ROBUST MEDIAN-PLANE BINAURAL SOUND SOURCE LOCALIZATION

Benjamin R Hammond, Philip J.B. Jackson

Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, GU2 7XH, UK

ABSTRACT

For a sound source on the median-plane of a binaural system, interaural localization cues are absent. So, for robust binaural localization of sound sources on the median-plane, localization methods need to be designed with this in consideration. We compare four median-plane binaural sound source localization methods. Where appropriate, adjustments to the methods have been made to improve their robustness to real world recording conditions. The methods are tested using different HRTF datasets to generate the test data and training data. Each method uses a different combination of spectral and interaural localization cues, allowing for a comparison of the effect of spectral and interaural cues on median-plane localization. The methods are tested for their robustness to different levels of additive noise and different categories of sound.

Index Terms— Binaural, Localization, Median-Plane

1. INTRODUCTION

Sound source localization using binaural microphones provides an abundance of opportunities for audio augmented reality [1, 2]. Machine based binaural sound source localization could also accompany head mounted display visualizations for the deaf and hard of hearing [3].

Some binaural sound source localization methods assume that the sound source lies on the frontal azimuthal plane [4, 5]. Other methods are designed to localize a sound source on the full-sphere i.e. from any direction of arrival (DOA) around the listener [6, 7]. Some localization methods exploit the movement of the sound source in relation to the listener's head [8, 9]. For non-moving sources, the main localization cues are the interaural cues and spectral cues. In anechoic conditions, for a theoretically symmetrical head, a sound source should produce identical signals at both ears when the direction of arrival of the sound source lies on the median-plane. As such the interaural signal differences will be absent [10, Chapter 2.3]. Although this is the case, some localization methods still implicitly use interaural cues for localization of sound sources that lie on the median-plane [11, 7]. As the interaural signal differences are absent on the median-plane, localization of sound sources on the median-plane using these methods results in high localization errors [6]. As a step towards robust full-sphere binaural sound source localization, there should be a particular focus on the robustness of localization of sound sources that lie on the median-plane.

To generate training data, localization methods make use of a head related transfer function (HRTF) dataset recorded using the same head that recorded the binaural test sound signal. The HRTF describes the frequency based filtering effect of the listener's head, pinna, and torso at the listener's ear canal from a point in space. The time domain equivalent is the head related impulse response (HRIR). A HRTF dataset consists of a collection of measured HRTFs at different DOAs around the listener for both ears [12, Chapter 1]. The spectral cues change between listeners, as the morphology of each listener is different. Therefore, in order to localize a non-moving sound source on the median-plane, the listener's unique HRTF dataset is needed, from which the spectral cues are derived. Methods are rapidly developing for the fast and accurate measurement of personal HRTF datasets in the home environment [13, 14, 15, 16]. When recording HRIRs, unwanted measurement artefacts can be introduced into the recordings. These artefacts are associated with the acoustic environment, the measurement procedure and the post-processing of the data [17, Chapter 8].

Many median-plane localization methods were tested by generating their test data and training data synthetically using the exact same HRTF dataset [18, 19]. This is referred to as the matched condition throughout this paper. With this testing condition, the exact same measurement noise is present in both the test data and training data. For methods tested using this condition, this can actually result in the measurement noise providing an additional localization cue [6]. Methods that are designed and tested only using the matched condition often suffer from overfitting of the training data. To test for robustness, the mismatched condition and off-center condition are also tested. The mismatched condition refers to the case of testing a method using different HRTF datasets to generate the training and test data. For the mismatched condition, the HRTF datasets are captured using the same model of dummy head, but in different rooms, using different measurement equipment. As such the measurement noise is different in the HRTF datasets used to generate the test data and training data, so unlike the matched condition, the measurement noise cannot be negated or used as an additional localization cue. The mismatched condition is a useful testing condition, as for median-plane binaural sound source localization to be possible in real world conditions with the use of a pre-measured HRTF dataset, the method must be robust to additive and convolutive noise provided by the recording equipment. An additional test condition, referred to as the off-center condition uses HRTF templates which have DOAs on the median-plane to localize test sounds generated using HRIR pairs with a lateral angle of 5° to 10° away from the median-plane. This condition is included to test the localization methods' robustness to positioning errors of the loudspeakers and microphones. It is important to test this condition, as even small positioning errors can have a large effect on the interaural cues around the median-plane, especially at higher frequencies.

