FURTHER INVESTIGATION OF THE EFFECTS OF SOURCE COLOURATION OF VOCAL VOWELS ON DISCRIMINATION OF FORWARD AND REARWARD MOTION IN VIRTUAL AUDITORY IMAGES

Sally-anne Kellaway SID 4300029204

Research Report, ARCH9031, Semester 2 2014
Masters Degree in Design Science (Audio and Acoustics)
Faculty of Architecture, Design and Planning, The University of Sydney

ABSTRACT

Using individualized HRTFs and a subject pool of 15 subjects, the results of two experiments are compared to further investigate the impact of source spectra on the listener's ability to determine forward and rearward virtual auditory movement. With three additional vowels added to this investigation, the impact of first or second formants will be compared to determine which has the highest impact on HRTF based spectral cues. Stimuli are convolved with individualized HRTFs and played back to the listener at either a frontward or rearward location through headphones. The results of both listening experiments are subject to a Signal Detection Theory analysis to determine the significance of interactions and biases imparted on the HRTF spectra from the source spectra.

1. INTRODUCTION

A significant history of research characterizes the various spectral elements that allow for human localization of sound sources. As discussed in a previous study by the author (Kellaway, S.), results of highly situational studies using narrow band, time restricted stimulus, are designed to describe the elements of the auditory system that facilitate localization. These controlled methods are not specifically designed to contribute to an understanding of more real-world localization studies, where a multitude of additional factors interact with our ability to localize sound. This approach and view is supported by Manor and Martens (2014), who raise that using stimuli with complex spectral details contributes towards the creation of spatial auditory images (regardless of the intention of the study). These spectrally complex stimuli are, regardless of the nature of interaction, more appropriate as a use case for the vast array of binaural technologies than short bursts of white noise by virtue of the opportunity for constructive and destructive interactions between stimuli and HRTF spectra. These interactions are better aligned to replicating a use case than the previously described lab style experiments, but both have their purposes.

There are numerous mathematical models of the human auditory system and localization predictor modules that have been developed and published, which also contribute to an increasing the understanding of the foundational localization ability of humans. However, it was recently raised by Xie (2013) that exclusively physical or mathematical evaluation methods cannot fully reflect actual auditory perception. This ideology is one that is reflected strongly throughout this study, as we seek to deepen the exploration of the interactions between the system that allows for auditory localization, and more "natural" sound stimulus that a normal human listener may happen across in daily listening. This is achieved by extending the range of vowel sounds that were investigated in the previous study.

Xie expands on his statement in his 2013 text by discussing at great length the ways by which researcher can design experiment and analysis methods to explore their theories. Discussing the importance of subjective assessment of virtual auditory displays, Xie states that evaluating or validating certain characteristics of a virtual auditory display is suited to experiments where subjective comparison tasks are employed. This qualifies the importance of a physical evaluation task of any form of virtual auditory display.

As such, we are able to explore questions that would garner a range of highly subjective responses, such as our major research question - how do we account for changes in accuracy of localization when the spectra of the source manipulates our HRTFs? Are incremental changes to our research model (such as the addition of further source sounds) something that we are able to predict through models, or is the conduction of the listening experiment, key to characterising the interaction between coloured source spectra and the accuracy of localization.

There is previous history of research that demonstrates the influence of manipulated spectra on localization; Macpherson and Middlebrooks (2003) present a correlation between the range in which there is a maximum impact on accuracy of localization and the range in which low-frequency tone modulation is perceptible by humans. Best, Carlile, Jin, and van Schaik (2005) explore the impact of manipulation or exclusion of high frequency content in speech sounds on perception of course direction.

This study continues in the direction of testing Blauert's (1969) hypothesis of timbre differences impacting on localisation ability, attempting to elaborating further on Blauert's characterization of the pinna's spectral cues by varying the source spectra and HRTF spectra in the experiment condition. The spectra of the source can be thought of as a filter, in the same way that the pinna imparts a filter on to an incoming sound. This filter from the source imparts a layer of spectral colouration, which can be systematically manipulated to determine the impact of this filter on the listener's localization ability. Blauert's approach is systematic in the manner that source locations are varied between three selected locations throughout the three experiments. This then creates a controlled variation in the timbre differences between each source location, where results of a statistical analysis from a perceptual listening task can be attributed to a source by the employment of systematic variations.

A closely related study by Manor and Martens (2014) also implemented a systematic approach to source spectra manipulation for perceptual identification of source elevation. Employing a systematic approach to the research question, Manor and Martens were able to apply a sophisticated analysis method to draw together several elements of the vocal vowel stimuli to support their conclusions. While this study of frontward-rearward movement bias imparted by vocal vowels uses a different set of methods, it takes a similar approach to researching and characterising bias.

In the previous, introductory study - which set out to investigate the impact of source spectra from three vowels on individual's front-to-back movement localization ability - Kellaway and Martens (2013) determined that there were significant spectral interactions between vocal vowel sounds and individual's recorded HRTFs. Within the context of the previous study, and this extension study, the systematic manipulation is established by the addition of more source sounds - spectrally profiling common vowels to maximize elements such as formant similarity or difference, overall tone colour, or theorized impact.

The previous study revealed conclusive evidence for the initial vowel set. When considering only front to rear plane movement in virtual auditory display, the signal detection theory analysis of the 24 participant study revealed that vowel formants had significant ability to colour HRTF spectra and manipulate key spectral cues.

This study is an expansion on the 2013 study, where a similar method is used on a new set of vocal vowel stimuli. This allows for a deeper investigation in to the interaction between source spectra and HRTF spectra. This report will primarily deliver results of an analysis of the contribution from primary or secondary formants. This will be executed by combining data from the previous study and data from this study, with aim to conduct a Signal Detection Theory analysis. This will allow the comparison of hit and false alarm rates observed when vowels with strong primary or secondary formants are played to the listener, and reveals statistical biases that originate from the interplay between source and HRTF spectra.

This approach will be elaborated further, and is informed from a survey of literature in the area surrounding the characterization of perception of auditory source direction.

2. LITERATURE REVIEW

2.1. Context

It is important to note that this study calibrates its context to the point beyond the duplex theory. It is well acknowledged by many, and well tested, documented and discussed (including Kistler and Wightman (1997)) that accurate localization is achieved by the use of monoaural spectral cues. It is recognized that ITD and ILD do have an impact on localization to some extent (particularly away from the 'cone of confusion' regions), but research such as Kistler and Wightman (1997) conclusively demonstrate that if one ear is plugged with an ear plug and covered with a muting cover, localization is still possible. It is raised in this study that some level of localization assistance is delivered through an imperfect neutralization of stimuli arriving at the contralateral ear, and bone conduction.

Regardless of the depth and breadth of research investigating monaural localization, it is necessary to recognize that there are two perspectives when investigating the role of spectra in localisation. One is the investigation of the human element of the system – the pinna system and subsequent interpretation of localization cues from this system. The other perspective is from the source spectra, and the way by which this element can impact and colour on perception and localization.

Understanding directional bias involves changing perspectives from looking at the way in which the auditory system spectrally filters sounds to 'encode' a direction, to 'decoding' what exaggerated versions of these spectral filters tell listeners in terms of direction. The spectral filters of certain sound sources can be described as exaggerated within certain contexts, especially in experiments such as this, where the impact of certain elements of the spectra are investigated not as arbitrary interferences, but as systematic variations. Systematic variations selected for the purpose of this experiment were designed to impact on certain localization cues. These localization cues were determined by surveying many existing studies that

characterize the directional 'qualities' of narrow band stimuli.

2.2. Traditional Approaches to Understanding Localisation

A more traditional perspective on spatial analysis of spectra influence is achieved from the angle of spectral analysis of spatially captured data (by capture of a Head Related Transfer Function (HRTF)). This style of analysis has delivered a wealth of data that correlates to deliver spectral cues that are critical for accurate perception of auditory source direction. Even when considering one singular discrimination – Front to Back – there is wide range of documentation - see, for example, Kistler and Wightman (1997) or Zahorik, Bangayan, Sundareswaran, Wang, and Tam (2006) document the importance of the frequency bands 1kHz for Rearward and 3.5kHz for Frontal sources.

This style of analysis does in fact contribute to this study – specifically denoting the importance of spectral bands for certain localisation tasks. Identifying them as features of a listener's HRTF also raises the question that there might be an impact to the accuracy of localisation if these discrete cues are manipulated. This effect was documented in the preceding study, and will be continually investigated in the current study.

2.3. Spectral-Directional Colouring and Bias

This course of study has been investigated to an introductory level previously by the authors of this study. As stated previously, the preceding investigation of vowel spectra concluded that the spectra of the selected vowels did interact with the participant's individualised HRTFs. As a result of this, the next step in this study is to investigate which (primary or secondary) vocal formats impact on accuracy of localisation, and which have the most severe impacts.

There is significant amount of literature that creates a context for this study. Blauert introduces the theory that the direction of a source sound can be altered simply by altering the spectrum of the signal at the eardrum (1969), and hence altering the spectrum of the signal can be manifested in either minute or immense ways.

Testing the limits of the listener ability to detect these differences, Zakarauskas and Cynader (1993) implement local smoothing of spectra in their computational approach. The computational approach allowed Zakarauskas et al to follow previously established computational exploration methods that are usually targeted at defining functional tasks that allow researchers to move towards understanding perceptual mechanisms. Zakarauskas et al set out to understand the way by which we extract and functionally use HRTF-based cues. They go about discovering the perceptual assumptions listeners make to localize sounds by calculating the mathematical transforms necessary for a listener to arrive a correct localization of a sound source.

