vectors in the vicinity of the pinna were calculated in [13] and [14], whereas the intensity vectors in the vicinity of an upscaled pinna model were measured in [15]. The focus of [13, 14] and [15] are pinna-related transfer functions, where are also highly personalised. The aim of this paper is to formulate a generic head-related intensity-based transfer function (HRIBTF) dataset, which can be applied to binaural reproduction methods without using personal anthropometric data.

2. EXPERIMENT SETUP

The experiment measures the intensity vectors in the vicinity of the KEMAR's ears as a function of the body rotation angle. An excitation sequence is played by a

loudspeaker and the intensity vectors are calculated by the three MEMS microphone pairs located next to the KEMAR's outer ears. The KEMAR's ears and the centre of the loudspeaker are on the same two-dimensional (2D) plane. Therefore, altering the body rotation angle is equivalent to changing the azimuth angle of the loudspeaker.

2.1 The loudspeaker, the KEMAR and its body rotation angle

The experiment utlises the Type 45 BA KEMAR from GRAS Sound & Vibration and the Tannoy System 600 nearfield monitor [16]. The distance between the centre of the KEMAR's head and the loudspeaker is 1.1 m. When the body rotation angle is zero, the KEMAR is facing directly towards the centre of the loudspeaker. The body rotation angle increments as the KEMAR turns counterclockwise. The head and the torso of the KEMAR are always facing the same direction (i.e., there are no relative rotations between the head and the torso). The body rotation is achieved by placing the KEMAR on a single axis rotary table driven by a servo motor. Body rotations are performed in 10^{o} intervals. Figure 1 shows the sampling of the body rotation angle. The red arrow shows the rotation direction.

2.2 MEMS microphone pairs

The intensity vectors are measured by the MEMS microphone pairs shown in Figure 2. Each pair consists of two ICS-43432 MEMS microphones [17] spaced 1.5 cm apart. The microphone pairs are placed in the vicinity of the KEMAR's ear on three Cartesian axes. Figure 3 shows the microphone placement and the setup of the Cartesian axes. For each ear, the positive x direction is pointing forward along the cheek, the positive y direction is pointing vertically up towards the roof, and the positive z direction is pointing horizontally out of the ear.

2.3 The measurement environment

Figure 4 shows the measurement environment. Only the loudspeaker circled in red is used to play the excitation sequence. The room is not ideal due to the presence of reflections. Although reflections interfere with the measurement results, they are prevalent in most listening environments such as the living room and many exhibition space. Therefore, the dataset presented in this paper is relevant to practical listening environments.

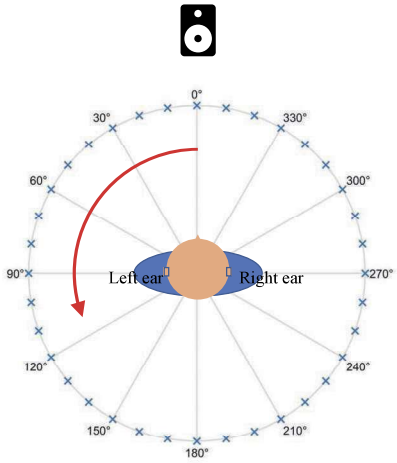


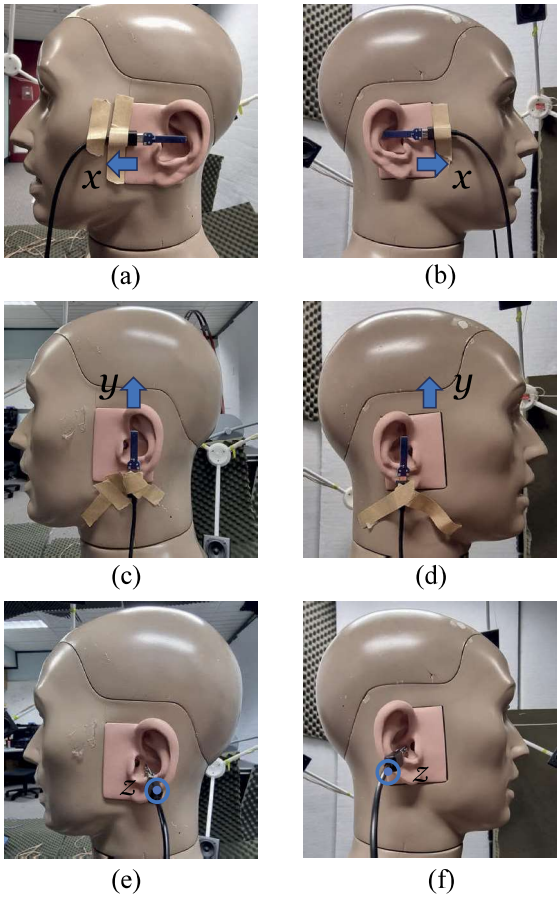
Figure 1. Sampling of the body rotation angle. The KEMAR is at 0 degrees body rotation and is facing the loudspeaker, The red arrow shows the rotation direction.

