The Perceptual Contribution of Pinna Related Transfer Function Attributes in the Median Plane

By

JADE RAINE CLARKE

Supervisor: Dr. Hyunkook Lee

ABSTRACT

This thesis describes the work carried out to investigate the perceptual effects of pinna notches in median plane sound localisation. Literature regarding sound localisation and the effects of the pinnae is outlined before a thorough description of the measurement procedure to obtain individualised HRTFs (head related transfer function) is given. HRTFs of three subjects were recorded at seven different positions in the median plane (0˚, 30˚, 60˚, 90˚, 120˚, 150˚ and 180˚). Two experiments were carried out using the measurements. The first consisted of reducing the magnitude of, and removing pinna related notches in the HRTFs to identify the perceptual effects of notch manipulation in both virtual reverberant and pseudo-anechoic conditions. Results for the first experiment show a great deal of variation between subjects, although it can be said that pinna notch filling is most detrimental to median plane localisation in the BRIR (binaural room impulse response) condition and often results in hemispheric reversals and localisation inconsistencies. The second experiment compared localisation abilities of binaurally presented sound sources in reverberant conditions to that of binaural pseudo-anechoic conditions, using real room loudspeaker localisation as a reference. Results from the latter experiment show that virtual localisation in the median plane is better in the presence of reverberation, and that subjective experience in a listening room may influence this result.

ACKNOWLEDGEMENTS

The author would first and foremost like to thank Dr. Hyunkook Lee for all of his time, insight, support and guidance over the years, as well as the Music Technology studio technicians at the University and the members of the Applied Psychoacoustic Laboratory (past and present) for their additional support. Most notably, particular appreciation to Dale Johnson and Maksims Mironovs for sparing a great deal of their time to take part in the experiments. The author would also like to thank Alexandra Bartles for her time and expertise in recommending the procedure to obtain individualised ear impressions, and additionally for taking the time to create the impressions at the university.

THE PERCEPTUAL CONTRIBUTION OF PINNA RELATED TRANSFER FUNCTION ATTRIBUTES IN THE MEDIAN PLANE I

ABSTRACT II

ACKNOWLEDGEMENTS II

TABLE OF FIGURES V

LIST OF TABLES VII

1. SOUND LOCALISATION 1

1.1. DIRECTIONAL BANDS 2

1.2. SPECTRAL CONTENT FOR VERTICAL LOCALISATION 3

1.3. INDIVIDUAL DIFFERENCES IN HRTFS 4

1.4. HRTF MODELLING AND ENHANCEMENT 5

1.5. HEAD AND TORSO REFLECTIONS 6

2. THE PINNAE 7

2.1. ANTHROPOMETRY 7

2.2. PINNA RELATED CUES 8

2.3. PINNAE MODELS 8

2.4. PINNAE OCCLUSION 9

2.5. PINNAE ASYMMETRY 10

2.5.1. BINAURAL ASYMMETRY 10

2.5.2. ANTHROPOMETRIC ASYMMETRIES 12

2.6. SUMMARY 13

3. LOCALISATION OF SOUND IN ROOMS 14

3.1 LOCALISATION IN REVERBERANT ROOMS 14

3.1. ROOM DIVERGENCE EFFECT 16

3.3 LOCALISATION OF REAL SOURCES VERSUS VADS 16

3.2. SUMMARY 18

4. HRIR MEASUREMENTS 19

4.1. PARTICIPANTS 19

4.2. EAR IMPRESSIONS 19

4.3. HRTF ACQUISITION 20

4.4. SPEAKER POSITIONS 20

4.5. HEAD ALIGNMENT 21

4.6. HEADPHONE MEASUREMENTS 22

5. NOTCH FILLING EXPERIMENT 23

5.1. HRIR PROCESSING 25

5.1.1. COMPUTATION OF INVERSE FILTERS 26

5.1.2. IDENTIFICATION AND FILTERING OF NOTCHES 27

5.2. EXPERIMENTAL PROCEDURE 29

5.2.1. STIMULI 30
5.3. RESULTS 31
5.3.1. BRIR LOCALISATION 32
5.3.2. ANECHOIC LOCALISATION 36
5.4. DISCUSSION 39
5.4.1. REVERBERANT CONDITION 40
5.4.2. ANECHOIC CONDITION 41
5.4.3. HEMISPHERIC REVERSALS 41
5.4.4. OVERALL EFFECT OF NOTCH FILLING 42
5.4.5. LIMITATIONS 44
6. LOCALISATION UNDER DIFFERENT LISTENING CONDITIONS 45
6.1. EXPERIMENTAL PROCEDURE 46
6.1.1. STIMULI 47
6.2. RESULTS 47
6.2.1. HL RESULTS 49
6.2.2. MM RESULTS 52
6.2.3. DJ RESULTS 53
6.2.4. FRONT-BACK REVERSALS 55
6.3. DISCUSSION 55
6.3.1. LIMITATIONS 58
7. CONCLUSION 59
7.1. FURTHER WORK 60
8. REFERENCES 61
9. APPENDIX 71
9.1. IDENTIFIED NOTCHES 71
9.2. NF100 NOTCH CONDITIONS 73
9.3. NF50 NOTCH CONDITIONS 80
9.4. NOTCH FILLING CODE 84

Table of Figures

Figure 1: Approximation of the dimensions of the pinna. Redrawn from Gilkey & Anderson (2014, p. 32)..7

Figure 2. Ear impressions inserted into the ear of subject DJ before microphone insertion. 19

Figure 2. Ear impressions inserted into the ear of subject DJ before microphone insertion. 20

Figure 3. Distances of speaker positions in relation to the position of the participant's head 21

Figure 4. Head alignment with XY lasers for BRIR measurements of subject DJ 21

Figure 5. Processing steps involved in filtering pinna notches from the HRTF spectrum. Where the 'raw' part represents the original BRIR recording and the 'true' part represents the same recording with a diffuse filter created from measurements of the microphones in the listening room applied to it 25

Figure 6. Example of subject HL's filters used for headphone equalisation. The left hand figure represents the frequency response of the average of all of the headphone measurements for the left and right ear and the right figure is the resulting headphone filter created after Kirkeby regularisation is applied. The average measurement (left) represents the transfer function of the headphones, the microphones and the path between the two transducers at the time of measurement, and was removed from the resulting signal once the filter (right) was created and applied to the stimulus 26

Figure 7. Notch filling in the most extreme condition (NF100) for subject HL at 0˚. The plots show frequency manipulations on both the left and right ears. For each ear, there are two graphs. The top plot represents the actual filtering applied to the raw pinna response (pink), against the 'true' pinna measurement (black) and the raw measurement (blue) that was filtered. The plot below represents the 'target' magnitude for each condition (pink). Effectively, how the true pinna is being filtered in each condition. 28

Figure 8. Notch filling in the least extreme condition (NF50) for subject HL at 0˚ to demonstrate the preservation of the spectral shape of the notches as their magnitudes are reduced. The plots show frequency manipulations on both the left and right ears. For each plot at each condition, there are two graphs. With the top representing the actual filtering applied to the raw pinna response (pink), against the 'true' pinna measurement (black) and the raw measurement (blue) before filtering. The plot below represents the 'target' magnitude for each condition (pink). 29

Figure 9. Diagram of familiarisation test used for the experiment. Responses were obtained after the subject placed the red circle on the outside of the circle, which had a resolution of 1˚. Each of the stimuli, which could be played by simply clicking on a circle, were randomised for each sitting of the test 29

Figure 10. Bubble plots of localisation results for notch filling experiment in the reverberant room condition. Where larger bubbles represent more responses at that particular angle. NF0 represents the original HRTF, NF50 is the HRTF with the pinna related notch filled at 50% and NF100 is the spectral notch filled 100%, being flat at 0dB. 32

Figure 11. Bubble plots of localisation results for notch filling experiment in the anechoic condition. Where larger bubbles represent more responses at that particular angle. NF0 represents the original HRTF, NF50 is the HRTF with the pinna related notched filled at 50% and NF100 is the spectral notch completely filled, 100%, being flat at 0dB 38

Figure 12. PRTF spectra of subject DJ, comparing the 60˚ and 120˚ angles for each ear for each step of the notch manipulation (NF0, NF50 and NF100) 40

Figure 13. Frequency response of HL and MM, comparing 150˚ and 180˚ NF0 conditions to determine if localisation of the two stimuli were similar due to similarities in the response. 43

Figure 14. Listening test interface for the real room speaker localisation. This test made use of a head tracker, which would indicate on the right in red which way the subject should move their head in order to realign. During feedback the test would stop 47

Figure 15. Bubble plots for localisation tests completed in all three listening conditions for subjects HL, DJ and MM. With the x axis representing the target position and the y axis the perceived angle by each subject. The larger bubbles represent more responses at that particular angle. The red, green and blue colours represent the real room (RR), VAD R and VAD A, respectively. Results from NF0 in the notch filling experiment are displayed in the VAD conditions 49

Figure 15. Bubble plots for localisation tests completed in all three listening conditions for subjects HL, DJ and MM. With the x axis representing the target position and the y axis the perceived angle by each subject. The larger bubbles represent more responses at that particular angle. The red, green and blue colours represent the real room (RR), VAD R and VAD A, respectively. Results from NF0 in the notch filling experiment are displayed in the VAD conditions 49

Figure 16. RMS localisation error for subject HL at each response method at each angle tested. The bars represent the RMS error in responses compared to the target position. These calculations removed any

outliers by discarding any results which exceed 1.5 interquartile ranges above the upper quartile or below the lower quartile. 51

Figure 17. RMS localisation error for subject MM at each response method at each angle tested. The bars represent the RMS error in responses compared to the target position. These calculations removed any outliers by discarding any results which exceed 1.5 interquartile ranges above the upper quartile or below the lower quartile. 52

Figure 17. RMS localisation error for subject MM at each response method at each angle tested. The bars represent the RMS error in responses compared to the target position. These calculations removed any outliers by discarding any results which exceed 1.5 interquartile ranges above the upper quartile or below the lower quartile. 52

Figure 18. RMS localisation error for subject DJ at each response method at each angle tested. The bars represent the RMS error in responses compared to the target position. These calculations removed any outliers by discarding any results which exceed 1.5 interquartile ranges above the upper quartile or below the lower quartile. 54

List of Tables

Table 1 Median of results and whether the results significantly differ to target angle after carrying out Wilcoxon Signed-Ranks test: * $p < 0.05$, ** $p < 0.01$. Bimodality of data tested and significant bimodalities identified: ^ $p < 0.05$, ^^ $p < 0.01$. Bold font indicates statistical testing show that the results are considered to be accurate. 34

Table 2. Percentage of reversals for each subject in each of the notch filling conditions tested in the reverberant room condition. * *only* frontal target positions reversed (front to back). ^ *only* rear targets reversed (back to front). 90˚ target was not included in these calculations. 35

Table 3. Median of anechoic notch filling results and whether the results significantly differ to target angle after carrying out Wilcoxon Signed-Ranks test: * $p < 0.05$, ** $p < 0.01$. Bimodality of data tested and significant bimodalities identified: ^ $p < 0.05$, ^^ $p < 0.01$. Bold font indicates statistical testing show that the results are considered to be accurate. 37

Table 4. Percentage of reversals for each subject in each of the notch filling conditions tested in the anechoic condition. 90˚ target was not included in these calculations. For these conditions, reversals were perceived in both the anterior and posterior of the median plane 39

Table 5. Median of results and whether the results significantly differ to target angle after carrying out Wilcoxon Signed-Ranks test: * $p < 0.05$, ** $p < 0.01$. Bimodality of data tested and significant bimodalities identified: ^ $p < 0.05$, ^^ $p < 0.01$. Bold font indicates statistical testing show that the results are considered to be accurate. For the purpose of comparison, results from NF0 for both room conditions in the notch filling experiment are displayed in the VAD conditions 48

Table 6 Percentage of reversals for each subject in each of the three conditions tested. * *only* frontal target positions reversed (front to back). ^ *only* rear targets reversed (back to front). 90˚ target was not included in these calculations. 55

Table 6 Percentage of reversals for each subject in each of the three conditions tested. * *only* frontal target positions reversed (front to back). ^ *only* rear targets reversed (back to front). 90˚ target was not included in these calculations. 55

1. Sound Localisation

Immersive three-dimensional (3D) audio is fast becoming more accessible for the average consumer and 3D auralisation is readily available via headphone playback in many modern technologies. To 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 (Lord Rayleigh, 1907) describes interaural time difference (ITD) and interaural level difference (ILD); the low frequency time and high frequency level difference in input signals reaching the left and right ears (Hans Wallach, 1939). However, effective sound reproduction in the vertical plane requires a deeper understanding of the auditory system and its physiology.

Previous studies have attempted to isolate individual spectral cues based on anthropometry in order to understand the upper body's contribution to the HRTF (head related transfer function) (Avendano et al. 1999; Gardner, 1973; Wanrooij & Opstal, 2007). Others have created structural HRTF models based on anthropometry (Algazi et al., 2002; Algaziet al., 2001). HRTFs are the acoustical transfer functions from a particular sound source located in space to the tympanic membrane.

Auditory localisation in the median plane, where ITD and ILD is not apparent, depends primarily on cues produced by sound reflecting from various parts of the upper body such as the HAT (head and torso) and the pinnae (the external part of the outer ear). Sound diffraction is frequency dependent, with effects only appearing when the wavelength is less than or equal to the size of the diffracting object (Gilkey & Anderson, 2014, pp. 49–75). The HAT is thought to account for low frequency elevation cues below 3 kHz (Avendano et al., 1999).

1.1. Directional Bands

Certain spatial locations in the median plane are known to correspond with different boosts in the frequency spectrum. Blauert (Blauert, 1997, pp. 108–111) broadly named these boosts ‘directional bands’, corresponding to 1 kHz and 10 kHz behind the listener, 300-500 Hz and 3 kHz in front and 8 kHz above. Hebrank & Wright (1974) conducted an experiment in which subjects were asked to indicate subjective impression of the direction in the median plane of white noise with LP (low pass), HP (high pass), BP (band pass) and BS (band stop) filters applied. The authors reported that sources containing frequency components between 4-16 kHz were necessary for vertical localisation. They found that behind cues contain a small peak at 10-12 kHz with a decrease in energy above and below the peak. Frontal cues contain a one-octave notch with lower cut-off of 7-9 kHz and high cut-off of 10 kHz and increasing the lower cut-off frequency of the one-octave notch is reported to increase the perceived frontal elevation.

Chun et al. (2011) conducted a study in which spectral notches were filtered and directional bands were boosted in loudspeaker localisation to further elevate sound sources from a given loudspeaker angle. The authors first created a structural HRTF based on head and pinna models and applied it to a complex mono signal. The stimuli used were male and female speech, a guitar, violin and a pop music sample. Spectral notch filtering and directional band boosting was then applied to the signal via QMF (quadrature mirror filter) analysis and synthesis. Their experiment resulted in sounds seeming to be elevated in space and around 20° Φ (elevation) when emanating from a loudspeaker located at 0° Φ.

Different tones are known to have a different spatial characters, as established by Pratt (1930), who found associations between pitch and vertical localisation. This was confirmed and investigated further by Roffler & Butler (1968b), who linked associative cues with vertical sound localisation i.e. a bird sound, which predominantly consists of high frequency content, is usually located above a listener, therefore humans learn by experience that high tones are from above. Additionally, results were consistent when tested on young children who had not yet learnt these associations. The authors found that instead of being located with regard to their actual position in space, listeners located band-passed tone bursts along a vertical axis where high frequency tones appear higher than their low frequency counterparts.

It was hypothesized by Blauert (1997, p. 101) that broadband sound signals contain more information about the source's position in space than that of narrowband signals, therefore correlation of the source's direction and the perceived auditory event should occur most with broadband signals. This was also found by a number of researchers, including Asano et al. (1990), Avendano et al. (1999), Butler & Belendiuk (1977), Hebrank & Wright (1974), Iida et al. (2007), Macpherson & Sabin (2013) and Middlebrooks et al. (1989).

1.2. Spectral Content for Vertical Localisation

To understand the necessary resolution of the spectral content in HRTFs, Asano et al. (1990) simplified the spectrum by means of frequency smoothing in order to identify how microscopic low frequency (below 2 kHz) and macroscopic high frequency characteristics affect median plane localisation. This method of frequency smoothing was implemented due to the resolution of the auditory system, which is higher at low frequencies. A source was virtually presented via headphone playback using HRTFs with different degrees of frequency smoothing applied. The authors found that frequencies below 4-5 kHz had little directional dependency, however, a notch existed at 6-10 kHz, which increased in frequency as the source elevation increased. Additionally, they identified a peak moving down in frequency at 11-14 kHz as the elevation angle increased. The results are consistent with that of Blauert (1997, pp. 109–111) and Hebrank & Wright (1974). The authors concluded that instead of recognising minute spectral patterns, the auditory system evaluates information regarding power in certain bands around 5-10 kHz relative to the power in bands above and below these frequencies.

Altering the magnitude of the spectrum has previously been studied by Macpherson & Sabin (2013), who did not pay attention to specific peaks and notches, but instead reduced the magnitude of peaks and notches or reduced or increased the spectral contrast across the entire frequency spectrum, modelling the level effect in the case of notch filling and the negative level effect in the peak levelling condition. They found that as peaks were levelled off and notches were filled in, elevation localisation accuracy decreased. Their experimental results suggest that peak levelling is less detrimental to localisation than notch filling. Furthermore, increasing spectral contrast had little effect on perceived elevation, however the perceived location appeared to depend on the relative rather than absolute frequency shape of the spectrum.

The level effect was explored by Vliegen & Van Opstal (2004), who found that vertical plane elevation localisation performance deteriorates for short duration wideband sounds presented at moderate to high levels. Their experiment presented wideband stimuli ranging from 3 – 100 ms presented at 26 – 73 dB SPL in two-dimensional space. Subjects responded with rapid head movements towards the perceived location. Elevation localisation was found to be strongly affected by both stimulus level and duration. Elevation gain was found to increase with increasing sound duration, until a plateau was reached above 30 ms. This effect had the most effect at the lowest and highest sound levels.

Roffler & Butler (1968a) first carried out localisation experiments in which broadband noise, LP filtered noise (<2 kHz), HP filtered noise (>2 kHz and >8 kHz), 600 Hz tone bursts and 4800 Hz tone bursts were played back to subjects at 4 loudspeaker locations in the median plane, at -13°, -2°, 9° and 20° Φ. The first experiment suggested that listeners could not localise the tonal stimuli or that of the LP noise. In their second experiment, 13 LP filtered noise stimuli with cut-off frequencies between 500 Hz – 12 kHz were present to subjects in a random order; performance significantly improved when energy between 7-8 kHz was included in the stimuli. Their final experiment explored the importance of the pinnae in vertical localisation, where 8 listeners were asked to locate broadband noise and 8 kHz HP filtered noise. One test was with the pinnae intact and another was with the pinnae strapped against their head. The latter test resulted in the most localisation error in the median plane. Authors 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.

1.3. Individual Differences in HRTFs

When discussing HRTFs, the question of whether individual measurements are necessary arises. Due to the unique, individual shape of the pinnae, each individual’s pinna-related cues will differ, consequently altering their HRTF in a different manner. Therefore, if non-individualised pinnae are used to obtain the listener's HRTF, median plane localisation is subject to error (Butler & Belendiuk, 1977). This was confirmed by Møller et al. (1996), who presented listeners with sounds through their own HRTF and those of other people's, the non-individual recordings result in very poor localisation accuracy, especially with regards to the median plane, distance and front-back confusion.

Itoh et al. (2007) examined whether subjects' individual frequencies in directional bands differ, and whether directional bands change according to stimuli bandwidth. Their experiment presented 1/3 and 1/6 octave band noise lasting for 1.2s with 100ms onset and offset ramps and 4.8 s inter-stimulus intervals to subjects in front, above and behind in the median plane. The authors confirm that directions bands significantly differ between subjects, although they occur behind at 0.8-1.6 kHz and 10-12.5 kHz centre frequencies, in front at 2-5 kHz and above at 6.3-8 kHz.

1.4. HRTF Modelling and Enhancement

Once the necessary features of the HRTF are understood, it is possible to attempt to model and enhance them. This has been done by a number of researchers in recent years. To understand the HRTF further, Algazi et al. (2001) performed a structural analysis by regarding it as a cascade of two filters; one for the HAT and one for the pinnae. Because each filter block is related to specific features of the anatomy, filter parameters can be related to anthropometry, allowing adaptation of generic HRTF models to suit individuals. The authors decomposed HRTFs by comparing measurements of isolated pinnae and the HAT with the pinnae removed (either by removing the pinnae of the KEMAR or strapping listeners' pinnae to their head), it was then shown that simple compositional rules can produce good correspondence between the measured and composite HRTFs.

Brungart & Romigh (2009) devised a novel HRTF enhancement technique that exaggerates the directionally dependent spectral cues for elevation localisation using both individual and generic HRTFs without impacting azimuth localisation. The enhancement procedure directly affected the magnitude of the frequency components of the HRTF by increasing the salience of the spectral variation, without distorting the interaural difference cues. This was done by separating what they called the 'lateral' and ‘vertical’ spectral components of the HRTF and only affecting the ‘vertical’ part. Their results show that over-enhancement of individual HRTFs may distort the spectrum too much, contradicting listeners’ experience of their own ears, however, dramatic improvements in vertical localisation were obtained, resulting in as much as 33% reduction in localisation error.

Iida et al. (2007), created a parametric HRTF (pHRTF) model based on extracted spectral cues obtained from individualised HRTFs (mHRTF). To achieve this, they first identified the influence of each spectral peak and notch in the spectrum by recomposing HRTFs through introducing each peak or notch from the mHRTF, until they were all present in the pHRTF. Listening tests were done in order to identify the point at which median plane localisation was accurate. Intuitively, pHRTFs composed of all spectral peaks and notches provide the same localisation accuracy as that of mHRTFs. In addition, they found that pHRTFs recomposed of the first peak and the first two notches in the spectrum provide almost the same accuracy as that of mHRTFs. Although the centre frequency, level and sharpness significantly changes with source elevation for notches, the peak, which is located at around 4 kHz, is not influenced by source elevation. It was hypothesised that the purpose of this peak is to act as a reference to analyse the directionally dependant notches. The work is consistent with that of Hebrank & Wright (1974), who concluded that frontal cues have a one octave notch with a lower cut-off between 4-8 kHz (which corresponds with the first notch) and increased energy above 13 kHz, that behind cues have a small peak between 10-12 kHz with a decrease in energy above and below the peak (corresponding to notch one and two). Their work plays a large role in the simplification of HRTFs and in understanding the specific spectral cues necessary for effective 3D localisation.

1.5. Head and Torso Reflections

As stated, the HAT accounts for frequencies in the HRTF below 3 kHz, its effects appear to be more prominent away from the median plane and it compliments high frequency monaural pinna cues. Localisation tests with low frequency sources show that multi path head-diffraction effects as well as shoulder and torso reflections may cause low frequency binaural elevation cues. The maximum time delay caused by the shoulders at 90° Φ is about 0.8ms (Avendano et al., 1999).

2. The Pinnae

The monaural characteristics of ear input signals provide the most significant cues in localising the vertical position of a sound source, mainly due to a series of directional dependent delays (100-300μ seconds) to the source by the pinnae cavities, creating peaks and notches in the spectrum (Batteau, 1967; Blauert, 1997, pp. 304–310; Gardner, 1973; Spagnol et al., 2013; Wanrooij & Opstal, 2007; Wright et al., 1974). The human hearing mechanism is extremely good at detecting these time delays, the just noticeable difference in delay being 20μ seconds (Wright et al., 1974). The ear's input signal could be regarded as a series of delay paths, whose inverse transform describes the source's location in space (Batteau, 1967; Geronazzo et al., 2010).

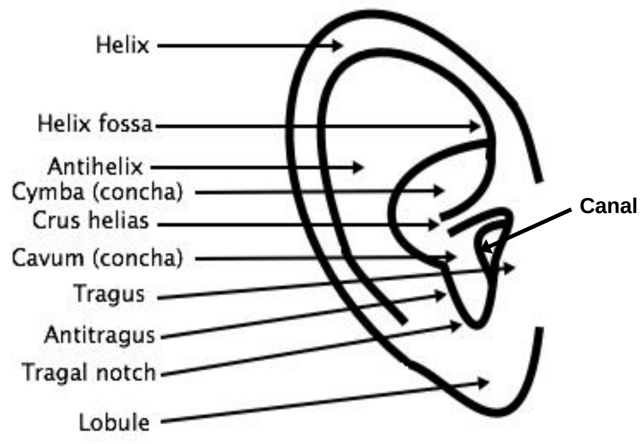


Figure 1: Approximation of the dimensions of the pinna. Redrawn from Gilkey & Anderson (2014, p. 32)

2.1. Anthropometry

The pinna (figure 1) has a fixed tubular canal, approximately 7.5mm in diameter and 22.5mm in length, vertically terminated by the tympanic membrane. The precise canal length is difficult to define as there is no clear boundary between it and the concha. The concha is a shallow cavity roughly 4cm^3 and partially divided by the crus helias (Gilkey & Anderson, 2014, p. 31). The authors believe that the cavum and cymba (lower and upper sections of the concha, respectively) have specific acoustical attributes that work independently to modulate the input signal, as opposed to the helix, anti-helix and lobule, which together are believed to produce a flange effect (Gilkey & Anderson, 2014, p. 32).

2.2. Pinna Related Cues

High frequencies with wavelengths smaller than the pinna dimensions reflect off its contours, typically the concha wall and rim, making high frequencies necessary for accurate vertical localisation. According to Gilkey & Anderson (2014, p. 28), the ear's primary resonance is 2.6 kHz, however, the most significant frequency modulations occur above 5 kHz. Asano et al. (1990) and Raykar et al. (2005) confirm that the modulations above 5 kHz, which are a result of the pinnae, are important for vertical localisation. Algazi et al. (2001) additionally certified that these modulations are a result of pinna reflections by comparing the spectrum of a pinna-less head with that of a headless pinna using both the KEMAR (Knowles Electronics Mannequin for Acoustic Research) and human listeners. They found that the high frequency HRTF components came from the PRTF (pinna related transfer function).

Geronazzo et al. (2010) modelled the characteristics of the PRTF by separating the resonant and reflective parts of the spectrum. They found two areas of high magnitude in the spectrum; the first one with a centre frequency of 4 kHz, spanning all elevations among most subjects in the CIPIC HRTF database, its bandwidth increasing with elevation. The second area of high magnitude varies in shape and amount among all subjects, however it is reported to be most prominent at low elevations between 12-18 kHz. This is consistent with results from Iida et al. (2007) and from Gilkey & Anderson (2014, p. 37), who identified the 6 modes of the concha under the blocked meatus condition. The results show the first mode to be omnidirectional at 4.2 kHz, a quarter the wavelength of the concha depth. The second and third modes were termed the 'vertical' pair, and modes four, five and six called the 'horizontal' triplet, corresponding to around 12-18 kHz.

2.3. Pinnae Models

Spectral cues produced by the pinnae are vital for effective 3D auralisation in the median plane, where binaural cues are not dominant. As a result, many researchers have sought to identify and model its various effects. Raykar et al. (2005) automatically extracted the frequencies of the notches in a PRTF, distinguishing them from other features in the HRTF (HAT/knee reflections). The estimated spectral notches were related to the physical dimensions of the pinna: for each extracted notch, a corresponding distance was plotted on an image of the pinna, thus allowing specific pinna-related peaks and notches to be identified. Their method removes all torso and knee reflections by applying a 1ms half Hann window to the beginning of the signal, this is because torso reflections occur after around 1.6ms and knee reflections after around 3.2ms, depending on the size of the subject (Iida et al., 2007; Raykar et al., 2005).