Reflections from the pinnae create peaks and notches in the spectrum of the sound source as it arrives at the entrance of the ear canal. For different directions of arrival, these peaks and notches occur at different frequencies and have varying degrees of sharpness. The relative level between successive peaks and notches also differs. These peaks and notches predominantly occur in the high-frequency range; approximately above 5kHz [20]. The spectrum

EPSRC Programme Grant S3A: Future Spatial Audio for an Immersive Listener Experience at Home (EP/L000539/1).

of the sound source, as well as additive or convolutive noise provided by the recording equipment may produce confounding peaks and notches in the spectrum of the binaural test sound signal. As such, the localization methods should be tested for their robustness to a variety of different sound sources with different spectral shapes. Ultimately, a localization method should aim to be robust to different recording environments and the presence of reverberation in the binaural recording. However, as the common binaural sound source localization cues are contained within the direct component of the sound, it is first prudent to test the median-plane localization methods in anechoic conditions, i.e. containing only the direct component of the binaural sound signal. As such, the localization methods will be tested using only anechoic conditions in this paper. In other areas of research, accessibility of large scale datasets has made it possible to develop deep learning methods [21]. However there are currently few publicly available HRTF datasets. Each of the methods tested in this paper use only one HRTF dataset to use as templates or generate training data. These methods then have the advantage of performance with a small amount of training data.

2. MEDIAN-PLANE BINAURAL SOUND SOURCE LOCALIZATION

The methods in this paper have been selected to give a diverse representation of the state of the art median-plane sound source localization methods. The first method to be tested is the Peak & Notch Frequency method [22], which estimates the location of a sound source by comparing the estimated frequencies of the peaks and notches in the HRTF of the binaural test sound signal with the estimated frequencies of the peaks and notches in each HRTF of the template HRTF dataset. The peaks and notches are defined as relative maxima and minima in a spectrum respectively. Small spectral fluctuations in the spectrum of the HRTFs and each time frame of the log-magnitude spectrogram of the binaural test sound signal are smoothed using a Gaussian filter. For each time frame of the log-magnitude spectrogram of the binaural test sound signal, a DOA is estimated by a comparison of the frequencies of the peaks and notches in the time frame of the spectrogram to the peaks and notches of each HRTF pair in the training HRTF dataset. The direction of arrival is then estimated as the DOA assigned to the most time frames. For this last step, as there is a large amount of HRTF pairs on the median-plane in the training dataset, it was decided instead to create a PDF as a function of the DOA using a Gaussian kernel smoothing function. The DOA of the sound source is then given by the DOA at the maximum of the PDF. It was found that the frequency range of 4kHz - 18kHz produced the best localization estimates for the mismatched condition.

The second tested method is the Speech Prefilter method [18]. Firstly, a voice activity detector is used to detect voiced speech frames in the binaural sound signal. Time frames that do not have voiced speech detected in them are removed [23]. The prefilter is in the form of cepstral coefficients learned from a training set of speech samples. The received binaural sound signal at the left and right ears are transformed to cepstrograms. Truncated cepstral coefficients are averaged over each time frame of these cepstrograms. The Fourier transform of the summation of the prefilter and the truncated cepstral coefficients yields the estimate of the magnitude of the HRTF for each ear. The estimated magnitude spectrum of the HRTF from the binaural test sound signal for the left and right ears are concatenated, and the magnitude spectrum of the HRTF templates for the left and right ears are also concatenated. Cross-correlation coefficients between these two concatenated spectra are estimated. The DOA of the sound source is estimated by the entry in the database that yields the maximum correlation coefficient. For this paper, the prefilter is trained from approximately one hour of speech sounds from the CSTR VCTK Corpus [24]. The speech samples have been selected to give an equal balance of male and female voices and a diverse set of accents. These training speech samples are not included in the test data. The original method used a narrow frequency range of 3.5kHz - 7.5kHz. For the mismatched condition, it was found that a frequency range of 3.5kHz - 18kHz produced much better localization estimates, as there are more prominent peaks and notches in this range. Additionally, the log-magnitude spectrum is used instead of the magnitude spectrum in all cases.