Figure 2. MEMS microphone pair used in the experiment.

3. INTENSITY CALCULATION

This section outlines the steps to calculate the intensity vectors from the microphone pair measurements. First, the impulse responses between the loudspeaker and the pair of microphones are measured by using the sine sweep method. In this experiment, the sampling frequency is 16 kHz. The measured impulse responses are 89999 samples long and only 256 samples from index 4300 to 4555 are kept. The indexes are carefully chosen to reduce the influence of the wall reflections and compensate the delay of the measurement equipment. Next, the impulse responses are transformed into the frequency-domain to obtain the transfer functions. To reduce errors, the transfer functions are averaged across five measurement runs. In this experiment, system response due to the measurement equipment such as the DAC and the amplifier are not compensated. This is because in a realistic listening environment, it is challenging to separate the (unwanted) distortions due to the measurement equipment from the raw recordings.

Let $p(x, \omega)$ and $p(x+\Delta x, \omega)$ denote the transfer functions measured at a pair of microphones in the x direction,



(a) (b) (c) (d) (e) (f)

Figure 3. Placement of the MEMS microphone pairs for the left ear shown in (a), (c) and (e), as well as the right ear shown in (b), (d) and (f).

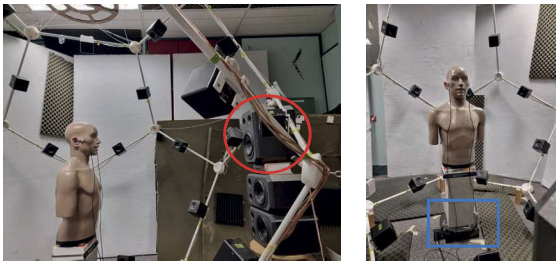
the active intensity at angular frequency ω [18]

$$I_a(x,\omega) = \frac{1}{\rho_0 \Delta x \omega} \mathrm{Im}\{p(x,\omega)p^*(x+\Delta x,\omega)\} \quad (1)$$

in which the spacing between the microphones $\Delta x = 0.015$ m, $\rho_0 = 1.2042$ kg/m^3 is the density of air at 20^o Celsius, $(\cdot)^*$ denotes conjugate and $\mathrm{Im}\{\cdot\}$ denotes the imaginary part. Similar equations are used to calculate the intensity vectors in the y and z directions.

4. MEASUREMENT RESULTS

Figure 5 shows the active intensity vectors in the x direction. Up to 7 kHz, for both the left ear and the right ear, the intensity vectors are mostly negative when the body rotation angle is between 0 and 70 degrees, as well as between 300 and 350 degrees. This is because the KEMAR is facing the loudspeaker and the sound waves are projected from the front. When the body rotation angle is between 90 and 250 degrees, the intensity vectors are predominantly positive for both the left ear and the right ear. This is because the loudspeaker is located on the rear side of the KEMAR and sound waves travel from the back following the positive x direction in Figures 3(a) and (b).



(a) (b)

Figure 4. Measurement environment. (a) The loudspeaker in the red circle plays the excitation signal. (b) The single axis rotary table in the blue rectangle controls the body rotation angle of the KEMAR.

Figure 6 shows the active intensity vectors in the y direction. Although the centre of the loudspeaker and the centre of the KEMAR's head are on the same 2D plane, the intensity vector exhibits variations when the ear is ipsilateral. For instance, the intensity vector at the left ear shows variations when the body rotation angle is between 250 and 350 degrees; whereas the intensity vector at the right ear shows variations when the body rotation angle is between 0 and 120 degrees. The variation could be due to imperfect microphone pair placement, pinna reflections and wall reflections.

Figure 7 shows the active intensity vectors in the z direction. As expected, the intensity vector is negative when the ear is ipsilateral. For instance, large negative amplitude can be observed between 3 kHz and 4 kHz at the left ear when the body rotation angle is around 270 degrees, as well as at the right ear when the body rotation angle is about 90 degrees.

5. LINK TO DATASET

The dataset is available at `https://github.com/FJWang01/Head-Related-Intensity-Based-Transfer-Function`.

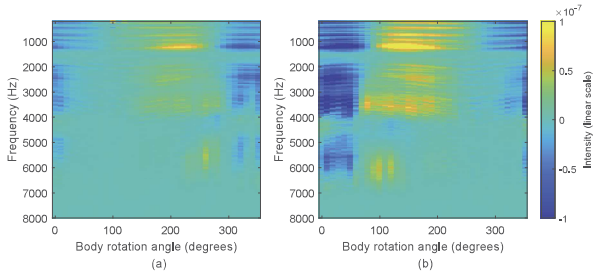


Figure 5. Active intensity vector in the x direction in the vicinity of (a) the left ear and (b) the right ear.