The conclusion was reached that listeners assume the HRTF spectra does not vary from their normal pattern when localizing sounds to specific locations. This is relevant to our hypothesis that an unknown colouration of the spectra arriving at the ear may influence inconsistent identification of source direction. By Zakarauskas' findings, it may continue to be the case that listeners report direction inconsistently because there are variances between their HRTFs as they normally perceive them, and their HRTFs as presented during this experiment.

This finding is carried across many studies that investigate the range of simplifications that can occur before spectral cues are too degraded to provide sufficient localization cues. Manor and Martens (2014) acknowledge the assumptions made by a listener in order to localise sounds – all spectral variations are assigned to either the source spectra or HRTF spectra in order to resolve a location. Through a stepwise series of regression analysis, Manor and Martens conclude that localization tasks that rely on spectral features are not dependent on determination of the origin of trial-to-trial spectral variation. This method of analysis and predictive analysis determined that participants did not resolve the origin of trial-to-trial spectral variation in the experiment as either source or HRTF spectra.

Another interesting perspective offered by Zakarauskas et al is that accurate localization is still possible when locally smoothed HRTF data is used during localization tasks. This implies the potential that smaller modulations or micro peaks or notches that disrupt the overall structure of major peaks and notches in HRTFs do not have a major role in localization and do not represent large enough degradation of spectral cues.

A critical proposal from Butler was the potential that some directional areas will facilitate simplifications more readily than others (1990). This can be derived from assessing the correlation between the number of "covert peaks" in a given area and frequency of listener localization to that area. Areas that a reported to have fewer "covert peaks" will not support radical simplifications, as simplifications will likely degrade the magnitudinal profile for that direction.

The Covert Peak Area theory is a highly prominent and influential element to this study. Butler's contributions on Covert Peak Areas (CPAs) were proposed in 1984 and elaborated through a series of studies in 1987, 1988, 1990 and 1992. CPAs are explained by Butler as the place in space where one narrow frequency band (or segment) of the recorded sound spectrum is amplified more than it is from any other spatial location. Butler characterised a "more amplified band" as a 1kHz-wide band having a recorded magnitude near maximum levels (Rogers & Butler, 1992).

These peaks are identified as 'covert' because they are not immediately clear when magnitude is plotted against frequency. Only when the dB gain is plotted against the location of the sound source do these 'covert' peaks appear across the areas measured.

These CPAs are not just measured peaks; they are also detectable and localisable features. Butler and Rogers delve deeply into extracting CPAs for the frontal hemisphere, identifying that the CPAs in this area more adequately account for monaural localization performance of Vertical-Plane positioned sounds (Butler et al., 1990). Measured results indicate that there are an abundance of CPAs to allow for localization across the extremes of the coronal plane. One aspect that raises questions about the CPA theory from Butler is what spectral information is used to localize sounds when there is an absence CPAs in an area. Butler notes a certain range of elevation locations where CPAs are judged to be sporadic, and hence fail to explain accurate localization (by this method alone).

Studies focussing on the identification of spectral-directional colouring and bias for elevated locations commenced with Blauert's first study in 1969. Conducting a series of experiments using sources positioned on the median plane, Blauert identified timbral cues that allowed humans to localise sounds for forward, rearward and elevated locations. Blauert openly denounced theories that stated that bone conduction and visual/tactile and vestibular cues were key for median plane localization, and used these first experiments to open a new perspective for localisation studies. Blauert's experiments aimed to draw more conclusions in to how timbre differences related to head and pinna filtering impacted or aided sound localization – prior to this, there was extremely limited understanding of the role of spectra in localisation.

Blauert proposed that with a fixed head position, the dominating frequency of the source sound has a significant impact on localization. Three experiments were designed and executed to characterise and explore the phenomenon. Blauert was able to characterise both "directional" and "boosted" bands by playing narrow-based noise to participants. It was then concluded that "directional bands" in the binaural signals enabled the discrimination of the direction of Frontward, Rearward and Elevated (above ear-level) sources across the total pool of participants. Directional bands are defined by an extremely heightened level of localization of narrow-band noise to the Frontward, Rearward and Elevated areas.

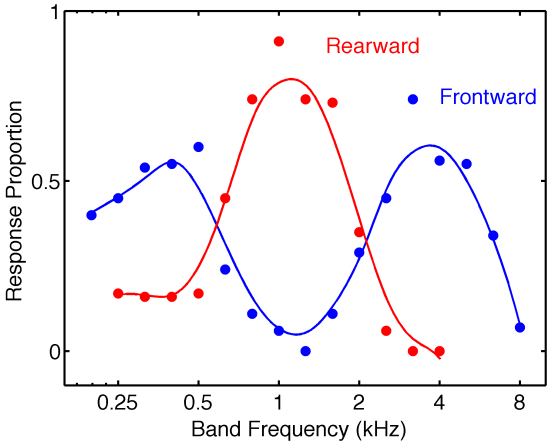


Figure 1 - Blauert's Directional Bands of observer response proportions are charted against frequency to reveal where narrow band noise is localized to. The dominant directions are highlighted as bands. Blue = Front, Red = Rear (Blauert, 1969)

While there are results presented include elevation, Blauert focuses on Front to Rear differences in his 1969 study. A more detailed characterisation on the horizontal planes can be delivered by delving in to more recent studies, such as Martens' 1987 study on spatial energy.

Martens (1987) developed a series of detailed spatial energy plots that demonstrate the inherent spatial qualities of various frequency bands, and used the same measurements to calculate localisation ability as a function of two Principal Components. Taking individual measurements from a range of participants at 36 discrete locations surrounding each listener, Martens filters the data received to generate critical band measurements. Each critical band is plotted separately with normalised magnitude values as a function of azimuth. Most significantly, these measurements are subject to a principal components analysis (PCA) to determine two "rotational weighting" filters that are applicable (to different degrees) to each location on the 360-degree azimuthal plane.

The data from one participant in the study (MDL) was published in detail, as the PCA results from MDL were highly applicable to the other subjects in the study. The First Principal Component (PC1) described the spectral shadowing between the two ears highly adequately, but it is the second Principal Component (PC2), which shows a significant spectral variance between 0 (front) and 180 degrees (rear), as detailed in Figure 2.

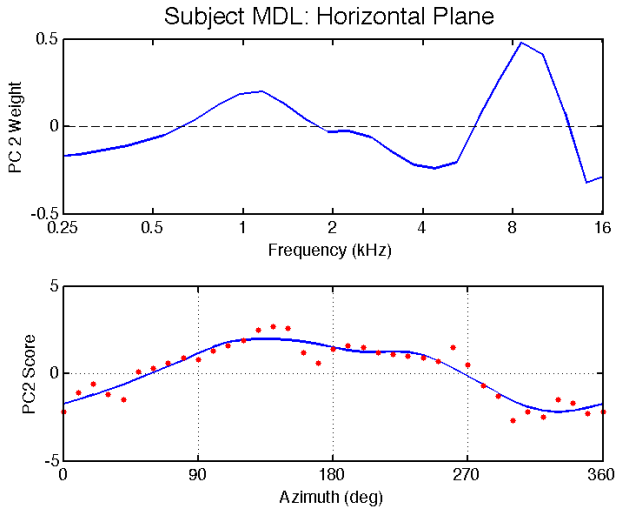


Figure 2 –Rotation Weighting (upper panel) and Score (lower panel) from Martens et al 1987 paper for subject MDL. The rotation weighting is one of the two filters obtained from PCA in the study, and the Score describes the weight of the filter over the 360 degree azimuth range.(Martens, 1987)

The upper panel of Figure 2 shows the PC2 rotational weighting filter that was calculated from the PCA. This weighting filter is combined with the Score from the second panel to arrive at the total weighting of the filter for each location around the 360 azimuthal range captured for the study. To note the most significant details for this study, at frontward locations (0 degrees), the 250Hz and 4kHz bands become positively weighted and correlate with Blauert's directional bands. Rearward cues that were features in Blauert's directional bands (1kHz particularly), are up-weighted by the PC score to feature a significant peak at 1kHz.

Even though the data presented is for one participant, it serves as a demonstration of the magnitude of cue dominance over the lateral directional range. One phenomenon that Martens points out is that when peaks are at their highest energy, the opposing coronal plane spectrum significantly lacks energy, which is evidenced in Blauert's response proportions as well.

For specific studies on spectral-directional bias such as these studies from Blauert, Butler and Martens, it can be considered that these are three presentations of similar datasets. All three are spatially varying visualizations of magnitude as a function of frequency. They all demonstrate clear correlations between each dataset, which describes a pattern of spatial movement across the frequency spectrum.

It is also possible to detect a change in localization ability across the frequency spectrum. When discussing King and Oldfield findings (1997), Letowski and Letowski (2012) discuss the concept of a "center of gravity" in the frequency spectrum of the source. This effect describes a (frequency) point at which the difference in accuracy in localizing high and low frequency sounds is at a peak (maximum difference in accuracy). This centre of gravity theory aligns with previous discussions of spectral-directional bias because the stimulus sounds that are used in the experiment are highly resonant, formant-based vocal vowel sounds. These sounds do have very clear differences in tone colour, and hence each have a strong "centre of gravity" where the majority of the spectral energy is focused (the first formant is the most noticeable peak of magnitudinal energy).

Additionally, the "center of gravity" theory specifically relates to Blauert's "Boosted Bands" findings, which demonstrates the maximum difference across the frequency spectrum and how these bands correlate to the CPAs. This theory aligns with the findings from Zakarauskas and Cynader (1993), where they describe that localization is a function of the centre frequency of the stimuli by virtue of the nature of any specific frequency to be 'suited' to localization to any specific spatial location.

