The Effects of Decreasing the Magnitude of Elevation-dependent notches in HRTFs on Median Plane Localisation

JADE CLARKE and HYUNKOOK LEE

Applied Psychoacoustics Laboratory (APL), University of Huddersfield, Huddersfield, U.K.
e-mail: u1263399@unimail.hud.ac.uk, h.lee@hud.ac.uk

September 23rd 2016

Abstract

Accurate sound localisation in the median plane is known to require certain spectral cues caused by the pinnae and upper body; which provides important information about a sources location in space, with pinna cues in particular producing considerably prominent peaks and troughs in the spectrum. Iida et al. (2007) suggest that the first peak and the first two pinna related spectral notches (> 5 kHz) provide enough information for the source to be accurately localised. The importance of the magnitude of spectral peaks and notches in median plane localisation is thought to depend on the relative rather than absolute distribution across the frequency spectrum (Macpherson and Sabin, 2013). The present experiment tested three subjects and explored whether a necessary magnitude of pinna related spectral notches exists, if so, what happens to the perceived source position once the magnitude is reduced. The results generally showed that for elevated sources, as pinna related notches were reduced in magnitude, the source position moved upwards in space. This seems to be due to the overall frequency spectrum becoming closer to that of sources located higher in space. For a non-elevated frontal source in the median plane on the other hand, an increase in front-back confusion occurred as a result of notch magnitude manipulation. This again is considered to be due to the altered frequency distribution mimicking that of the HRTF for the opposite direction (i.e., relative dominance of frequencies around 1 kHz over those around 4 kHz).

1 Introduction

Three-dimensional (3D) audio has become a topic of particular interested of late, especially for the purpose of virtual reality technologies and immersive high-definition media displays. This has meant that there has been a recent surge in the acquisition of HRTF (head-related transfer function) datasets and binaural research. A major limitation to these datasets however is that a very large number of measurements are to be made for complete 360° immersion, resulting in a great deal of data and storage being required.

This has lead the authors to believe that a certain degree of HRTF simplification based on structural models could reduce the size of these data sets, and potentially increase sound localisation accuracy for the purpose of real-time binaural interactive audio displays. These datasets could be used in conjunction with head tracking devices, which could aid in providing full 3D audio immersion.

In order to accurately and effectively binauralise audio in 3D space, it is important to consider the temporal and spectral cues necessary for human sound localisation. In the horizontal plane, Lord Rayleigh's Duplex Theory [1] describes the roles of interaural time difference (ITD) at low frequencies and interaural level difference (ILD) at high frequencies In the vertical plane, it is generally accepted that spectral distortions caused by the body provide effective cues for localisation. These distortions include reflections from the knees when a person is sat down, the head and torso (HAT) and the pinnae.

Sound diffraction is frequency dependent, with effects only appearing when the wavelength is less than or equal to the size of the diffracting object. The HAT is thought to account for low frequency elevation cues below 3 kHz [2]. Roffler and Butler [3] found that accurate localisation of stimuli in the vertical plane requires the stimulus to be complex, to include frequencies above 7 kHz and for the pinnae to be present.

Previous studies have attempted to isolate individual spectral cues based on anthropometry in order to understand the upper body's contribution to the HRTF [2, 4, 5]. Others have created structural HRTF models based on anthropometry [6, 7]. Brungart and Romigh [8] enhanced existing spectral cues in an attempt to improve vertical localisation, as did Chun et al. [9] albeit by means of stereo reproduction.

A study of particular interest is by Iida et al. [10], who created a HRTF model by first identifying significant, elevation dependent, spectral peaks and notches in subjects' measured HRTFs and then recomposed what they called a 'parametric' HRTF of only extracted peaks and notches. This enabled them to identify exactly which spectral cues were necessary for vertical localisation.