Furthermore, Ahuja & Hegde (2014) modelled the pinna related spectral notches using linear prediction residual cepstrum, claiming that their computational model of the frequencies of spectral notches was better than those obtained by anthropometric means, as done by Raykar et al. (2005), as their method provides finer pinna contours when related to anthropometry.

2.4. Pinnae Occlusion

Wanrooij and Opstal (2007) conducted listening tests in which one of the pinnae was perturbed. Results show that localisation in the horizontal plane shifts towards the unperturbed pinna and that accuracy drastically suffers under these conditions. The authors suggest that the contralateral monaural input is required for accurate localisation. However, they found that localisation accuracy in the median plane was not significantly disturbed, indicating that elevation localisation can be resolved with monaural cues only.

Blauert (1997, p. 99) reported a number of studies where the pinnae were modified in various ways, usually bypassed by means of varying sized tubes stuck in the ear canals of subjects, extending beyond the ends of the outer ears. These found that the auditory event appeared to come from behind, regardless of the sound’s actual position in space, which indicated that the pinnae are necessary for front back resolution, as well as vertical localisation.

The importance of pinnae cavities were additionally investigated by Gardner & Gardner (1973). The authors created moulds of the cavities of their own pinnae and completed localisation tests gradually occluding more of the pinnae. They found that increasing occlusion progressively decreases localisation ability, though localisation was still somewhat better in the anterior as opposed to the posterior section of the median plane. Authors also found that without occlusion, localisation was better for broadband noise and that localisation abilities for 8 kHz and 10 kHz bands were similar to that of wideband noise.

A follow on experiment was completed with five participants completing localisation tests with varying levels of pinna occlusion in binaural localisation in the median plane (Gardner, 1973). In this experiment, subjects were asked to verbally indicate which of the nine possible loudspeakers the sound was coming from. These were positioned in front of them between ±18˚ ϕ. Four conditions were tested in the experiment; 1) no occlusion, 2) cavities of one pinna occluded, 3) cavities of both pinnae occluded and 4) cavities and passageway to canal of one pinna occluded. The latter condition was not included in analysis of binaural pinna effects. Similar to the authors previous experiment, more localisation error occurred with more occlusion. However due to the discrete response method, hemispheric reversals could not be accounted for in this experiment.

2.5. Pinnae Asymmetry

Various studies have explored the contribution of asymmetries and monaural versus binaural listening conditions on localisation and many authors have sought to understand the differences between the left and right ears and its contribution to localisation, finding that one ear can be better than the other at processing spatial information. These studies have extended into the field of anthropometry where the physical features of the pinnae have been measured and compared.

2.5.1. Binaural Asymmetry

Searle et al. (1975) studied binaural pinna disparity by comparing the individualised left and right HRTFs of subjects as a function of elevation angle, ϕ. The authors found that broad notches exist at around 8.5kHz from 15˚ to 30˚ ϕ, which gradually disappear as ϕ increases. This becomes obvious when elevation angle increases from 90˚-120˚, where spectra are much smoother and therefor notch variations due to asymmetry are less noticeable. Listening tests carried out under dichotic (original asymmetric HRTF) and diotic (left copied to right ear or vice versa) listening conditions revealed that diotic conditions provide less accurate results. The authors suggest this could be due to the subject being deprived of one or two independent monaural cues or deprived of their binaural pinna disparity.

The contribution of asymmetries in the vertical plane was previously studied by Ivarsson et al. (1980) where subjects' binaural performance and monaural left or right only was tested. As expected, authors found that subjects were much better at localising sources that were presented binaurally; additionally, they found that performance using the left ear was better than the right ear. From these results they suggested this could imply the existence of cerebral asymmetries for auditory localisation; as each hemisphere of the brain processes the contralateral ear signal, and speech is processed by the left cerebral hemisphere and spatial awareness the right (this is also true for visual cues), a right side cerebral dominance may be present when analysing spatial attributes of an auditory scene.

This was supported by Butler (1994) who performed localisation tests on 10 subjects in the median plane as well as various other directions using three conditions: binaural, monaural left and monaural right using real loudspeakers positions and blocking one of the ear canals for the monaural conditions. They found that left ear localisation is significantly more accurate than the right (p = .048) and there is less displacement from the midline (p = .059) when the left ear is presented with a source in the median plane. The authors suggest that a right cerebral dominance exists with respect to spatial awareness. Burke et al. (1994) and Abel et al. (1999) found a right ear advantage when processing linguistic material and left ear dominance with regards to spatial information and Savel (2009) also noticed a left ear dominance, however the latter observation was limited to the horizontal plane.

The contribution of interaural spectral differences (ISD) in terms of localisation was explored by Jin et al. (2004). They presented broadband stimuli binaurally to subjects and presented a 'flat' frequency spectrum between 300-14kHz at the left ear, while preserving the ITD and ILD cues by means of a finite impulse response (FIR) filter and additionally preserving the ISD between the two ears. They presented subjects with two conditions: veridical interaural (VI) and veridical right (VR). From this experiment it was established that the ISD cue alone is not sufficient to provide accurate information regarding the source's location. However only the true right ear spectrum was tested and there was no investigation into ISD with respect to left-right dominance.

Zhong et al. (2013) investigated the left-right asymmetry of 52 subject's HRTFs in both the horizontal and median planes and derived 'asymmetry coefficients' by means of cross-correlation analysis on equivalent rectangular bands (ERB), scaled in accordance with the human hearing resolution. The authors found that HRTFs are fairly left-right symmetric below 5.5kHz but at higher frequencies more asymmetry occurs as a result of the pinnae. Additionally, more asymmetries were found from 90˚-135˚ ϕ. It was suggested that this could be due to anatomical asymmetries in the head shape, hairstyle, pinnae etc. Authors also suggest that as ϕ increases from -30˚ - 75˚, asymmetry moves towards higher frequencies due to pinna related notches. In addition to these objective measurements, subjective psychoacoustic tests were done binaurally with six subjects. Four stimuli (white noise and low-passed white noise with 4kHz, 8kHz and 12kHz cut-off frequencies) were presented to subjects with either their individualised original binaural, left copied to right side and right copied to left side. Testing revealed that the amount of asymmetry increased with frequency, except at 60˚ ϕ and for 4kHz low-passed stimuli, the amount of asymmetry lies in a region of 'indiscrimination' except at 0˚, 30˚ and 60˚ ϕ. It was concluded that median plane HRTFs have much less asymmetry and can be considered symmetric below 5.5kHz, with asymmetry becoming noticeable at around 6.2kHz.

2.5.2.Anthropometric Asymmetries

Pinnae asymmetry has previously been explored with respects to anthropometry and biometrics, a study of particular interest is by Claes et al. (2015) who studied the intersubject and intrasubject variation in computer tomography (CT) scans of the left and right pinnae across 340 subjects (178 male, 162 female). They found that the most asymmetric parts of the pinnae were located in the lobule, tragus and posterior part of the helix, and this could be a result of the protruding nature of these physical features, which are susceptible to 'developmental instabilities'. Directional asymmetries were additionally found in the crus of the helix, antitragus and the lower posterior part of the helix. A greater intrasubject symmetry was observed for the concha, antitragus and antihelix, however more intersubject variations exist with these parts.

Asymmetry with regards to front-back localisation and anthropometry was recently explored by Bomhardt & Fels (2017), who studied 3D ear models obtained by CT scans and their corresponding HRTF data sets from the Institute of Technical Acoustics at Aachen University (Bomhardt & Fels, 2016). The authors compared the ipsilateral and contralateral ear response of six different directions in the horizontal plane (-10˚, -30˚, -60˚, -70˚, -135˚ and -165˚). Their work is consistent with that of Searle et al. (1975) in that clearly pronounced notches gradually disappear as elevation angle increases. Front-back confusion rate was much larger for symmetric than asymmetric HRTFs and this was larger for lateral directions (-60˚ and -70˚) than for frontal and rear directions.

The literature indicates that left – right asymmetries in HRTFs are important for accurate localisation and efficient externalisation with regards to binaural reproduction of a sound source. It has so far been established that a right ear dominance exists with respect to accurate localisation and that the perturbing physical features of the pinnae are more susceptible to asymmetries, which can be observed in the PRTF spectra. These asymmetries become more prominent and move to higher frequencies as ϕ increases.

2.6. Summary

The literature reviewed in this chapter provides information regarding the role of the pinnae in sound source localisation. The anthropometry was described before various aspects of the pinnae's contribution to localisation, both binaurally and monaurally, were outlined. It is evident from the literature that the contours of the pinna produce certain directionally dependent peaks and notches in the HRTF spectra, that in turn provide cues about the sound source's location in space. The directionally dependent notches in the HRTF spectra were explored further in chapter 5.

3. Localisation of Sound in Rooms

In addition to sound localisation abilities depending on source spectrum, localisation ability is known to differ under various listening conditions, whether that be in real room listening or using VADs (virtual auditory displays), in both reverberant and anechoic conditions. And even in anechoic conditions, the question of the necessary length of the HRIR to reproduce sufficient information about the vertical location of a sound source has been explored. Iida & Oota (2018) completed an experiment whereby the length of anechoic HRIRs were shortened and localisation accuracy of various IR lengths were compared. The authors used the full IR length, and IRs truncated to 2ms, 1ms, 0.5ms and 0.25ms and convolved these with 1.2s long Gaussian white noise, band-passed at 200 Hz to 17 kHz. Shorter HRIRs effectively removed some of the physical characteristics from the response, such as knee and shoulder reflections and additionally smoothed out the HRTF frequency response, removing some of the peaks and notches in the spectrum which the same authors previously identified as necessary for elevation localisation (Iida et al., 2007). The experiment concluded that the full length HRIR and the 1ms and 2ms truncated IRs provided approximately the same vertical localisation performance, while a significant increase in mean localisation error was observed between the full length HRIR and the shorter HRIRs (0.5ms and 0.25ms). Though there appeared to be a large number of inconsistencies and errors with all IR lengths.

3.1 Localisation in Reverberant Rooms

Localisation of real sound sources in rooms which allow for reflections and reverberations can improve localisation abilities. In an experiment by Shinn-Cunningham (2001), who tested anechoic versus room localisation in two dimensional space (horizontal and vertical plane), noted that spatial perception was found to be critically affected by room acoustics. Subjects were found to better localise sound in reverberant environment than in the anechoic condition, and they began to ‘learn’ the room after repeated testing in the same space, resulting in more consistent localisation after multiple testing. The effect of ‘room learning’ was also found by a number of studies, including Kop & Shinn-Cunningham (2002), Santarelli et al. (1999), Shinn-Cunningham (2000), Shinn-Cunningham et al. (2005).

In a study addressing learning reverberation in two dimensional localisation, Shinn-Cunningham (2000) reviewed evidence that listeners are able to adapt to reverberation in a room, and that although reverberation can degrade perception of source direction, performance can improve after some experience in a given room. The study recognised that *a priori* expectations of localisation cues in an anechoic space fits a specific source position, however once room reverberation patterns are learnt, disruptive spatial cues created by the reverberations can be suppressed.

A study by Sinker & Shirley (2016) investigated the perceptual consequences of altering the length of BRIRs (binaural room impulse response) recorded using a Neuman KU-100 dummy head and convolved with four anechoic sources (pink noise bursts, male speech, female speech and a short music excerpt). They used a MUSHRA style test and asked subjects to rate externalisation of each stimulus. IR lengths ranged from 20ms-100ms, varying in 10ms increments. They additionally added a 200ms (fully auralised) reference and 0ms (completely dry) anchor. The same stimulus set was tested in both the room in which the IRs were recorded and that of a different room. The authors found that IR lengths up to 40ms significantly improved externalisation, however above this value, although externalisation increased, it was not perceptually significant. The study noted no significant difference in externalisation was found between matched and mismatched rooms. A similar IR boundary length for externalisation was determined by Crawford-Emery & Lee (2014), though their study used non-anechoic musical stimuli, which already had reverberation present in the stimuli.

Another study examining the effect of IR length on externalisation of binaural audio was carried out by Völk (2009). The experiment explored both distance and height judgement related to externalisation. Experimenters played 200ms broadband noise stimuli convolved with a generic HRIR located at 0˚ in the median plane, truncated at 13.6ms, 15.2ms, 17ms, 19.3ms, 23.2ms, 46.4ms, 92.9ms and 185.8ms, separated by first order reflections. Results showed that the higher the order of reflection, the larger the distance judgements by subjects, although reverberation after 100ms did not influence externalisation. Additionally, height judgments were not significantly influenced by IR length.

3.1. Room Divergence Effect

Experiments have found that listening to a virtualised sound in a different room to that which the original IR has been recorded can create a room divergence effect, which consequently negatively effects externalisation of the sound source (Werner & Klein, 2014). In a study by Sporer et al. (2016) the effect of DRR (direct-to-reverberation ratio) was explored in the context of externalisation with regards to room divergence. Experimenters allowed subjects to adjust the DRR based on individual expectation alone, in order to achieve perceptual congruence between sound image synthesis and the listening room used. This method pushed synthesis towards the room and increased perceived externalisation for the subject.

3.3 Localisation of Real Sources Versus VADs

Localisation of virtual sources, especially in relation to the median plane, would be expected to be as accurate as that of real sources, if individualised HRIRs are used and correctly recorded. However, this is often not the case. In localisation experiments by Wightman & Kistler (1989a), individualised HRIRs were recorded in an anechoic chamber at various azimuth and elevation angles. Experimenters then carried out a psychophysical validation test of the recorded IRs, comparing subjects' virtual localisation abilities to that of real sources located in the same position in the room (Wightman & Kistler, 1989b). The subjects were asked to report the apparent spatial position of 250ms Guassian noise bursts, with a 20ms onset and offset ramp and 300ms inter-stimulus interval presented at 70dB SPL and band-passed between 200 Hz and 14 kHz. Measurements in the median plane included 54°, -18°. 0°, 18° and 36° Φ. Subjects were trained for 10 hours in the free-field prior to testing. They were asked to indicate the perceived source position of either the loudspeaker or virtual sound source by absolute judgement and calling out the apparent azimuth and elevation angle to the experimenters using a standard spherical coordinate system. For example, they would call out "0, 0", if the sound was perceived to be directly in front of them or "0, 90" if the sound appeared to be directly above. A large limitation of this response method was individual differences in position estimation, with subjects having no visual association to the estimated angle which was called out. Results for the experiment displayed a substantial individual difference in localisation, but for all subjects, identification of elevation was poorest at high elevations in the rear. Additionally, there was a significant increase in front-back confusion in headphone listening.

In fact, the percentage of front-back reversals were almost double for virtual sources than for real sources (11% and 6%, respectively).

Two localisation tests carried out by Bronkhorst (1995) compared virtual listening with that of real sources, one with long stimuli, allowing head movements to discern the location of the sounds and the other with short noise bursts to eliminate the use of head movements. Virtual sources were created using both individualised and generic HRIRs recorded in an anechoic room. Stimuli used in the experiment were harmonic signals with a fundamental frequency of 250 Hz and upper frequencies ranging from 4 kHz – 15 kHz with a 2ms rise and decay time. Results for the experiment indicate that virtual sources with individualised HRIRs can be localised almost as accurately as real sources if head movements are allowed and the source is continuous. However, in general, vertical localisation is somewhat poorer for virtual sources, confusion patterns for virtual and real sources were completely different by a factor of about 2. When comparing localisation of individual HRIRs to generic HRIRs, confusion rates of inexperienced listeners were more than double (30% and 64%, respectively). Additionally, results show that stimuli with frequencies below 7 kHz are similar between real and virtual sources, but when frequencies up to 15 kHz are added virtual source localisation becomes poorer than that of the real counterparts. Authors hypothesize that poor localisation of virtual sources may be down to incorrect simulation of high frequency spectral cues created during HRIR measurements.

Rychtáriková et al. (2009) carried out localisation experiments in which real loudspeaker localisation was compared to that of virtual localisation in both anechoic and reverberant conditions. HRIRs were recorded using a CORTEX® MK2 artificial head in an anechoic room and reverberant room, with the same loudspeaker setup. Virtual sources were played to subjects in both anechoic and reverberant conditions. A reverberant room was additionally virtualised using the ODEON® software using a model of the reverberant room containing the loudspeakers to create virtual reverberant stimuli. Localisation experiments of the real loudspeaker were completed in a reverberant and anechoic room and loudspeakers were positioned in 15° increments in the frontal horizontal plane. A total of five localisation tests were completed, 1) anechoic, real loudspeaker, 2) virtual anechoic, 3) reverberant, real loudspeaker, 4) virtual reverberant and 5) virtual simulated reverberant. Three stimuli were included in the testing; high frequency sound (3150 Hz), middle frequency sound in the 500 Hz band and the sound of ringing telephone. During stimulus playback subjects were asked to select which speaker, out of the 13 possible positions, the sound was emanating from, for both the loudspeaker and both VAD conditions. The experiment concluded that there was no significant difference between any of the listening scenarios, however less errors were noted in the real loudspeaker tests than with the VAD conditions, and the simulation based stimuli resulted in the most errors. Additionally, errors increased in the presence of reverberation and with increasing loudspeaker distance, though as ILDs and ITDs are more apparent in the horizontal plane, overall localisation was quite accurate.

Another experiment comparing real loudspeaker localisation with that of virtualised sources using non-individualised HRIRs was by Pedersen & Jorgensen (2005). The authors tested 9 different elevation angles in the median plane, as well as various other azimuth angles. Loudspeaker localisation and BRIR recordings were carried out in a large, reverberant television studio room. They tested white noise in 250ms bursts and 2s long continuous bursts, with the latter allowing head movements, making use of a head tracker. Stimuli was played to subjects at 75 dB SPL (A). The response method used in this experiment was a toy gun with a tracking device. Subjects were stood up on a small platform surrounded by an acoustically transparent curtain and were able to move their entire bodies in the direction of the apparent source direction. They were then asked to point the gun at the perceived location. Intuitively, localisation was better for longer stimuli presented in the free field than short or virtually presented stimuli. Uncertainty in terms of elevation doubled in virtual presentation of the sound source compared to the real source, with the errors in judgement being 12° for real and 24° for virtual sources. In the case of front-back confusions, the percentage was more than double for shorter real sources than their longer counterparts (4.2% and 9.1%, respectively), and quadrupled in the case of virtual sources (5.1% for 2s stimuli and 21.3% for 250ms stimuli).

3.2. Summary

This chapter has described literature relating to sound source localisation in various settlings. A great deal of the studies compare real loudspeaker localisation to that of its virtual counterpart. However, with the exception of Rychtáriková et al. (2009), who used non-individualised HRIRs in a simulated room, no other studies have directly compared anechoic against reverberant localisation to understand the importance of reflections and reverberations on median plane localisation. Chapter 6 investigates this further.

4. HRIR Measurements

It has been well established in the literature that in order to accurately localise sound sources in the median plane, certain spectral cues, predominantly caused by an individual's pinnae, head and torso, are of great importance (Algazi et al., 2002b; Avendano et al., 1999; Roffler & Butler, 1968a). The experiments outlined in this work seek to explore the effects of the pinna in median plain localisation. To thoroughly examine these effects, it is necessary to obtain individualised HRTF measurements in order to identify subjective attributes that contribute to localisation. This section outlines the measurement procedure undertaken to obtain these measurements.

4.1. Participants

Three normal hearing participants, aged 25 – 40, took part in the experiment. These consisted of postgraduate research students and a faculty member at the University. None reported any hearing damage. All three participants had extensive experience in listening tests, including localisation tests, but only one had experience in localisation in VADs. These will be referred to as follows, HL (participant with the most experience in this type of test), DJ and MM (participants with little experience in this type of test but with a great deal of listening test experience).

4.2. Ear Impressions

Prior to acquiring the HRTF measurements, individualised ear impressions were made of each test participant's left and right ears. In order to acquire these measurements, participants were first required to undergo an otoscopy by a hearing professional to ensure that no obtrusions such as cerumen were blocking the pathway to the tympanic membrane. Impressions were then created by a professional audiologist. These were made out of polyvinyl siloxane, a viscous liquid that hardens once its two components are mixed. The substance was placed in the ear and extended to 1cm inside the ear canal from beyond the limits of the concha. Once set, an incision was carefully made to remove the silicone beyond the beginning of the concha, so that only the start of the ear canal and meatus section was intact. A 3mm hole was drilled through the ear canal part of the impressions for the microphones to be mounted. Care was taken to ensure that the holes for the left and right ear impressions were situated in the same place.

Figure 3. Ear impressions inserted into the ear of subject DJ before microphone insertion.

4.3. HRTF Acquisition

Measurements were made at the University of Huddersfield in the Applied Psychoacoustics Laboratory's Critical Listening room (6.2m x 5.6m x 3.8m; T30 = 0.25s), an ITU-R BS.1116-2 compliant room. Exponential sine sweeps were presented through Genelec 8040A loudspeakers. Recordings were made using Knowles FG 23329-D65 miniature electret microphones running through a Merging Technologies Horus audio interface at 96kHz sampling frequency. An Apple Mac computer controlling HAART (Huddersfield Acoustics and Audio Research Toolbox), an impulse response (IR) acquisition toolbox for Max MSP (Johnson et al., 2015) recorded the resulting signal. Sweep durations lasted 10 s (seconds) emitting a frequency range of 48 Hz – 20 kHz at 70 dB SPL.

4.4. Speaker Positions

Loudspeaker used for measurements were carefully aligned in space to ensure that they were at the correct elevation angle with regards to the centre of the listener's head. Due to the limitations of the room, it was not possible to place them in an equidistant sphere around the subject's head. Considerations were taken into level alignment, however this was not done as it was thought to have little effect on externalisation and elevation perception. Individual levels of each speaker, from the head and exact distances of each speaker can be seen in figure 3.

Due to the relatively dampened characteristics of the measurement room, the difference in distances from the loudspeaker to the subjects' heads was negligible. A study by Bronkhorst & Houtgast, (1999) investigated auditory distance perception in rooms. The authors found that after up to 8 dB level increase in sound source, the perceived distance shifts less the 0.5m in a dampened room (T60 = 0.1s).

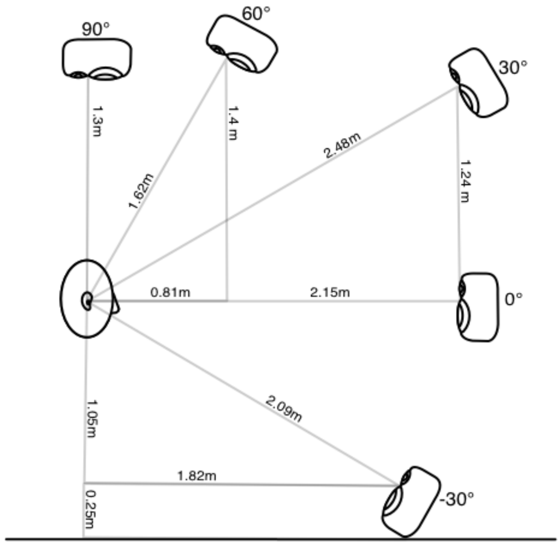


Figure 4. Distances of speaker positions in relation to the position of the participant's head

4.5. Head Alignment

Each participant's head was aligned using two self-levelling XY laser levels. This was done so that the XY crossover of laser A went through the centre of the 0° speaker position when the participant was absent, and when the participant was placed in position the X line of laser A went through the centre of both ears and the Y line went straight down the centre of the participant's head. The height of laser B was the same as that of laser A (1.3m) and was positioned 90°, perpendicularly from laser A. The XY crossover point of laser B was in the participant's ear canal, with the X line crossing the tip of the nose.

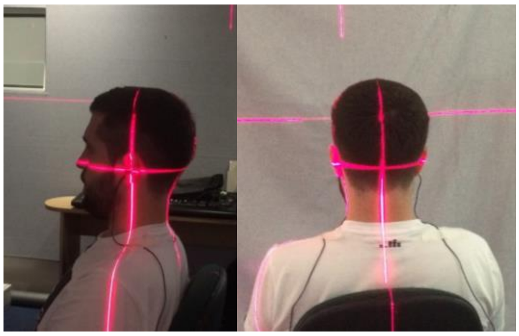


Figure 5. Head alignment with XY lasers for BRIR measurements of subject DJ

Participants were seated and asked to strictly keep their head very still and to hold their gaze on a black cross marked on an acoustic curtain directly in front of them. This was to avoid interaural differences that could be captured by head movements. The inside of a builders' hat was placed around participant's heads and attached to a stand to further reduce movement and the author monitored any additional movements that may have occurred. To reduce strong floor reflections, foam diffusers were place on the floor around the participant at the time of measurement. In order to avoid bias from participant seeing the speaker positions, acoustic curtains blocked the view of all of the loudspeakers.

Initially, frontal directions were measured (i.e. -30˚, 0˚, 30˚, 60˚ and 90˚), once these were acquired, listeners were turned around by 180˚ and the same measurements were made for the rear (i.e. 120˚, 150˚, 180˚ and 210˚). Measurements were repeated a minimum of two times and checked in Adobe Audition to ensure there were no interaural time differences. Negative elevations (-30˚ and 210˚) were excluded from all testing as some measurement errors were found after BRIR recordings had been made.

4.6. Headphone Measurements

Once the BRIRs were measured, Sennheiser HD650 open-back headphones were placed on subjects' heads, with the microphones still in place. Additional sine sweep measurements were then made for the headphone transfer functions (HpTFs). The sweeps emitted from the headphones were at 72dB SPL. Participants were asked to place the headphones on their heads so that a measurement could be made. They were then asked to remove the headphones and to place it on their head again for another measurement to be made. This process was repeated 10 times so that an average headphone response could be calculated. The average measurement represents the transfer functions of the headphones, the microphones and the path between the two pieces of equipment. Later, an inverse filter was created with the measurement, which removed the HpTF from the measured HRIRs. This ensured that the resulting signal played back to the listener via headphones did not contain any of the characteristics of the experimental apparatus and only that of the path from the loudspeaker to the ear canal entrances.

In order to create the headphone filter, Kirkeby regularisation was applied to the average of the headphone measurements. The filter was created in the same way as that of Kearney and Doyle (Kearney & Doyle, 2015) and creates a minimum phase filter which can be described as producing a 'one-sided' response, which compensates magnitude with the minimal group delay possible. An important parameter for the inverse filtering is the amount of regularisation that is applied; this parameter takes two arguments, first the 'range' which determines the dynamic range of which frequency components will be inverted.

Secondly the ‘dampening factor’ which dictates the amount of ‘dampening’ outside of said range. This is a very important parameter as it determines the resulting dynamic range of the filter and therefor ‘how much’ frequencies will be filtered.

5. Notch Filling Experiment

As outlined in the literature, the auditory system relies on various high frequency, pinna related peaks and notches in the spectrum to discern the apparent location of a sound source in space, especially with regards to sounds in the median plane (Asano et al., 1990; Batteau, 1967; Gardner, 1973; Geronazzo et al., 2010; Hebrank & Wright, 1974; Iida et al., 2007; Macpherson & Sabin, 2013; Roffler & Butler, 1968a, 1968b; Spagnol et al., 2013; Wright et al., 1974, 1974).

Some researchers have sought to identify exactly which pinna-related attributes contribute to accurate elevation localisation. A study by Iida et al. (2007) did just that. The experimenters created a ‘parametric’ HRTF based on notable peaks and notches in the spectrum. Their results concluded that the first peak and the first two high frequency notches were responsible for elevation perception, and recreating a HRTF with just these features can give almost the same localisation ability as the original measurement.

The experiment described in this chapter identifies and gradually removes pinna related notches in the HRTFs of three participants in the median plane. Seven loudspeaker positions in the anterior and posterior of the median plane were chosen for the experiment. These were located between 0˚ and 180˚, spaced in 30˚ increments. The identification criteria for a pinna notch was that it was above 5 kHz in the PRTF spectrum and that it’s depth exceeded -6 dB (Iida et al., 2007). Once notches. Once the notches were identified, they were reduced in magnitude to 50% and 100% of their original magnitude, with the latter being flat at 0 dB.