This theory is carried in to Manor and Martens (2014)'s study as the "centroid" element of the analysis in the study of vocal vowel sounds on elevation localisation. Manor and Martens conduct a Spectral Moment Analysis of all stimuli used in their experiment to provide adequate data for a localisation prediction model, for which one element is the calculation of the spectral Centroid. The Centroid element of the Spectral Moment calculation involves the calculation of the mean of the critical band distribution. The other elements (in the calculation) provide more detail surrounding the dispersion surrounding the centroid, asymmetry of the distribution and width of peaks.

This spectral "center of gravity" could also be considered related to Best et al. (2005) research in to the impact of Low-pass filters on the localization of speech stimulus. Using a static low-pass filter point, it was demonstrated that participants with smaller or larger ears demonstrated different levels of performance with the localization task presented. It is possible that a "centre of gravity" for spectral cues may vary for listeners depending on what their range for spectral cues is at the time of the experiment, as size of the pinna does impact on the filter to incoming sound sources.

2.4. High frequency content and speech localization

The link between high frequency content and localization errors has been well established for some time. Hebrank and Wright (1974) establish through experimenting with white noise bursts the impact of both Low-Pass filtering and High-pass filtering that localization accuracy on the median plane. The impact they discover is very strongly dependent on the spectral range of the stimulus and hence what cues are available to the listener. For white noise, an improvement from 50% to near 100% localization accuracy is noted as a Low-pass filter is raised from 8

kHz to 15kHz for all measurements taken in a 240 degree arc around the listener.

Expanding on this research, Best applies a similar experiment to speech by ranking all speech elements used in the study by high frequency content, and then chart polar angle error against the High Frequency ranking. For this experiment, the correlation between increased high frequency content and decreased polar angle error was not noted as being perfectly smooth, however, there was an average decrease to error when there was more high frequency content in the source. It was also noted that the mean polar angle error was significantly lower for stimulus presented with full frequency range (as opposed to low-passed stimuli). This implies that the stimuli presented with more high-frequency content in the 8kHz+ region are better localized in experiment conditions. Best offers the observation that speech is a broadband stimulus, a fact that is overlooked or oversimplified by many pieces of literature. Experiments and analyses that low-pass stimuli to remove or reduce the high frequency content will be limiting the participant ability to accurately localize sounds.

Best et al (2005) test the limit of low-passing stimulus in the second experiment of their 2005 study. Incorporating stimulus that has 20, 40 and 60 dB reduced high frequency information in to a standard localization task allows us to isolate a rough threshold where localization of speech stimulus becomes compromised by low passing. In doing so, they isolate two possibilities that explain the results achieved through the experiment. They propose that the high frequency content of the stimulus is varied and hence the incremental volume low-pass filter will have different levels of impact on stimulus with different levels of high-frequency information. This interpretation is linked to the finding that stimulus that includes low-volume high-frequency information, the auditory system is still able to use this information regardless of volume attenuation because it was not crucial for localization to begin with.

This view is supported by Manor and Martens (2014), when they assert that the impact of a low-pass filter is a pertinent issue whose impact is separate to the consistency of localisation that is dependent on pitch and spectral centroid. These two issues can be tied together by creating a circumstance where significant localisation cues are present in the high frequency reduced, hence the low pass reducing the ability for localisation where pitch and centroid are impacted, but all three colluding together to form heavily degraded localisation cue conditions.

Best's results build upon previous studies that focus on the impact of stimulus volume masking in various environments. For broadband noise stimulus, Best raises the work of Harris (1998), where it was reported that accurate elevation localisation and front-back reversal errors increased when stimuli were presented at lower sound pressure levels. Specifically for speech stimulus, Abouchacra et al (1998) demonstrate that speech in diffuse noise is more easily localized when the signal-to-noise level increased. While these supporting examples do not feature manipulations restricted to the same bandwidths as Best's study (or even this study), they demonstrate that localization is dependent on level of the signal and the integrity of the spectral cues. There are further correlations that can be drawn to more traditional localization studies – a simple visual inspection of Blauert's Directional Bands reveals that there are discrete directional cues that exist in the 8kHz+ range. This simple observation would imply that restricting or exclusion of this high-frequency content would indeed impact on the localization ability of participants.

2.5. Systematic Variation vs Introduced Interference

In the same way that specific, or systematic alterations of the spectra can be observed to have an impact on localization ability, random spectral interference has been studied within the context of localization as well. Kistler and Wightman (1997) investigate the impact a range of factors within the monaural localization paradigm, with the results of the second experiment within that particular study of particular interest to this study.

Kistler and Wightmann conduct an experiment where each critical band was subject to randomized scrambling of magnitude at a 40dB range in both monaural and binaural listening conditions. Spectrum scrambling in this presentation can be described as an insertion of interference to the normal spectra received by the listener. The major impacts of the spectrum scrambling that were noted from the experiment were a significant increase in front-back confusion and a strongly diminished ability to detect elevation. It was noted that the results were hugely varied across all six of the participants in the experiment.

It was noted that in monaural conditions, up-down and left-right localization was impacted less (as opposed an increased impact within the binaural condition). This suggests the importance of the monaural spectral cues by outlining the difficulty imposed on localization by two scrambled signals delivered to either ear in the binaural condition. In this condition, conflicts between the levels at each frequency band would have made binaural localization more challenging during the experiment.

An interesting distinction raised within the Kistler and Wightmann study was that while spectral uncertainty negatively influences localization ability, 'normal' listening conditions with 'normal' sounds do not present as constant, consistent spectra. Any number of variations can occur with the source or environment that may create interferences with the spectra arriving at the ear. Hence results of any conditional experiment with random interference cannot be generalizable to all listening conditions. This introduces the role of experimentation with systematic variations in spectra within localization

tasks. Systematic variations can help to characterize targeted elements that impact on localization, and by their nature may be more salient to replicated experimentation across many experiment condition (through the virtue of reproducibility).

2.6. Experimental conditions and their impact on perceptual localization tasks

For perceptual tasks such as the experiment undertaken for this study, experiment conditions must be considered, designed and controlled very carefully to suit the contents to be collected and evaluated. This concept is a major point of discussion throughout Xie's (2013) text on the design methods for perceptual evaluation of virtual auditory displays. This draws into consideration several other texts that discuss the perceptual impacts of various elements of experiment design.

Source familiarity

Particularly for speech stimulus, familiarity with the source spectrum may lead to poorer performance in localization tasks. This effect was discussed in Best et al. (2005) and was raised as part of their considerations for their selected method. For particular tasks where a change in vocal formant or tone colour of the voice was of more importance, an approach similar to Manor and Martens (2014) becomes necessary, where an analysis of the source spectra involves analysing the four elements of the Spectral Moment – Centroid, Bandwidth, Skewness and Kurtosis. For our purposes (testing of the impact of the colouration from the selected vocal vowels), isolating the point of variation to the vowels sounds required a single voice to be used. A single voice allows the participant to become more familiar with the voice and identify differences between each of the stimulus vowels.

3. METHOD

3.1. Listeners

Audio and Acoustics students from the University of Sydney comprised the subjects for the two listening experiments. The resulting number of participants for both of the experiments was fifteen. All subjects were self-assessed to have normal hearing, and were included based on willingness and ability to attend all sessions of 30 minutes each. Participant number was fairly restricted as the data was captured using the subject's individualized HRTFs, which were recorded in a session that was run for a previous experiment.

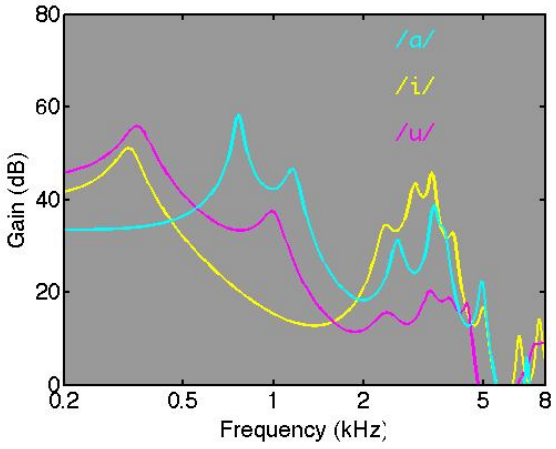
3.2. Stimulus Preparation

Individualised HRTFs were recorded for a total pool of 24 students from the introductory study. HRTFs were recorded for frontward (40°) and rearward (140°) locations at the University of Sydney (UoS) anechoic room. The method for the recordings was equivalent to studies such as Møller et al. (1995) and Farina (2000). Details of the HRTF capture were included in the Previous study (2013).

Ten seatings of the impulse response of the AKG K1000 Earspeakers were also recorded, so that an averaged correction filter could be developed in the Mathworks software MATLAB. Earspeaker Transfer Functions (ETF) were developed at the same time by averaging the ten seatings recorded. The Earspeaker Transfer Function "cancels out" the Earspeakers used in the test, and ensures that colouration from the delivery method does not compound with and influence the listener perception of the source and HRTF spectra.

The speech sounds used as stimulus were American English consonant-vowel (CV) syllables spoken by a male, digitally recorded using an omni-directional microphone placed one metre from the speaker. Each CV began with /h/ and in the original recordings, terminated with one of six vowel sounds. The vowels used across both the 2013 preliminary study and this study were; /ɑ/ as in 'hot', /i/ as in 'heat', /u/ as in 'hoot', /e/ as in 'hate', /æ/ as in 'hat' and /o/ as in 'haute' and each of these vowels were separated from the 'h' and used separately.