To understand the necessary resolution of the spectral content in HRTFs, Asano et al. [11] simplified the spectrum by means of frequency smoothing in order to identify how microscopic low frequency (below 2 kHz) and macroscopic high frequency characteristics effect median plane localisation. Frequency smoothing was implemented in this manner due to the resolution of the auditory system, which is considered to be higher at low frequencies. A source was virtually presented via headphone playback and HRTFs with different degrees of smoothing were applied. They found that frequencies below 4-5 kHz had little directional dependency, however, a notch exists that increases in frequency as source elevation increases at 6-10 kHz and a peak that moves down in frequency as elevation angle increases is present at 11-14 kHz. The results are consistent with that of Blauert [12, pp. 109–111] and Hebrank and Wright [13]. The authors concluded that instead of recognising minute spectral patterns, instead the auditory system evaluates information regarding power in certain bands around 5-10 kHz relative to the power in bands above and below.

Altering the magnitude of the spectrum has previously been studied by Macpherson and Sabin [14], who did not specifically pay attention to pinna related notches (>5 kHz), but instead reduced the magnitude of either all peaks, all notches or reduced or increased the spectral contrast across the entire frequency spectrum. Their experiment also differed in that they did not limit their testing to the median plane but instead looked at localisation in general. They found that as peaks were levelled off and notches were filled in, elevation localisation accuracy decreased. Also, increasing spectral contrast had little effect on perceived elevation; the perceived location appeared to depend on the relative rather than absolute frequency shape of the spectrum.

A pilot study has been conducted to determine whether the entire dynamic range of pinna-related spectral notches (>5 kHz), as identified by Iida et al. [10], is required for accurate median plane localisation through binaural playback. The consequences of these spectral manipulations with regards to perceived source position is explored and compared to previous research. The manipulated notches are specifically related to pinna anthropometry, where pinna reflections cause spectral filtering; as the reflection angles change due to a change in source direction, different spectral distortions occur [3, 15]. These distortions are caused by minute time delays (100-300μs) caused by the contours of the pinnae [16].

Individualised HRTFs were measured in the upper, frontal median plane and only the magnitude of relevant pinna related notches altered. Localisation tests determined the point at which the source was no longer perceived to be around the target position.

2 Methodology

Three listeners (one female and two males), consisting of two final year undergraduate students and one academic staff member from the University of Huddersfield took part in the experiment. They will be denoted as S1, S2 and S3. All subjects had previous experience in localisation tests, however S3 had never done a binaural localisation test before. All listeners reported normal hearing.

2.1 Impulse Response Acquisition

For each subject, three speaker positions were measured in the frontal median plane: 0°, 30° and 60° from the listener's ear position. This was done using the 'blocked-meatus' method [17] whereby Knowles FG-23329-D65 miniature in-ear microphones (2.59mm in diameter) were fixed in subjects' ears using a small amount of silicone putty, so that the diaphragm of the microphones were situated at the entrance of each ear canal and the entrance of the ear canal was blocked.

Measurements were made in the Applied Psychoacoustics Laboratory's Critical Listening room (6.2m x 5.6m x 3.8m; T30 = 0.25s) at the University of Huddersfield, an ITU-R BS.1116-2 compliant room suitable for 3D loudspeaker reproduction. Stimuli were presented through Genelec 8040A loudspeakers and recorded using a Merging Technologies Horus audio interface recording at 44.1 kHz sampling frequency into an Apple MacBook Pro laptop running HAART (Huddersfield Acoustics and Audio Research Toolbox) [18]. The software emitted an exponential sine sweep lasting 10 seconds from 48 Hz – 28 kHz at around 85dB(A), measured at the ear position of the subject. The recorded signals were deconvolved and equalised appropriately in order to acquire impulse responses.

Subjects were asked to strictly keep their head very still and to hold their gaze to the centre of the loudspeaker at 0° (directly in front of them) to reduce interaural differences. A small headrest was placed against the back of subjects' heads to further reduce movement. To reduce floor reflections, foam diffusers were place on the floor around the listener at the time of measurement.