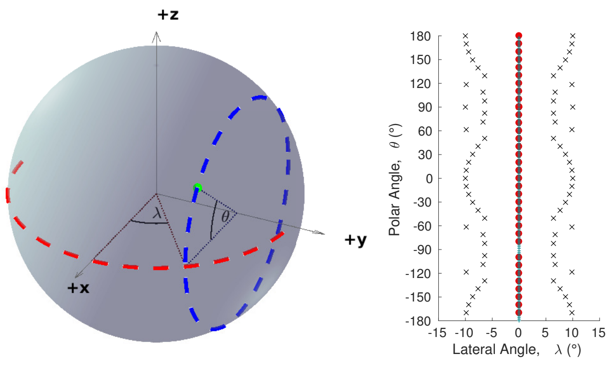
The third tested method is the Cross-Convolution method [25, 19]. For this method, the binaural test sound signal received at the left and right ears are filtered with each contralateral HRIR pair in the training HRTF dataset. Cross-correlation is used to determine the similarity of the left and right ear's filtered observations. The correlation coefficient is calculated using each HRIR pair in the dataset and the maximum correlation coefficient yields the DOA estimate.

The fourth tested method is the MUSIC (MUltiple SIgnal Classification) Signal Subspace method [11]. For this method, the directional information is extracted in narrow subbands from a binaural test sound signal. In order to estimate the location of a sound source, a composite estimator based on signal subspace decomposition is used with a set of HRTF templates from the training HRTF dataset. In order to improve the performance of this method for the mismatched condition, the frequency range is extended to 18kHz. It was additionally found that results improved by using the frequency bins from the STFT of the binaural sound signal instead of filtered narrow subbands. The best results were yielded with an STFT with a window length of 512 samples and a hop size of 128 samples, using a sampling rate of 48kHz.

3. TESTING PROCEDURE

In order to test the robustness of the localization methods, a diverse range of monaural sound sources are used to generate the binaural test sound signals. 10 of these sounds are taken from the environmental sound corpus in [26]. They have been selected for their diversity in spectral characteristics. Namely, they are: Waves crashing, electric saw cutting, water pouring, train moving, chopping wood,typing on keyboard, ice dropping into glass, bells chiming, cars honking and sheep baaing. Additionally, as speech is one of the most important everyday sounds, it is tested under its own category. As such, 10 speech samples have been chosen from the CSTR VCTK Corpus [24]. The speech samples have been selected to give an equal balance of male and female voices and a diverse set of accents. White noise and pink noise are additionally used for testing, giving a total of 22 monaural sound sources used in this paper.

The interaural-polar coordinate system describes the direction of arrival of a sound source with the lateral angle, $\lambda \in [-90°, 90°]$, and the polar angle $\theta \in [-180°, 180°)$ [27], as shown in Figure 1 (a). The HRTF dataset measured in [28] contains 178 HRIR pairs that have DOAs which are approximately evenly spaced on the median-plane. This dataset is referred to as the TH Koln dataset throughout this paper. These 178 HRIR pairs are used for training data by all of the tested methods to estimate the DOA of the sound source. The HRTF dataset used to generate the binaural test sound



(a) (b)

Figure 1: (a) The interaural-polar coordinate system. For a listener at the origin, +x extends directly ahead of the listener with a lateral and polar angle of: $\lambda = 0°$, $\theta = 0°$. The lateral angle range is displayed by the dashed red line. The polar angle range is displayed by the dashed blue line. (b) DOA of HRTFs used for training data and to generate test data. Blue plus sign: DOA of HRTFs used for training data. Red circle: DOA of HRTFs used for the mismatched condition. Black cross: DOA of HRTFs used for the off-center condition.