For the source stimuli, a set of three mid-vowels were selected from a library of pre-recorded vowel sound vocalisations existing within the University of Sydney. To contrast from the 2013 (Kellaway & Martens) (where "ɑ", "i" and "u" were used), the mid-vowels selected were "o", "ae" and "e". These vowels were then processed with the ETF for the AKG K1000.



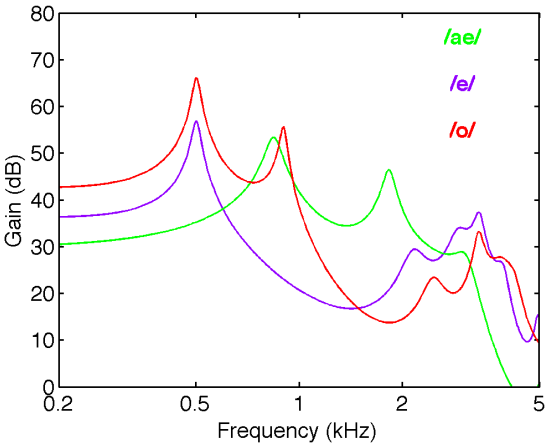


Figure 3 - Spectrum of the six stimuli as presented in the prior and current studies. Plot ranges from 200Hz to 5kHz and show the formants of each vowel as clear peaks in magnitude.

For the purpose of clarity, each vowel spectrum was subject to a long-time-average 50-pole LPC analysis to show the resonant structure of the vowels. This allows the major features (formants) of the vowels to be visually compared without the fine spectral detail that may be a result of periodic glottal modulation.

While these spectral energy plots only report the magnitude ranging between 200Hz and 5kHz, there was measureable energy in the 5-10kHz range for each stimulus sound. This energy was averaged for each stimulus sound to deliver a flat level of energy, delivering the values in the table below.

Study Number	Stimulus	Averaged 5kHz-10kHz Energy
(Study 1 2013)	i	-48.53
	u	-54.49
	ɑ	-62.30
Study 2	e	-56.44
	o	-63.79
	ae	-54.96

Table 1 - Resulting magnitude of energy in the 5-10kHz range used in the experiments.

As previously discussed in the Literature Review section of this study (during the discussion of the Best et al. (2005) research), this high frequency content has a significant role in the localisation of speech sounds. Without this high frequency content, the results of the localisation test would be compromised, as the spectrum the listeners would be unknowingly listening for would be missing.

The AKG K1000 Earspeakers were used again in this experiment. With further discussion and consideration of High Frequency content forming a more significant component of this study, it is important to note that the Earspeakers are a highly acceptable and appropriate choice for accurate speech reproduction. Their frequency response is largely flat (+/- 2dB) in the 200Hz to 5kHz range (and beyond to the 50Hz-10kHz range) that is critical for the vocal formants that we are investigating, and for localisation of vocal stimulus (according to (Best et al., 2005). This consideration, and additionally the removal of the Earspeaker response with the recording and usage of a Transfer Function (ETF), qualifies the continued usage of the AKG K1000s.

To present consistently within the experiment structure, each vowel was processed with both the Frontward and Rearward HRTFs for the AKG K1000 Earspeakers to arrive at two virtual sources for all 3 vowels (a result of 6 possible individual stimuli). A similar GUI was implemented (updated to reflect the new vowels) and ensures that no additional difference was detectable by the participants.

3.3. Procedure

The procedural methods for the listening experiment were equivalent to the previous study conducted by Kellaway and Martens (2013). This means that the same Two Interval, Two Alternative Forced Choice (2I-2AFC) test was conducted with 30 trials over 3 seatings for each participant, with an exploratory session between the last two seatings. This method is equivalent to the method suggested by Xie (2013) for auditory comparison and discrimination tasks.

The two sources (e, ae or o) were convolved with the same Forward or Rearward HRTF for each individual, and the six stimuli were randomised and presented in pairs. As per the first study, the participant selected "forward" or "rearward" to nominate the direction of movement of the source after each pair was presented in an equivalent GUI in the MATLAB software.

First Interval	Second Interval		
	e	ae	o
e	e-e	e-ae	e-o
ae	ae-e	ae-ae	ae-o
o	o-e	o-ae	o-o

Table 2 - Details the nine possible scenarios that occurred during the listening test.

The experiments were executed in the Anechoic Room at the University of Sydney. This room has near-anechoic qualities, which does not impact on the performance on the task because the delivery of stimulus is through the AKG K-1000 Earspeakers, hence any reflections would be of such a low magnitude that they are unlikely to be perceived at a level which may interfere with the task (Xie, 2013).

3.4. Data Analysis

Data analysis was conducted in a similar method to the previous study to ensure comparability to previous results. This included application of a time alignment, head shadowing extraction, and removal of the average participant HRTF response to allow for analysis of direction-varying elements exclusively.

This data was then subjected to the same Signal Detection Theory analysis, the methods for which are covered in both Kellaway and Martens (2013) and Wickens (2001) "Elementary Signal Detection Theory".

4. RESULTS

Results of the first formant study were presented as part of the Kellaway and Martens (2013) study. There will be the presentation of the second formant results, followed by an in-depth SDT analysis of both the first and second formant data to ensure that the impact of the second formant is adequately explored. The method for presenting significant results remains the same – contrasting the ability of the listener to determine consistent front-back movement "hits" against inconsistencies "false alarms" for all trial cases. The spectral interactions between source and HRTF will also be considered, as well as an in-depth statistical SDT analysis.

4.1. Current Study Results

First Interval	Second Interval		
	e	ae	o
e	0.93	0.82	0.69
ae	0.93	0.91	0.93
o	0.98	0.93	0.89

Figure 4 - Hit Rates achieved through the experiment.

For Figure 4, the "Hit" rates reported are equivalent to a 'consistent' response from the first study. A consistent response remains as a response where the listener nominates the movement of the virtual source is forward when the HRTFs presented in sequence are a 'rearward' to a 'forward' HRTF. The opposite presentation of HRTFs will only allow for a consistent response of 'rearward'. Figure 4 contains results from all cases including all Source and HRTF sequences. A presentation where same source was used in both temporal intervals allows us to examine the interaction between HRTF and source spectra. Notice the generally high Hit rates across all cases, with the minor exceptions of e-ae, e-o and o-o.

First Interval	Second Interval		
	e	ae	o
e	0.22	0.09	0.11
ae	0.22	0.07	0.13
o	0.38	0.22	0.16

Figure 5 - False Alarm rates recorded through the experiment.

False Alarm rates reported during the experiment reveal the 'inconsistent' response rates. There proportions are relatively low, with key points of interest in early analysis being the lowest (ae-ae) and highest (o-e) results. Reasons for these interactions between hit and false alarm rates will be revealed with an SDT analysis, which will show the bias and changing listener criterion that impacted on listener performance.

4.2. Analysis of e-ae-o dataset

After collection and collation of the Hit and False Alarm rate proportions has been completed, an SDT analysis of these data allows us to determine the significance of bias from source colouration, and how deeply it impacted on the sensitivity criteria of the listener.

This involves calculating the Sensitivity (d') of the listener to the shifting HRTFs, the Criterion (c) of the listener when listening for Source Direction and the linearised bias (lnβ) due to source.

4.2.1. Sensitivity

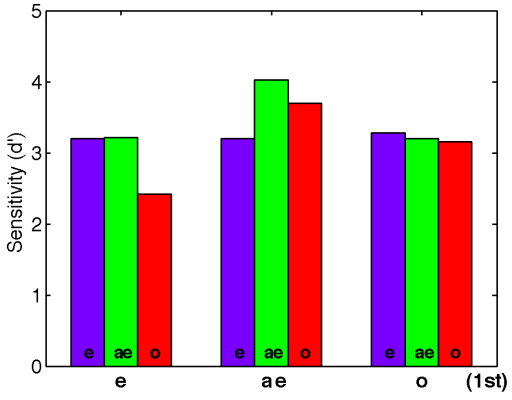


Figure 6 - Sensitivity to Shifting HRTFs. The first internal is listed along the X-axis, with each second interval listed as a function of each first interval.

Figure 6 shows the d' values that were calculated from the experiment. Overall, the sensitivity values achieved were quite high, indicating that the participants could detect the direction of the source regardless of colouration from the source (a result that was reflected in the raw hit rates). The results are also fairly static, which

indicates that the Hit rates and False Alarm rates rose and fell congruently. It must be noted that the overall sensitivity achieved as a result of this experiment, across all cases, was slightly lower than in the previous study. This study achieved an overall d' value of 3.27, whereas the previous study achieved 3.72. The reasons for this could include the lower d' score achieved for the e-o trial combination, which we predict is unlikely to be reproduced if the experiment was to be repeated.