Once all HRTFs were measured, Sennheiser HD650 open-back headphones were placed over subjects' ears, with the microphones still in place, and additional measurements were obtained for the HpTFs (Headphone transfer functions). Once a measurement was made, the headphones were removed from the subject's head, and then placed on the head again. This process was repeated three times and an average of all measurements was later calculated. Each of these measurements represent the transfer functions of the headphones, the microphones and the path between the two, and were to be used to calculate an inverse filter to ensure that the resulting signal played back to the listener via headphones does not contain any of the characteristics of the experimental apparatus and only that of the path from the loudspeaker to the ear canal entrances.

2.2 Impulse Response Processing

Each IR (impulse response) was analysed using Adobe Audition CS6 software and all of the pinna related notches (> 5 kHz) identified. The prerequisite to constitute a notch in the spectrum was that the lower dip boundary exceeded 5 kHz and the notch depth, in relation to the upper and lower boundaries exceeded 6dB. All notches were then reduced in magnitude to around half the original magnitude in relation to the level at the upper and lower frequency boundaries of the notch. This process was repeated, with the exact same FFT filter settings applied to the file to reduce the notch further.

Each HRTF measurement had 3 conditions; no EQ, half of the notches reduced and full notch reduction. These are denoted as EQ0, EQ1 and EQ1 respectively.

To remove the characteristics of the headphones and other experimental apparatus from the measured impulse responses, an inverse filter was created from the headphone measurements made at the time of HRTF acquisition. This was done using Kirkeby regularisation in the same manner as that of Kearney and Doyle [17].

Once processed and headphone equalised, all impulse responses were convolved with 200ms white noise bursts with a 1ms rise and fall time.

2.3 Listening Test

Listening tests were completed in a quiet listening room in the University of Huddersfield and stimuli were played back through Sennheiser HD650 headphones – the same that were used for the inverse filter measurement. It was designed to be as simple as possible, consisting of a circle on a screen representing the median plane. This was a custom adaptation of HULTI-GEN in Max MSP [19].

Subjects were asked to move a pointer to the perceived location on the circle, ignoring if the sound was slightly off centre. Each test consisted of 45 trials, three EQ conditions of three different directions, each repeated 5 times during the test. Each test was repeated 3 times and prior to each test a familiarization test was to be completed which consisted of only the unaffected HRTFs (i.e. EQ0 conditions).

3 Experimental Results

Data from the listening tests were analysed using the SPSS Statistics 22 software. The data was found to be non-parametric; additionally results for some directions were bimodal due to the large amount of front-back ambiguities often present in median plane localisation, especially for non-elevated sources. This can be seen in the bubble plots in Figure 1.

3.1 Non-elevated Sources

The data indicates that at 0° for subject S1 there is a significant difference in source position between EQ0 and EQ1 (p = .009) and between EQ0 and EQ2 (p = .003), with both pairs of conditions resulting in a moderately large effect size (r = -.543 and r = -.602 respectively). Results for this direction, for this particular subject, show a strong shift towards the source being perceived to be behind the listener as the level of notch reduction was increased. This suggests that notch equalisation does in fact effect median plane localisation for subject S1 at 0° in that the image shifts towards the rear.