A similar study to the present experiment was conducted by Macpherson & Sabin (2013), who modelled the Level Effect (Vliegen & Van Opstal, 2004) by measuring HRTFs of five subjects in an anechoic chamber and filling in the high frequency notches in the HRTF spectra. In doing this, they defined floor and ceiling levels to which the notches would be flattened. Unlike the current study, they did not reduce notch magnitudes based on the magnitude of each individual notch (for example, in the current experiment if the notch's lower bounds are at -16 dB in original condition, it would be at -8 dB in the 50% notch filled condition, with the original spectral shape of the notch still intact). Instead, their experiment flattened HRTF notches to the specified floor levels, which were -5, 0, +5, +10 and +15 dB, with the higher levels often resulting in hemispheric reversals by participants. The increase in floor value was intended to model a reduction in stimulus level. Their experiment used 100ms random-phase, wideband (0.6 – 16 kHz) noise with a 20ms raised cosine onset and offset ramps presented at 55 dB SPL. The test locations compromised 26 near-midline locations in the front, 26 in the rear and 26 lateral locations, split between the left and right side of the listener. The 52 near-midline locations were spaced between -60 and +60 degrees in the vertical polar angle and -30 and +30 degrees laterally for frontal positions and between 120 and 240 degrees laterally for rear locations. Of the 52 frontal and rear positions tested, none included any median plane positions. Once participants were presented with the stimulus over headphones, they indicated the apparent location of the target by turning their body and tilting their head to orient their face in the perceived direction. Any responses that fell into the incorrect hemisphere (reversals) were excluded entirely from analysis. The experimenters found that the perceived location depended on the relative rather than absolute frequency shape of the spectrum.

The purpose of the experiment undergone in this report was to identify the effects of altering the magnitude of high frequency (>5 kHz) pinna related notches, as determined by Iida et al. (2007), while still maintaining their spectral shape and not merely flattening them at a floor level like in the experiment by Macpherson & Sabin (2013). This effectively reduced their spectral contrast before completely removing them. Listening tests were carried out to determine how localisation is affected when they are reduced and then removed, that is, whether the source position appears to move up or down, change hemispheres (e.g. front to back) or whether it is no longer possible to consistently discern the source position in their absence. This method of notch filling was carried out using both BRIRs, where 400ms long impulses were used, from the measurements described in section 4, and using pseudo-anechoic HRIRs, acquired by windowing the initial 5ms of the recorded BRIRs. The consequences of these spectral manipulations, with regards to median plane localisation will be explored and compared to previous research.

5.1. HRIR Processing

To identify the effects of the pinna related notches in HRTFs, a filtering method was finalised which automatically finds any notches in the spectra above 5 kHz whose depth exceeded -6dB, similar to identification criteria noted by Iida et al. (Iida et al., 2007). Identified notches were reduced by either 50% or 100% of their original depth, the latter being flat at 0dB from the notch's lower and upper boundary. Only the first millisecond of the BRIR was filtered in the frequency domain. Prior to filtering, a diffuse field filter was applied to the data for analysis purposes so that the 'true' HRTF spectra could be analysed for notch identification. The notch filling process is outlined in Figure 5 and will be described in detail in this section. The nature of the filtering used a similar technique to that of Courtois et al. (2016), whereby the magnitude of individual frequency bins are altered and the phase information of the original file is copied to the new, altered response. This process ensures that only the magnitude components are affected, and the original phase information is preserved. The filtering resulted in three versions of each measured direction in each of the BRIR and pseudo-HIRR conditions; original IR, pinna notches 50% filled and 100% filled, denoted as NF0, NF50 and NF100, respectively.

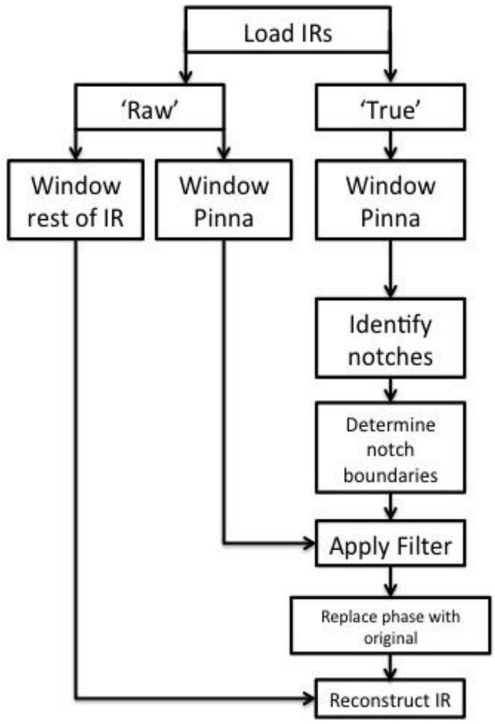


Figure 6. Processing steps involved in filtering pinna notches from the HRTF spectrum. Where the 'raw' part represents the original BRIR recording and the 'true' part represents the same recording with a diffuse filter created from measurements of the microphones in the listening room applied to it

Once notches had been filled, headphone equalisation filters were applied to each stimulus. These were obtained for each subject using the HpTFs acquired at the measurement stage (section 4.6).

5.1.1. Computation of Inverse Filters

This experiment required the computation of two different inverse filters. The first one was created for each individual subject and represented the inverse of the headphone and microphone responses for use in the stimulus presentation during listening tests. As noted in section 4.6, a total of 10 headphone measurements were taken for each subject. An average of the frequency spectrum was calculated for each ear. Then an inverse filter was computed, from 20 Hz to 20 kHz, with a maximum boost/cut of 10 dB and 1/3 octave smoothing. The inverse filter design used Kirkeby regularisation (Kirkeby et al., 1998). An example of the resulting filter can be seen in figure 6.

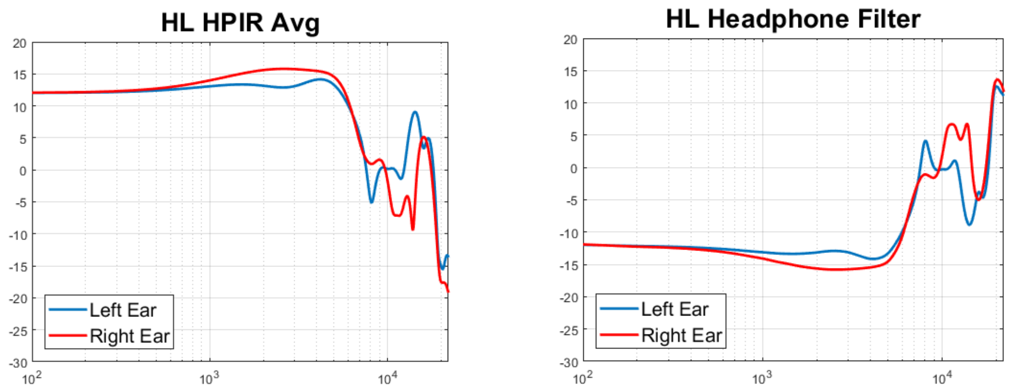


Figure 7. Example of subject HL’s filters used for headphone equalisation. The left hand figure represents the frequency response of the average of all of the headphone measurements for the left and right ear and the right figure is the resulting headphone filter created after Kirkeby regularisation is applied. The average measurement (left) represents the transfer function of the headphones, the microphones and the path between the two transducers at the time of measurement, and was removed from the resulting signal once the filter (right) was created and applied to the stimulus

Secondly, an inverse filter was created with the diffuse field response, this was created by taking the average spectrum for all measured points for each ear, with each microphone for each of the directions measured. Processing for this filter was conducted in the same way as that of the headphone equalisation filter. This filter was applied by means of convolution to each of the seven directions tested (0˚, 30˚, 60˚, 90˚, 120˚, 150˚, 180˚) in this experiment. This process obtained the ‘true’ pinna spectrum, used for analysis purposes to identify pinna notches, their upper and lower bounds, depth and frequency shape.

5.1.2.Identification and Filtering of Notches

Before any notch analysis was possible, the diffuse field filter, described above, was applied to each subject’s BRIRs. This removed the frequency response of the microphones, the loudspeaker and the acoustic properties of the room from the measurements. Once the filter was applied to all measurements, the first 1 ms of the resulting signals were windowed to isolate the pinna part only.

Once the pinna section of the ‘true’, diffuse filtered HRTF was obtained, notches fitting the pinna notch criteria were found. This was achieved by searching through all of the frequency bins in spectrum between 5 kHz and 20 kHz. If any bin values were less than -6 dB within the frequency range, their index and magnitude would be stored. The index was used to determine where to filter the ‘raw’ pinna measurement and the magnitude was used to determine how much filtering would be required to remove the notch.

In order to analyse the frequency components in the PRTF, fast Fourier transform (FFT) was applied to the pinna section of the signal. First the pinna was windowed in the time domain. A rectangular window was applied to the first 1ms starting from the highest peak in the time domain signal of the measured BRIR. The a 0.5ms half Hann window either side of the rectangular window was applied to the signal, and the three windowed components were concatenated. This effectively created a Tukey window, preserving the pinna portion of the HRIR, and tapering off the sections of the signal which were not of interest. The resulting signal was 192 samples in length. An FFT size of 4096 was chosen, which resulted in the pinna signal being heavily zero padded to increase the frequency resolution of the PRTF.

Following this, the notch boundaries were obtained by noting the index of the first and last frequency bin in each identified notch. These boundaries begin and end at the point in which the notch reaches -6 dB, so to find the point at which the notch reaches 0 dB, frequency bins above and below the notch boundaries were checked to find the closest bin to 0 dB. A final check to ensure notch boundaries did not overlap was then carried out. If boundaries did overlap, the consecutive notches were joined and treated as a single notch. The notch filling code can be seen in appendix 2. Start and end frequencies of each notch fitting the criteria where stored, along with the depth of each notch (appendix 1).

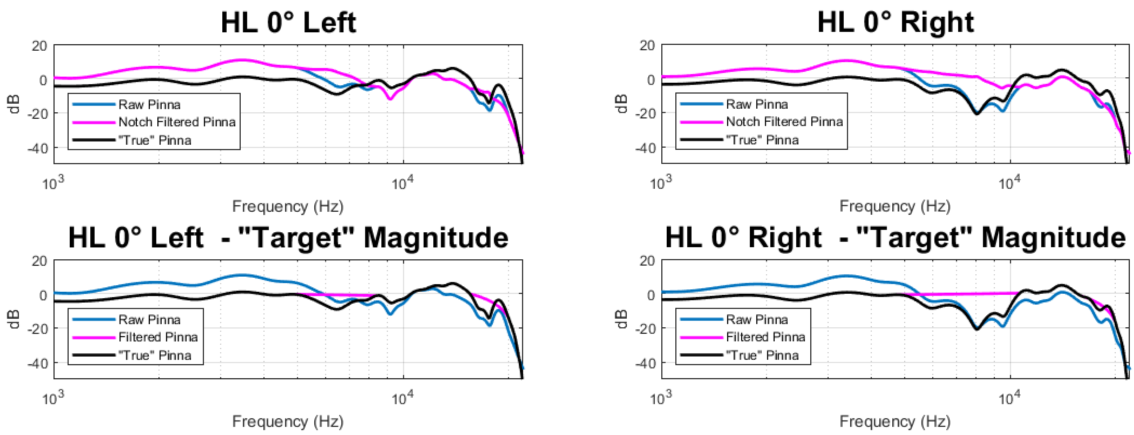


Figure 8. Notch filling in the most extreme condition (NF100) for subject HL at 0˚. The plots show frequency manipulations on both the left and right ears. For each ear, there are two graphs. The top plot represents the actual filtering applied to the raw pinna response (pink), against the 'true' pinna measurement (black) and the raw measurement (blue) that was filtered. The plot below represents the 'target' magnitude for each condition (pink). Effectively, how the true pinna is being filtered in each condition.

Once notch boundaries and the corresponding magnitude of each bin in the notch were obtained from the diffuse pinna measurement, filtering was applied to corresponding frequency domain bins in the first 1ms of the raw BRIR measurement. This was achieved by creating a 'target magnitude', the necessary magnitude required to create a flattened notch in the 'true' pinna, which was used to filter the raw measurement. Figure 7 shows the effective 'true' pinna filtering for subject HL at 0˚, and the corresponding effect the filtering had on the raw measurement. Once the spectral manipulation was complete, the original phase information from the raw pinna section was applied to the new filtered pinna (Courtois et al., 2016). Finally, in the time domain, the filtered pinna section then replaced the original pinna section from the raw measurement, so that all of the other anthropometric characteristics and early reflections were preserved and unaffected by filtering. Figure 8 shows the NF50 condition for subject HL at 0˚, where pinna notch compression can be observed, still keeping the spectral shape intact.

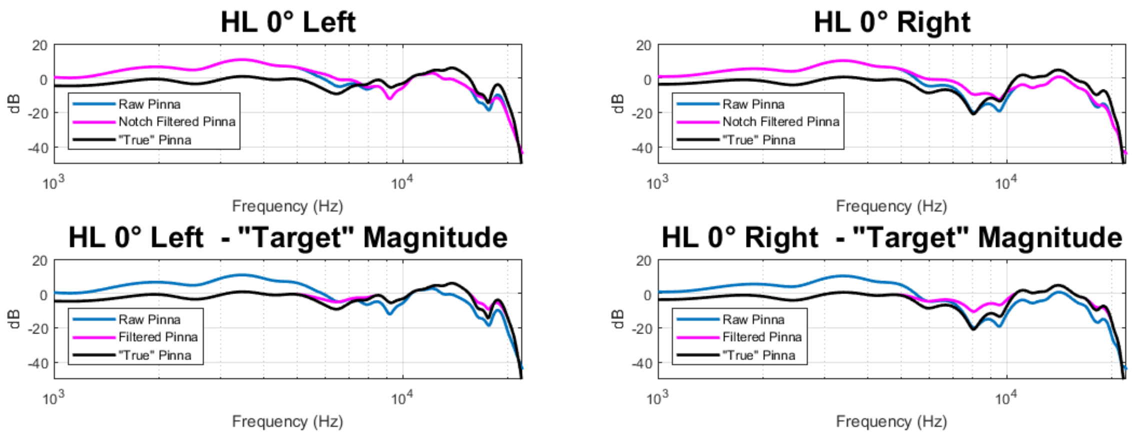


Figure 9. Notch filling in the least extreme condition (NF50) for subject HL at 0˚ to demonstrate the preservation of the spectral shape of the notches as their magnitudes are reduced. The plots show frequency manipulations on both the left and right ears. For each plot at each condition, there are two graphs. With the top representing the actual filtering applied to the raw pinna response (pink), against the 'true' pinna measurement (black) and the raw measurement (blue) before filtering. The plot below represents the 'target' magnitude for each condition (pink).

5.2. Experimental Procedure

Two listening tests were completed for this experiment. A pseudo-anechoic condition, which only used the first 5ms of the measured BRIR. And a 'reverberant' condition, using 400ms of the measured BRIRs, which provided much more externalisation, and therefore more resolution to discern the source's vertical location (Völk, 2009). All tests were conducted in the same listening room as that of the BRIR acquisition and acoustic curtains concealed all loudspeakers in the room during all listening tests to prevent any visual bias during subjects' responses.

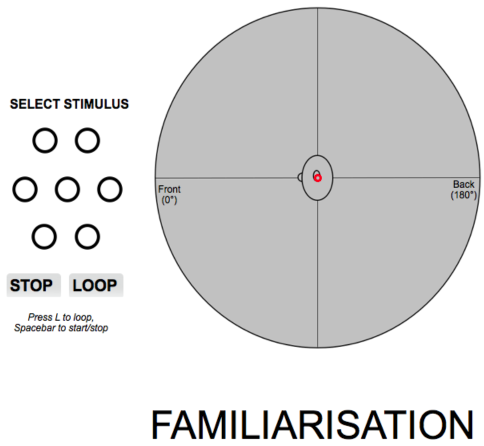


Figure 10. Diagram of familiarisation test used for the experiment. Responses were obtained after the subject placed the red circle on the outside of the circle, which had a resolution of 1˚. Each of the stimuli, which could be played by simply clicking on a circle, were randomised for each sitting of the test

A familiarisation test was carried out before each test session to get participants used to the test and the response method and could be seen as informal training to allow participants to discern where they perceive the recording to be located in space. In the familiarisation interface (figure 9) subjects could select any of the stimuli available, and compare them side by side. Only the original stimuli (NF0) were presented in this familiarisation test and were in a random order. The test did not indicate which direction each stimulus represented.

In the actual notch filling test, the response interface was exactly the same as that of the familiarisation test, however, only the circular interface was visible, along with a button to start and stop the audio. Subjects were only able to proceed onto the next stimulus once they had listened to the current stimulus and indicated their response.

Subjects were seated in a darkened room (the same room as the BRIRs were measured in) and instructed to align their gaze to a visual marker placed on the acoustic curtain before each stimulus was presented to ensure correct alignment of their heads. They were additionally instructed to close their eyes during stimulus playback to further avoid any potential visual bias. They were then asked to listen to the stimulus and place the marker around the outside of the circle in accordance with where they perceive the sound to be coming from. The judgement had to be made during stimulus playback to avoid any shift in response if a judgement was to be made some time after the stimulus had stopped playing. The test interface ensured a 3s gap between trials to avoid successive trials impacting each other. It has been noted in previous research by Pike et al. (2014) that a silent period exceeding 250 ms is required for accurate perception, otherwise consecutive stimuli may affect each other. This is known as the retention interval (RI) and was explored in terms of timbral discrimination accuracy, where an increased sensitivity to onset spectra were inversely affected by RI time.

5.2.1.Stimuli

200ms white noise bursts were used with a 1ms rise and fall time and a 400ms inter-stimulus silence interval. 200ms was chosen as it has been reported by Blauert that the effect of dynamic localisation cues occur after 200ms (Blauert, 1997, p. 95), and although the listening tests were conducted over headphones and with subjects' heads fixed in position, it eliminated the psychological expectation of source movement if subjects move their head slightly during testing. Broadband white noise was chosen as it contains all elements of the frequency spectrum and spectral differences are more apparent.

The noise bursts were convolved with each notch filled IR (both the IRs created in the reverberant condition and in the pseudo-anechoic condition), for each direction for all subjects. Prior to convolution of the noise stimulus and the IRs in all conditions, the headphone equalisation filter for each subject (described in section 5.1.1) was applied to their respective BRIRs and HRIRs, in all notch filling conditions. This step ensured that the response of the headphones and microphones were removed from the final stimuli. All stimuli were presented at 70 dB SPL in the headphone playback. This was the same level at which the BRIRs were recorded at.

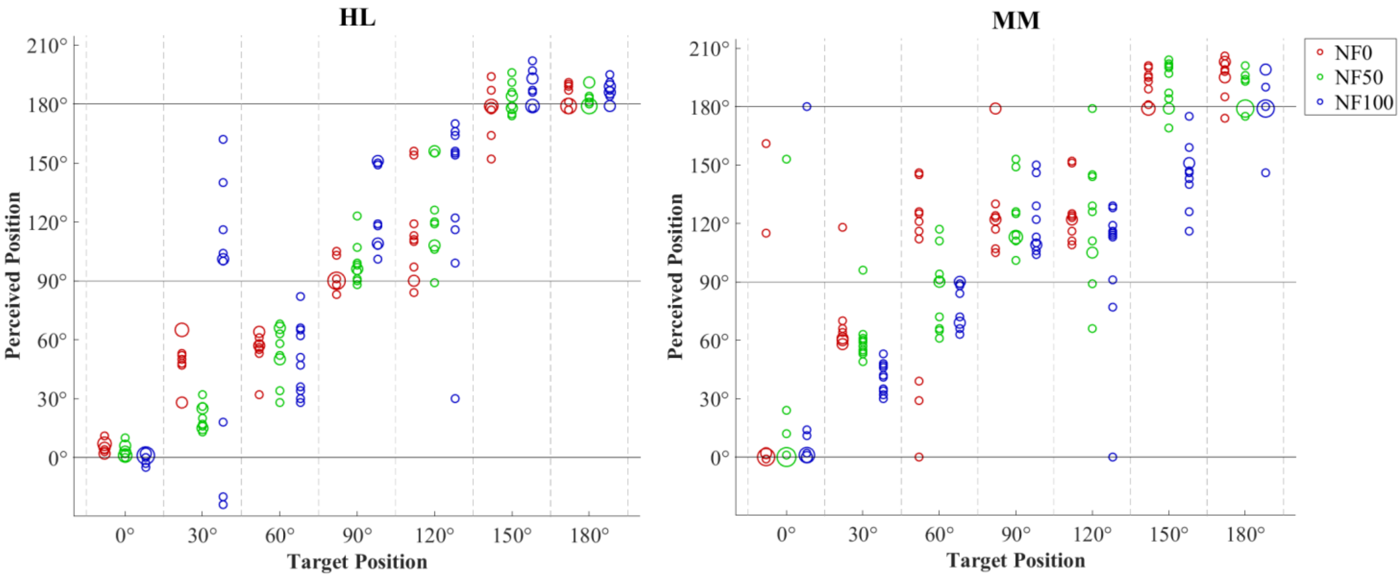
Six tests were completed for each condition (HRIR and BRIR), with each stimulus repeated 10 times, resulting in 35 stimuli per test. A total of 21 stimuli were tested for each IR length (seven directions with three notch filling conditions) and were randomised between all six tests. Randomisation was achieved in Max MSP, where the 21 different stimuli being tested were indexed and repeated 10 times in a list, which was then split across 6 tests in a random order. This meant that stimuli were completely random in each test, and could be repeated a number of times per test or possibly not presented at all in one or more of the tests. Test times varied for each subject, with subject HL spending the most time per test (between 4 and 7 minutes per test), subject DJ spending the least amount of time (between 2.5 minutes and 4 minutes per test) and subject MM spending between 3.5 and 5 minutes per test.

5.3. Results

Box and bubble plots were obtained before the median and standard deviation of each condition were calculated. A Shapiro-Wilk test confirmed that the distribution of the data was non-parametric. A one-sample Wilcoxon test was performed on each condition, testing participants' results against the hypothesized mean, which in this case was the 'target' or 'original' loudspeaker angle. After this, Friedman tests were performed on each group of EQ conditions for each angle to determine if notch manipulation has any effect on perceived direction. If results from the Friedman test displayed significant results, post hoc testing in the form of a Wilcoxon signed ranks pairwise comparison would be executed, showing which pairs of conditions result in a significant difference in source position.

5.3.1.BRIR Localisation

Results for this test varied a great deal between subjects, with subject experience being evident in terms of consistency and accuracy. HL (most experienced subject) proved to have the most accurate and consistent results. Before any effects of notch manipulation were assessed, the data was checked for multimodality. This was done in Matlab using Hartigan's Dip test of unimodality (Hartigan & Hartigan, 1985). If any data was found to be bimodal, and such bimodalities occurred on opposite hemispheres, this would indicate FBC (front-back confusion). Some of the data was found to be bimodal for all subjects, however these bimodalities occurred between angles in the same hemisphere, merely shifting up or down in space but not to the front or back.



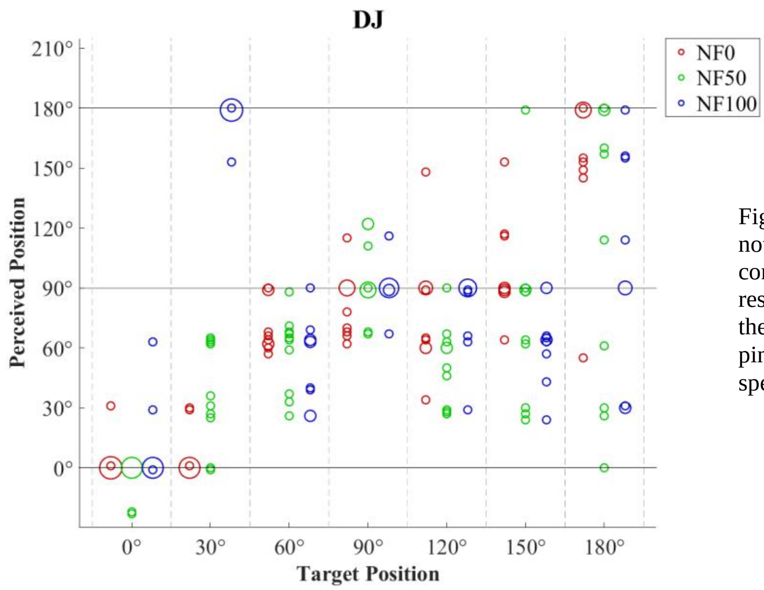


Figure 11. Bubble plots of localisation results for notch filling experiment in the reverberant room condition. Where larger bubbles represent more responses at that particular angle. NF0 represents the original HRTF, NF50 is the HRTF with the pinna related notch filled at 50% and NF100 is the spectral notch filled 100%, being flat at 0dB.

5.3.1.1. Localisation of Original BRIRs

Localisation of the original BRIRs were first examined to determine any shifts or confusion in localisation. In the case of HL, the most experienced listener, localisation was reasonably accurate and consistent, with some confusion between 30˚ and 60˚ targets as well as between 90˚ and 120˚ targets. The latter confusion is expected and is known to be a region where discrimination is low, the JND in localisation is up to three times larger than at 0˚ (Searle et al., 1975).

For subject MM, localisation of the original BRIR conditions were fairly consistent for most directions but they were not often localised at the target position, except for 0˚ and 120˚. Additionally, most responses for 150˚ and 180˚ were perceived to be behind and below the subject, with 60˚, 90˚ and 120˚ mostly appearing to originate from around the 120˚ region. Subject DJ also localised the raw conditions fairly consistently. Little elevation was perceived for 30˚, which was consistently localised at 0˚ and generally speaking, localisation accuracy and consistency declined somewhat in the rear in comparison to the front, with a great deal of hemispheric reversals perceived for rear targets. A more comprehensive analysis of the original conditions can be seen in the next experiment (chapter 6).

5.3.1.2. Individual Subject Analysis of Notch Manipulation

As previously mentioned, subject experience had a large impact on the consistency and accuracy of the listening test results. In the case of subject HL, effects of notch manipulation were evident, and it can be said that especially for elevated sources (from 30° - 150° Φ), eliminating pinna notches from the HRTF spectra, reduced consistency in localisation. At 0° however, notch manipulation improved localisation accuracy and consistency. Wilcoxon signed ranks pairwise comparison of the 0° results confirms that notch manipulation significantly affected localisation when the entire notch was removed (p = .006) in that it was much more consistently localised and shifted to the target position. However, this was the only case in which localisation improved after notch manipulation. At 30˚ and 90˚ in particular, notch filling significantly affected localisation after the entire notch was removed (p = .015, p = .015, respectively), where localisation became very inconsistent and mostly shifted to the back. Similar inconsistencies were observed at 60˚ and 150˚, however these changes in localisation were not significant.

Subject		0˚	30˚	60˚	90˚	120˚	150˚	180˚
HL	NF0	5°**	**^	**57˚**	**90˚**	**110.5˚**	178.5°**	**180˚**
	NF50	2.5°**	18.5°**	**55˚**	96°*	^	181.5°**	**180.5˚**
	NF100	**1˚**	101°*	**49˚**	**^^	154.5˚	186.5°**	186.5°*
MM	NF0	**0˚**	61°**	118.5°*	122.5°**	**122.5˚**	191°**	198.5°**
	NF50	**0˚**	58°**	90°**	113.5°**	**118.5˚**	192°**	**179˚**
	NF100	1°*	41.5°**	**^	111.5°**	**114.5˚**	**146.5˚**	**179˚**
DJ	NF0	**0˚**	0˚ **	*^	**84˚**	** ^^	89.5°**	**^^
	NF50	**0˚**	^	**64.5˚**	**89˚**	55°**	76°**	135.5°**
	NF100	**0˚**	179°**	**63˚**	**90˚**	89.5°**	64°**	**^^

Table 1 Median of results and whether the results significantly differ to target angle after carrying out Wilcoxon Signed-Ranks test: * p < 0.05, ** p < 0.01. Bimodality of data tested and significant bimodalities identified: ^ p < 0.05, ^^ p < 0.01. Bold font indicates statistical testing show that the results are considered to be accurate.