4.2.2. Listener Criterion and Bias Scores

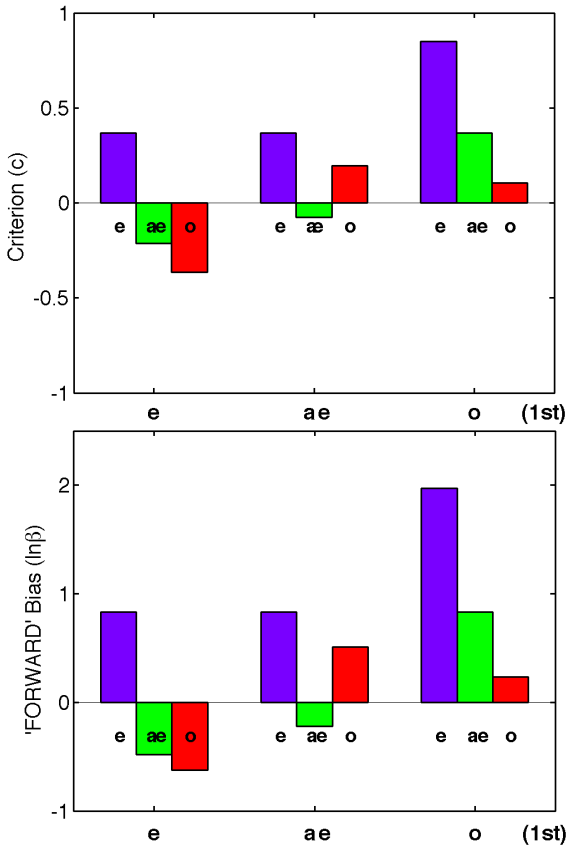


Figure 7 - Changing Sources Changes Criterion. Listener criterion scores are charted per trial combination, where the first interval is listed on the X-axis. A 0 score denotes an unbiased listener criterion (where the listener is using only the HRTFs to determine source direction), a positive score denotes a forward bias and negative a rearward bias. Criterion (c) is shown in the first chart, and the linearized values in the second.

Further analysis allows us to see the manner by which the listener criterion was manipulated by the source spectra in different trial combinations. There are a number of key observations that are worthy of highlighting from this analysis. A near-0 criterion score accompanies the high sensitivity score of the ae-ae case, indicating that the spectra of the ae stimulus may impart a minimal number of interactions with the HRTF spectra.

The e-o trial combination shows the strongest negative criterion bias, followed by the e-ae case. These two cases, in combination with the e-e catch trial demonstrate the strong forward bias imparted on the HRTF spectra by the e stimulus. Cases where e was the second temporal interval stimulus demonstrate extreme forward bias.

In trial combinations where the o stimulus was in the first temporal interval, all following stimuli were perceived as being more forward, regardless of whether the HRTF presentation was forwards or rearwards. The spectral analysis of why this may be the case will be discussed in the following Discussion section.

4.3. Analysis of "ɑ", "i" and "u" vowel set

An in-depth SDT analysis was not presented for the "ɑ", "i" and "u" vowel set conducted during the Kellaway and Martens (2013) study. This presents the immediate opportunity to conduct this analysis on the First Formant Study to directly compare the impact of source colouration on listener sensitivity to forward and rearward virtual auditory source movement.

4.3.1. Sensitivity

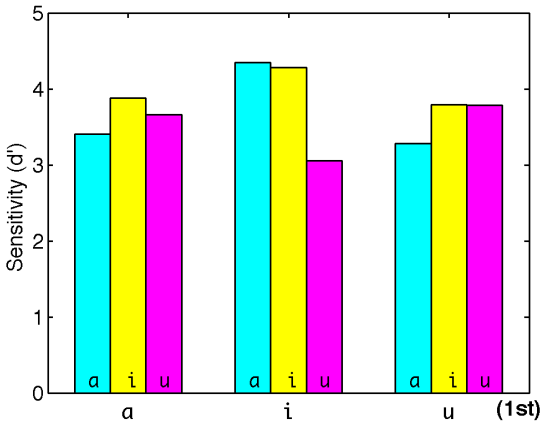


Figure 8 - d' Sensitivity scores for the earlier study on the ɑ-i-u vowel set.

As stated previously, the previous demonstrated a higher overall sensitivity to auditory movement, and generally higher d' sensitivity scores. There is less deviation in this set of results, which follow the higher proportion of direction-consistent "hit" responses that were achieved in the first study. Regardless, the deviation between the highest and lowest d' scores from this dataset is larger than in the e-ae-o study, indicating that the task of consistently identifying auditory source movement was more difficult in some tasks due to the interactions between source and HRTF spectra (i-u, u-ɑ). It is worthwhile noting the high sensitivity and consistency scores achieved in the i-i catch trial case. Interestingly, the continuing level of consistency deviates from other studies where it is suggested that 'bright' and

'narrow' bandwidth sources are harder to localise (Manor & Martens, 2014). This effect may only be relevant to elevation localisation tasks.

4.3.2. *Listener Criterion and Bias Due to Source*

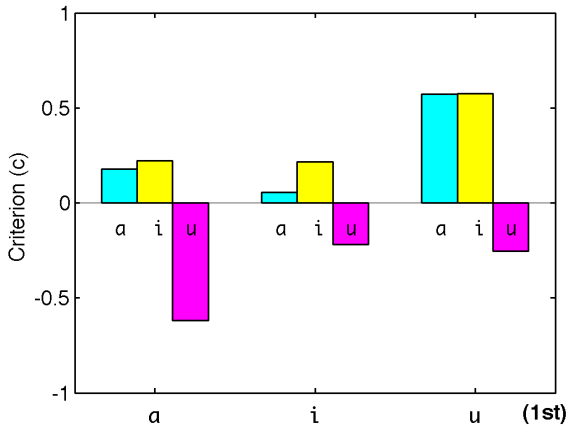


Figure 9 - Linearized bias values for the a-i-u source colouration study. Listener criterion scores are charted per trial combination, where the first interval is listed on the X-axis, as per Figure 7.

Listener Bias analysis from the previous study unveils a stronger rearward bias in some trial cases. Comparatively, less trial combinations demonstrated a strong bias forwards or rearwards, but the rearward bias from the a-u trial combination, and the strong forward biases from the u-i and u-a offer trial pairs to compare spectral features and differences. This assists us in understanding whether source colouration is a function of vocal formant, or more purely, spectral colouration only.

5. DISCUSSION

Use of the SDT method within the context of a follow up study allows an opportunity for a deeper, comparative investigation in to the impact of a range of vocal formants on localisation in virtual auditory displays. We can achieve this by assessing the Hit and False Alarm rates to calculate the Criterion and Bias across the datasets. These statistics point us to further analysis of various elements of the spectrum of both the source and the participant HRTFs to attempt to characterise the impact of interactions between source and HRTF spectra.

SDT analysis also allows us to follow up by calculating and comparing the operating characteristics across all participants. The SDT approach negates concerns raised by researchers such as Xie (2013), where his proposal of a Hit Rate and Standard Deviation analysis produced the suggestion that listeners might too easily characterise any audible difference as the stimulus that meets their discrimination criteria.

5.1.1. *Consistency of Listener Performance*

All cases across both the current and previous study experienced a very high level of listener performance. This is evident from the high sensitivity scores from both datasets, additionally featuring a small amount of deviation between highest and lowest sensitivity. These observations can be complimented by an analysis of the Respondent Operating Characteristic, which shows each trial combination as a function of Hit rate and False Alarm rate.

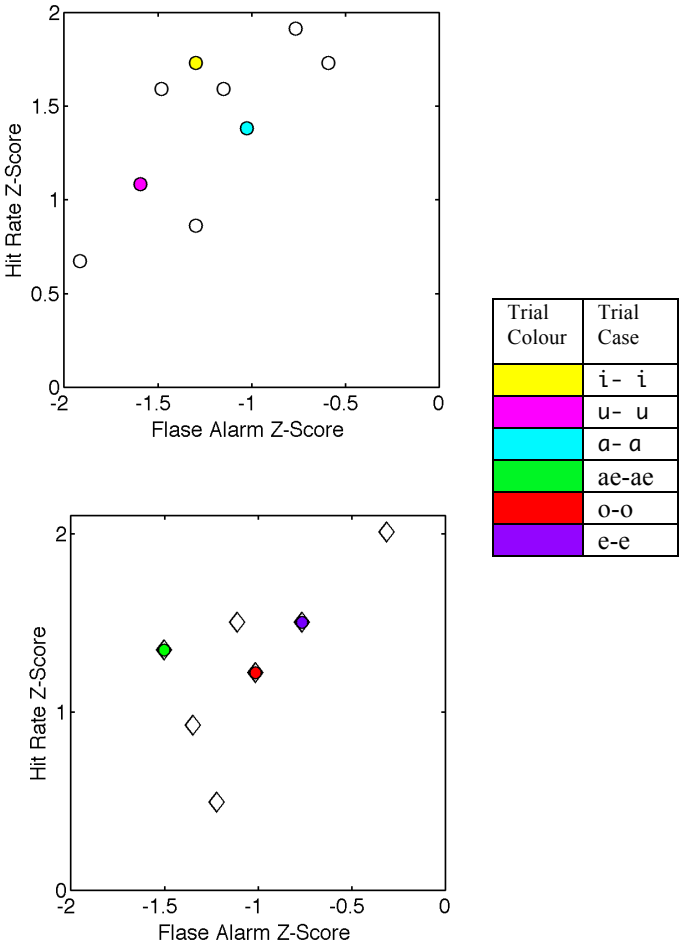


Figure 10 – The upper panel represents the prior a-i-u study, and lower panel represents the e-ae-o currently study. Linearised Respondent Operating Characteristic for all cases, showing the high consistency of responses in the participant pool. Overlaid diamonds and catch-trial cases (as detailed in the Legend, right panel) in the lower panel demonstrate trial cases that achieved the same hit rate/false alarm score ratio.