signals for the mismatched condition and the off-center condition is the dataset measured at RIEC, Tohoku University as part of the "Club Fritz" project [29]. The binaural test sound signals are created synthetically by convolving the HRIR pairs at each of the test positions with each of the 22 monaural sound sources. Stereo uncorrelated pink noise is added to the binaural sound signals to give signal-to-noise ratios (SNRs) from 0dB to 30dB in 10dB steps. For the mismatched condition, 35 HRIR pairs are used, all of which lie on the median-plane and are spaced in 10 degree increments, with the exception of a HRIR pair at $\theta = -90°$, which is absent. For the matched condition, the binaural test sound signals are generated synthetically using the HRIR pairs in the TH Koln dataset with a lateral angle of $\lambda = 0°$ and the nearest polar angle to the DOAs used for the mismatched condition. For the off-center condition, the DOAs of the HRIR pairs used to generate the test data have a lateral angle between and including $\lambda = -10°$ and $\lambda = -5°$, and also between and including $\lambda = 5°$ and $\lambda = 10°$. Within these two lateral angle ranges, the HRIR pairs with a polar angle nearest to the polar angle of the test positions in the mismatched condition are used, giving a total of 70 HRIR pairs used for the off-center condition, as shown in Figure 1 (b).

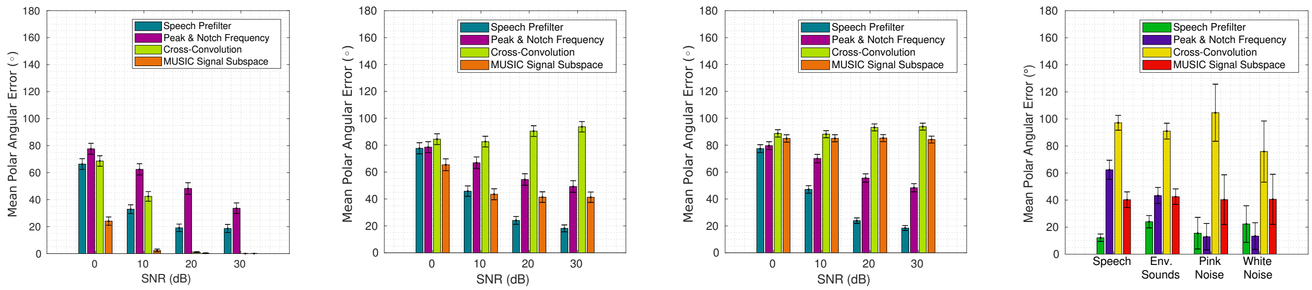
4. RESULTS AND DISCUSSION

Figure 2 (a-c) shows the mean polar angular error for localization on the median-plane. For each testing condition, the polar angular error is the polar angle between the ground truth test position and the estimated position of the sound source, from the point of view of the listener. The results are shown for each of the methods and each of the test conditions. The Speech Prefilter method performs at a similar level for the different testing conditions (matched, mismatched, off-center), though it performs slightly better in the matched condition. Predictably it performs worse with decreasing SNR. The Peak & Notch Frequency method also performs similarly for the different testing conditions. As these two methods only use spectral cues for localization, the interaural cues do not provide any additional benefit for the matched condition, nor do they provide confounding information for the mismatched or off-center conditions. Overall the Speech Prefilter method is better at localization than the Peak and Notch Frequency method. It could be the case that it performs better as it additionally implicitly considers the sharpness and relative level of the peaks and notches, as well as the frequency at which the peaks and notches occur.

For a symmetrical head, the HRTF magnitude spectrum for the right ear should be identical to that of the left ear and the interaural cues should be zero throughout the frequency range for all locations on the median-plane. However a slight error in positioning of the dummy head relative to the loudspeakers results in the peaks and notches for one ear occurring at slightly different frequencies to the those of the other ear. This positioning error results in non-zero values for the interaural cues, with the largest values occurring around the frequencies of the peaks and notches. For the matched condition, this exact same measurement noise exists in the interaural cues for both the binaural test sound signal and HRTF templates, and as such, for this testing condition, the methods that use interaural cues actually exploit this measurement noise as a localization cue. Therefore both the Cross-Convolution method and the MUSIC Signal Subspace method perform very well in the matched condition. However, the exact same measurement noise due to positioning error would not exist in both the binaural test sound signal and HRTF templates in a real world setting and so this demonstrates why developing a median-plane binaural sound source localization method for use with the matched HRTF condition should be avoided. For the mismatched and off-center conditions, the Cross-Convolution method performs very poorly, as the measurement noise in the interaural cues is now different for the binaural test sound signal and HRTF templates. The MUSIC Signal Subspace method implicitly uses both spectral cues and interaural cues. The interaural cues result in the method performing well in the matched condition, however they hinder the method in the mismatched condition and off-center condition.