Subject MM's results differed somewhat in that for some directions notch manipulation reduced consistency and accuracy of localisation, particularly at 0˚ and 120˚. For elevated sources in the frontal median plane, as well as lower rear locations, notch manipulation improved consistency somewhat and drew the median responses closer to the target position. This was especially apparent for 150˚, where the median of the results drew closer to the target angle (146.5˚) and more elevation was perceived after notch removal. Notch manipulation at 60˚ moved the perceived position into the correct hemisphere, though when the entirety of the notch was filled (NF100), the results became significantly bimodal (p = .04), with confusion in the perceived angle being between 60˚ and 90˚. As with HL, there was some confusion in localising 90˚ and 120˚ and regardless of notch manipulation.

In contrast to HL, notch manipulation somewhat improved localisation in terms of accuracy and consistency at 90˚ for DJ. The subject, as mentioned in the previous section, consistently localised the 30˚ original condition at 0˚, however after notch removal (NF100), this shifted to 180˚. For the 0˚ direction, full notch filling did not affect localisation, however, filling 50% of the notch significantly affected the perceived location (p = .040) in that the position shifted down slightly as consistency decreased. At 60˚, localisation results for the original condition (NF0) were significantly bimodal (p = .034), with confusion between around 60˚ and 90˚, though after notch manipulation, results were a great deal more spread in the upper frontal region. At this angle results did not differ a great deal between the two notch manipulations (NF50 and NF100) and were somewhat similar to all three conditions at 120˚. Elevated rear targets for subject DJ were predominantly perceived in the frontal hemisphere, though the results were relatively spread. And although the 180˚ NF0 condition was statistically bimodal (p = .039), the perceived positions were either at 180˚ or around 150˚, however after notch manipulation, results spread a great deal, with a lot of hemispheric confusion and, in the NF100 condition, resulting in significantly multimodal results (p = .030).

5.3.1.3. Front-Back Reversals

Statistically significant bimodalities have been indicated in Table 1. Bimodalities indicate significant confusions between two angles, and as observed from the bubble plots, this is not necessarily two angles in different hemispheres in space. It has been well established in the literature that it is a common occurrence in median plane localisation for front-back reversals to take place (Abel et al., 1999; Wightman & Kistler, 1989b, 1999; Zhang & Hartmann, 2010). Overall percentage of reversals have been calculated and displayed in table 2.

Reversals (%)	*HL*	*DJ*	*MM*
NF0	1.7%^	20%^	16.7%*
NF50	1.7%^	35%^	13.3%
NF100	13.3%	43.3%	5%

Table 2. Percentage of reversals for each subject in each of the notch filling conditions tested in the reverberant room condition. * *only* frontal target positions reversed (front to back). ^ *only* rear targets reversed (back to front). 90˚ target was not included in these calculations.

Table 2 demonstrates that for subjects HL and DJ, filling in pinna notches results in an increase in hemispheric reversals. Both subjects only reverse rear targets before notch manipulation and in the least extreme notch filling condition. Then when the notches are entirely filled in, mean reversal rates dramatically increase and begin to occur for both front and rear targets.

The same cannot be said for subject MM, who experienced less reversals as notches were removed from the HRTF spectra. In the original condition (NF0) this subject only experienced reversals for frontal targets, however as notches were removed, although some reversals were perceived in both hemispheres, overall mean reversal rates declined.

5.3.2. Anechoic Localisation

Localisation in the anechoic condition was somewhat less accurate and consistent than that of the reverberant room condition, regardless of notch manipulation. Results from this test have undergone the same analysis as results for the reverberant room condition for notch filling. Bimodalities were tested and notch filling conditions were tested against each other to determine whether a significant change in localisation occurs after filling notches, and if this significant change is only apparent in a certain hemisphere i.e. front or back. Results from the testing can be seen in table 3.

5.3.2.1. Localisation of Original HRIRs

For subject HL, similarly to their results for the BRIR experiment some confusion occurred between results for the 30˚ target and 60˚, and much like the BRIR results, 150˚ and 180˚ was often localised below 180˚. Additionally, none of the 90˚ stimuli were correctly localised, with most responses located at around 180˚ with some confusion towards the lower frontal hemisphere. These results were somewhat similar to that of MM, who likewise, could not accurately localise the 90˚ target, and in the same manner would localise it at around 180˚ and perceived 150˚ and 180˚ above 180˚ (below the midline).

Subject DJ on the other hand, was fairly consistent with localisation judgement in the anechoic condition. Results for their 0˚ HRIR shifted to around -30˚ and 30˚ shifted to 0˚, and this was consistently localised at those positions. Results for 120˚ were similar to that of 60˚ and similarly to HL and MM, 150˚ and 180˚ were generally perceive to be around 180˚ and in the bottom hemisphere. As with the reverberant room condition of the notch filling test, a more comprehensive analysis of the original conditions can be observed in section 6.2.

Subject		0°	30°	60°	90°	120°	150°	180°
HL	NF0	**0.5°**	63°**	**62°**	177.5°*	^	187.5°**	**184.5°**
	NF50	**-3°**	67°*	**63°**	179°	**169°**	192°**	189.5°*
	NF100	^^	78°*	170.5°**	180°*	172°**	194°**	190.5°*
MM	NF0	0.5°*	122.5°**	^^	179°	166.5°	201°**	195°**
	NF50	**0°**	*^	1.5°	^^	^^	205°**	190.5°*
	NF100	*^	94°	33.5°	0.5°	1°**	203°**	190.5°*
DJ	NF0	-20°	0°	**^	**89°**	90°	179.5°	*^
	NF50	^	**50°**	90°**	46.5°	90°	**^	**179°**
	NF100	*^^	77.5°**	**^	**^^	78°	^^	**179°**

Table 3. Median of anechoic notch filling results and whether the results significantly differ to target angle after carrying out Wilcoxon Signed-Ranks test: * $p < 0.05$, ** $p < 0.01$. Bimodality of data tested and significant bimodalities identified: ^ $p < 0.05$, ^^ $p < 0.01$. Bold font indicates statistical testing show that the results are considered to be accurate.

5.3.2.2. Individual Subject Analysis of Notch Manipulation

In contrast to the effect of notch manipulation for HL at 0˚ in the reverberant conditions, in the anechoic condition, localisation suffered considerably after manipulation at this angle in that it resulted in significant front-back confusion (p = .00 for NF100 at 0˚). Similarly, at 30˚ and 90˚, although results were not close to the target position, consistency decreased after notch filtering. At the 60˚ target, results were fairly similar for NF0 and NF50, with many responses around the target, and some confusion towards the back. However, a significant difference in localisation was observed between the aforementioned conditions and NF100 (p = .029 and p = .041, respectively), where the median moved to around 170˚ and the source was consistently perceived at around that position. A similar effect was observed at 120˚, where a significant change in localisation was perceived between NF0 and NF100 (p =.029); here the source went from being significantly bimodal (p = .048) in the NF0 condition, in that front-back confusion was perceived, to the target being consistently localised at around 172˚.

For subject DJ, consistency did not change a great deal after notch manipulation, and bimodalities in the data were found for most angles, some in the original condition and some after notch filling. At 30˚ and 120˚, notch filling resulted in an increase in inconsistent results, with a great deal more hemispheric reversals. However, notch filling improved consistency at 90˚, albeit shifting the perceive location down, away from the target to around the 20˚ region. For this subject the only angle in which notch filling resulted in a significant change in localisation abilities was at 30˚ between the original condition and

NF100 (p = .012). At this angle, notch manipulation resulted in some front-back reversals and made responses very inconsistent.

MM exhibited no major change in localisation after notch filling. Results were relatively spread for all conditions and in one case (120˚), all that notch filling achieved was to shift the perceived location from 180˚ to around 0˚, though this was a statistically significant change (p = .012 for NF0 and NF100). In the same way as subject HL, all of the 150˚ and 180˚ stimuli, regardless of notch manipulation, were perceived to be above 180˚ (lower in space from the midline) and little elevation was perceived at all for the 90˚ target, regardless of notch filling. Additionally, for this subject, a great deal of reversals were observed for all angles in the frontal hemisphere and 120˚.

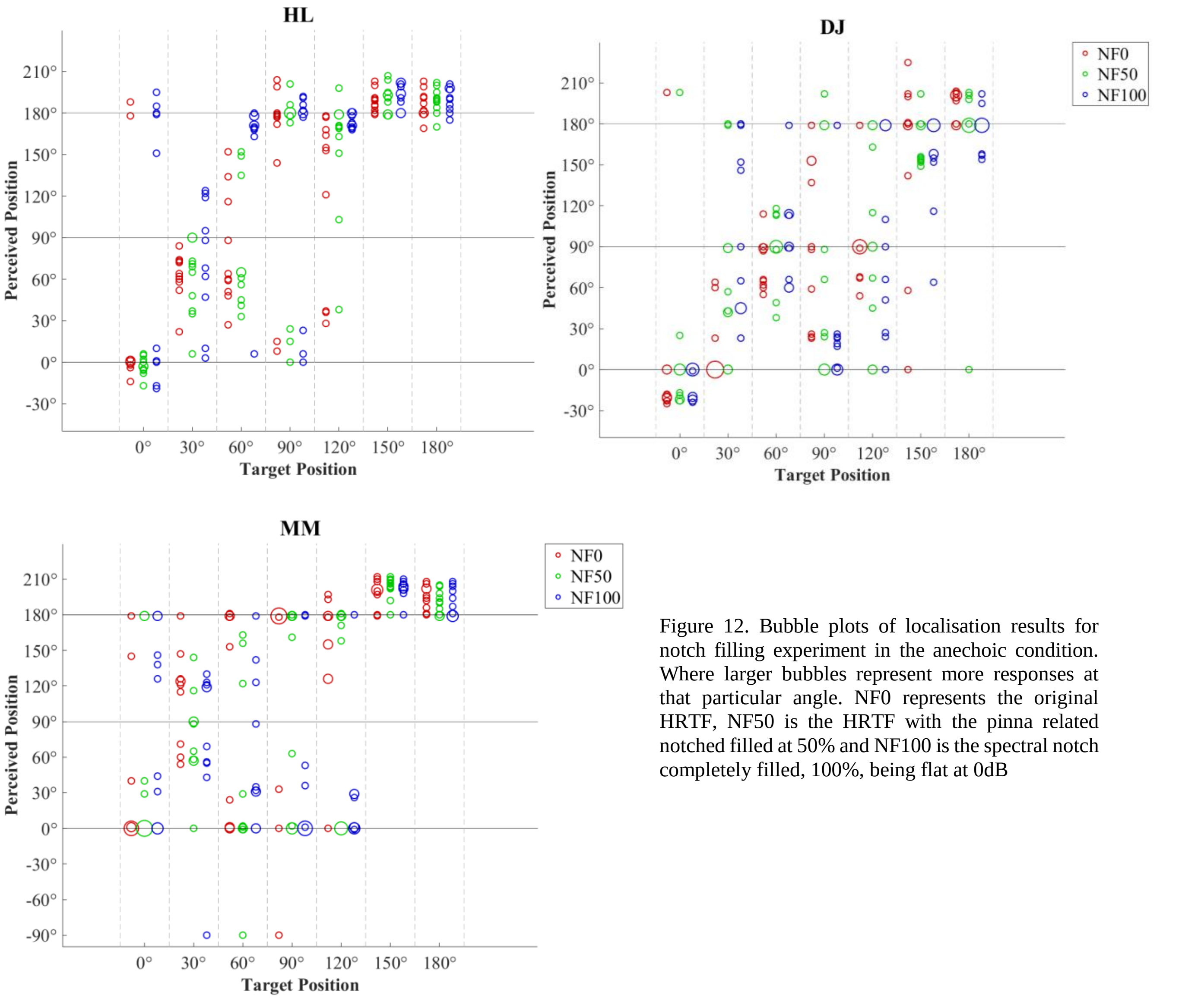


Figure 12. Bubble plots of localisation results for notch filling experiment in the anechoic condition. Where larger bubbles represent more responses at that particular angle. NF0 represents the original HRTF, NF50 is the HRTF with the pinna related notched filled at 50% and NF100 is the spectral notch completely filled, 100%, being flat at 0dB

5.3.2.3. Front-Back Reversals

Results for the anechoic VAD localisation test were checked for any bimodalities in that data, which are indicated in table 3. Results from this experiment also displayed much more front-back confusion than that of the BRIR test. Subject HL displayed significantly bimodal data in the form of front-back confusion for 0˚ NF100 (p = .00) and 120˚ NF0 (p = .048), and additionally 90˚ NF50 and NF100 and 120˚ NF0, though these results were not statistically significant (p = .092, .094, .089 respectively).

MM perceived significant front-back confusion for 0˚ NF100, 60˚ NF0, 90˚ NF50 and 120˚ NF50 (p = .006, .000, .000, .000, respectively). Subject DJ displayed significant bimodalities in the form of front-back confusion for 60˚ NF100 (p = .036) only. Other bimodalities in the results were found for this subject but they were not bimodal between different hemispheres. Other clear front back confusions, though not significant, included 30˚ NF100, 60˚ NF100, 90˚ NF0, 90˚ NF50, 120˚ NF50 and 120˚ NF100. It can be said, that except for 90˚, front-back confusions tended to occur after notch manipulation.

Reversals (%)	*HL*	*DJ*	*MM*
NF0	13.3%	13.3%	25%
NF50	6.7%	18.3%	20%
NF100	30%	23.3%	38.3%

Table 4. Percentage of reversals for each subject in each of the notch filling conditions tested in the anechoic condition. 90˚ target was not included in these calculations. For these conditions, reversals were perceived in both the anterior and posterior of the median plane

It can be observed in table , that in a similar manner to the BRIR testing of notch filling, the most extreme notch filling conditions result in the highest percentage of reversals for all subjects. Although for subject HL and subject MM, reversals were reduced from NF0 to NF50, with HL resulting in nearly half as many reversals in the NF50 condition than the original condition (NF0).

5.4. Discussion

In this experiment, two localisation tests were carried out in order to identify the effects of pinna related notches on median plane localisation. Experiments for notch filling were completed using both 400ms BRIRs containing all room reflections and reverberations and 5ms long, ‘pseudo-anechoic’ HRIRs. The results from the tests show that regardless of pinna notch manipulation, localisation is not very accurate, and hemispheric reversals are common. It has been reported by Blauert, (1997, p. 105) that observations of headphone localisation made by Von Békésy stated that hemispheric reversals may occur based on the expectations or state of mind of the listener. The human ability to focus on a particular attribute of a signal and suppress the rest allows the cognitive reversal of a perceived direction.

5.4.1.Reverberant Condition

Results from this experiment varied a great deal between subjects. For subject HL, removal of pinna notches in the HRTF spectrum reduced localisation accuracy and consistency for elevated sources, but not for the 0˚ source. For subject MM however, accuracy and consistency became somewhat better for elevated sources in both hemispheres, though not at 90˚. Subject DJ on the other hand, experienced no notable change in localisation after notch manipulation for most directions, except at 30˚, where removal of pinna notches resulted in complete hemispheric reversal. For this subject, all responses for 60˚ and 120˚ for all notch conditions were very similar, even in the original condition. Figure 12 compares the PRTF of the 60˚ and 120˚ angles to identify whether the subject localised these angles in similar positions due to the frequency spectra being similar. Analysis of the HRTF spectra of the 'true' pinna for this subject at these angles indicate major similarities in the high frequencies above 10 kHz in the NF100 condition (figure 12). However, the similarities are not as prominent in the NF0 and NF50 conditions, although these only represent the PRTF and none of the other reflections present in the full BRIR.

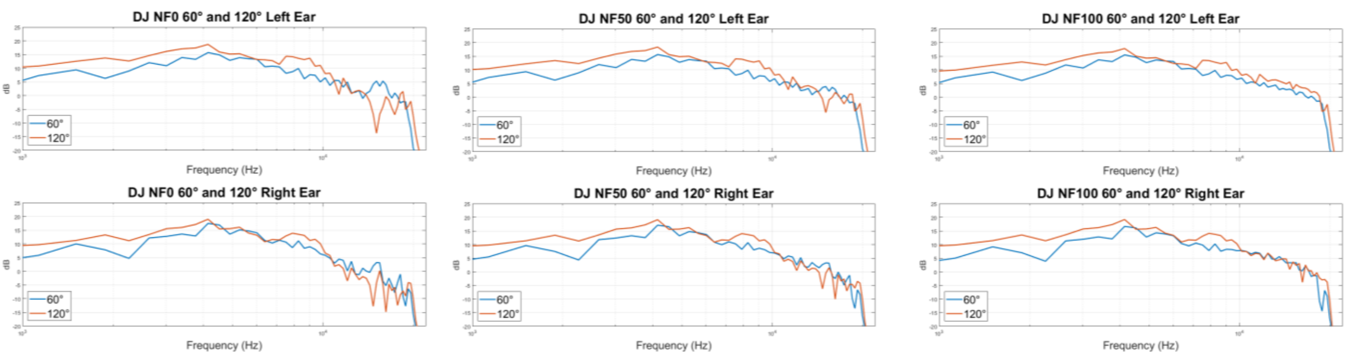


Figure 13. PRTF spectra of subject DJ, comparing the 60˚ and 120˚ angles for each ear for each step of the notch manipulation (NF0, NF50 and NF100)

5.4.2. Anechoic Condition

Results from this experiment were relatively inconsistent, and notch filling appeared to have little effect on the perceived location of the sound source. This could be down to the lack of externalisation due to the short length of the impulses used, which made it very difficult to adjust the elevation angle to the source, even with respect to the raw HRIRs (Völk, 2009).

Observations from the anechoic listening test has revealed that notch manipulation did not have a large effect on localisation in anechoic conditions in the median plane. For all subjects who took part in the test, no change in localisation occurred after any sort of notch filling in the posterior directions at 150˚ and 180˚. Additionally, very little change was observed at 0˚, 30˚ or 90˚. This could be due to the extreme lack of externalisation due to the use of such short impulse responses for testing. For example, all subjects failed in accurately localising 90˚, regardless of notch manipulation and experienced little elevation at all for this angle, often perceiving it to be at either 0˚ or 180˚.

5.4.3. Hemispheric Reversals

Front-to-back and back-to-front reversals were common in both the anechoic and the reverberant conditions of this test. For subject HL in both conditions (reverberant and anechoic), reversals predominantly occurred after notch manipulation. A possible explanation of the front-to-back reversals after notch filling is that after high frequency notches in the spectra are removed, especially if these notches' upper bounds exceed 10 - 12.5 kHz, this could initiate directional bands for behind (Blauert, 1997, pp. 108–111; Itoh et al., 2007). Hebrank & Wright (1974) established that behind cues have a small peak at 10 – 12 kHz, with a decrease in energy above and below it. This may also explain the high reversal percentages in the NF50 condition.

Similar reversals, and subjective variance in altering the HRTF spectra was observed by Langendijk & Bronkhorst (2002), who carried out localisation tests of Gaussian noise bursts (band-passed at 200 Hz – 16 kHz) and virtually presented them to subjects in the median plane after flattening the spectra to various degrees. Their experimental results indicated that in the one-octave band flattening between 6-16 kHz, one subject perceived all frontal targets in the rear, and another perceived most rear targets in front. The authors also suggest that the most important up-down cues are present in the 5.7 – 11.3 kHz region. In their experiment, of the eight subjects that were tested, two of them could not correctly localise sound in the baseline (original) condition for some positions. They explain that the most likely reason for this is that the listeners failed to learn how to map the available cues to the correct position.

5.4.4. Overall Effect of Notch Filling

When comparing localisation of the original BRIRs and HRIRs, one would assume that localisation is much more accurate and consistent with the original (NF0) condition, as it is a much more natural way of hearing and localising sound. However, this was not necessarily the case. It is clear from comparison of the NF0 in both BRIR and HRIR conditions, that localisation is generally better in reverberant conditions and that less hemispheric reversals and overall error is experienced with the presence of reverberations, if the test participants are familiar with the room used in the experiment. However, in the anechoic conditions of this pinna notch filling experiment, although errors exist in the original HRIR condition, notch filling does not seem to have an effect of localisation and comparable errors occur after notch manipulation.

It can be observed for subjects HL and MM that regardless of notch manipulation and room condition, both the 150˚ and the 180˚ targets were consistently localised below the midline (180˚+ ϕ), with the exception of MM at 150˚ NF100 in the reverberant condition. Further examination of the HRTF spectra in the anechoic condition for these directions demonstrate that both subjects do not display notable similarities in the high frequencies which may cause the different stimuli to be located at similar positions.

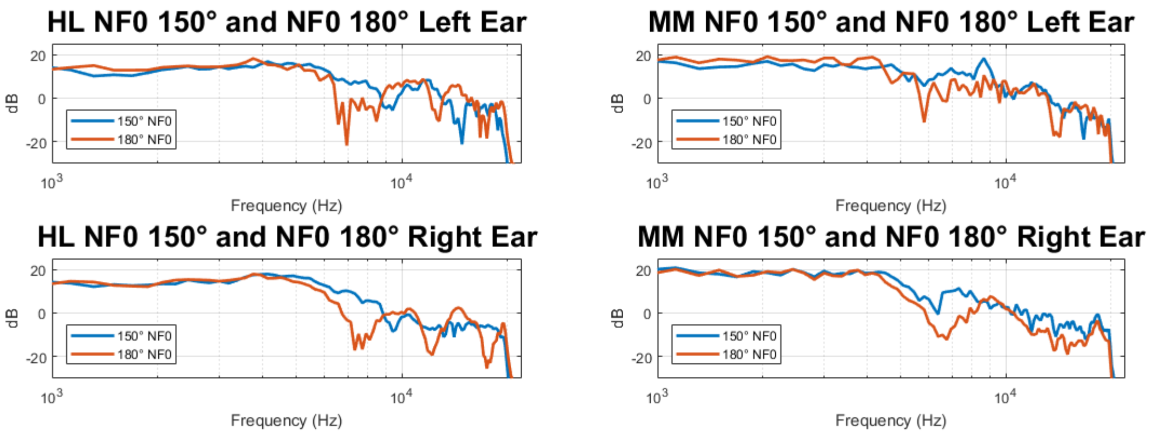


Figure 14. Frequency response of HL and MM, comparing 150˚ and 180˚ NF0 conditions to determine if localisation of the two stimuli were similar due to similarities in the response.

When examining the effects of pinna notch manipulation for the three listeners, there was no agreement among all subjects for any angle, both in the anechoic and room condition, on the effects of notch manipulation. For every angle tested in both conditions, at least two subjects would display the same trend and the third would show the opposite. Overall, this experiment has observed no major change in localisation due to notch manipulation, and generally speaking, effects due to manipulation appear relatively random when taking all subjects into account.

Results from the experiment by Macpherson & Sabin (2013) established that as notches were filled in, an increase in hemispheric reversals occurred, which was also the case in the current experiment, with the exception of subject MM in the reverberant condition, who overall experienced less reversals the more notches were manipulated. Their experiment also showed more back-to-front confusions than front-to back. This was not the case in the current experiment. In every condition, both anechoic and reverberant, there were always more front-to-back reversals for subjects HL and MM, with a great deal of the rear targets perceived to have little or negative elevation in the rear. This was not the case, however, with subject DJ, who experienced mostly back-to-front reversals in both room conditions.

5.4.5.Limitations

Perhaps the largest limitation in this experiment was the small sample size, of both repetitions in the tests and subjects tested. This was largely to do with the time constraints of the listeners and listeners available for testing. A great deal of time was required of each listener: they had to get an otoscopy, ear impressions made, and HRTFs measured before sitting a total of 12 tests for this experiment alone. And the otoscopy and ear impressions were a time sensitive part of the experiment as the audiologist had very limited time available to examine and measure participants. Additionally, the small sample size of the data per subject made the statistical analysis less reliable (only 10 data points per condition). Due to the very large subjective variance in the results, all subject's results were not combined for analysis.

Another limitation, related to the small subject size was also that all three subjects tested had considerable experience in the listening room, having already 'learnt' the room reverberations, which could be used as cues for localisation (B. Shinn-Cunningham, 2000), and could explain the better overall localisation judgements in the reverberant conditions.

6. Localisation Under Different Listening Conditions

It became evident in the previous experiment (Chapter 5) that binaural sound localisation can be poor regardless of spectral manipulations. The second experiment in this study investigates the difference in vertical localisation abilities of three participants in the median plane in virtual conditions compared to that of real loudspeakers. Localisation tests were carried out in a listening room using loudspeakers and over headphones using VADs in both listening room and pseudo-anechoic conditions. The listening room used was the same as that of the room used for BRIR measurements, described in section 4.3.

The purpose of the experiment was to determine how localisation differs in VAD anechoic conditions compared to VAD of a listening room in the median plane. An additional listening test of real loudspeakers in the listening room was carried out and used as a reference in which to compare subjects' natural listening capabilities to that of the VAD counterparts.

A similar experiment was previously completed by Rychtáriková et. al. (2009), however instead of using individualised BRIR recordings, they recorded a CORTEX MK2 manikin artificial head in anechoic and reverberant conditions in the horizontal plane. The authors tested the same three conditions tested in this experiment; real loudspeaker, BRIR and HRIR conditions. However, their experiment explored the horizontal, not the vertical, plane and instead of a continuous scale in which participants could indicate the perceived location of the sound, they were asked to indicate which loudspeaker, out of the 13 positions used, was the one that the sound emanated from.

Furthermore, Pedersen & Jorgensen's (2005) experiment tested localisation performance in the median plane, comparing real loudspeakers to that of nonindividualised (generic) BRIRs recorded in a reverberant room. Their experimental results showed that localisation of virtual sources were twice as bad as that of real sources. A similar experiment, though completed in anechoic conditions, was carried out by Wightman & Kistler (1989). Their results indicate that virtual sources cause an increase in elevation and FBC compared to that of real sources.

Al Saleh (2011, Chapter 5) completed localisation experiments in the horizontal plane in reverberant and anechoic listening rooms. The author found that a higher mean error in localisation existed in reverberant conditions compared with anechoic, however this result was not significant. Comparable experiments have been carried out in the literature, however, none have compared individualised HRIRs in VAD anechoic and VAD reverberant listening conditions in the median plane, in reference to natural listening abilities of real loudspeakers in the same listening room.

6.1. Experimental Procedure

The subjects were provided with a graphical user interface (GUI) created using Max7, where a side-view circle representing the median plane was shown. Each trial page of the GUI presented a sound stimulus and subjects were asked to indicate the perceived elevation angle on the circle during stimulus playback. The GUI implemented a 3 second gap between stimuli to ensure that consecutive stimuli had no effect on each other. The GUI design was the same for all three tests.

Results acquired from the NF0 condition in both the BRIR and pseudo-HRIR tests from the notch filling test (Chapter 5) were used in the binaural localisation comparison for this experiment. For the sake of this experiment they will be referred to as VAD R and VAD A, respectively.

To understand each participant's natural localisation abilities, the third listening test compared participants' binaural results to that with the real loudspeakers utilised for the measurements, this test will be referred to as 'RR'. For this test, all of the conditions outlined in the binaural experiment were kept the same as best as possible. The same seven loudspeaker positions were used, in the same room and at the same level that the HRTFs were recorded. A response interface very similar to the one used for the binaural experiment was used (figure 14), it only differed in that it included the use of a head tracker which allowed 1° of movement for either pitch, roll or yaw rotations before stopping the audio and asking the participant to realign their head. This ensured that participants did not move their head during testing.