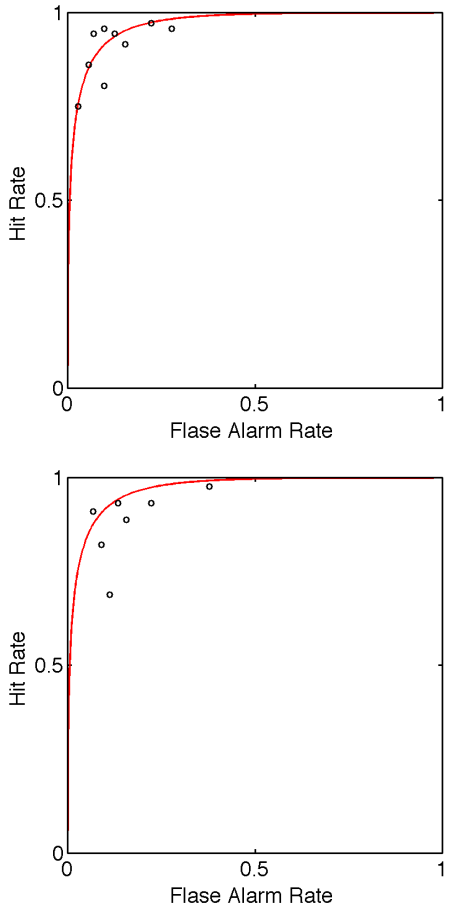


Figure 11 – The upper panel represents the prior a-i-u study, and the lower panel represents the e-ae-o current study. Respondent Operating Characteristic for all cases, showing the high consistency of responses in the participant pool.

Through a Z-score transform, we are able to assess the core cluster of cases from both studies. While there were more outlier cases in this current study of the e-ae-o vowel set, the core cluster of catch trials (same vowel in both temporal intervals eg. e-e) remains with high Hit rates and low False Alarm rates. It is worth noting that the core cluster pictured in Figure 11 (lower) features a more clustered distribution than 11 (upper). This implies that the spectral interactions caused by the second vowel set caused more difficulty for participants in responding to the task accurately.

5.1.2. Analysis of Spectral Features Causing Bias

The literature review section established expansively the amount of research that has been conducted in the investigation of spatial effects associated with spectral bands (or vice versa). Long standing research by Blauert, Butler and Martens et al indicate that there is a correlation between certain frequency bands and consistent localisation to specific spatial areas. Blauert demonstrated that there were a number of specific frequency bands that were associated with forward and rearward localisation, which sets a significant platform for developing the investigation to probe the extent of forward or rearward bias imparted by our stimuli.

We will seek to uncover the origin of these biases (as revealed in the statistical analysis of the experiments), and how evidence of this bias sits within the context of relevant literature. Understanding the origin of the bias will come from comparing the peak gain difference from our most heavily biased trial combinations to the Rear-to-Front magnitude difference evident in the mean HRTFs. These findings will be set within the context of literature by comparing the data collected in this study to results from related studies.

5.1.3. Origins of Spatial Effect and Bias

It is well established that the spectral filter imparted on incoming source sounds is a direction-dependent filter. It is through the difference of one location's HRTF filter from any other that we are able to localize sounds in a 3D field. A simple subtraction of the HRTF for a rearward location and the HRTF for a frontward location makes it possible to view the difference function between the two locations, where charting the result as a function of frequency makes any peaks and notches easy to locate. Deep notches can indicate that there is a frequency difference that functions as an indicator for source direction, which is why comparing the spectrum difference of the most heavily frontward or rearward biased stimuli may reveal to us similar spectral features.

The most significant frontward/rearward bias combination that was noted for the 'e-ae-o' vowel set was noted for the o-e and e-o trial combinations (forward and rearward bias respectively). Comparing the difference of the o and e spectrums to the Rear to Front gain difference from the average HRTF spectrum will reveal that were are common spectral shapes and features. These spectral features present as a bias in the results, due to the listener being unable to separate the spectrum of the source from the spectrum of their own HRTF by virtue of assuming invariance in their HRTF (Zakarauskas & Cynader, 1993).

This style of spectral bias is also noted in other perceptual localisation studies that analyse results on the front-rear axis. Zahorik et al. (2006) note in their study that a wide band from 3-7kHz was a reliable spectral cue for frontward locations, and that after a period of training their listeners with non-individualised HRTFs, their localisation performance improved. It was hypothesised that the magnitude of cues such as the 3-7kHz region was one of the main catalysts for this improvement.

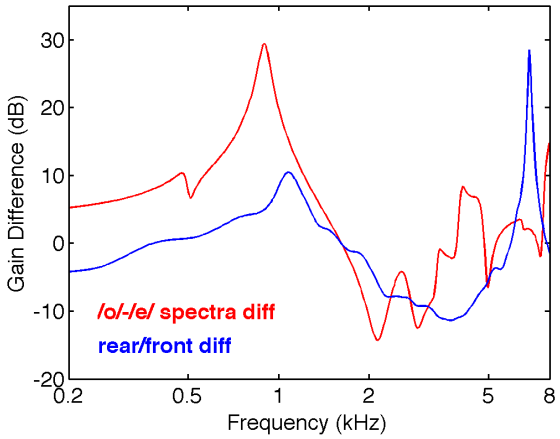


Figure 12 – Plot of gain difference between the 'o' to 'e' vowels spectra and the gain difference between the rear to front mean HRTFs. Note the +30dB peak in difference that aligns with the 1kHz Rear to Front cue.

Figure 12 shows the gain difference for the o-e stimuli with the difference between the rear to front mean HRTFs recorded for the total group of participants. Plotted in this manner, it possible to display a rearward skew with a positive magnitude difference, and frontward skew with a negative magnitude difference. It is clear that these two curves follow largely the same path, with a dramatic +25dB peak in the lower 1kHz rearward cue region, and a smaller -10dB notch pattern in the 2kHz frontal cue region (which is consistent with the Zahorik et al findings).

If we plot the strongest recorded biases from the first vowel set (ɑ, i and u) against the rear to front HRTF difference we observe a similar effect.

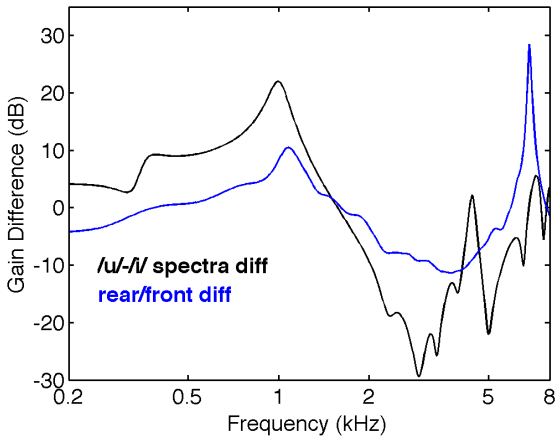


Figure 13 - Plot of gain difference between the 'u' to 'i' vowels spectra and the gain difference between the rear to front mean HRTFs. Note the +20dB peak in difference that aligns with the 1kHz Rear to Front cue.

To analyse a strong rearward bias in a similar manner, we look at the i-u trial combination. This leads us to compare the spectrum of the i and u stimuli. We can immediately note that the u spectrum includes a significant amount of energy at (approximately) 1kHz, which is typical of frontward sources, and the i spectrum includes a huge amount of energy at 3kHz, which is an important frontward cue. This results in a spectral difference plot that shows exaggeration of both rearward 1kHz (+20dB) and frontward 3kHz (-30dB) cues.

What Figures 12 and 13 tell us is that there is a measureable level of narrow-band energy being added by the source spectra. It also tells us that this narrow-band energy belongs to the same frequency range as key HRTF cues for frontward or rearward localisations. These narrow band colourations are then compounded to the HRTF, which creates cancellations and exaggerations of various frontward or rearward cues, resulting in a limitation to the localisation ability of the listener.

These visual analyses are supported by Blauert's Directional Bands theory (Blauert, 1969). The vowels that were noted as heavily skewing direction judgment in both the SDT analyses, and the visual HRTF gain difference analysis also follow Directionl Bands as outlined by Blauert.

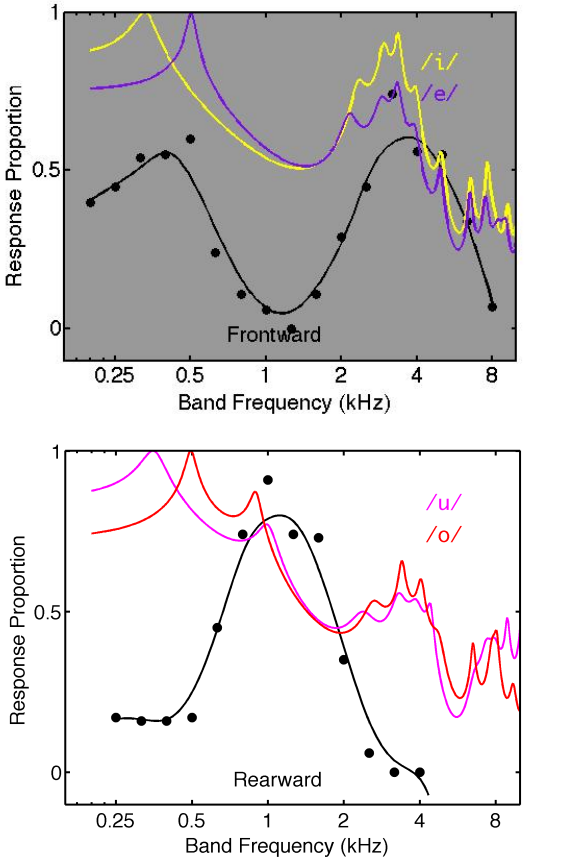


Figure 14 – Blauert's Directional Bands, plotted as Frontward (upper panel) and Rearward (lower panel) alongside relevant normalized vowel spectra. Observer response proportions are plotted against frequency to reveal where narrow band noise is localized. (Blauert, 1969)

It is clear to note the correlation between the normalized vowel spectra and response proportions as captured by Blauert. There is a stronger correlation between the overall shape of the Frontward CPAs and Frontward-biasing vowels. These vowels follow the peak-notch-peak shape that forms between 0.5 and 3.5 kHz, which supplies further reasoning for why these two vowels achieved such high biasing values throughout the two experiments. In this experiment, the /e/ vowel achieved a bias score that was more than double that of i from the first experiment. In this presentation, we can easily see the alignment of the first formant with the 1kHz peak, implying that the spatial effects noted by Blauert were an indicator of spatial biasing or offset in vocal sounds, as well as in white noise.