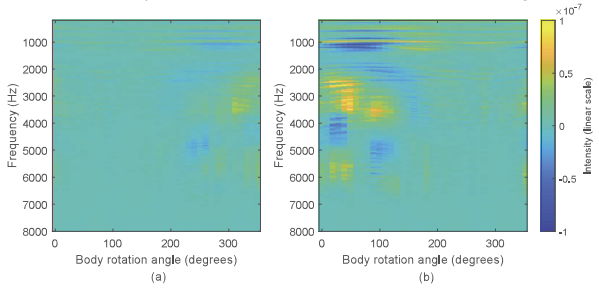


Figure 6. Active intensity vector in the y direction in the vicinity of (a) the left ear and (b) the right ear.

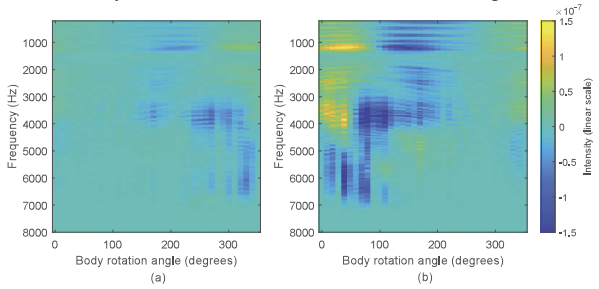


Figure 7. Active intensity vector in the z direction in the vicinity of (a) the left ear and (b) the right ear.

6. CONCLUSION

This paper analysed the intensity vectors in three Cartesian directions in the vicinity of the KEMAR's ear in a reverberant environment. The intensity vectors were measured using MEMS microphone pairs, which are significantly cheaper than intensity probes in the market. The intensity vectors exhibited variations as a function of KEMAR's body rotation angle. Future work will incorporate the intensity vectors measured by the MEMS microphone pairs in sound field reproduction and source localisation algorithms.

7. REFERENCES

[1] V. R. Algazi and R. O. Duda, "Headphone-based spatial sound," *IEEE Signal Processing Magazine*, vol. 28, no. 1, pp. 33–42, 2011.

[2] Z. Ben-Hur, D. L. Alon, R. Mehra, and B. Rafaely, "Binaural reproduction based on bilateral ambisonics and ear-aligned hrtfs," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 901–913, 2021.

[3] S. Li and J. Peissig, "Measurement of head-related transfer functions: A review," *Applied Sciences*, vol. 10, no. 14, 2020.

[4] J. Reijniers, B. Partoens, J. Steckel, and H. Peremans, "HRTF measurement by means of unsupervised head movements with respect to a single fixed speaker," *IEEE Access*, vol. 8, pp. 92287–92300, 2020.

[5] J. S. Andersen, R. Miccini, S. Serafin, and S. Spagnol, "Evaluation of individualized HRTFs in a 3D shooter game," in *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, pp. 1–10, 2021.

[6] 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 *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4468–4472, 2014.

[7] J. He, W.-S. Gan, and E.-L. Tan, "On the preprocessing and postprocessing of HRTF individualization based on sparse representation of anthropometric features," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 639–643, 2015.

[8] R. Miccini and S. Spagnol, "A hybrid approach to structural modeling of individualized HRTFs," in *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 80–85, 2021.

[9] H. Zuo, P. N. Samarasinghe, and T. D. Abhayapala, "Intensity based spatial soundfield reproduction using an irregular loudspeaker array," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 1356–1369, 2020.

[10] H. Zuo, T. D. Abhayapala, and P. N. Samarasinghe, "3D multizone soundfield reproduction in a reverberant environment using intensity matching method,"

in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 416–420, 2021.

[11] J.-W. Choi and Y.-H. Kim, "Manipulation of sound intensity within a selected region using multiple sources," *The Journal of the Acoustical Society of America*, vol. 116, pp. 843–852, 08 2004.

[12] D. Arteaga, "An ambisonics decoder for irregular 3d loudspeaker arrays," in *The 134th AES Convention*, 01 2013.

[13] P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Visualization of acoustic pressure and velocity patterns with phase information in the pinna cavities at normal modes," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8217–8221, 2014.

[14] P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Vertical normal modes of human ears: Individual variation and frequency estimation from pinna anthropometry," *The Journal of the Acoustical Society of America*, vol. 140, pp. 814–831, 08 2016.

[15] P. Rönkkö, "Measuring acoustic intensity field in upscaled physical model of ear," Master's thesis, Aalto University, 2019.

[16] Tannoy Ltd., "Tannoy system 600 nearfield monitor." `https://www.fullcompass.com/common/files/2709-System600SpecSheet.pdf`. Accessed: 2023-05-15.

[17] InvenSense, "ICS-43432." `https://invensense.tdk.com/products/digital/ics-43432/`. Accessed: 2023-05-15.

[18] M. Möser, *Technische Akustik*. Springer-Verlag GmbH Berlin Heidelberg, 2015.