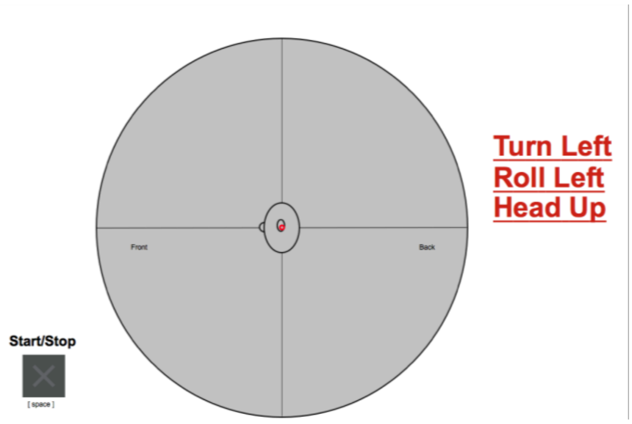


Figure 15. Listening test interface for the real room speaker localisation. This test made use of a head tracker, which would indicate on the right in red which way the subject should move their head in order to realign. During feedback the test would stop

6.1.1. Stimuli

Stimuli for the VAD section of this test have been described in chapter 5.2.1. The same stimuli were emitted from the loudspeakers located at the seven loudspeakers in the listening room. This used 200ms with 1ms cosine rise and fall time and a 400ms inter-stimulus silence interval.

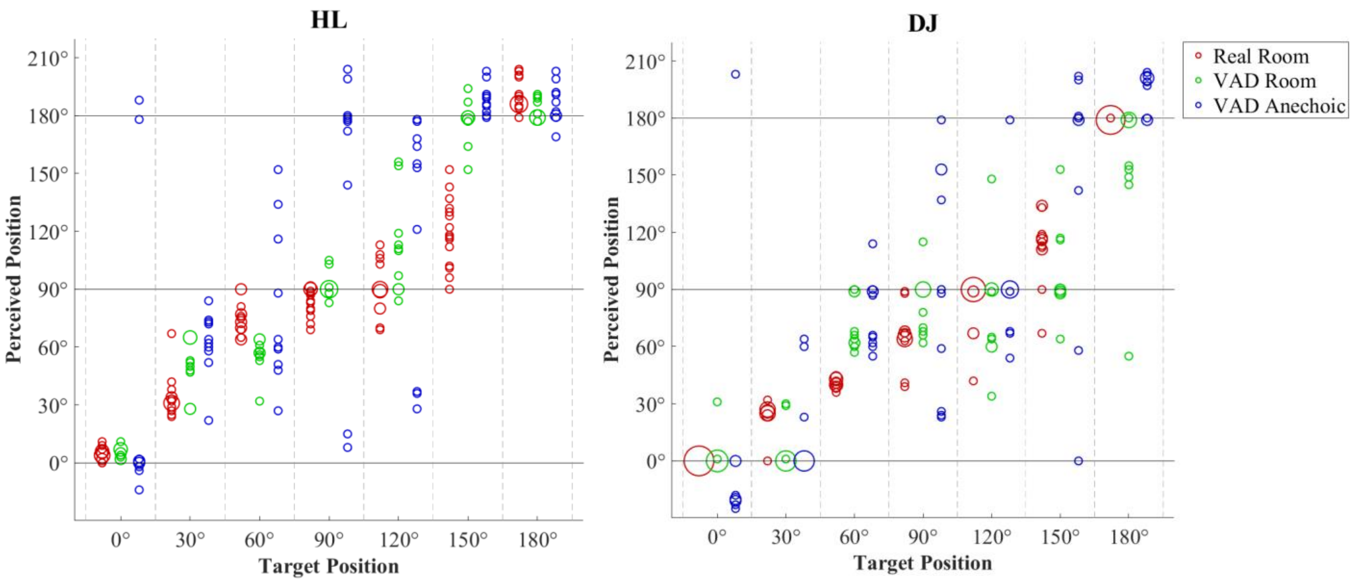
6.2. Results

This section presents the data retrieved from the listening tests, bubble plots of the data are presented in figure 15. A comparison of localisation accuracy and consistency was made between the three different conditions for each subject. This determines how localisation varies under the different conditions, which method provides the best localisation in terms of accuracy and consistency and finally how the source image deviates in different conditions. After completing a Shapiro-Wilk test on the data, it was decided to use non-parametric statistical tests as the data was not normally distributed.

Subject	Method	0°	30°	60°	90°	120°	150°	180°
HL	Real	5°**	**31°**	73°**	87°**	90°**	118°**	186°**
	VAD Room	5°**	**^	**57°**	**90°**	**110.5°**	178.5°**	**180°**
	VAD Anechoic	**0.5°**	63°**	**62°**	177.5°*	^	187.5°**	**184.5°**
MM	Real	**0°**	41°**	** ^	**90°**	90°**	141°**	179°**
	VAD Room	**0°**	61°**	118.5°*	122.5°**	**122.5°**	191°**	198.5°**
	VAD Anechoic	**0.5°**	122.5°**	^^	179°	**166.5°**	201°**	195°**
DJ	Real	**0°**	26°**	41°**	66°**	90°**	116°**	179°**
	VAD Room	**0°**	0° **	°*^	**84°**	** ^^	89.5°**	**^
	VAD Anechoic	**-20°**	**0°**	*^^	**89°**	90°	**179.5°**	*^

Table 5. Median of results and whether the results significantly differ to target angle after carrying out Wilcoxon Signed-Ranks test: * $p < 0.05$, ** $p < 0.01$. Bimodality of data tested and significant bimodalities identified: ^ $p < 0.05$, ^^ $p < 0.01$. Bold font indicates statistical testing show that the results are considered to be accurate. For the purpose of comparison, results from NF0 for both room conditions in the notch filling experiment are displayed in the VAD conditions

In order to determine if a significant change in localisation was present in the three conditions, a Friedman test was carried out to examine the main effect of the reproduction method on each subjects' test results for each angle. Following this, if any angles were found to contain significant differences in the data, a Wilcoxon pairwise comparison test with Bonferroni correction was carried out to determine if these playback conditions result in a significant change in localisation perception.



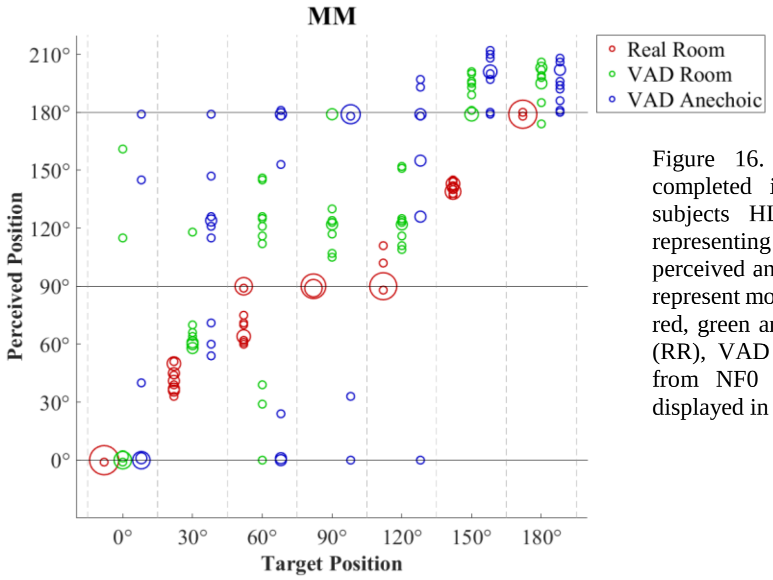


Figure 16. Bubble plots for localisation tests completed in all three listening conditions for subjects HL, DJ and MM. With the x axis representing the target position and the y axis the perceived angle by each subject. The larger bubbles represent more responses at that particular angle. The red, green and blue colours represent the real room (RR), VAD R and VAD A, respectively. Results from NF0 in the notch filling experiment are displayed in the VAD conditions

To determine whether significant FBC were experienced under any conditions, Hartigan's Dip test was carried out on all data. Bimodalities in the data do exist, however not between front and back but different locations within the same hemisphere. For example, the results for subject DJ at 120˚ in the VAD R condition displayed significantly bimodal results (p <0.01), but this bimodality was between 60˚ and 90˚. This is with the exception of participant MM, who experienced FBC at 60˚ under the VAD A condition (p = .00). The participant experienced no elevation under this condition and confused the source image between around 0˚ and 180˚

6.2.1.HL Results

Friedman tests carried out on the experimental results indicate a significant difference in localisation between the three conditions for sound sources elevated at 0˚, 30˚, 60˚, 90˚ and 150˚. Post-hoc Wilcoxon pairwise comparison tests revealed that all but 0˚ resulted in a significant change in localisation between the different test conditions. These significant changes were usually between the real and both VAD conditions. At 30˚ a significant difference in localisation observed between real and VAD room, and real and VAD anechoic, p = .014, p = .004, respectively, and similarly, at 150˚ where p = .002 and p = .002. When comparing localisation of the two VAD conditions, the only angle resulting in a significant difference in localisation ability was 150˚ (p = .016). Observations of the data has shown that for non-elevated sources (0˚ and 180˚) localisation does not differ significantly between the two VAD conditions ($p > 0.05$). Responses were fairly consistent at 0˚, with the source being perceived to be slightly elevated in both RR and VAD R conditions and shifting to below 0˚ in the VAD A condition. Similarly, for 180˚, RR and VAD A localisation were fairly similar, with the VAD R condition being the most consistent.

A much greater variance in responses existed in the results for elevated stimuli, which were statistically significant at 30˚, 60˚ and 150˚. In the 30˚ condition, localisation was accurate and consistent in the real room localisation, however when this angle was presented in the form of VADs, the source shifted upwards and in the case of the VAD R, some confusion in perception existed between 30˚ and 60˚, with the 30˚ angle resulting in a bimodal distribution (p = .0139). This change in localisation was significant between both RR and VAD R, and RR and VAD A (p = *.0411* and p = .0117, respectively).

At 60˚ a significant change in localisation was perceived between the RR and VAD R conditions (p =.006). For this angle the RR condition was perceived to be slightly more elevated than the target position, spread from 60˚ to around 80˚. VAD A results for this angle were very inconsistent and some confusion towards the back was perceived, though this was not significantly bimodal. Statistical results for 90˚ responses did not display any significant results, however the data suggests that localisation was far more consistent and accurate in the VAD R condition than the other two, with the RR condition being perceived in the upper frontal region and the VAD A condition in the 180˚ region. In the RR condition for this angle the source was predominantly perceived to be around 90˚, with some shifts toward the upper rear region. A great deal of confusion existed in the 150˚ RR condition, with all results spread from 150˚ and 90˚. Consistency was much higher in the VAD conditions for this angle, however, elevation was lost, with responses perceived at around 180˚ and shifting slightly downwards.

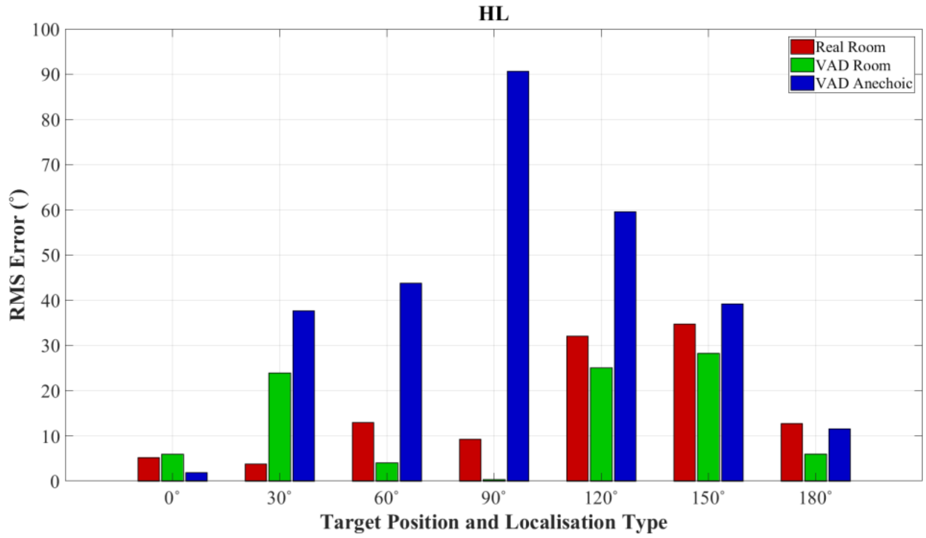


Figure 18. RMS localisation error for subject HL at each response method at each angle tested. The bars represent the RMS error in responses compared to the target position. These calculations removed any outliers by discarding any results which exceed 1.5 interquartile ranges above the upper quartile or below the lower quartile.

The mean localisation error of each condition, compared to each target angle has been calculated and displayed in figure 16. Results indicate the localisation in the VAD anechoic condition is worse than that of VAD room for all directions tested. Additionally, localisation in the VAD room condition results in less error than real room localisation for all elevated directions in the rear and 90˚. Localisation in the VAD A condition can be seen getting worse as elevation increases in both the front and rear hemisphere. The bubble plot for this subject (figure 15) supports this; as the source moves to higher elevations, some confusion becomes apparent (though not statistically significant) and the sense of elevation is lost somewhat.

6.2.2.MM Results

Results for this participant indicate a significant change in localisation for 30˚, 90˚, 120˚, 150˚ and 180˚. A Wilcoxon pairwise comparison of the different presentation methods revealed that all of these angles have at least one significantly different pair of results. For non-elevated sources, localisation did not differ considerably between the two VAD conditions. RR localisation at these angles were very consistent and accurate, and this did not differ at 0˚ in the VAD conditions, however at 180˚, the VAD conditions were both consistently localised below the target. At 30˚, similarly to the results displayed by participant HL, localisation was fairly consistent in the RR condition, however the perceived source

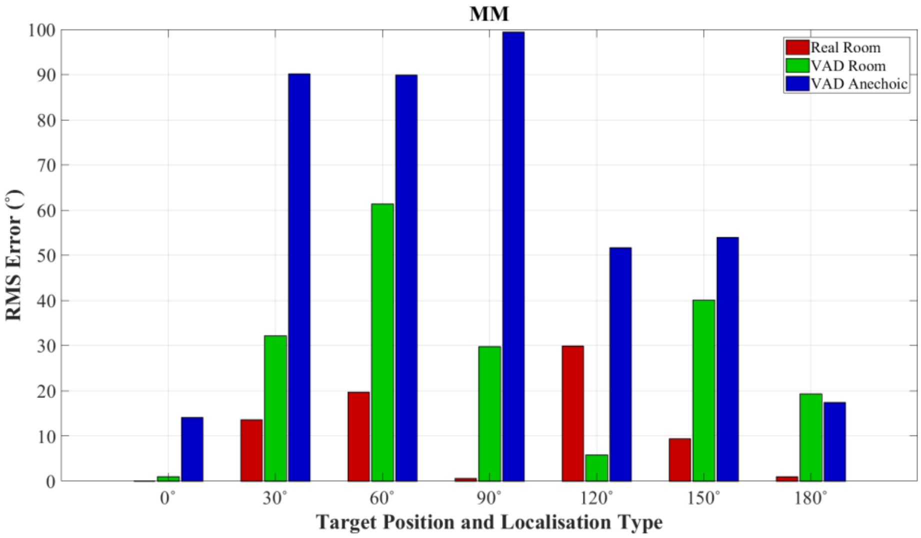


Figure 19. RMS localisation error for subject MM at each response method at each angle tested. The bars represent the RMS error in responses compared to the target position. These calculations removed any outliers by discarding any results which exceed 1.5 interquartile ranges above the upper quartile or below the lower quartile.

position moved upwards somewhat in the VAD R condition and some confusion occurred in the VAD A condition, though this was not significant. The 60˚ RR condition displayed some confusion in localisation between 60˚ and 90˚ (p = .000), and when presented to the participant in VAD format, this confusion remained (p = .000), with the VAD R predominantly moving to the upper rear section of the median plane and the VAD A condition displaying some front-back confusion along with no elevation. At 90˚ the RR localisation was very accurate and consistent, however this decreased in both VAD conditions, with the VAD R localised at around 120˚ and the VAD A moving to the non-elevated rear. Similar to results from participant HL, localisation became poorer in the rear hemisphere. At the 120˚ target position, the RR was perceived to emanate from around 90˚, however, VAD R was correctly localised around the target, additionally some confusion was perceived for this angle in the VAD A condition as to whether it came from 120˚ or 150˚. The 150˚ target was consistently localised slightly higher than the original target (around 160˚) in the RR condition, however the sense of elevation was completely lost in both VAD conditions, with most responses being below the midline.

RMS error plots in figure 17 makes it evident that a great deal more error in localisation was present in both VAD conditions compared to that of the real room condition. None of the conditions for 0˚ were significantly different to the target (p = 1., p = .125, p = .063 for RR, VAD R and VAD A, respectively), however more error was perceived for the VAD A condition.

6.2.3.DJ Results

A significant change in localisation, as determined by a Friedman test, revealed that 0˚, 60˚ and 180˚ was affected by playback method. This participant displayed very accurate and consistent results for both of the non-elevated sources in the RR conditions. This accuracy remained somewhat unchanged during the VAD R conditions, with some upward shifts in the rear direction. At 180˚, a significant change in localisation existed between the RR and VAD A condition and between the VAD R and VAD A condition (p = .047 and p = .002, respectively). The latter was the only direction in which localisation performance of VAD R compared to that of VAD A was statistically significantly different. The VAD A condition for these angles shifted downwards in space, though results were reasonably consistent.

For this participant, localisation of the RR at 30˚ was consistent, and close to the target, however unlike the other two participants, this listener perceived the VAD condition to have no elevation, for both VAD R and VAD A conditions at this angle. These were both consistently localised at 0˚, with some outliers.

At 60˚, the participant localised the RR condition slightly lower than the target, at around 40˚, although this was reasonably consistent. Localisation of the VAD R and VAD A results significantly differed to that of the RR condition (p = .006 and p = .006, respectively), though the two VAD conditions did not greatly differ from each other. They were both localised, fairly consistently, at the target position, with some confusion as to whether the sound was emanating from directly above at 90˚.

RR results at 90˚ were mostly localised at around 60˚, with some outliers below and above at 30˚ and 90˚. In the case of the VAD R localisation, some confusion occurred between 60˚ and 90˚. In the case of the VAD A localisation, a great deal of confusion can be observed, with the overall spread of data spanning almost the entire upper hemisphere.

Results for the rear directions are similar to that of participants HL and MM in that they are fairly inaccurate and inconsistent. The participant localised all conditions of the 120˚ direction at around 90˚, edging toward the upper frontal region at around 60˚. For 150˚, the RR condition was fairly consistently localised at 110˚-120˚, however this consistency completely disappeared in both VAD conditions.

As was the case with subjects HL and MM, the most error in localisation was at 90˚ in the VAD A condition. This subject differed however in localisation abilities with VAD R compared with VAD A in that for rear directions, VAD A resulted in less error than VAD R.

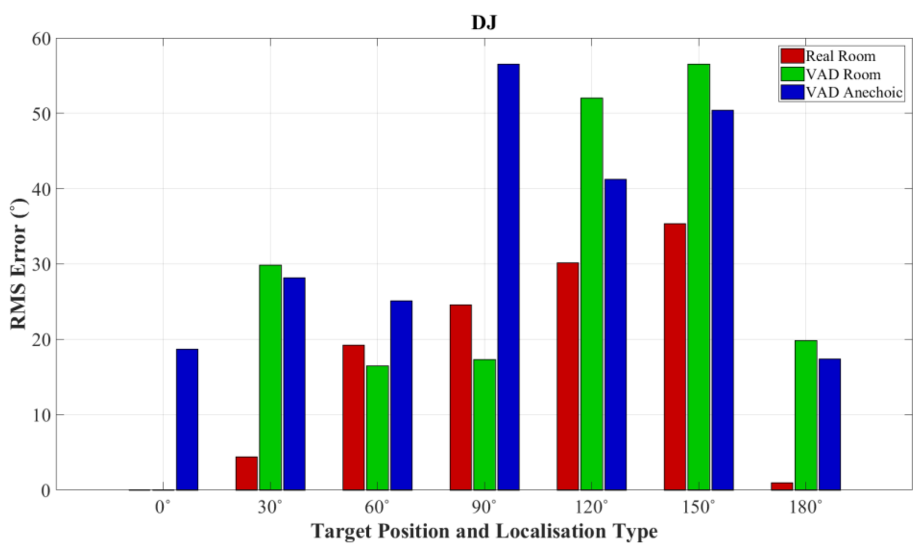


Figure 21. RMS localisation error for subject DJ at each response method at each angle tested. The bars represent the RMS error in responses compared to the target position. These calculations removed any outliers by discarding any results which exceed 1.5 interquartile ranges above the upper quartile or below the lower quartile.

6.2.4. Front-back Reversals

It is common in controlled localisation tests for front-back reversals to be perceived, even in real loudspeaker localisation tests (Mironovs & Lee, 2018). In these particular localisation tests, front-backreversals were found in almost every test condition for each subject. It is clear from table 6, which shows the overall percentage of reversals for each subject in each test condition, that the most reversals were perceived for the VAD A condition for all subjects.

Reversals (%)	*HL*	*DJ*	*MM*
VAD A	13.3%	11.7%	25%
VAD R	1.7%^	11.7%^	16.7%*
RR	4.5%^	4.5%^	0%

Table 6 Percentage of reversals for each subject in each of the three conditions tested. * *only* frontal target positions reversed (front to back). ^ *only* rear targets reversed (back to front). 90˚ target was not included in these calculations.

The extent and difference in reversals are not the same for each subject. HL resulted in confusion at 120˚ in all three test conditions, and only at this angle for VAD R and RR conditions. With VAD A resulting in 37.5% of overall reversals at 120˚. Similarly, subject DJ experienced most reversals for target positions in the rear, with only rear targets perceived in the front for VAD R and RR conditions. Participant MM, on the other hand, predominantly perceived reversals for frontal directions, often perceiving them to be located at the rear. Only 6.7% of reversals were in rear target positions (back to front). For this subject, reversals were only observed for virtual test conditions.

6.3. Discussion

When comparing localisation of the original VAD R, VAD A and RR, one would assume that localisation is much more accurate and consistent with real loudspeakers (RR), as it is the natural way of hearing and localising sound. However, this was not necessarily the case. Participant HL, for example, exhibited less error and localised more consistently in the VAD R conditions than in the RR test for all rear directions and at 90˚, though their results in the VAD A conditions were considerably less accurate and consistent than in both of the other tests. This observation for HL was not the case for subjects MM and DJ. Not much statistical significance was observed between localisation performance of VAD R and VAD A. Although it can be observed that subject MM’s VAD A localisation appears to be worse than the reverberant counterpart, and between 60˚ and 120˚, the sense of elevation was absent, with most responses being somewhat confused between 0˚ and 180˚. Alternatively, RR responses for these angles were mainly perceived to be at around 90˚, the angle at which no VAD stimuli were perceived. Overall, results from the VAD R condition are much more accurate, consistent and, with the exception of the rear directions for subject DJ, results in far less localisation errors than that of the VAD A condition.

Bad localisation performance in the rear was observed by Makous & Middlebrooks (1990), who completed localisation tests of short broadband (150ms) and continuous sounds in the free-field using both the horizontal and vertical planes. Their results showed that localisation is better in the front and attributed this to the sound being more attenuated from behind due to the shadowing effects of the head and pinnae, resulting in the loss of some high frequency components of the spectrum. Greater errors for behind locations were also found by Brungart et al. (1999) and Letowski & Letowski, (2011).

Although there did not appear to be a great deal of statistically significant bimodalities in the results data, causing FBC in particular, there was a large number of hemispheric reversals in the VAD conditions, this can be seen in table 5 in chapter 6.2.4. This result is supported by Pedersen & Jorgensen (2005), who found that FBC rates went from 9.1% for real sources to 21.3% for virtual sources when using 250ms wideband stimuli. Experiments by Letowski & Letowski (2011) also found that FBC approximately doubles in virtual conditions compared to real loudspeaker localisation.

A possible explanation for poor localisation ability in the VAD A condition as well as FBC in both VAD conditions could be the lack of head movements allowed in the test (Wallach, 1940; Wightman & Kistler, 1999). In localisation experiments comparing real and virtual sources in anechoic conditions, Bronkhorst (1995) found that virtual sources created with individualised HRIRs could be localised almost as accurately as real sources, only if head movements could be made and the sound was played continuously. The experimenter stated that localisation of short broadband sounds without head movements, as used in this experiment, is far less accurate for virtual than for real sources. Often in those cases sound was localised in the wrong quadrant, as was similarly the case in the current experiment. This finding is supported by Pedersen & Jorgensen (2005).

Localisation without the presence of reverberation can be known to suffer. In an experiment judging distance perception in anechoic and reverberant listening conditions, Shinn-Cunningham (2000) found that anechoic perception was near chance, however, the same author, in an experiment localising near-field sources in a reverberant room found that overall localisation in a reverberant room is worse than in anechoic conditions (Santarelli et al., 1999).

In an experiment by Rychtáriková, et al. (2009), anechoic VAD condition resulted in the least mean error, although there was no significant difference in localisation between virtual and anechoic conditions for any of the stimuli tested. Dissimilarly, results for this experiment showed the greatest error in localisation was from the anechoic VAD condition. Though the current experiment largely differs to theirs in that they limited directions to the frontal horizontal plane only, thus making use of ITD and ILD cues.

It is possible that poor virtual localisation could be down to incorrect simulation of high frequency spectral cues, created during BRIR measurement (Bronkhorst, 1995). Though, an experiment by Kulkarni & Colburn (1998) smoothed individualised HRTF spectra of subjects to determine how much spectral detail is required to be able to discriminate between 'real' and 'virtual' sound sources. Their results show that HRTFs can significantly be smoothed without affecting the perceived location of the sound.

It can be observed for all subjects that the VAD A condition is worse than VAD R for all directions. Results show a greater percentage of reversals and a higher mean error for VAD A compared to VAD R, with the exception of subject DJ, who displayed more error in the rear when presented with VAD R than with VAD A. A possible reason for better localisation, in general, in the VAD R condition is that the reflections present in the BRIR contribute to increase the perception of distance. This in turn provides a higher resolution in which the subject can discern the location of the sound source.

6.3.1.Limitations

A major consideration in this test is that all of the participants had extensive experience in the listening room, both in terms of general listening and in completing various localisation tasks since the room was built. Some studies have identified that this familiarity with a room can enable listeners to 'learn' certain localisation cues, that may not be present in other listening environments, and this learning process helps them to better localise sounds in the space (Shinn-Cunningham, 2000; Shinn-Cunningham, 2001; Shinn-Cunningham et al., 2005). This 'room learning' effect could attribute to why in general, VAD R resulted in less errors and more consistency that VAD A.

In addressing FBC, some studies in the literature resolve these issues in the data by coding the responses to the 'correct' hemisphere. A confusion in the data will be 'reflected' into the expected hemisphere before analysis. This procedure is known to overestimate the confusion rate and underestimate the error rate for target positions near 90˚ and could be misleading when analysing overall localisation judgements (Wightman & Kistler, 1989b).

A possible limitation to both experiments was the response method used. This did not include any visual markers around the circle representing the median plane. Only markers at 0˚, 90˚ and 180˚ to indicate various quadrants of the plane. This could result in a couple of degrees' variation in response for each subject. Additionally, as it is an absolute judgement test, position estimation for each subject may shift between each test in translation of perceived position to the position indicated on the interface.

7. Conclusion

The first experiment outlined in this work sought to identify the effects of pinna notch manipulation and removal. The biggest limitation in the testing was the small sample size, predominantly in terms of number of subjects tested. In nearly every condition, two of the three subjects would show similar results after notch manipulation, and the remaining subject would demonstrate opposing results.