The comparison of the rearward CPAs against the vowels that demonstrated a rearward bias is less straightforward to the eye. It is possible to consider the proportion of energy that falls within the 800Hz-2kHz region that surrounds the 1kHz Rearward CPA as significant. Impacts of the high levels of energy that lie outside the major 1kHz peak might not be as simple as applying further CPAs associated with Elevation or Frontward locations. The results obtained through this study and analysis clearly indicate a Rearward bias, therefore no other direction is dominant. The significant amount of lower-range energy (sub-500Hz) could indicate that there may be a timbral or tonal indicator that listeners were triggered by in addition to a Rearward spatial bias.

5.1.4. Spectral features of Movement in VADs

Source spectrum introduces a range of colourations, significantly from vocal formants, and these colourations have a significant impact on localization ability, as evidenced in the experiment data collection and analysis. These colourations have a measureable impact on localization ability because some of the colourations interfere with significant spectral cues that humans use to determine source direction and movement. For this study, evaluating the similarity of HRTF cues for the forward/rearward directional plane (specified as being either 40 and 140 degrees in the method), to the source spectral cues is paramount to understanding and attributing the colouration and change in localization ability recorded.

The most salient feature that was observed across all participants was a peak in gain difference at the (Approximate 1kHz) mark. Charting the difference between the Rear and Front HRTFs for all listeners arrives at a relatively small R/F difference as an average, but if individuals are charted, the significance of the cue is maintained. Using a cross-correlation function, we are able to determine the frequency at which most participants' R/F Peak Gain data is most similar and scale-shift each participant to align the data on the log frequency scale.

This result of these operations are shown in Figure 12, where each individual participants frequency shifted and smoothed R/F difference is charted along with the mean (as the dashed black line). It is clear that a +4dB peak in the data collected from all participants demonstrates that there is a 1.1kHz cue that allows for the discrimination of rearwards movement.

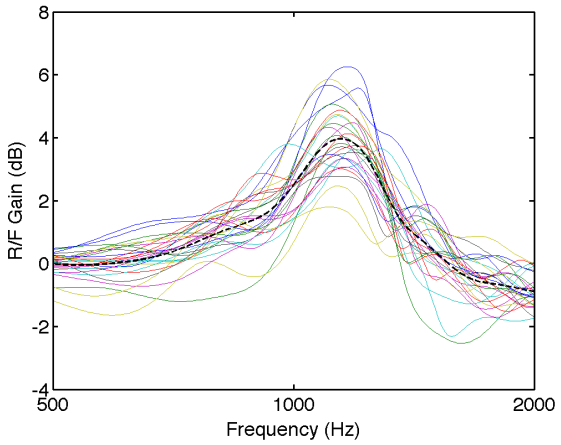


Figure 15 – Peak Gain Frequency chart. Rear/Front gain difference charted between 500Hz and 2kHz after spectral alignment and smoothing across 24 subjects. The mean is shown as the black dotted line.

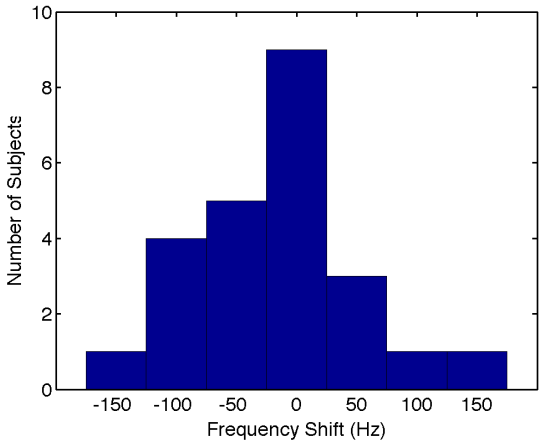


Figure 16 – Chart showing the Peak Gain Frequency shift from Figure 12, demonstrating that none are greater than 150 Hz along the log frequency axis, and many subjects (9/24) needed a shift of less than 25 Hz to align their peak gain with the mean peak near 1100Hz.

The presentation of this cue from our experimental data is not unexpected or surprising, the existence of a cue for rearwards localization was proposed as early as 1969 in Blauert's work (as previously discussed). According to literature discussed in the literature review section, there are several distinct directional effects that are imposed by the use of specific frequencies on human listeners. The vowels that demonstrate this spectral feature are the vowels, which demonstrate the strongest rearward biases,

especially when compared to vowels that have significantly less energy in the 1.1kHz region.

For example, the o, u and ae vowel spectra feature significant energy at 1.1kHz and all demonstrate rearward bias when compared to i, a and e across the two experiments. The analysis of the spectrum from these rearward bias examples were touched on previously, and demonstrate the impact that source colouration can have on source movement identification.

5.1.5. An extension of the Principal Components Evaluation

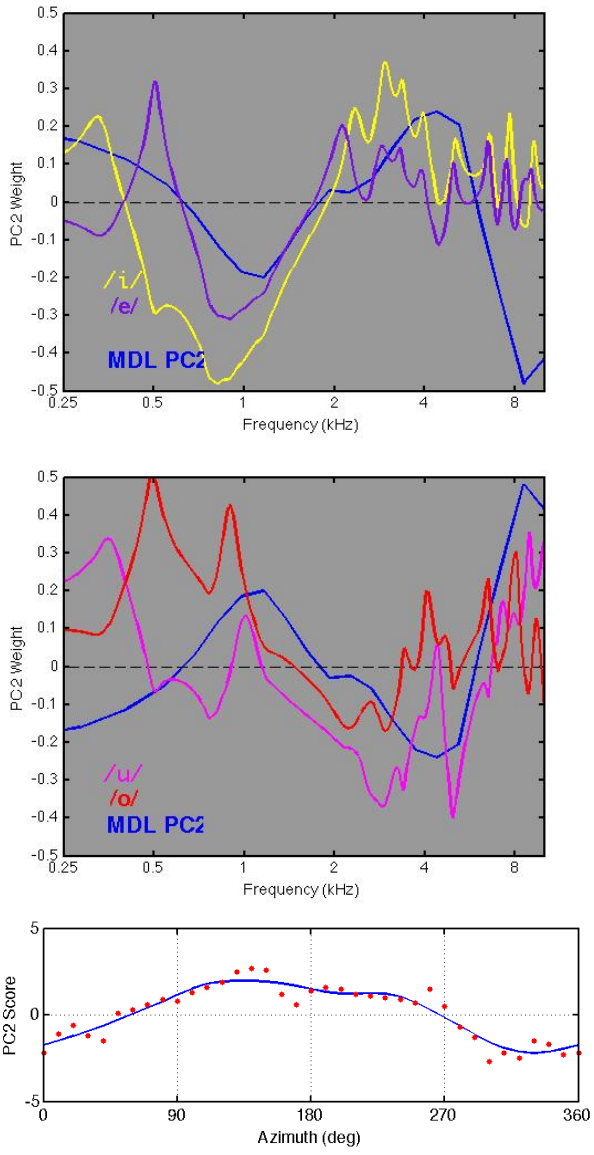


Figure 17 - Martens' Principal Component 2 data for subject MDL plotted with the frontward biasing vowels (Uppper panel) and against the rearward biasing vowels (lower panel) (1987)

We can also compare the principal component that was discussed as being most relevant to this study with the vowels used in both experiments. For these plots, the PC2 is shown for frontward incidence in the upper panel and rearward incidences in the middle panel – simply inverted to demonstrate the difference between frontward and rearward weightings. The lower panel is the original weightings, which can be referred to, to clarify the transition between frontward and rearward weighting functions. The vowels are prepared by subtracting the mean of all 6 vowels to show only varying spectral peaks and notches that deviate from 0dB.

Grouping the vowels in to frontward bias and rearward bias groups, we can see how the first formants of the highest bias vowels in the upper and middle panels align with the peak in the PC2 weighting score. The upper panel shows the alignment of the 2.5-4kHz centered "frontward" directional band (as identified in Blauert (1969)) with the cluster of formants in the frontward biased vowels. The shape of the spectral energy difference for these frontward vowels also follows the general high-low-high shape between 250Hz and 4kHz. This formant-like shape is followed more by the i vowel than the e vowel, which is expected as the i vowel achieved a higher bias score overall.

The rearward vowels shown in the middle panel also follow the generic shape of the PC2 weighting function. In this case, the opposing high-low-high shape of the weighting function is followed quite closely by the vowels (with minor variations). The higher amount of energy in the 250-1000Hz region is due to the first formants of the vowels. With the overall trend of these highly biased vowels to follow the general shape of PC2 weighting function, we can see a correlation between these datasets and further resonance between the other spectral-directional theories (such as Blauert (1969)).

5.1.6. An extension of Magnitude Ripple Analysis

Identifying the spectral features that impact on localization is an important element to understanding the impact of source colouration within this context. Equally as important is identifying the magnitude range that these colourations need to meet to impact on the listener. So far, there has been an implied magnitude range used to determine significant impact (which has been repeatedly supported by bias analysis from the experiments). As per the introductory study, we are able to use research from the Macpherson and Middlebrooks (2003) study in to ripple disturbance in source spectra used in localization tasks to explore whether the vowels used in the current study impart a significant magnitude disturbance.