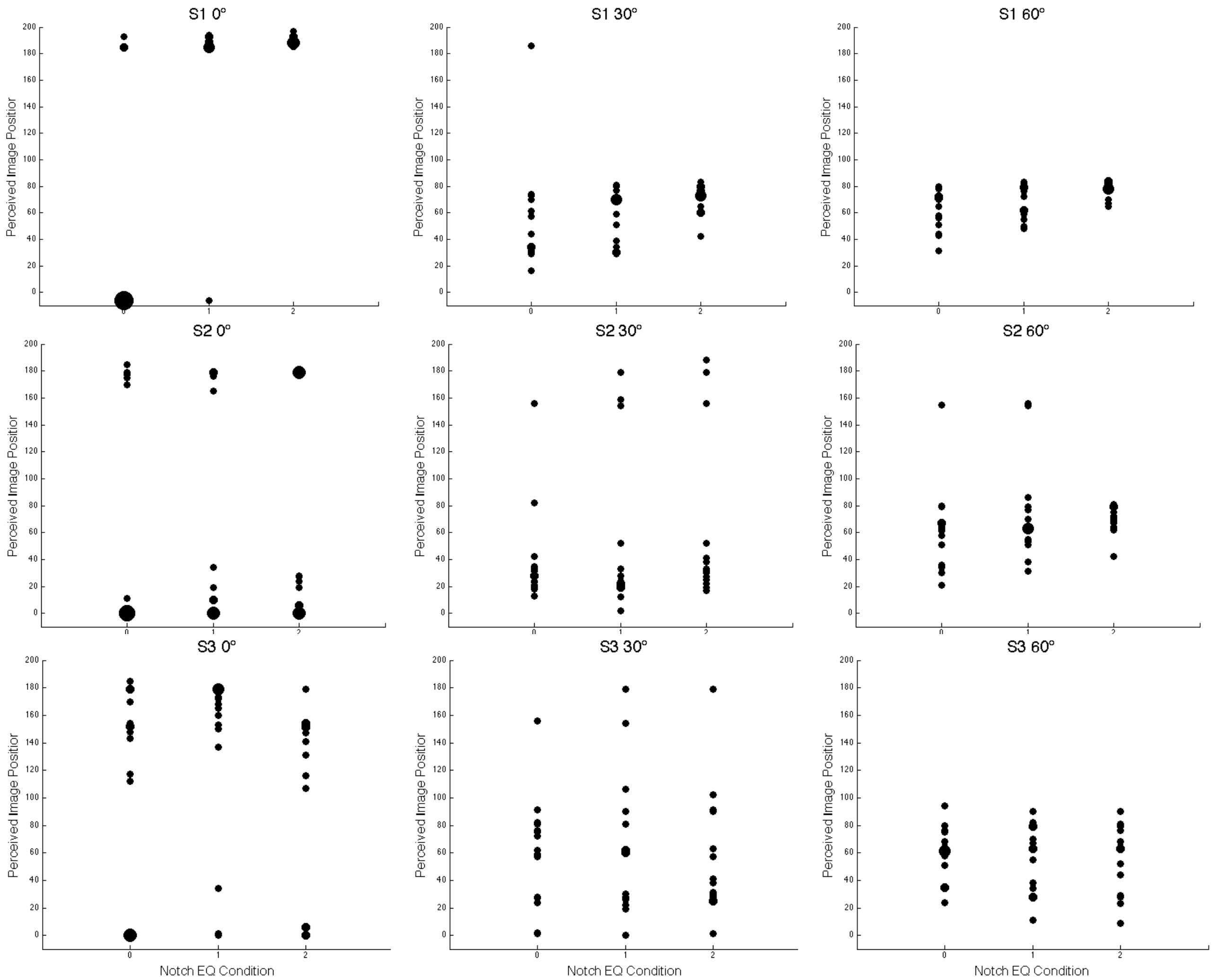


Figure 1. Bubble plots representing listening test responses, where the bubble size is proportional to the number of responses perceived at that particular location. Each graph represents the responses for the indicated subject at a certain angle in the median plane, where the X axis represents the EQ condition (EQ0, EQ1, EQ2) and the Y axis is the perceived source location in the median plane. The increase in front-back confusion after notch reduction with non-elevated sources can be observed for all subjects. Additionally, the general rise in source position (especially at 60˚ for S1 and S2) can also be noted

However, for subjects S2 and S3, localisation regardless of notch manipulation, appeared to generally be inaccurate. Reducing notch magnitude seemed to have no significant impact on perceived source position or localisation accuracy.

3.2 Elevated Sources

Statistical analyses imply that generally, for elevated sources (30° and 60°), particularly for subjects S1 and S2 (listeners with experience in these kinds of listening tests), localisation deviates from the target position as pinna notches are reduced in magnitude.

For S1 at 30° a significant difference between EQ0 and EQ2 exists (p = .033); producing a medium effect size (r = -.467). The mean source position taken from the descriptive statistics of the Wilcoxon test suggests that this effect is in fact a perceived rise in source position, increasing from 53.6° to 71.5°.