In the reverberant condition, subject HL displayed a reduction in accuracy and consistency for all elevated sources after notch removal, except at 0˚ where it improved. On the other hand, subject MM displayed more accurate results after notch removal, except at 90˚ where no major change was observed. Subject DJ on the other hand, displayed no notable change in localisation at any angle except at 30˚, where a hemispheric reversal was perceived.

The results from the pseudo-anechoic tests however display no significant change in localisation after notch manipulation at 150˚ and 180˚, and show very little change at 0˚, 30˚ and 90˚. Additionally, none of the subjects could correctly localise the 90˚ target, regardless of notch manipulation.

Further investigation highlighted that removing notches can sometimes make the PRTF spectra of different angles very similar in the high frequency region, and perhaps mimic the features of different directions. Removing notches has also displayed an increase of hemispheric reversals, which in the same manner as the previous point, could initiate or remove certain directional bands for behind and in front. Overall, localisation was better in the 'reverberant' condition, this could be due to better externalisation in this condition and the possible effect of room learning as a consequence of the subjects selected for the experiment.

The second experiment outlined in this work took the results obtained from experiment one in the NF0 condition (measured BRIRs) and compared them to subjective localisation abilities of real loudspeakers, in the same room that the BRIRs were measured. The test compared VAD A, VAD R and RR results of the same white noise stimulus, presented at the same seven loudspeaker angles in the median plane.

Contrary to this authors expectation, although the RR condition was localised more consistently, it was not necessarily more accurate for all conditions. For example, subject HL displayed less error and more consistency in the VAD R condition than in the RR at 90˚ and rear directions.

Two of the subjects (MM and DJ) displayed no significant difference in localisation when comparing VAD A and VAD R. Although subject MM's results were worse in the VAD A condition compared to VAD R. Additionally, this subject did not perceive any of the test stimuli in either VAD condition at 90˚, however in the RR condition, most of the responses for angles 60˚ - 120˚ were perceived around 90˚.

7.1. Further Work

Following the studies undergone in this thesis, it would be beneficial to repeat the notch filling experiment with more subjects, at more median plane locations with a greater number of repetitions. This would help to gain a better understanding of pinna notch reduction and removal. The results are not consistent between subjects in these experiments and more data might identify effects that may not have been picked up in this study.

Following the second experiment (chapter 6), further investigation into the role of early and late reflections, as well as reverberation, in median plane localisation would be of interest. It is clear from the literature that stronger reflections (i.e. higher direct-to-reverb ratio (DR ratio)) provide a higher spatial resolution in which sound sources can be localised. It would be of interest to investigate whether there is an optimal IR length to provide enough reflections to accurately discern the spatial location of the sound. Additionally, whether too many reflections can be detrimental. Following this, the effect of altering the DR ratio on median plane localisation would be of interest to this study.

8. References

Abel, S. M., Giguère, C., Consoli, A., & Papsin, B. C. (1999). Front/Back Mirror Image Reversal Errors and Left/Right Asymmetry in Sound Localization. *Acta Acustica United with Acustica*, *85*(3), 378–386.

Ahuja, C., & Hegde, R. M. (2014). Fast modelling of pinna spectral notches from HRTFs using linear prediction residual cepstrum. *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* 4458–4462. https://doi.org/10.1109/ICASSP.2014.6854445

Al Saleh, H. (2011). *Effects of reverberation and amplification on sound localisation* (Phd, University of Southampton). Retrieved from https://eprints.soton.ac.uk/333290/

Algazi, V. Ralph, Duda, R. O., Duraiswami, R., Gumerov, N. A., & Tang, Z. (2002a). Approximating the head-related transfer function using simple geometric models of the head and torso. *The Journal of the Acoustical Society of America*, *112*(5), 2053–2064. https://doi.org/10.1121/1.1508780

Algazi, V. Ralph, Duda, R. O., Duraiswami, R., Gumerov, N. A., & Tang, Z. (2002b). Approximating the head-related transfer function using simple geometric models of the head and torso. *The Journal of the Acoustical Society of America*, *112*(5), 2053–2064. https://doi.org/10.1121/1.1508780

Algazi, V.R., Duda, R. O., Morrison, R. P., & Thompson, D. M. (2001). Structural composition and decomposition of HRTFs. *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on The*, 103–106. https://doi.org/10.1109/ASPAA.2001.969553

Asano, F., Suzuki, Y., & Sone, T. (1990). Role of spectral cues in median plane localization. *The Journal of the Acoustical Society of America*, *88*(1), 159–168. https://doi.org/10.1121/1.399963

Avendano, C., Algazi, V. R., & Duda, R. O. (1999). A head-and-torso model for low-frequency binaural elevation effects. *1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 179–182. https://doi.org/10.1109/ASPAA.1999.810879

Batteau, D. W. (1967). The Role of the Pinna in Human Localization. *Proceedings of the Royal Society of London B: Biological Sciences*, *168*(1011), 158–180. https://doi.org/10.1098/rspb.1967.0058

Blauert, J. (1997). *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT Press.

Bomhardt, R., & Fels, J. (2016, May 26). *Mismatch between Interaural Level Differences Derived from Human Heads and Spherical Models*. Presented at the Audio Engineering Society Convention 140. Retrieved from http://www.aes.org/e-lib/browse.cfm?elib=18221

Bomhardt, R., & Fels, J. (2017, May 11). *The Influence of Symmetrical Human Ears on the Front-Back Confusion*. Presented at the Audio Engineering Society Convention 142. Retrieved from http://www.aes.org/e-lib/browse.cfm?elib=18590

Bronkhorst, A. W. (1995). Localization of real and virtual sound sources. *The Journal of the Acoustical Society of America*, *98*(5), 2542–2553. https://doi.org/10.1121/1.413219

Bronkhorst, A. W., & Houtgast, T. (1999). Auditory distance perception in rooms. *Nature*, *397*(6719), 517. https://doi.org/10.1038/17374

Brungart, Douglas S., Durlach, N. I., & Rabinowitz, W. M. (1999). Auditory localization of nearby sources. II. Localization of a broadband source. *The Journal of the Acoustical Society of America*, *106*(4), 1956–1968. https://doi.org/10.1121/1.427943

Brungart, D.S., & Romigh, G. D. (2009). Spectral HRTF enhancement for improved vertical-polar auditory localization. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2009. WASPAA '09*, 305–308. https://doi.org/10.1109/ASPAA.2009.5346479

Burke, K. A., Letsos, A., & Butler, R. A. (1994). Asymmetric performances in binaural localization of sound in space. *Neuropsychologia*, *32*(11), 1409–1417. https://doi.org/10.1016/0028-3932(94)00074-3

Butler, R. A. (1994). Asymmetric performances in monaural localization of sound in space. *Neuropsychologia*, *32*(2), 221–229.

Butler, Robert A., & Belendiuk, K. (1977). Spectral cues utilized in the localization of sound in the median sagittal plane. *The Journal of the Acoustical Society of America*, *61*(5), 1264–1269. https://doi.org/10.1121/1.381427

Chun, C. J., Kim, H. K., Choi, S. H., Jang, S.-J., & Lee, S. (2011). Sound source elevation using spectral notch filtering and directional band boosting in stereo loudspeaker reproduction. *IEEE Transactions on Consumer Electronics*, *57*(4), 1915–1920. https://doi.org/10.1109/TCE.2011.6131171

Claes, P., Reijniers, J., Shriver, M. D., Snyders, J., Suetens, P., Nielandt, J., … Vandermeulen, D. (2015). An investigation of matching symmetry in the human pinnae with possible implications for 3D ear recognition and sound localization. *Journal of Anatomy*, *226*(1), 60–72. https://doi.org/10.1111/joa.12252

Courtois, G., Marmaroli, P., Rohr, L., Lissek, H., Oesch, Y., & Balande, W. (2016). Dynamic Range Limiting of HRTFs: Principle and Objective evaluation. *Journal of the Audio Engineering Society*, *64*(10), 731–739.

Crawford-Emery, R., & Lee, H. (2014, April 25). *The Subjective Effect of BRIR Length on Perceived Headphone Sound Externalization and Tonal Coloration*. Presented at the Audio Engineering Society Convention 136. Retrieved from http://www.aes.org/e-lib/browse.cfm?elib=17191

Gardner, M. B. (1973). Some monaural and binaural facets of median plane localization. *The Journal of the Acoustical Society of America*, *54*(6), 1489–1495.

Gardner, Mark B. (1973). Some monaural and binaural facets of median plane localization. *The Journal of the Acoustical Society of America*, *54*(6), 1489–1495. https://doi.org/10.1121/1.1914447

Gardner, Mark B., & Gardner, R. S. (1973). Problem of localization in the median plane: Effect of pinnae cavity occlusion. *The Journal of the Acoustical Society of America*, *53*(2), 400–408. https://doi.org/10.1121/1.1913336

Geronazzo, M., Spagnol, S., & Avanzini, F. (2010). *Estimation and modeling of pinna-related transfer functions*. 6–10.

Gilkey, R., & Anderson, T. R. (2014). *Binaural and Spatial Hearing in Real and Virtual Environments*. Psychology Press.

Hartigan, J. A., & Hartigan, P. M. (1985). The Dip Test of Unimodality. *The Annals of Statistics*, *13*(1), 70–84.

Hebrank, J., & Wright, D. (1974). Spectral cues used in the localization of sound sources on the median plane. *The Journal of the Acoustical Society of America*, *56*(6), 1829–1834. https://doi.org/10.1121/1.1903520

Iida, K., Itoh, M., Itagaki, A., & Morimoto, M. (2007). Median plane localization using a parametric model of the head-related transfer function based on spectral cues. *Applied Acoustics*, *68*(8), 835–850. https://doi.org/10.1016/j.apacoust.2006.07.016

Iida, K., & Oota, M. (2018). Median plane sound localization using early head-related impulse response. *Applied Acoustics*, *139*, 14–23. https://doi.org/10.1016/j.apacoust.2018.03.027

Itoh, M., Iida, K., & Morimoto, M. (2007). Individual differences in directional bands in median plane localization. *Applied Acoustics*, *68*(8), 909–915. https://doi.org/10.1016/j.apacoust.2006.08.001

Ivarsson, C., De Ribaupierre, Y., & De Ribaupierre, F. (1980). Functional ear asymmetry in vertical localization. *Hearing Research*, *3*(3), 241–247. https://doi.org/10.1016/0378-5955(80)90050-7

Jin, C., Corderoy, A., Carlile, S., & Schaik, A. van. (2004). Contrasting monaural and interaural spectral cues for human sound localization. *The Journal of the Acoustical Society of America*, *115*(6), 3124–3141. https://doi.org/10.1121/1.1736649

Johnson, D., Harker, A., & Lee, H. (2015, May). *HAART: A new impulse response toolbox for spatial audio research*. Presented at the 138th Annual Audio Engineering Society AES Convention, Warsaw, Poland. Retrieved from http://www.aes.org/e-lib/browse.cfm?elib=17625

Kearney, G., & Doyle, T. (2015). A Virtual Loudspeaker Database for Ambisonics Research. *ICSA 2015 045*.

Kirkeby, O., Nelson, P. A., Hamada, H., & Orduna-Bustamante, F. (1998). Fast deconvolution of multichannel systems using regularization. *IEEE Transactions on Speech and Audio Processing*, *6*(2), 189–194. https://doi.org/10.1109/89.661479

Kop, N., & Shinn-Cunningham, B. G. (2002). *Localization in a Reverberant Room: Effects of Nearby Walls and Room Learning*.

Kulkarni, A., & Colburn, H. S. (1998). Role of spectral detail in sound-source localization. *Nature*, *396*(6713), 747–749. https://doi.org/10.1038/25526

Langendijk, E. H. A., & Bronkhorst, A. W. (2002). Contribution of spectral cues to human sound localization. *The Journal of the Acoustical Society of America, 112*(4), 1583–1596. https://doi.org/10.1121/1.1501901

Letowski, T., & Letowski, S. (2011). Localization Error: Accuracy and Precision of Auditory Localization. *Advances in Sound Localization*. https://doi.org/10.5772/15652

Lord Rayleigh, J. W. S. (1907). XII. On our perception of sound direction. *Philosophical Magazine Series 6*, *13*(74), 214–232. https://doi.org/10.1080/14786440709463595

Macpherson, E. A., & Sabin, A. T. (2013). Vertical-plane sound localization with distorted spectral cues. *Hearing Research*, *306*, 76–92. https://doi.org/10.1016/j.heares.2013.09.007

Makous, J. C., & Middlebrooks, J. C. (1990). Two-dimensional sound localization by human listeners. *The Journal of the Acoustical Society of America*, *87*(5), 2188–2200. https://doi.org/10.1121/1.399186

Middlebrooks, J. C., Makous, J. C., & Green, D. M. (1989). Directional sensitivity of sound-pressure levels in the human ear canal. *The Journal of the Acoustical Society of America*, *86*(1), 89–108. https://doi.org/10.1121/1.398224

Mironovs, M., & Lee, H. (2018, May 14). *On the Accuracy and Consistency of Sound Localization at Various Azimuth and Elevation Angles*. Presented at the Audio

Engineering Society Convention 144. Retrieved from http://www.aes.org/e-lib/inst/browse.cfm?elib=19469

Møller, H., Sørensen, M. F., Jensen, C. B., & Hammershøi, D. (1996). Binaural Technique: Do We Need Individual Recordings? *Journal of the Audio Engineering Society*, *44*(6), 451–469.

Pedersen, J. A., & Jorgensen, T. (2005). *Localization Performance of Real and Virtual Sound Sources*. Retrieved from AM3D A/S AALBORG (DENMARK) website: https://apps.dtic.mil/docs/citations/ADA454835

Pike, C., Mason, R., & Brookes, T. (2014, April 25). *The Effect of Auditory Memory on the Perception of Timbre*. Presented at the Audio Engineering Society Convention 136. Retrieved from http://www.aes.org/e-lib/browse.cfm?elib=17175

Pratt, C. C. (1930). The spatial character of high and low tones. *Journal of Experimental Psychology*, *13*(3), 278–285. http://dx.doi.org.libaccess.hud.ac.uk/10.1037/h0072651

Raykar, V. C., Duraiswami, R., & Yegnanarayana, B. (2005). Extracting the frequencies of the pinna spectral notches in measured head related impulse responses. *The Journal of the Acoustical Society of America*, *118*(1), 364–374. https://doi.org/10.1121/1.1923368

Roffler, S. K., & Butler, R. A. (1968a). Factors That Influence the Localization of Sound in the Vertical Plane. *The Journal of the Acoustical Society of America*, *43*(6), 1255–1259. https://doi.org/10.1121/1.1910976

Roffler, S. K., & Butler, R. A. (1968b). Localization of Tonal Stimuli in the Vertical Plane. *The Journal of the Acoustical Society of America*, *43*(6), 1260–1266. https://doi.org/10.1121/1.1910977

Rychtáriková, M., Bogaert, T. V. den, Vermeir, G., & Wouters, J. (2009). Binaural Sound Source Localization in Real and Virtual Rooms. *Journal of the Audio Engineering Society*, *57*(4), 205–220.

Santarelli, S., Kopco, N., Shinn-Cunningham, B. G., & Brungart, D. (1999). Near-field localization in echoic rooms. *The Journal of the Acoustical Society of America*, *105*(2), 1024–1024. https://doi.org/10.1121/1.424899

Savel, S. (2009). Individual differences and left/right asymmetries in auditory space perception. I. Localization of low-frequency sounds in free field. *Hearing Research*, *255*(1–2), 142–154. https://doi.org/10.1016/j.heares.2009.06.013

Searle, C. L., Braida, L. D., Cuddy, D. R., & Davis, M. F. (1975). Binaural pinna disparity: Another auditory localization cue. *The Journal of the Acoustical Society of America*, *57*(2), 448–455. https://doi.org/10.1121/1.380442

Shinn-Cunningham, B. (2000). *Learning reverberation: Considerations for spatial auditory displays*. Retrieved from https://smartech.gatech.edu/handle/1853/50675

Shinn-cunningham, B. (2001). Localizing sound in rooms. *Proceedings of the ACM/SIGGRAPH and Eurographics Campfire: Acoustic Rendering for Virtual Environments*.

Shinn-Cunningham, B. G. (2000). *Distance Cues for Virtual Auditory Space*.

Shinn-Cunningham, Barbara G., Kopco, N., & Martin, T. J. (2005). Localizing nearby sound sources in a classroom: Binaural room impulse responses. *The Journal of the Acoustical Society of America*, *117*(5), 3100–3115. https://doi.org/10.1121/1.1872572

Sinker, J., & Shirley, B. (2016, May 26). *The Effect of Early Impulse Response Length and Visual Environment on Externalization of Binaural Virtual Sources*. Presented at the Audio Engineering Society Convention 140. Retrieved from http://www.aes.org/e-lib/browse.cfm?elib=18250

Spagnol, S., Geronazzo, M., & Avanzini, F. (2013). On the Relation Between Pinna Reflection Patterns and Head-Related Transfer Function Features. *IEEE Transactions on Audio, Speech, and Language Processing*, *21*(3), 508–519. https://doi.org/10.1109/TASL.2012.2227730

Sporer, T., Werner, S., & Klein, F. (2016, September 21). *Adjustment of the direct-to-Reverberant-Energy-Ratio to Reach Externalization within a Binaural Synthesis System*. Presented at the Audio Engineering Society Conference: 2016 AES International Conference on Audio for Virtual and Augmented Reality. Retrieved from http://www.aes.org/e-lib/browse.cfm?elib=18513

Vliegen, J., & Van Opstal, A. J. (2004). The influence of duration and level on human sound localization. *The Journal of the Acoustical Society of America*, *115*(4), 1705–1713. https://doi.org/10.1121/1.1687423

Völk, F. (2009). *Externalization in data-based Binaural Synthesis: Effects of Impulse Response Length*. 4.

Wallach, H. (1940). The role of head movements and vestibular and visual cues in sound localization. *Journal of Experimental Psychology*, *27*(4), 339–368. https://doi.org/10.1037/h0054629

Wallach, Hans. (1939). On Sound Localization. *The Journal of the Acoustical Society of America*, *10*(4), 270–274. https://doi.org/10.1121/1.1915985

Wanrooij, M. M. V., & Opstal, A. J. V. (2007). Sound Localization Under Perturbed Binaural Hearing. *Journal of Neurophysiology*, *97*(1), 715–726. https://doi.org/10.1152/jn.00260.2006

Werner, S., & Klein, F. (2014, August 26). *Influence of Context Dependent Quality Parameters on the Perception of Externalization and Direction of an Auditory Event*. Presented at the Audio Engineering Society Conference: 55th International Conference: Spatial Audio. Retrieved from http://www.aes.org/e-lib/browse.cfm?elib=17371

Wightman, F. L., & Kistler, D. J. (1989a). Headphone simulation of free-field listening. I: Stimulus synthesis. *The Journal of the Acoustical Society of America*, *85*(2), 858–867. https://doi.org/10.1121/1.397557

Wightman, F. L., & Kistler, D. J. (1989b). Headphone simulation of free-field listening. II: Psychophysical validation. *The Journal of the Acoustical Society of America*, *85*(2), 868–878. https://doi.org/10.1121/1.397558

Wightman, F. L., & Kistler, D. J. (1999). Resolution of front–back ambiguity in spatial hearing by listener and source movement. *The Journal of the Acoustical Society of America*, *105*(5), 2841–2853. https://doi.org/10.1121/1.426899

Wright, D., Hebrank, J. H., & Wilson, B. (1974). Pinna reflections as cues for localization. *The Journal of the Acoustical Society of America*, *56*(3), 957–962. https://doi.org/10.1121/1.1903355

Zhang, P. X., & Hartmann, W. M. (2010). On the ability of human listeners to distinguish between front and back. *Hearing Research*, *260*(1–2), 30–46. https://doi.org/10.1016/j.heares.2009.11.001

Zhong, X.-L., Zhang, F., & Xie, B.-S. (2013). On the spatial symmetry of head-related transfer functions. *ResearchGate*, *74*(6), 856–864. https://doi.org/10.1016/j.apacoust.2013.01.004

9. Appendix

9.1. Identified Notches

Each subject at each elevation angle for both ears. The tables indicate the subject name, elevation angle, ear, notch start frequency (Hz), notch end frequency (Hz) and depth (dB) of each notch that has been manipulated.

HL				
Angle	**Ear**	**S Freq. (Hz)**	**E Freq. (Hz)**	**Depth (dB)**
0	L	5555	7359	-9.2
		15773	18586	-14.3
		18914	20000	-14.2
0	R	5250	10477	-20.9
		16992	20000	-16.4
30	L	16570	20000	-19.9
30	R	7757	10008	-13.9
		10523	12000	-9.2
		18164	20000	-21.3
60	L	16148	20000	-19.3
60	R	10430	12281	-9.8
		17719	20000	-22.1
90	L	9234	13031	-8.7
		15094	20000	-17.9
90	R	12188	20000	-28.1
120	L	9563	13383	-11.9
		13664	20000	-22.6
120	R	9422	18281	-11.9
		18305	20111	-21.3
150	L	13617	15586	-22.2
		15984	17016	-6.6
		18328	20000	-24.9
150	R	9047	10055	-6.4
		10500	17789	-12.8
		18305	20000	-21.4
180	L	6328	8930	-31.9
		12211	13031	-6.1
		16430	18117	-13.9
		19078	20000	-12.3
180	R	6469	9117	-28
		11063	13805	-14.7
		16125	20000	-25.5

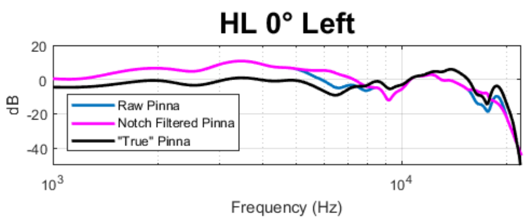
MM				
Angle	**Ear**	**S Freq. (Hz)**	**E Freq. (Hz)**	**Depth (dB)**
0	L	5930	11297	-23.1
		12773	20000	-29.2
0	R	5531	11297	-13.3
		18914	20000	-16.5
30	L	7336	8930	-11.2
		9938	14367	-21.1
		15375	18750	-16
		18797	20000	-18.5
30	R	5742	6563	-6.2
		7617	9703	-34.3
		19266	20000	-15.1
60	L	10008	20000	-28.8
60	R	5578	6609	-8.5
		7641	8648	-7.1
		10172	20000	-33.8
90	L	10172	20000	-31.1
90	R	5016	6844	-9.2
		7711	20000	-44
120	L	9633	20000	-50.1
120	R	5531	6656	-8
		9070	20000	-53.3
150	L	9680	11344	-8.9
		12773	20000	-24
150	R	5367	6961	-12.9
		8742	20000	-19.1
180	L	5484	7172	-35.3
		9000	20000	-22.9
180	R	5414	8578	-21
		9516	20000	-33.3

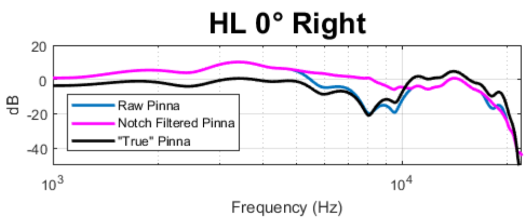
DJ				
Angle	**Ear**	**S Freq. (Hz)**	**E Freq. (Hz)**	**Depth (dB)**
0	L	6000	7125	-6.3
		7289	11648	-16.8
		15398	20000	-24.6
0	R	8789	11227	-15.4
		18633	20000	-17.2
30	L	8109	8977	-6.1
		10195	12633	-8.3
		18961	20000	-15.5
30	R	9539	12328	-12.1
		18492	20000	-19.6
60	L	12352	14391	-7
		15984	20000	-25.6
60	R	15328	20000	-35.4
90	L	12164	20000	-39.2
90	R	14484	20000	-22.7

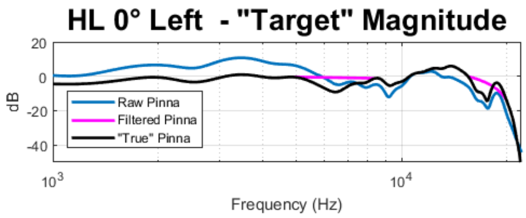
120	L	11977	20000	-24.8
120	R	11273	20000	-30.9
150	L	10055	11625	-8.2
150	R	10125	11578	-8.3
		16734	20000	-33
180	L	6023	10008	-11.3
		13594	17906	-14.6
		18281	20000	-19.4
180	R	14063	17180	-29.4
		18492	20000	-19

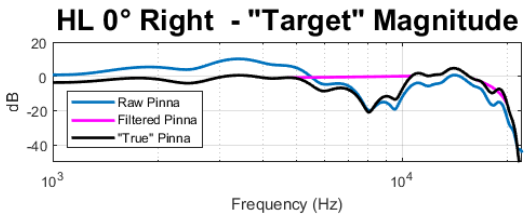
9.2. NF100 Notch Conditions

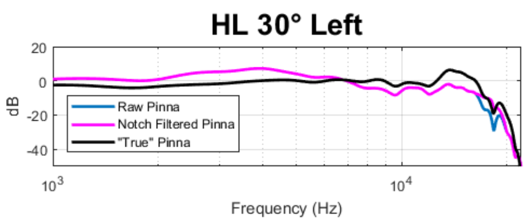
Notch filling in the most extreme condition (NF100) for all angles tested on all subjects. The plots show frequency manipulations on both the left and right ears. For each plot at each condition, there are two graphs. The top plot represents the actual filtering applied to the raw pinna response (pink), against the 'true' pinna measurement (black) and the raw measurement (blue) that was filtered. The plot below represents the 'target' magnitude for each condition (pink). Effectively, how the true pinna is being filtered in each condition.