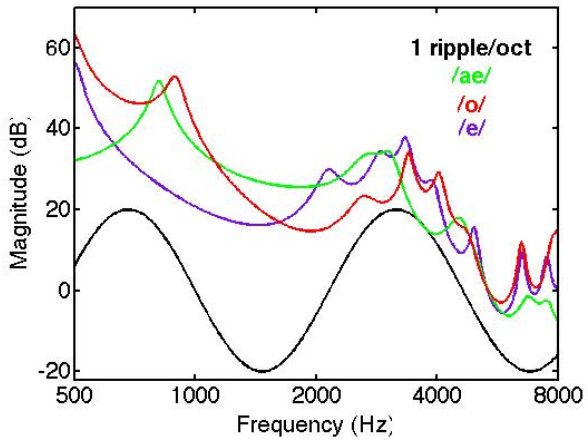


Figure 18 – Source Spectra plotted as before, with the Macpherson et al 1 ripple/oct at 20 dB included. Note that the source spectra notch and peak shapes roughly follow the shape of the ripple, especially between 1kHz and 8kHz.

Macpherson et al discussed at great length that a 'one ripple per octave at a magnitude of 20 dB' filter had the greatest impact on localization ability from their experiments with multiple densities of ripples. Comparing the ripple spectrum stimuli to the previous vowel set demonstrated that the notch and peak shapes of that vowel spectrum roughly followed the ripple shape at one ripple per octave. We can see a similar pattern emerge from this vowel set as well, with the formant spacing between the first and second formants falling in to the 1 ripple per octave pattern. However, the depth of the ripple evident in this set of vowels is only just short of the 40dB range which Macpherson and Middlebrooks identified as the most disruptive. While this minor shortfall was unlikely to have completely discounted a rippling-like effect, it is worthy noting that the bias values that were obtained for this second set of vowels was, overall, lower than the first set of vowels.

It is interesting to note that this effect was present in both sets of vowel spectra, as it not only supports the conclusion that there is a significant bias from source colouration, but also implies that the harmonic structure of vowel formants creates significant disturbance within itself. This implication may lead to further investigation – the boundaries of this effect inspires questions about the impact of time-variant stimuli (more like flowing speech) would be an interesting follow up.

Phase is another element that is investigated by Macpherson et al, and as discussed in the previous study, is an element where modulations are idiosyncratic and hence arrive a huge array of variance in experiment conditions. Macpherson identify that there is a large amount of disparity between results when modulating phase, and hence attribute this disparity to the idiosyncratic nature of Phase-based localization cues. The anthropometric features of the ear, vary significantly from participant to participant, and so vastly impact phase at the eardrum, as well as localization in general.

When considering the reasons for vast individual differences, it is important to note the role these individual differences have on not just Phase, but also high-frequency dependent localization. It has been demonstrated that the stimulus and filters used in this experiment resulted in high-frequency manipulations that correlate with the work of Best et al. (2005). A point raised in their discussion where individual differences were explained by factoring in conchae size, is relevant here too. Individuals with smaller conchae demonstrate their localization cues at higher frequencies (than participants with larger ears). For Best et al, this resulted in demonstrating that participants with smaller ears had spectral cues that were impacted more by low-pass cutoffs. For this study, it is suggested that ear size may have contributed to the results of this study as a function of the range of frequency cues.

5.1.7. Limitations of study

The main limitation of this study is tied to the small number of participants that we could draw in for the second sitting. The participant group that was used was 24 for the first study and for this follow-up, was only 15. The ability to process a greater range of data (from more participants) may have helped to differentiate more of the trial cases through the SDT analysis (recalling that there were 3 pairs of cases that all returned the same hit rate to false alarm Z-score ratio results).

The only other limitation would be the uncontrollable variables involved in the physical experiments. Seating-by-seating variances introduced by different arrangements of hair, seating and Earspeaker placement might have introduced some variances into the results. This effect was mostly negated by the averaging of ten HRTF recordings to arrive a the Earspeaker Transfer Function, however, the ETF could not have accounted for removal/replacement of Earspeakers mid-experiment (creating minor differences that may have required perceptual re-adjustment by the listener), or any changes in hair styling between the first experiment (2013) and the second (for this study).

5.1.8. Applications

The development of simplified filters which may improve localization is something that is discussed by Zahorik et al. (2006) and is largely application to studies such as this study as a-i-u vowel set study. As the analysis of these datasets represents a characterization of spectral cues that assist in evoking forward and rearward movement (possible from identifying bias imparted by the source spectra). These spectral cues could be reasonably adapted for use as spectral filters to assist localization on the sagittal plane for use in mediums where additional

localisation in virtual auditory displays is necessary – such as virtual reality or surround/high immersion video gaming.

6. CONCLUSION

This study has employed a perceptual approach with statistical analysis to probe the concept of source colouration impact on source movement discrimination in virtual auditory displays. Using the Signal Detection Theory method to analyse the recorded data from two listening experiments (including six different vocal vowel stimulus sounds) we were able to identify varying levels of bias across the stimulus set. It was noted that certain vowels (ɑ, ʉ, e, o) demonstrated statistically significant forward or rearward biases. Through a spectral analysis and comparison to other long-standing theories regarding spatial-spectral effects, it was demonstrated that strong formants that exist within vocal vowel sounds have a significant impact on localization ability when these formants align with localization and HRTF cues.

While the methods used in this study (and other related studies) identify significant impacts to localisation ability, it may be the case that we dedicate a significant amount of the analysis to seeking "overt" peaks in the several spectral analyses discussed, these "overt" peaks may only be fully operable in experimental conditions. As Manor and Martens discuss in the introduction to their study, it is entirely possible that seeking for "overt" peaks may actually miss the "covert" peaks that are responsible for localisation in some contexts (2014). This poses the challenge of applying the findings from this study to a more diverse VAD environment.

7. ACKNOWLEDGMENTS

The authors of this study would like to acknowledge the physical assistance of Manuj Yadav for setting up the experiment. We would also like to acknowledge the invaluable ongoing contribution of William Martens for supervision, guidance and MATLAB coding that was invaluable for the construction of the experiment and analysis of the data.

8. REFERENCES

Best, V., Carlile, S., Jin, C., & van Schaik, A. (2005). The role of high frequencies in speech localization. *The Journal of the Acoustical Society of America, 118*(1), 353. doi: 10.1121/1.1926107

Blauert, J. (1969). Sound Localization in the Median Plane (Vol. 22, pp. 205-213). Acustica.

Butler, R. A., Humanski, R. A., & Musicant, A. D. (1990). Binaural and Monaural Localisation of Sound in Two-dimensional Space. *Perception, 19*(2), 16.

Farina, A. (2000, February 19-22). *Simultaneous Measurement of Impulse Response and Distortion with a Swept Sine Technique.* Paper presented at the 108th AES Convention, Paris, France.

Hammershøi, D., Jensen, C., Møller, H., & Sørensen, M. (1995). Transfer Chracteristics of Headphones Measured on Human Ears. *Journal of the Audio Engineering Society, 43*(4), 14.

Hebrank, J., & Wright, D. (1974). Spectral Cues Used in the Localisation of a Sources on the Median Plane. *Journal of the Acoustical Society of America, 56*(6), 1829-1834.

Kellaway, S.-a., & Martens, W. L. (2013). *Effects of Source Colouration on Discrimination of Frontward Versus Rearwards Motion of Virtual Auditory Images*. Thesis. Architecture. University of Sydney. University of Sydney.

Kistler, D. J., & Wightman, F. L. (1997). Monaural Sound Localisation Revisited. *Journal of the Acoustical Society of America, 101*(2), 13.

Letowski, T., & Letowski, S. (2012). Auditory Spatial Perception: Auditory Localization: US Army Research Laboratory.

Macpherson, E. A., & Middlebrooks, J. C. (2003). Vertical-plane sound localization probed with ripple-spectrum noise. *The Journal of the Acoustical Society of America, 114*(1), 430. doi: 10.1121/1.1582174

Manor, E., & Martens, W. L. (2014). *Prediction of VirtualSound Source Elevation Improved by Including Input Source Spectral Bandwidth in the Prediction Equation*. Paper presented at the Internoise, Melbourne, Australia.

Martens, W. L. (1987). Principal Components Analysis and Resynthesis of Spectral Cues to Percieved Direction. *ICMC Proceedings*, 274-281.

Rogers, M. E., & Butler, R. A. (1992). The Linkage Between Stimulus Frequency and Covert Peak Areas as it Related to Monaural Localisation. *Perception & Psychophysics, 52*(2), 10.

Wickens, T. D. (2001). Forced Choice Procedures *Elementary Signal Detection Theory* (pp. 93-112). Los Angeles: Oxford University Press.

Xie, B. (2013). Psychoacoustic Evaluation and Validation of Virtual Auditory Displays (VADs). In N. Xiang (Ed.), *Head-Related Transfer Function and Virtual Auditory Display* (2 ed., Vol. 1, pp. 504). Florida, USA: J. Ross Publishing.

Zahorik, P., Bangayan, P., Sundareswaran, V., Wang, K., & Tam, C. (2006). Perceptual recalibration in human sound localization: Learning to remediate front-back reversals. *The Journal of the Acoustical Society of America, 120*(1), 343. doi: 10.1121/1.2208429

Zakarauskas, P., & Cynader, M. S. (1993). A Computational Theory of Spectral Cue Localisation. *Journal of the Acoustical Society of America, 94*(3), 9.