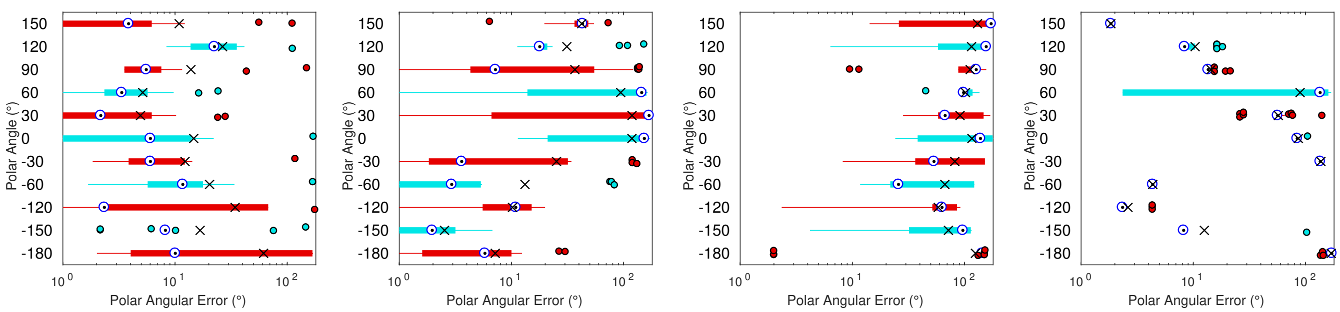
Figure 2 (d) shows the mean polar angular error, as a function of sound category for localization on the median-plane. As the Speech Prefilter method attempts to adjust the spectrum with respect to the average speech spectrum, it performs the best with speech sounds. Pink noise having a spectrum closely resembling speech is the next best at performance with this method, followed by white noise, and then environmental sounds, which have the most confounding spectral cues. The Peak & Notch Frequency method performs best with pink noise and white noise, as they have relatively flat spectra and as such are less prone to producing confounding peaks and notches in the spectrum of the binaural test sound signal. The MUSIC Signal Subspace method performs at a similar level with all sounds, and the cross-convolution method performs poorly in all conditions.

Figure 3 shows the box plot and mean for the polar angular error as a function of the polar angle, for localization on the median-plane. For each polar angle, the results are shown for binaural test sound signals synthetically generated at 30dB SNR with the mismatched condition. The Speech Prefilter method and the Peak & Notch Frequency method perform poorly at 120°, where the HRTF spectrum is relatively flat and has no discernable major peaks or notches. The Speech Prefilter method is also prone to front-back reversals for sound sources at the back of the head, where the first major notch occurs at a similar frequency to the first major notch in the HRTF spectra at the front of the head. The Peak & Notch Frequency method performs best for sound sources underneath the head, where



(a) (b) (c) (d)

Figure 2: Mean polar angular error (°) for localization on the median-plane, as a function of: (a-c) SNR, (d) sound category. The results are shown for binaural test sound signals synthetically generated using: (a-c) All of the monaural sound sources, (d) the monaural sound sources specified by the sound category in the abscissa. The methods shown are the Speech Prefilter, Peak & Notch Frequency, Cross-Convolution, and MUSIC Signal Subspace. The results are shown for: (a) matched condition, (b,d) mismatched condition, (c) off-center condition. The error bars correspond to 95% confidence intervals.