With S1 at 60°, a significant difference in localisation was found between EQ0 and EQ2, with a medium effect size (p = .027, r = .477). The mean values for these EQ conditions imply that the overall source position moves up in space as notches are reduced, going from 62.7 to 77.8 from EQ0 to EQ2 respectively. In addition to the overall source potion moving up for more extreme EQ conditions, the data is more accurately localised with these EQ conditions, with the inter-quartile range (IQR) moving from 21 to just 7 from EQ0 to EQ2 at 60°.

This tendency observed for S1 can also been seen for S2 at 60˚, not only does localisation become more consistent, but the source position moves upwards after notch reduction. The same cannot be said for S3; for this subject notch reduction does not seem to impact localisation for elevated sources. This could be due to the lack of experience in this type of listening test by the listener

4 Discussion

In this experiment, the effects of altering the depth of specific pinna-related cues in HRTFs in the median plane have been explored by acquiring individualised HRTFs for three subjects. After various processing was applied to the HRTF measurements in order to remove frequency characteristics of the experimental apparatus, pinna-related notches were identified and their magnitude reduced at two degrees of reduction.

Results from the pilot tests suggest that reducing pinna-specific spectral notches in individual HRTFs generally causes a deviation in localisation from the target source position. This result was expected, however, the exact extent of deviation was unknown. Although the experimental conditions were restricted and simplified, a trade-off between the necessity of the pinna-related spectral notches and the pitch-height effect [20] might be suggested from the data obtained from S1 and S2.

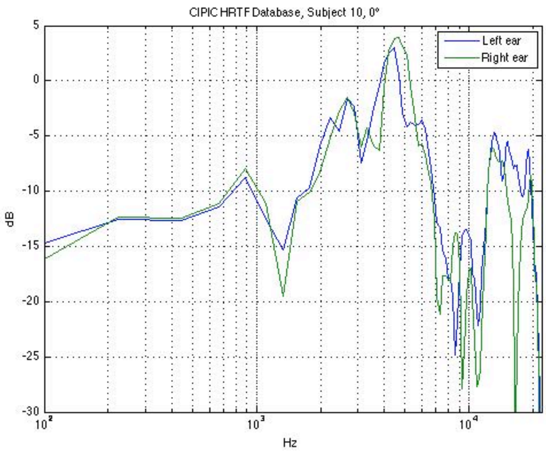
Despite the test results, it is important to consider that subjects may not perceive the source to be coming from the target position; even before notch equalisation is applied. They may naturally perceive the source to be coming from higher or lower in space than the original target angle or experience front-back confusion.

4.1 Non-elevated Sources

It can be said for non-elevated sources in the median plane (0° and 180°) that front-back confusion is common, even with unaffected HRTFs. But data from this experiment suggests that notch reduction can result in the image shifting entirely to the rear due to the change in high-low frequency ratio.

It has already been established that a high level of front-back ambiguities exists in median plane source localisation, especially with regard to binaural reproduction [12, p. 104]. As reported by Blauert, Von Békésy (1930) observed that localisation via headphone playback, even with the same source, can be made to appear from in front or behind depending on the expectations or state of mind of the listener. The ability to cognitively reverse the direction was based on the human ability to concentrate on a particular attribute of a signal and supress the rest.

(a)



(b)

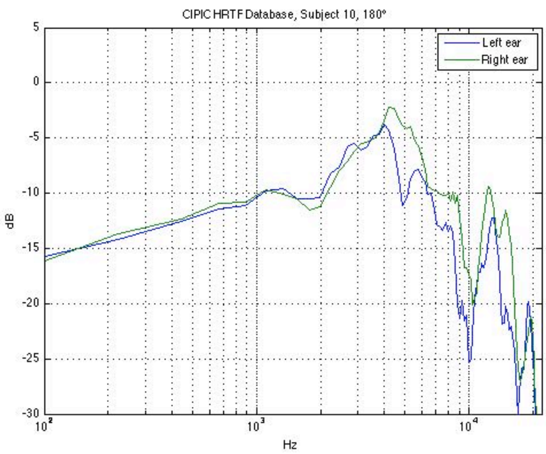
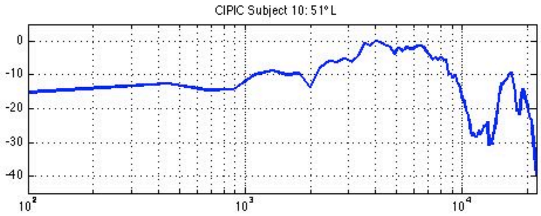