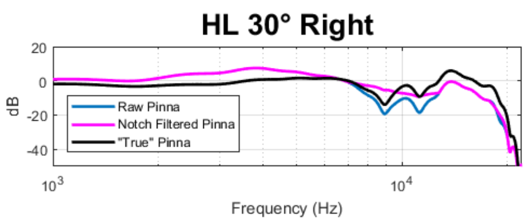


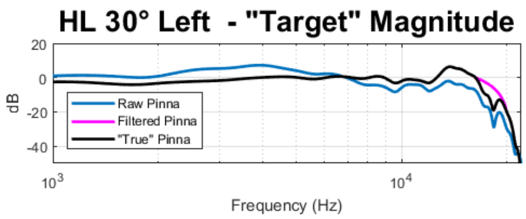


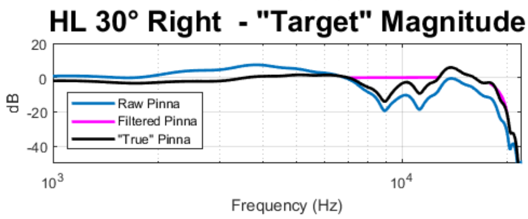


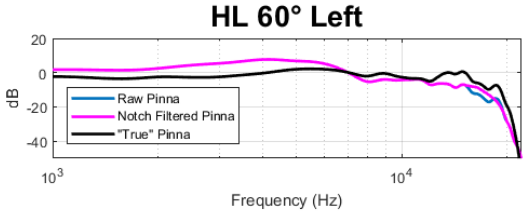


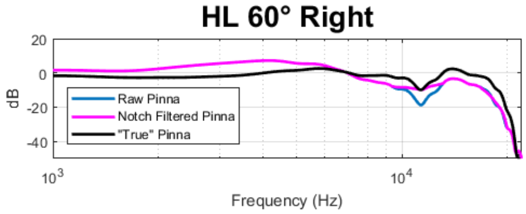


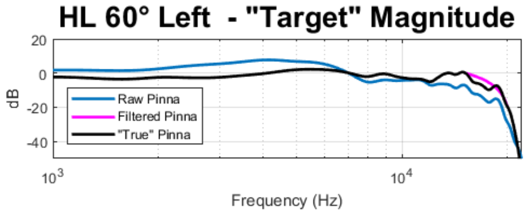


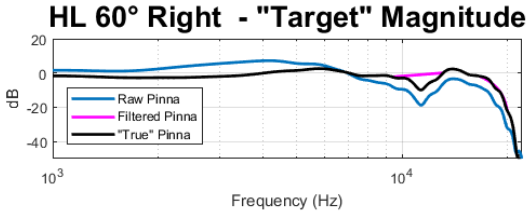


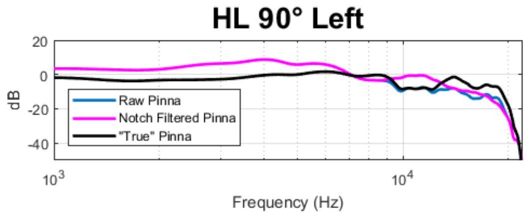


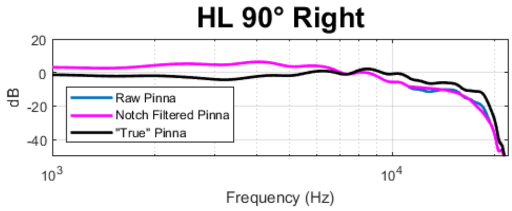

HL 60° Left
Raw Pinna
Notch Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

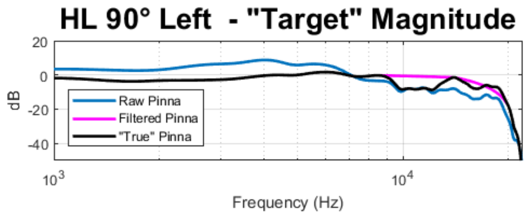

HL 60° Right
Raw Pinna
Notch Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

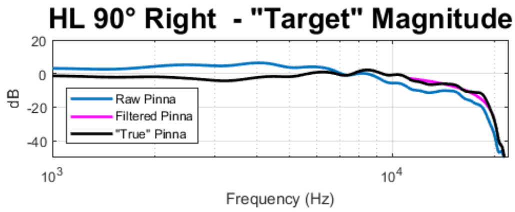

HL 60° Left - "Target" Magnitude
Raw Pinna
Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

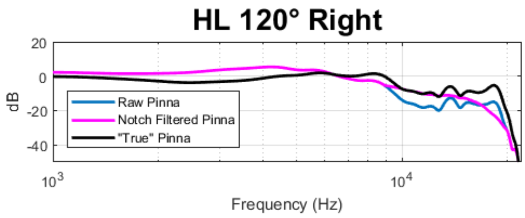

HL 60° Right - "Target" Magnitude
Raw Pinna
Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

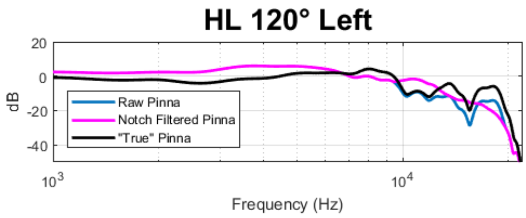

HL 90° Left
Raw Pinna
Notch Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

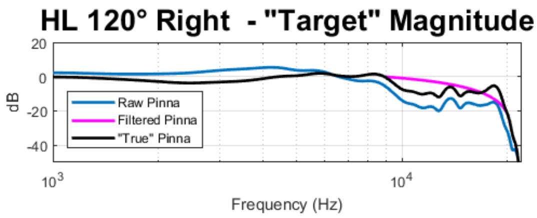

HL 90° Right
Raw Pinna
Notch Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

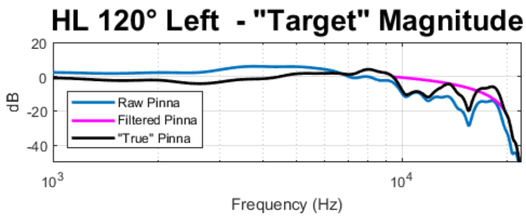

HL 90° Left - "Target" Magnitude
Raw Pinna
Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

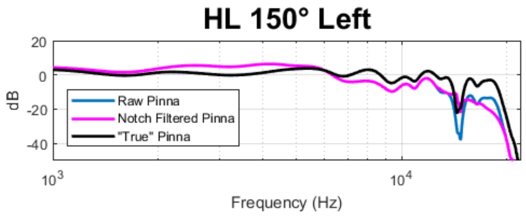

HL 90° Right - "Target" Magnitude
Raw Pinna
Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

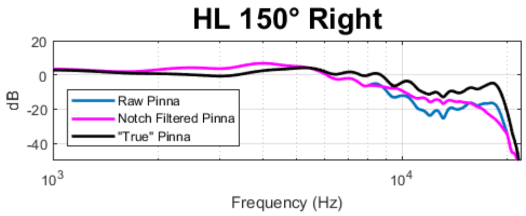

HL 120° Right
Raw Pinna
Notch Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

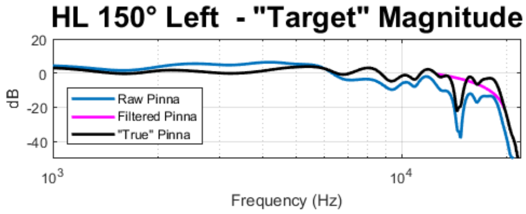

HL 120° Left
Raw Pinna
Notch Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

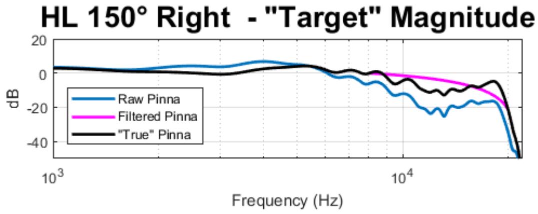

HL 120° Right - "Target" Magnitude
Raw Pinna
Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

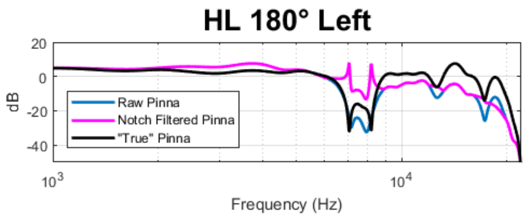

HL 120° Left - "Target" Magnitude
Raw Pinna
Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

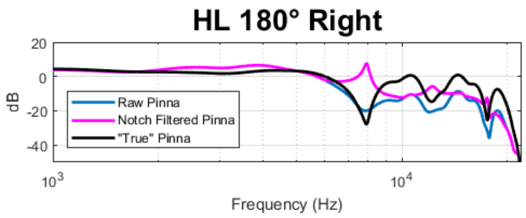

HL 150° Left
Raw Pinna
Notch Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

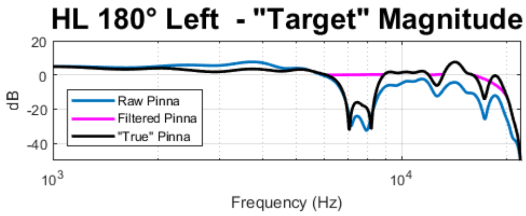

HL 150° Right
Raw Pinna
Notch Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

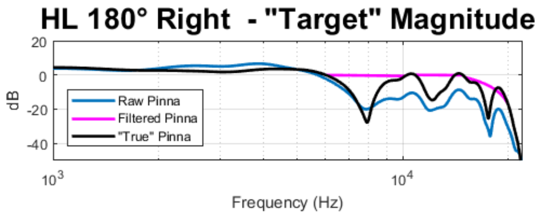

HL 150° Left - "Target" Magnitude
Raw Pinna
Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

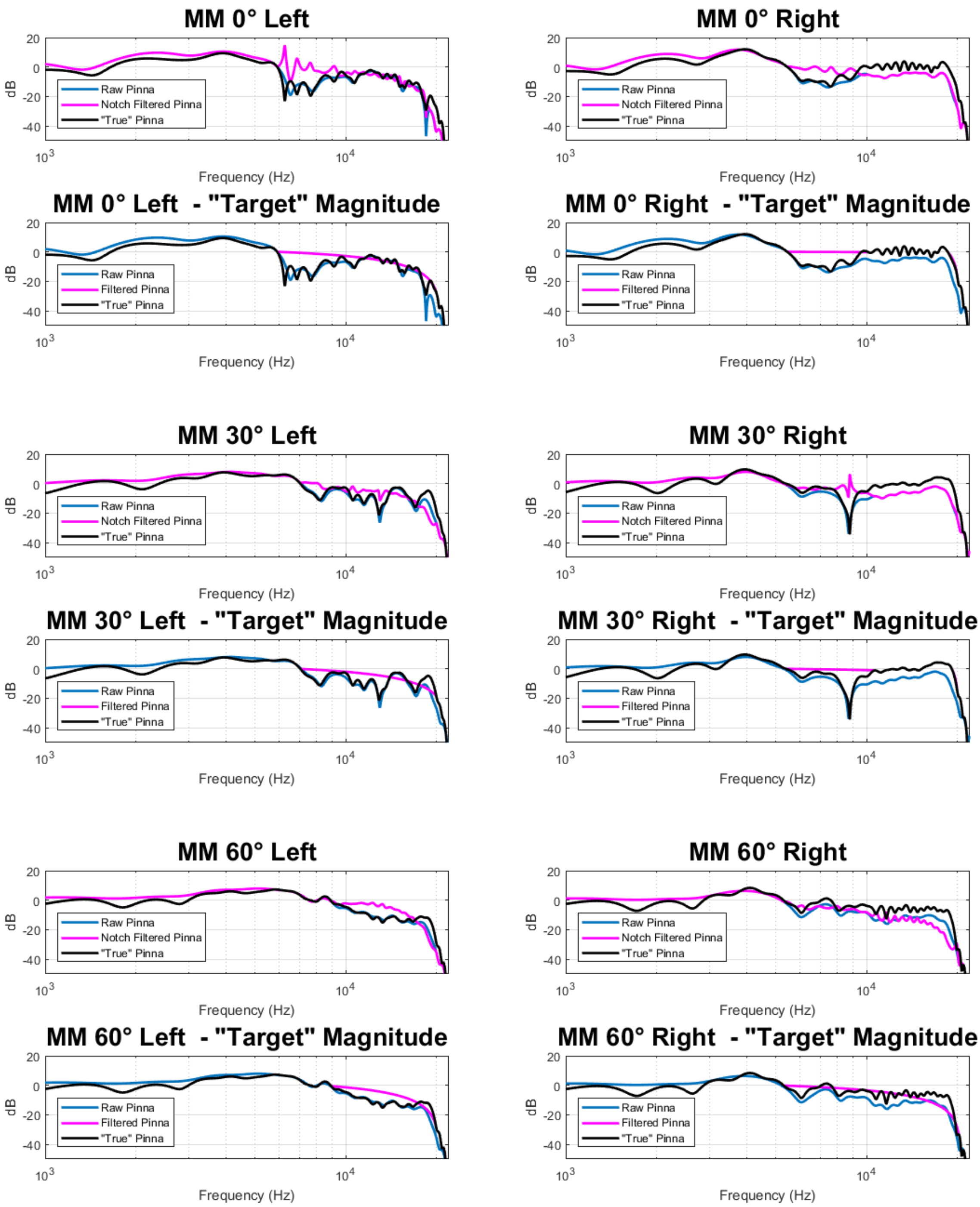

HL 150° Right - "Target" Magnitude
Raw Pinna
Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

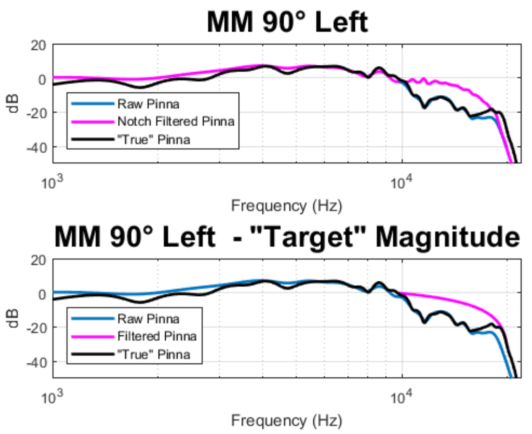

HL 180° Left
Raw Pinna
Notch Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

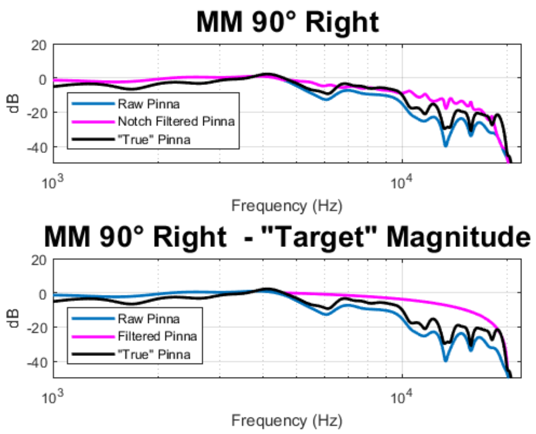

HL 180° Right
Raw Pinna
Notch Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

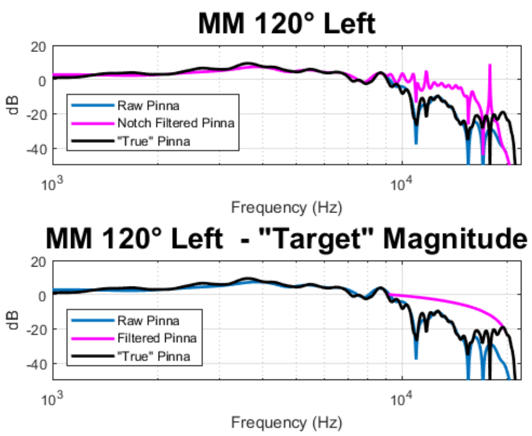

HL 180° Left - "Target" Magnitude
Raw Pinna
Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

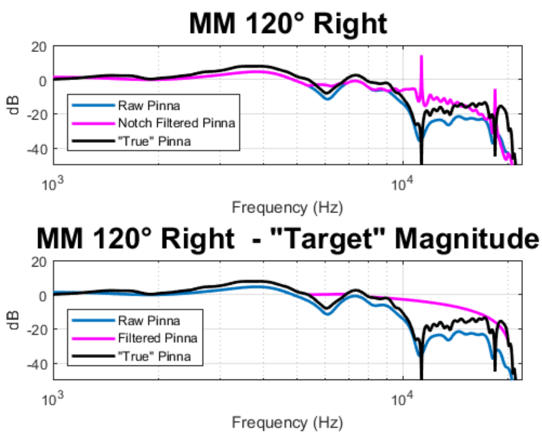

HL 180° Right - "Target" Magnitude
Raw Pinna
Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

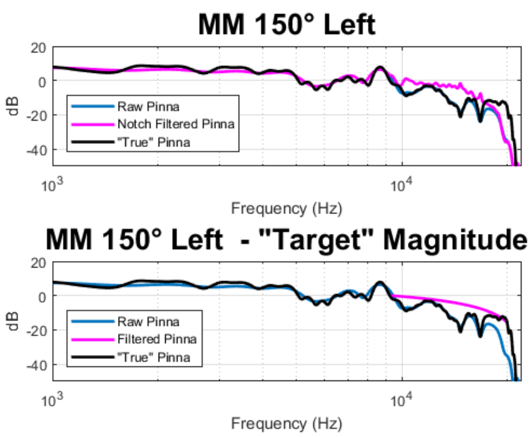

MM 0° Left
20
0
-20
-40
dB
Raw Pinna
Notch Filtered Pinna
"True" Pinna
10^3
10^4
Frequency (Hz)
MM 0° Right
Raw Pinna
Notch Filtered Pinna
"True" Pinna
Frequency (Hz)
MM 0° Left - "Target" Magnitude
Raw Pinna
Filtered Pinna
"True" Pinna
Frequency (Hz)
MM 0° Right - "Target" Magnitude
Raw Pinna
Filtered Pinna
"True" Pinna
Frequency (Hz)
MM 30° Left
Raw Pinna
Notch Filtered Pinna
"True" Pinna
Frequency (Hz)
MM 30° Right
Raw Pinna
Notch Filtered Pinna
"True" Pinna
Frequency (Hz)
MM 30° Left - "Target" Magnitude
Raw Pinna
Filtered Pinna
"True" Pinna
Frequency (Hz)
MM 30° Right - "Target" Magnitude
Raw Pinna
Filtered Pinna
"True" Pinna
Frequency (Hz)
MM 60° Left
Raw Pinna
Notch Filtered Pinna
"True" Pinna
Frequency (Hz)
MM 60° Right
Raw Pinna
Notch Filtered Pinna
"True" Pinna
Frequency (Hz)
MM 60° Left - "Target" Magnitude
Raw Pinna
Filtered Pinna
"True" Pinna
Frequency (Hz)
MM 60° Right - "Target" Magnitude
Raw Pinna
Filtered Pinna
"True" Pinna
Frequency (Hz)

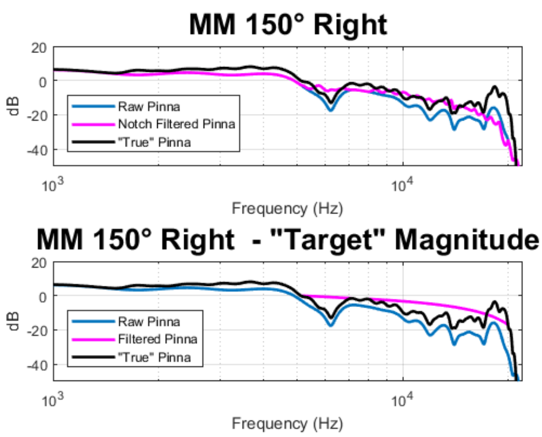

MM 90° Left
Raw Pinna
Notch Filtered Pinna
"True" Pinna
dB
Frequency (Hz)
MM 90° Left - "Target" Magnitude
Raw Pinna
Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

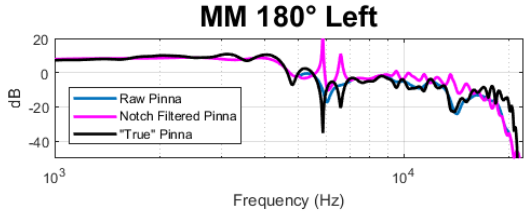

MM 90° Right
Raw Pinna
Notch Filtered Pinna
"True" Pinna
dB
Frequency (Hz)
MM 90° Right - "Target" Magnitude
Raw Pinna
Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

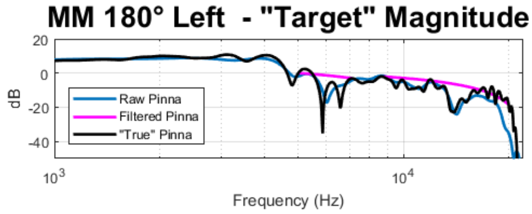

MM 120° Left
Raw Pinna
Notch Filtered Pinna
"True" Pinna
dB
Frequency (Hz)
MM 120° Left - "Target" Magnitude
Raw Pinna
Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

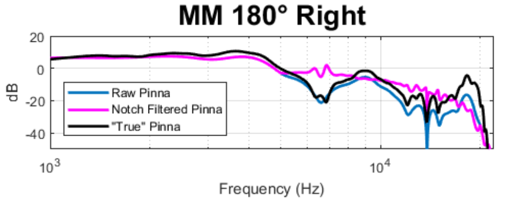

MM 120° Right
Raw Pinna
Notch Filtered Pinna
"True" Pinna
dB
Frequency (Hz)
MM 120° Right - "Target" Magnitude
Raw Pinna
Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

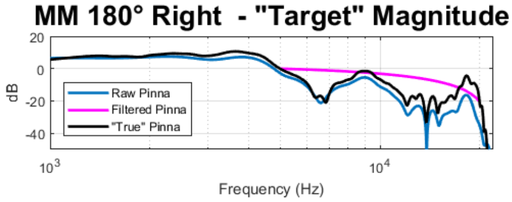

MM 150° Left
Raw Pinna
Notch Filtered Pinna
"True" Pinna
dB
Frequency (Hz)
MM 150° Left - "Target" Magnitude
Raw Pinna
Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

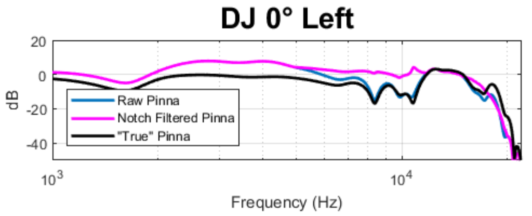

MM 150° Right
Raw Pinna
Notch Filtered Pinna
"True" Pinna
dB
Frequency (Hz)
MM 150° Right - "Target" Magnitude
Raw Pinna
Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

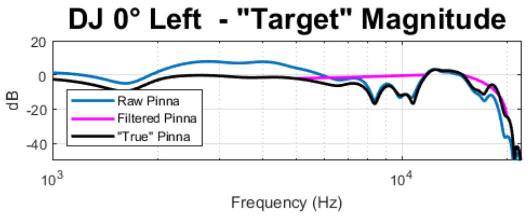

MM 180° Left
dB
Raw Pinna
Notch Filtered Pinna
"True" Pinna
Frequency (Hz)

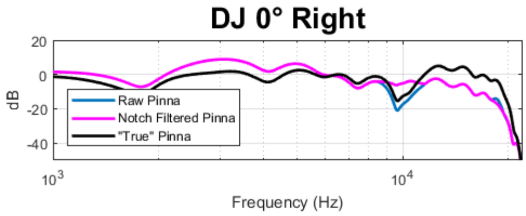

MM 180° Left - "Target" Magnitude
dB
Raw Pinna
Filtered Pinna
"True" Pinna
Frequency (Hz)

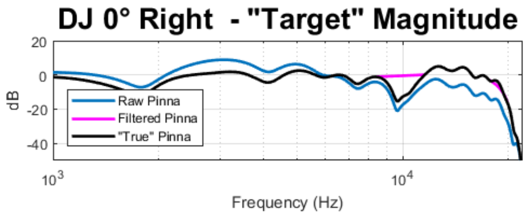

MM 180° Right
dB
Raw Pinna
Notch Filtered Pinna
"True" Pinna
Frequency (Hz)

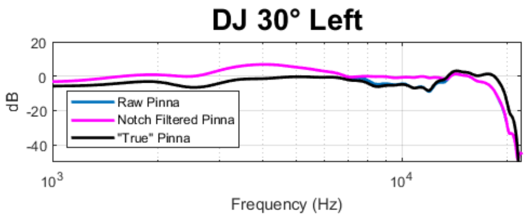

MM 180° Right - "Target" Magnitude
dB
Raw Pinna
Filtered Pinna
"True" Pinna
Frequency (Hz)

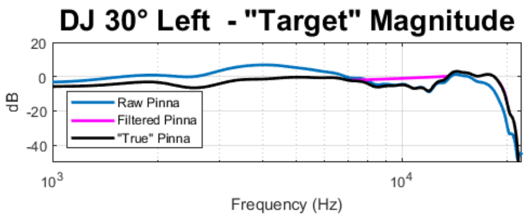

DJ 0° Left
dB
Raw Pinna
Notch Filtered Pinna
"True" Pinna
Frequency (Hz)

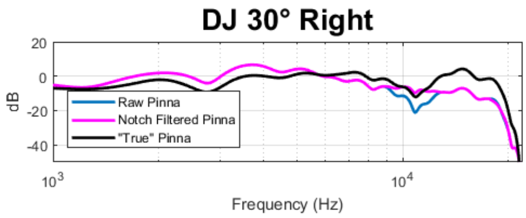

DJ 0° Left - "Target" Magnitude
dB
Raw Pinna
Filtered Pinna
"True" Pinna
Frequency (Hz)

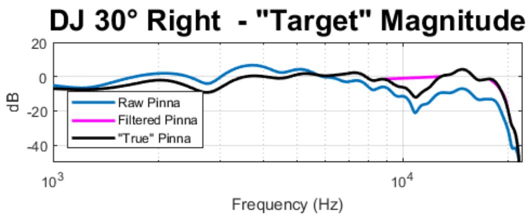

DJ 0° Right
dB
Raw Pinna
Notch Filtered Pinna
"True" Pinna
Frequency (Hz)

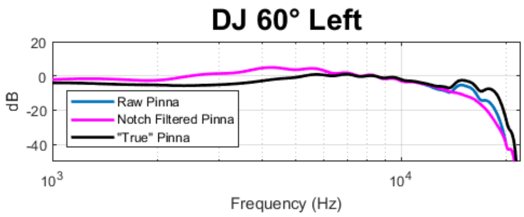

DJ 0° Right - "Target" Magnitude
dB
Raw Pinna
Filtered Pinna
"True" Pinna
Frequency (Hz)

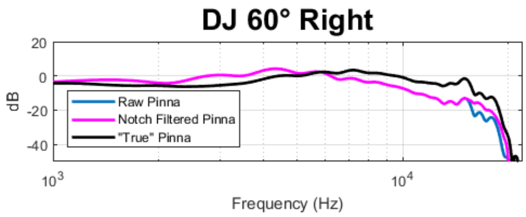

DJ 30° Left
dB
Raw Pinna
Notch Filtered Pinna
"True" Pinna
Frequency (Hz)

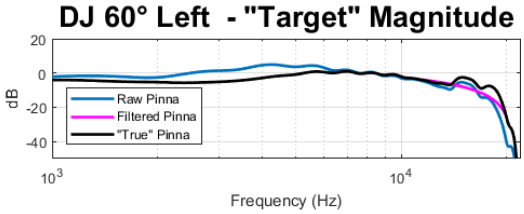

DJ 30° Left - "Target" Magnitude
dB
Raw Pinna
Filtered Pinna
"True" Pinna
Frequency (Hz)

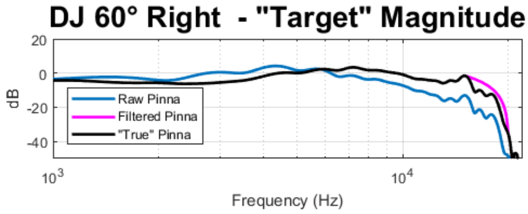

DJ 30° Right
dB
Raw Pinna
Notch Filtered Pinna
"True" Pinna
Frequency (Hz)

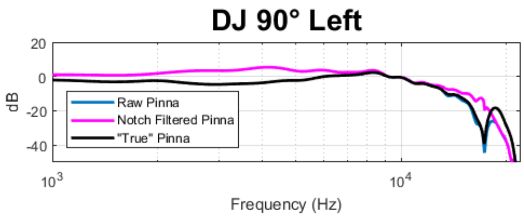

DJ 30° Right - "Target" Magnitude
dB
Raw Pinna
Filtered Pinna
"True" Pinna
Frequency (Hz)

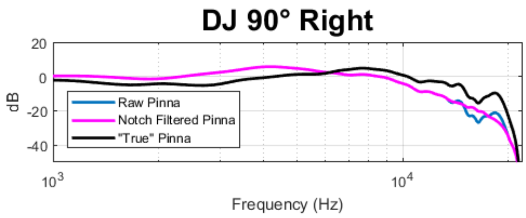

DJ 60° Left
dB
Raw Pinna
Notch Filtered Pinna
"True" Pinna
Frequency (Hz)

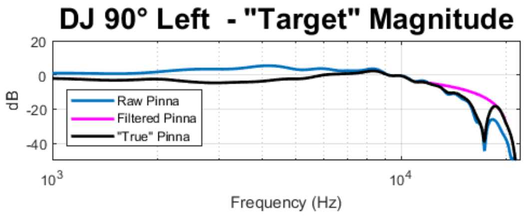

DJ 60° Right
dB
Raw Pinna
Notch Filtered Pinna
"True" Pinna
Frequency (Hz)