(a) (b) (c) (d)

Figure 3: Polar angular error (°), as a function of polar angle (°) for localization on the median-plane. The results are shown for binaural test sound signals synthetically generated at 30dB SNR. Results are shown for the mismatched condition. The methods shown are: (a) Speech Prefilter, (b) Peak & Notch Frequency, (c) Cross-Convolution, (d) MUSIC Signal Subspace. Cross: Mean; Black dot in white circle: Median; box: Inter-quartile range (IQR); whisker: Within quartile ± 1.5.IQR; outliers are shown as filled circles.

there are more peaks and notches in the spectrum caused by the HRTF. For the MUSIC Signal Subspace method, the variance is low, indicating that the algorithm is robust to the sound type. Additionally, in Figure 2 (a-c), the performance of the method is fairly consistent for SNRs of 10dB or higher. This would indicate that the method is also robust to additive noise. However, the method suffers in the mismatched condition due to its inherent utilization of interaural cues. The method performs well at angles where the interaural cues of the HRTFs used to generate the binaural test sound signal are similar to the interaural cues of the HRTF templates with the same DOA. However at angles where the interaural cues of the binaural test sound signal and the HRTF template with the same DOA are different, the interaural cues act as confounding cues and the DOA cannot be estimated well.

5. CONCLUSION

In this paper we compared four localization methods for their ability to localize a sound source on the median-plane. Appropriate adjustments have been made to the methods to make them robust to real world recording conditions. The Spectral Prefilter method uses the frequency, sharpness and relative levels of the peaks and notches in the spectrum for localization and it has outperformed the Peak & Notch Frequency method, which only uses the estimated frequencies of the peaks and notches for localization. For methods that use interaural cues for localization, there is a large disparity between the results for the matched and mismatched conditions. This is because the binaural test sound signal and the HRTF template with the same DOA contain the same measurement noise for the matched condition, resulting in the measurement noise being exploited as a localization cue. However, the binaural test sound signal and its corresponding HRTF template would not contain the exact same measurement noise in a real world setting, and so we make the case that future median-plane binaural sound source localization methods should not be designed for use with the matched condition. Furthermore, median-plane binaural sound source localization methods should use spectral cues only, and should not use interaural cues for localization. Future work should consider robustness of the median-plane binaural sound source localization methods to the presence of reverberation.

6. REFERENCES

[1] A. Harma, J. Jakka, M. Tikander, M. Karjalainen, T. Lokki, J. Hiipakka, and G. Lorho, "Augmented reality audio for mobile and wearable appliances," *J. Audio Eng. Soc*, vol. 52, no. 6, pp. 618–639, 2004.

[2] P. F. Hoffmann, F. Christensen, and D. Hammershi, "Insert earphone calibration for hear-through options," in *Audio Engineering Society Conference: 51st International Conference: Loudspeakers and Headphones*, Aug 2013.

[3] D. Jain, L. Findlater, J. Gilkeson, B. Holland, R. Duraiswami, D. Zotkin, C. Vogler, and J. E. Froehlich, "Head-mounted display visualizations to support sound awareness for the deaf and hard of hearing," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI '15. New York, NY, USA: ACM, 2015, pp. 241–250.

[4] M. Dietz, S. D. Ewert, and V. Hohmann, "Auditory model based direction estimation of concurrent speakers from binaural signals," *Speech Communication*, vol. 53, no. 5, pp. 592–605, 2011.

[5] J. Woodruff and D. Wang, "Binaural detection, localization, and segregation in reverberant environments based on joint pitch and azimuth cues," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 4, pp. 806–815, 2013.

[6] B. Hammond and P. Jackson, "Robust full-sphere binaural sound source localization," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2018)*. IEEE, 2018.

[7] M. Rothbucher, M. Durkovic, T. Habigt, H. Shen, and K. Diepold, "HRTF-based localization and separation of multiple sound sources," in *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 2012, pp. 1092–1096.

[8] N. Ma, T. May, H. Wierstorf, and G. J. Brown, "A machine-hearing system exploiting head movements for binaural sound localisation in reverberant conditions," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 2699–2703.

[9] C. Zhou, R. Hu, W. Tu, X. Wang, and L. Gao, "Binaural moving sound source localization by joint estimation of ITD and ILD," in *Audio Engineering Society Convention 130*. Audio Engineering Society, 2011.

[10] J. Blauert, *Spatial Hearing: the psychophysics of human sound localization*, revised ed. Cambridge, MA: MIT press, Cambridge, MA, 1997.