Figure 2. (a) CIPIC measurements for 0° (b) CIPIC measurement for 180° [21]

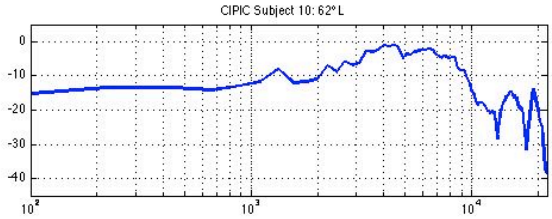
As measurements for the experiment were only made in the frontal median plane, it became apparent to the author that the frequency responses of behind directions could contribute to an understanding of why so much front-back confusion occurred for sources located at 0˚, especially after notch reduction. Measurements of subject 10 in the CIPIC HRTF database [21] at 0° and 180° in the median plane were examined (Figure 2). Looking at the frequency spectra of these measurements, it becomes evident that reducing the depth of the spectral notch located around the centre frequency of around 9 kHz at 0° would change the low-high frequency ratio, especially with regards to the area around 4 kHz, which is a directional band for in front [12, p. 110]. Therefore, as the notch level is reduced, the overall frequency spectrum is expected to become closer to that of the 180° measurement.

4.2 Elevated Sources

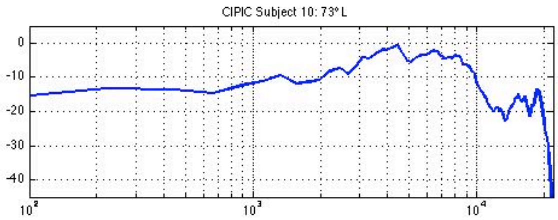
Based on the analysis of the data obtained from elevated sources, a general trend appears, especially with subjects S1 and S2, which implies that as the depth of notches are reduced, the source image shifts somewhat upwards in space. This upwards shift could be due to the pitch-height effect [20], which would become more noticeable with sources with relatively flat frequency spectra, such as the white noise source used in this experiment. Further testing with a range of source types and more subjects would provide better results.



(a)



(b)



(c)

Figure 3. CIPIC measurements for the left ear of Subject 10 at (a) 51°, (b) 62° and (c) 73° in the median plane [21]. It can be seen that the prominent high frequency, pinna related notches are reduced in magnitude as the elevation angle increases.

Due to the upwards shift in source position after notch reduction, it occurred to the author that perhaps these types of manipulations somewhat mimicked HRTFs situated higher in space. This can be seen when examining different median plane directions in the CIPIC HRTF database [21]. Generally as the source position moves upwards in space, so does the centre frequency of the most prominent high frequency notch, but additionally it would appear that the magnitude of the notch also decreases.

4.3 Further Work

The current pilot experiment provides provisional results and novel insights on the effect of spectral notch depth manipulation on vertical localisation. Further study is required to obtain adequate data with more subjects, covering a wider range of median plane directions; not limited to the frontal hemisphere. An extension of this would be to reduce individual notches; one at a time, instead of all of them together to understand which notch is of greater importance as well as the required magnitude the these specific notches.

Further work will also investigate the role of asymmetries between the left and right ears of subjects, as there appeared to be different numbers of notches, located in different frequency regions when comparing the two ears of subject. This did not only differ from subject to subject but also between different directions for the same subject.

References

[1] J. W. S. Lord Rayleigh, "XII. On our perception of sound direction," *Philos. Mag. Ser. 6*, vol. 13, no. 74, pp. 214–232, 1907.