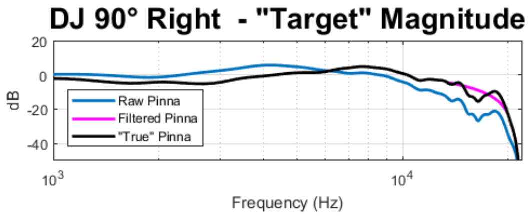

DJ 60° Left - "Target" Magnitude
dB
Raw Pinna
Filtered Pinna
"True" Pinna
Frequency (Hz)

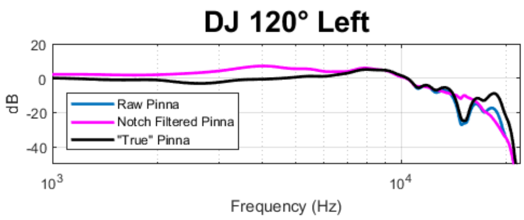

DJ 60° Right - "Target" Magnitude
dB
Raw Pinna
Filtered Pinna
"True" Pinna
Frequency (Hz)

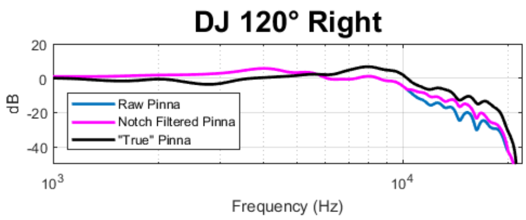

DJ 90° Left
dB
Raw Pinna
Notch Filtered Pinna
"True" Pinna
Frequency (Hz)

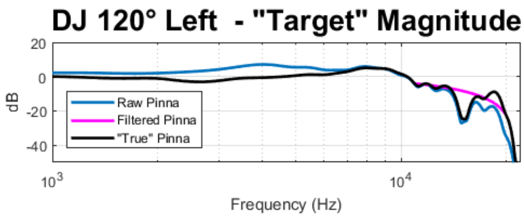

DJ 90° Right
dB
Raw Pinna
Notch Filtered Pinna
"True" Pinna
Frequency (Hz)

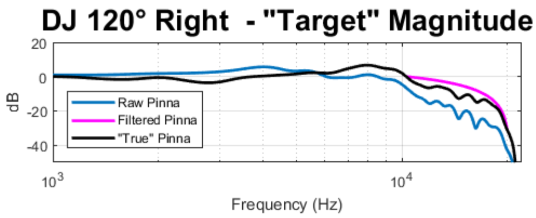

DJ 90° Left - "Target" Magnitude
dB
Raw Pinna
Filtered Pinna
"True" Pinna
Frequency (Hz)

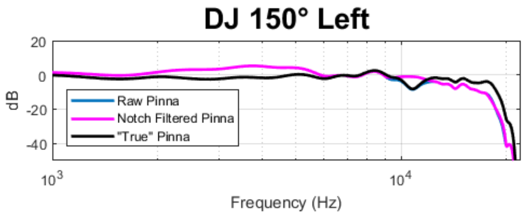

DJ 90° Right - "Target" Magnitude
dB
Raw Pinna
Filtered Pinna
"True" Pinna
Frequency (Hz)

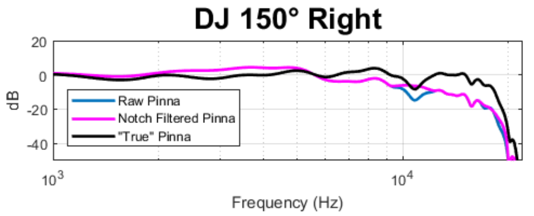

DJ 120° Left
dB
Raw Pinna
Notch Filtered Pinna
"True" Pinna
Frequency (Hz)

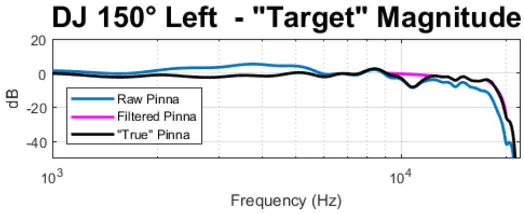

DJ 120° Right
dB
Raw Pinna
Notch Filtered Pinna
"True" Pinna
Frequency (Hz)

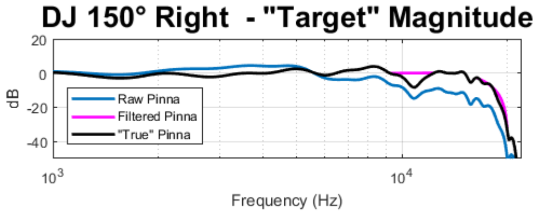

DJ 120° Left - "Target" Magnitude
dB
Raw Pinna
Filtered Pinna
"True" Pinna
Frequency (Hz)

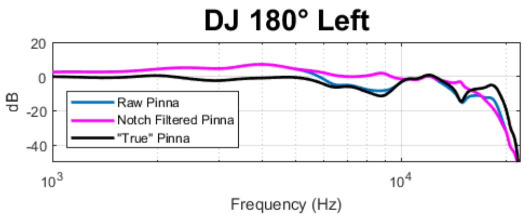

DJ 120° Right - "Target" Magnitude
dB
Raw Pinna
Filtered Pinna
"True" Pinna
Frequency (Hz)

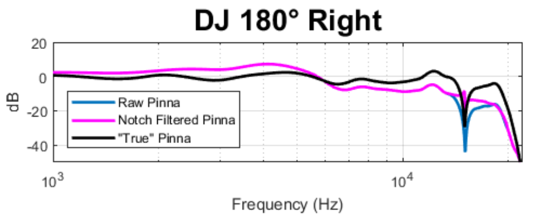


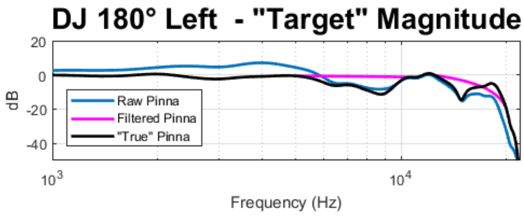


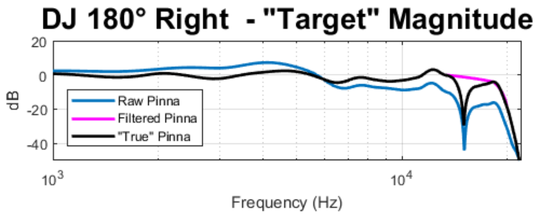






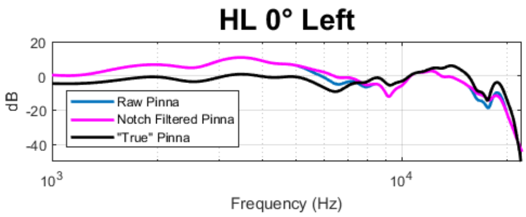


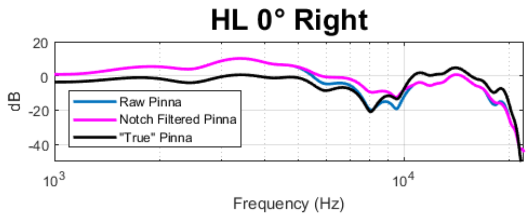


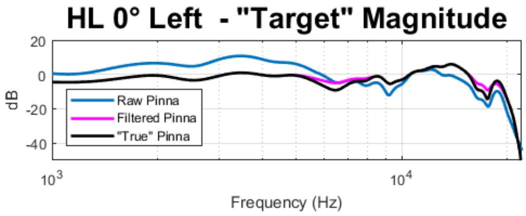


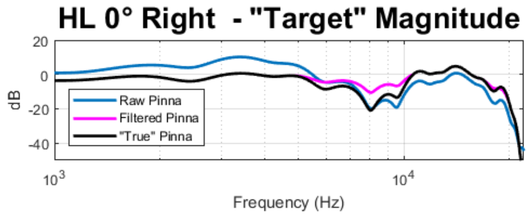
9.3. NF50 Notch Conditions

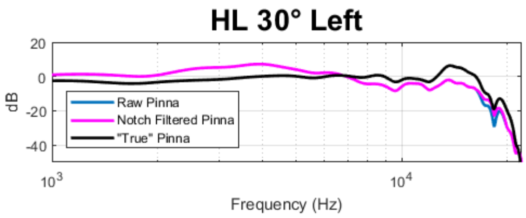
Notch filling in the least extreme condition (NF50) for all angles tested on subject HL to demonstrate the preservation of the spectral shape of the notches as their magnitudes are reduced. The plots show frequency manipulations on both the left and right ears. For each plot at each condition, there are two graphs. With the top representing the actual filtering applied to the raw pinna response (pink), against the ‘true’ pinna measurement (black) and the raw measurement (blue) that will be filtered. The plot below represents the ‘target’ magnitude for each condition (pink).

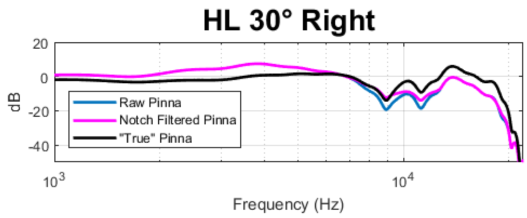

HL 0° Left
Raw Pinna
Notch Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

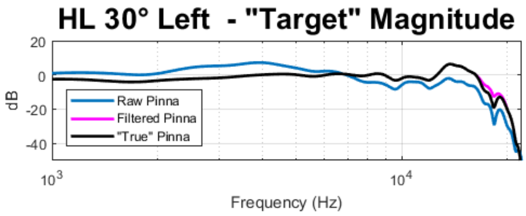

HL 0° Right
Raw Pinna
Notch Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

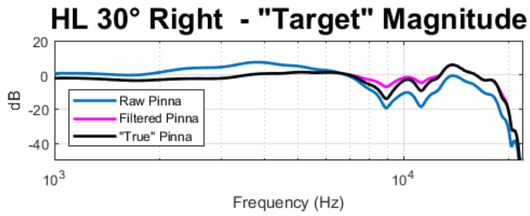

HL 0° Left - "Target" Magnitude
Raw Pinna
Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

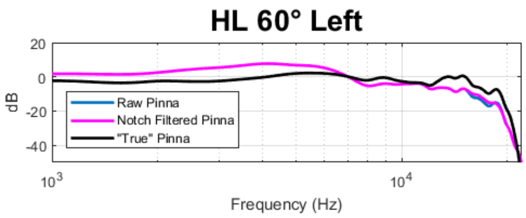

HL 0° Right - "Target" Magnitude
Raw Pinna
Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

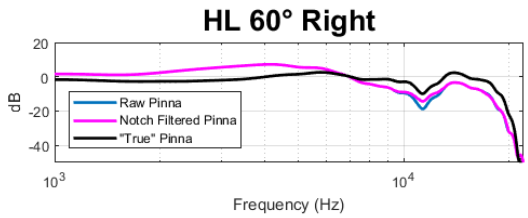

HL 30° Left
Raw Pinna
Notch Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

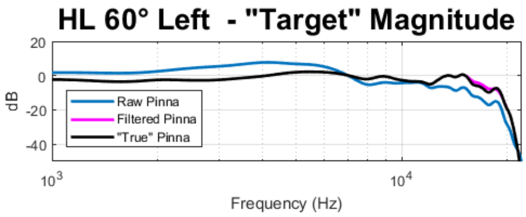

HL 30° Right
Raw Pinna
Notch Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

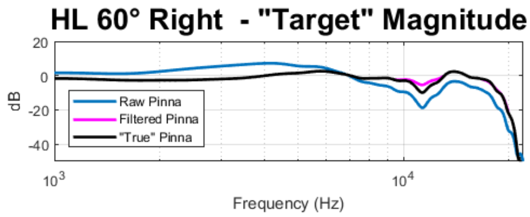

HL 30° Left - "Target" Magnitude
Raw Pinna
Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

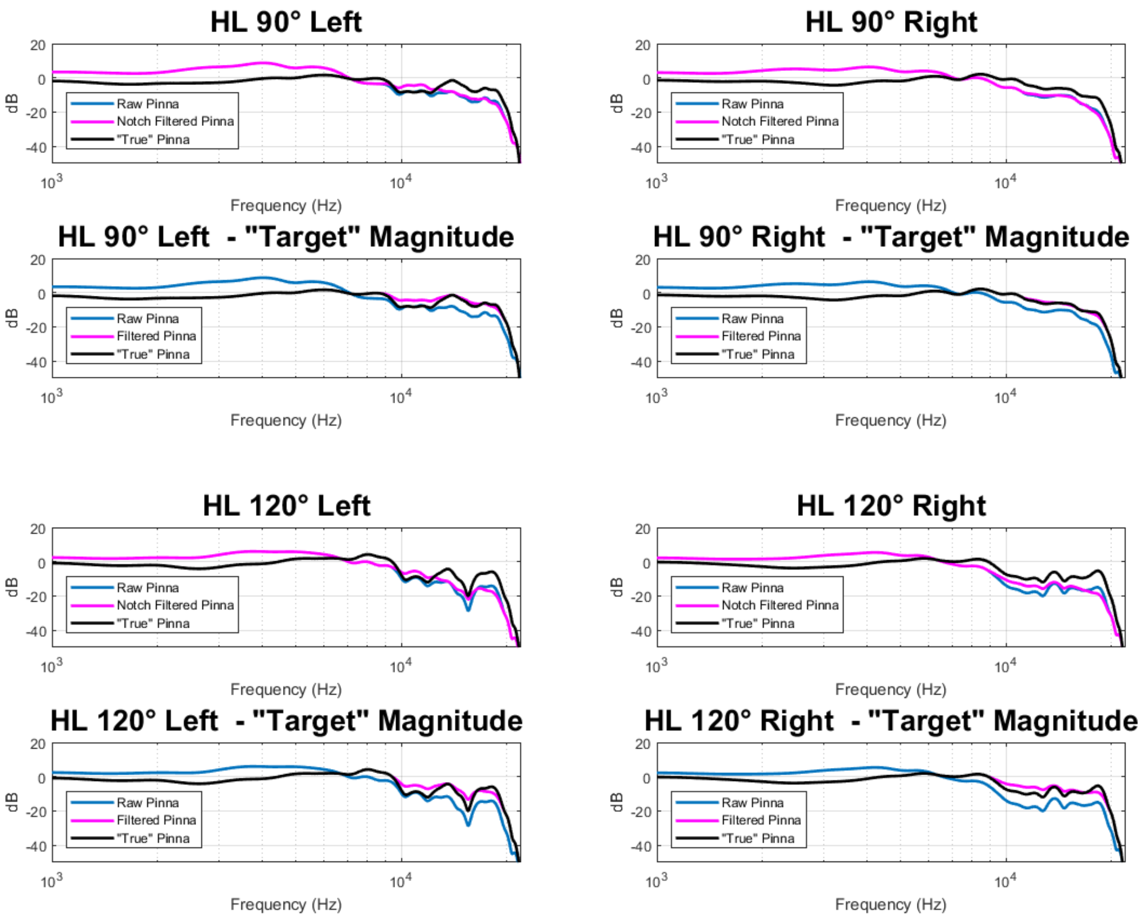

HL 30° Right - "Target" Magnitude
Raw Pinna
Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

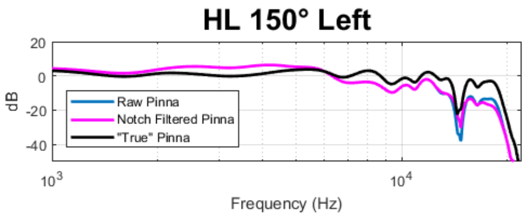

HL 60° Left
Raw Pinna
Notch Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

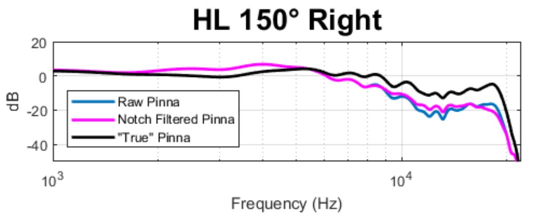

HL 60° Right
Raw Pinna
Notch Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

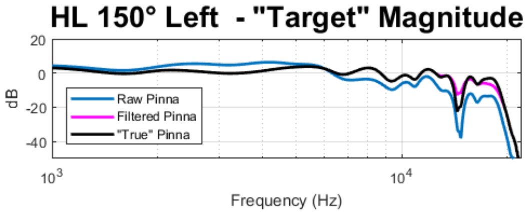

HL 60° Left - "Target" Magnitude
Raw Pinna
Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

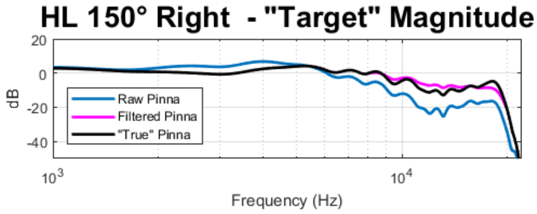

HL 60° Right - "Target" Magnitude
Raw Pinna
Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

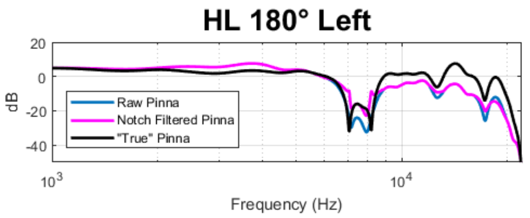

HL 90° Left
Raw Pinna
Notch Filtered Pinna
"True" Pinna
dB
Frequency (Hz)
HL 90° Right
Raw Pinna
Notch Filtered Pinna
"True" Pinna
dB
Frequency (Hz)
HL 90° Left - "Target" Magnitude
Raw Pinna
Filtered Pinna
"True" Pinna
dB
Frequency (Hz)
HL 90° Right - "Target" Magnitude
Raw Pinna
Filtered Pinna
"True" Pinna
dB
Frequency (Hz)
HL 120° Left
Raw Pinna
Notch Filtered Pinna
"True" Pinna
dB
Frequency (Hz)
HL 120° Right
Raw Pinna
Notch Filtered Pinna
"True" Pinna
dB
Frequency (Hz)
HL 120° Left - "Target" Magnitude
Raw Pinna
Filtered Pinna
"True" Pinna
dB
Frequency (Hz)
HL 120° Right - "Target" Magnitude
Raw Pinna
Filtered Pinna
"True" Pinna
dB
Frequency (Hz)

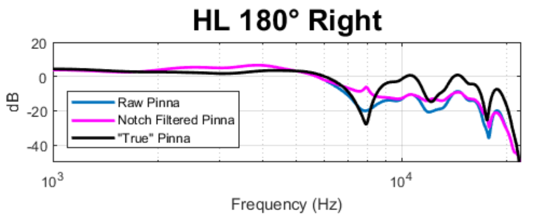

HL 150° Left
dB
20
0
-20
-40
Raw Pinna
Notch Filtered Pinna
"True" Pinna
10^3
10^4
Frequency (Hz)

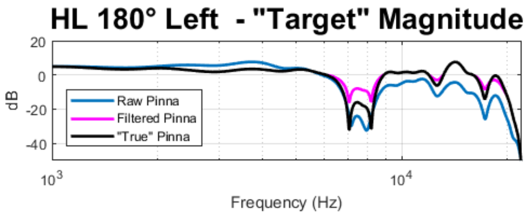

HL 150° Right
dB
20
0
-20
-40
Raw Pinna
Notch Filtered Pinna
"True" Pinna
10^3
10^4
Frequency (Hz)

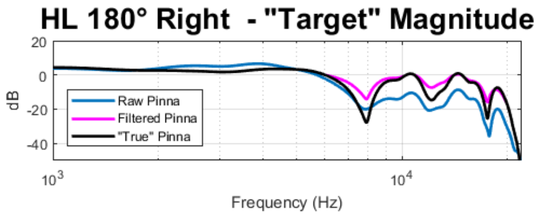

HL 150° Left - "Target" Magnitude
dB
20
0
-20
-40
Raw Pinna
Filtered Pinna
"True" Pinna
10^3
10^4
Frequency (Hz)


HL 150° Right - "Target" Magnitude
dB
20
0
-20
-40
Raw Pinna
Filtered Pinna
"True" Pinna
10^3
10^4
Frequency (Hz)


HL 180° Left
dB
20
0
-20
-40
Raw Pinna
Notch Filtered Pinna
"True" Pinna
10^3
10^4
Frequency (Hz)


HL 180° Right
dB
20
0
-20
-40
Raw Pinna
Notch Filtered Pinna
"True" Pinna
10^3
10^4
Frequency (Hz)


HL 180° Left - "Target" Magnitude
dB
20
0
-20
-40
Raw Pinna
Filtered Pinna
"True" Pinna
10^3
10^4
Frequency (Hz)


HL 180° Right - "Target" Magnitude
dB
20
0
-20
-40
Raw Pinna
Filtered Pinna
"True" Pinna
10^3
10^4
Frequency (Hz)

9.4. Notch Filling Code

```
angles =[0 30 60 90 120 150 180];
for u = 1:length(angles)
    for ch = 1:2
clearvars -except angles u ch

%% Init
subj = 'hyunkook';
ss = 'HL';
ang = num2str(angles(u))
len = 400;                          % final IR length (anechoic or room?)
EQcond = 1;                         % 0 = no fill, 1 = half, 2 = full


p_len = 1;                          % length of pinna section to be windowed (ms)
chop = 0.5;                         % length (ms) of window slope

searchBW = 60;                      % num bins to check  on either side of LF and
HF thresh to find notch boundary
thresh = -6;                        % notch threshold (dB)
Nfft = 4096;

% get gain ratio
if EQcond == 0
    ratio = 0;                          % notch filling percentage
elseif EQcond == 1
    ratio = 0.5;                        % notch filling percentage
elseif EQcond == 2
    ratio = 1;                          % notch filling percentage
end


IRdir = ['../' subj '/'];

%% Load and FFT

% Load 'RAW' IR to apply filter to
[aSt fs] = audioread([IRdir 'prep/' ang '.wav']);
a = aSt(:,ch);                            %process each channel individually

% Load IR with MIC INVERSE filter applied ('TRUE' pinna)
xst = audioread([IRdir 'micInv/' ang '_micInv.wav']);
x = xst(:,ch);                            % get relevant channel
Xtp = fft(x,Nfft);
magtp = abs(Xtp);                         % get frequency
dBtp = mag2db(magtp);                     % convert to db
Np = length(magtp);                       % size

%% Get Pinna part of raw IR to be Filtered
Lp = ceil(fs*(p_len/1000));               % window curve size in samples
Ncurve = ceil(fs*(chop/1000));            % window curve size in samples
win = hann(Ncurve*2);                     % Hann window for each end

%------RAW PINNA------%
[p pEnd] = chop_ir(a, chop-0.1, p_len, fs);     %window pinna part (chop = length
of slope in ms) <-----------1------------
```

```
P = fft(p,Nfft);                              % freq dom pinna part (large resolution)
mag_P = abs(P); % <<---------------------THIS PART WILL BE FILTERED--------------

mag_orgP = abs(P);
pL = length(p);                                 % save original pinna part
%------------------%

tailS = pEnd - Ncurve;                          % idx of start of 'reflections/reverb'
section
trect = a(tailS:pEnd-1);                        % first part to be windowed with first
half of hann - later to be mixed with end of pinna
tailSwin  =  win(1:Ncurve).*trect;                    %  mix  with  start  part  of  tail
<------------2------------

%------TAIL RECT------%
Nrect = ceil(fs*((len+1)/1000)) - Lp;          % size of 'tail' part (whole IR being
used, without pinna)
tailRect = a(pEnd:Nrect-Ncurve);                 % rectangular window of 'tail' part
<------------3------------
%----------------------%

%----TAIL fadeout-----%
endS = Nrect-Ncurve;                            %
efade = a(endS+1:endS+Ncurve);
fadeout = win(Ncurve+1:end).*efade;         %                                        <-
-----------4------------
%----------------------%


%% Find notch below certain threshold of PINNA part ONLY (MIC INV)

LFlim = round(((Np/2)/(fs/2))*5000);          % only look for notches after 5kHz
HFlim = round(((Np/2)/(fs/2))*20000);         % only look for notches up to 18kHz

% index of bins who's magnitude are below thresh
for i = LFlim:HFlim
    if dBtp(i) < thresh
        mark(i) = i;
    end
end

% Get first and last bin index of notch position
tog = 0; incS = 0; incE = 0;

if exist('mark','var') == 0 %                   % only perform if notches are found
    notchfill = a;
else
    for j = 2:length(mark)-1
        if (mark(j) == mark(j+1) - 1) && (mark(j-1) == 0)   % at least  2 consecutive
bins
            incS = incS + 1;
            Fstart(incS) = mark(j);                              % note index of first
bin in notch
            tog = 1;
        end
        if tog == 1
            if (mark(j) > 0 && mark(j+1) == 0) || mark(j) == mark(end-1)
                incE = incE + 1;
                Fend(incE) = mark(j);         % note index of last bin in notch
                tog = 0;
```

```
            end
        end
    end

    % Make sure same num of starts than ends
    if length(Fstart) ~= length(Fend)
        Fend(end+1) = mark(end);
    else
        if Fend(end) == 0
            Fend(end) = mark(end);
        end
    end

    % ------------Find notch boundaries of pinna part------------- %

    % get index of notch boundary (closest to 0dB)
    for i = 1:length(Fstart)
        [startMag(i),       indxStart(i)]       =       min(abs(0-dBtp(Fstart(i)-
searchBW:Fstart(i))));
        [endMag(i), indxEnd(i)] = min(abs(0-dBtp(Fend(i):Fend(i)+searchBW)));
    end
    Ns = Fstart - (searchBW - indxStart) -1;
    Ne = Fend + indxEnd;

    % check bounaries don't overlap - if they do, delete overlap
    w = 1;
    while w <= length(Ns)
        if w >= length(Ne)
            break
        else
            if Ne(w) >= Ns(w+1)
                Ns(w+1) = [];
                Ne(w) = [];
            else
                w = w + 1;
            end
        end
    end
    % calculate frequency of boundary bins from  mic inv pinna part
    for n = 1:length(Ns)
        NsFreq(n) = round(((fs/2)/(Np/2))*(Ns(n)));
        NeFreq(n) = round(((fs/2)/(Np/2))*(Ne(n)));
    end


    % read dB level of each notch for analysis
    for i = 1:length(Ns)
        [notchdB(i) notchPos(i)] = min(dBtp(Ns(i):Ne(i)));
        notchFc(i) = round(((fs/2)/(Np/2))*(notchPos(i)));
    end


    for m = 1:length(Ns)
        f = Ns(m):Ne(m);        %loop through all notches (of 'prepped' ir)

        %% Create target and convert to dB
        target_mag = linspace(magtp(f(1)),magtp(f(end)),length(f));
        target = 20*log10(target_mag)';

        original = 20*log10(magtp(f)); % save dB of 'raw' bins in notch
```

```matlab
        g = ratio*(target - original);  % gain coeff for filling in dB

        %Gain coef = ratio * bin values of 'true' pinna notch, this 'inverts'
        %the notch and adds it to the 'raw' data (+ in dB =/ * in mag)

        %% Apply filling to Original 'prepped' IR ('Y')

        L = length(mag_P);
        for i = 1:length(f)
            mag_P(f(i)) = 10^(g(i)/20) * mag_orgP(f(i));
            mag_P(L+2-f(i))= 10^(g(i)/20) * mag_orgP(L+2-f(i));
        end

        phase = angle(P); % phase angle of the original signal

        Z = mag_P.*exp(phase*sqrt(-1));  % complex number from mag and phase

        % get time signal
        Y2 = ifft(Z,'symmetric'); % modified pinna part (size = Nfft)



        %-------Mix parts back together-------%
        % re-window pinna part to remove 0s
        Y2 = chop_ir(Y2, 0.3, p_len, fs);
        % mix pinna decline and tail incline
        pt_xover = p(end-Ncurve:end-1) + tailSwin;
        % put together
        notchfill = vertcat(Y2(1:end-Ncurve-1),pt_xover, tailRect, fadeout);


        clear f
        clear g
        clear original
        clear target

    end

end


    end
 end
```