[11] D. S. Talagala, W. Zhang, T. D. Abhayapala, and A. Kamineni, "Binaural sound source localization using the frequency diversity of the head-related transfer function," *The Journal of the Acoustical Society of America*, vol. 135, no. 3, pp. 1207–1217, 2014.

[12] B. Xie, *Head-related transfer function and virtual auditory display*, 2nd ed. Florida, USA: J. Ross Publishing, Inc., 2013.

[13] F. Rumsey, "Spatial audio: Binaural challenges," *J. Audio Eng. Soc*, vol. 62, no. 11, pp. 798–802, 2014.

[14] K. Sunder, J. He, E. L. Tan, and W.-S. Gan, "Natural sound rendering for headphones: Integration of signal processing techniques," *Signal Processing Magazine, IEEE*, vol. 32, no. 2, pp. 100–113, March 2015.

[15] N. D. Hai, N. K. Chaudhary, S. Peksi, R. Ranjan, J. He, and W. S. Gan, "Fast HRFT measurement system with unconstrained head movements for 3d audio in virtual and augmented reality applications," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2017, pp. 6576–6577.

[16] R. Ranjan, J. He, and W.-S. Gan, "Fast continuous acquisition of HRTF for human subjects with unconstrained random head movements in azimuth and elevation," in *Audio Engineering Society Conference: 2016 AES International Conference on Headphone Technology*, Aug 2016.

[17] H. Kuttruff, *Room acoustics*, 5th ed. Abingdon, UK: Spon Press, 2009.

[18] D. S. Talagala, X. Wu, W. Zhang, and T. D. Abhayapala, "Binaural localization of speech sources in the median plane using cepstral HRTF extraction," in *2014 22nd European Signal Processing Conference (EUSIPCO)*, Sept 2014, pp. 2055–2059.

[19] M. Usman, F. Keyrouz, and K. Diepold, "Real time humanoid sound source localization and tracking in a highly reverberant environment," in *Signal Processing, 2008. ICSP 2008. 9th International Conference on*. IEEE, 2008, pp. 2661–2664.

[20] K. Iida, M. Itoh, A. Itagaki, and M. Morimoto, "Median plane localization using a parametric model of the head-related transfer function based on spectral cues," *Applied Acoustics*, vol. 68, no. 8, pp. 835 – 850, 2007.

[21] C. Rascon and I. Meza, "Localization of sound sources in robotics: A review," *Robotics and Autonomous Systems*, vol. 96, pp. 184 – 210, 2017.

[22] E. Blanco-Martin, F. J. Casajus-Quiros, J. J. Gomez-Alfageme, and L. I. Ortiz-Berenguer, "Estimation of the direction of auditory events in the median plane," *Applied Acoustics*, vol. 71, no. 12, pp. 1211–1216, 2010.

[23] Z.-H. Tan and B. Lindberg, "Low-complexity variable frame rate analysis for speech recognition and voice activity detection," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 5, pp. 798–807, 2010.

[24] C. Veaux, J. Yamagishi, K. MacDonald, *et al.*, "CSTR VCTK Corpus: English Multi-speaker Corpus for CSTR Voice Cloning Toolkit," 2017.

[25] M. Rothbucher, D. Kronmüller, K. Diepold, M. Durkovic, and T. Habigt, *HRTF sound localization*. INTECH Open Access Publisher, 2011.

[26] B. Gygi, G. R. Kidd, and C. S. Watson, "Similarity and categorization of environmental sounds," *Perception & psychophysics*, vol. 69, no. 6, pp. 839–855, 2007.

[27] R. Baumgartner, P. Majdak, and B. Laback, "Assessment of sagittal-plane sound localization performance in spatial-audio applications," in *The technology of binaural listening*. Springer, 2013, pp. 93–119.

[28] B. Bernschütz, "A spherical far field HRIR/HRTF compilation of the Neumann KU 100," in *Proceedings of the 40th Italian (AIA) Annual Conference on Acoustics and the 39th German Annual Conference on Acoustics (DAGA) Conference on Acoustics*, 2013, p. 29.

[29] A. Andreopoulou, D. R. Begault, and B. F. Katz, "Inter-laboratory round robin HRTF measurement comparison," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 5, pp. 895–906, 2015.