[2] C. Avendano, V. R. Algazi, and R. O. Duda, "A head-and-torso model for low-frequency binaural elevation effects," in *1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 1999, pp. 179–182.

[3] S. K. Roffler and R. A. Butler, "Factors That Influence the Localization of Sound in the Vertical Plane," *J. Acoust. Soc. Am.*, vol. 43, no. 6, pp. 1255–1259, Jun. 1968.

[4] M. B. Gardner, "Some monaural and binaural facets of median plane localization," *J. Acoust. Soc. Am.*, vol. 54, no. 6, pp. 1489–1495, Dec. 1973.

[5] M. M. V. Wanrooij and A. J. V. Opstal, "Sound Localization Under Perturbed Binaural Hearing," *J. Neurophysiol.*, vol. 97, no. 1, pp. 715–726, Jan. 2007.

[6] V. R. Algazi, R. O. Duda, R. Duraiswami, N. A. Gumerov, and Z. Tang, "Approximating the head-related transfer function using simple geometric models of the head and torso," *J. Acoust. Soc. Am.*, vol. 112, no. 5, pp. 2053–2064, Nov. 2002.

[7] V. R. Algazi, R. O. Duda, R. P. Morrison, and D. M. Thompson, "Structural composition and decomposition of HRTFs," in *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, 2001, pp. 103–106.

[8] D. S. Brungart and G. D. Romigh, “Spectral HRTF enhancement for improved vertical-polar auditory localization,” in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2009. WASPAA ’09*, 2009, pp. 305–308.

[9] C. J. Chun, H. K. Kim, S. H. Choi, S.-J. Jang, and S. Lee, “Sound source elevation using spectral notch filtering and directional band boosting in stereo loudspeaker reproduction,” *IEEE Trans. Consum. Electron.*, vol. 57, no. 4, pp. 1915–1920, Nov. 2011.

[10] 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,” *Appl. Acoust.*, vol. 68, no. 8, pp. 835–850, Aug. 2007.

[11] F. Asano, Y. Suzuki, and T. Sone, “Role of spectral cues in median plane localization,” *J. Acoust. Soc. Am.*, vol. 88, no. 1, pp. 159–168, Jul. 1990.

[12] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT Press, 1997.

[13] J. Hebrank and D. Wright, “Spectral cues used in the localization of sound sources on the median plane,” *J. Acoust. Soc. Am.*, vol. 56, no. 6, pp. 1829–1834, Dec. 1974.

[14] E. A. Macpherson and A. T. Sabin, “Vertical-plane sound localization with distorted spectral cues,” *Hear. Res.*, vol. 306, pp. 76–92, Dec. 2013.

[15] V. C. Raykar, R. Duraiswami, and B. Yegnanarayana, “Extracting the frequencies of the pinna spectral notches in measured head related impulse responses,” *J. Acoust. Soc. Am.*, vol. 118, no. 1, pp. 364–374, Jul. 2005.

[16] D. Wright, J. H. Hebrank, and B. Wilson, “Pinna reflections as cues for localization,” *J. Acoust. Soc. Am.*, vol. 56, no. 3, pp. 957–962, Sep. 1974.

[17] G. Kearney and T. Doyle, “A Virtual Loudspeaker Database for Ambisonics Research,” *ICSA 2015 045*, 2015.

[18] D. Johnson, A. Harker, and H. Lee, “HAART: a new impulse response toolbox for spatial audio research,” presented at the 138th Annual Audio Engineering Society AES Convention, Warsaw, Poland, 2015.

[19] C. Gribben and H. Lee, “Towards the development of a universal listening test interface generator in Max,” presented at the 138th Annual Audio Engineering Society AES Convention, Warsaw, Poland, 2015.

[20] C. C. Pratt, “The spatial character of high and low tones.,” *J. Exp. Psychol.*, vol. 13, no. 3, pp. 278–285, Jun. 1930.

[21] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, “The CIPIC HRTF database,” in *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, 2001, pp. 99–102.