Experimental investigations of auditory externalization and the application of head-movement information to hearing-aid signal processing

Alan William Boyd

MRC/CSO Institute of Hearing Research – Scottish Section
Centre for excellence in Signal and Image Processing,
Department of Electronic and Electrical Engineering
University of Strathclyde

A thesis submitted for the degree of
Doctor of Philosophy

2014

Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Alan William Boyd *Date*

Thesis Abstract

An externalized sound is one that is perceived outside the head, whereas an internalized sound is perceived inside the head. The effect of hearing aids on these phenomena was investigated through psychoacoustic experiments, a novel questionnaire and offline analyses. The importance of high-frequency pinna cues for externalization in normal-hearing and hearing-impaired listeners was investigated using open-ear simulations of different microphone placements, frequency responses, and the number of talkers. It was found that hearing-impaired listeners experienced a compressed or "flattened" perception of externalization in relation to pinna cues. The role of changes in the stimulus and source direction on the perception of externalization by hearing-impaired listeners with and without their hearing aids was also investigated. An effect of angle but no effect of hearing aids was found for hearing-impaired listeners. The effect of short-term acclimatization to hearing aids by normal-hearing listeners performing the same task was investigated; an effect of acclimatization was found. A questionnaire was developed to determine the prevalence of internalization in several situations. The prevalence of internalization increased with the number of hearing aids worn. The overall prevalence for any experience of internalization was 25% of the sample population. The effect of dynamic-range compression, signal type and listening environment on a potential indicator for internalization, the shape of the interaural-level-difference distribution, was analyzed. The analyses only indicated potential internalization under particular constraints related to listening environment and temporal resolution, not dynamic range compression.

Head movements have been shown to be crucial to the perception of internalization and externalization. Head-movement information can

also provide improvements to hearing-aid signal processing. Instantaneous head-motion information measured from a head-mounted gyroscope has been utilized to control the directionality of hearing-aid microphone arrays, based on user behaviour and assumed listening intent. Simulations show gain improvements over fixed and adaptive arrays during head turns in specific listener-source configurations. Tests show improved localization ability over simulated fixed arrays for the majority of listeners. The head-mounted gyroscope output was used to compensate for head movements during estimates of source direction of arrival (DOA). Estimates of DOA for sound sources were compensated for head movement detected by the gyroscope. This resulted in improved DOA estimates during head movements in comparison to the same system without head movement compensation. An improved head-motion detection system (combined gyroscope, accelerometer and magnetometer) was used to resolve sources into the front or rear hemifield biomimetically. Estimates of source DOA were shifted clockwise or anti-clockwise with head movement and aggregated in two DOA estimate histograms. The histogram with the largest peak robustly determined the hemifield in which the source was located. The acoustic scene has also been stabilized during head movements, resulting in increased timescales for direction of arrival (DOA) estimates to be aggregated and hence more robust DOA detection during head movements. The limits of DOA estimation using two microphones have been extended from a single front or rear hemifield to 360° using head movements biomimetically to resolve front-back confusions by comparing differentially rotated histograms.

Acknowledgements

Thanks to my fiancée Rachel, for sharing the 'eureka' moments, giving me perspective and support whenever I needed it, and above all for listening to me talk *at length* about auditory science for so long.

Thank you to my supervisor Prof. John Soraghan for his years of experience and knowledge. His great ability to offer useful observations at the right time would both spur me to new research areas and calm me down in equal measure.

Thanks to my supervisor Dr. Bill Whitmer. Through many lengthy discussions, I have learned an immeasurable amount about how to approach a new scientific question, both technically, philosophically and grammatically.

The unending support of my parents throughout my education cannot be overstated. More of this thesis is yours than you would ever accept, but it is true nonetheless.

Thanks to everyone at the MRC/CSO Institute of Hearing Research – Scottish Section for being the best friends and colleagues I've ever had the pleasure of working with.

Finally, thanks to the anonymous participants for their time and attention.

Author publications

- Boyd, A.W., Whitmer, W. M., Soraghan, J. J., Akeroyd, M. A. (2012) Auditory externalization in hearing-impaired listeners: The effect of pinna cues and number of talkers. *J. Acoustic Soc. Am.* **131**(3):EL268-EL274
- Boyd, A.W., Whitmer, W. M., Akeroyd, M. A. (2013) Improved estimation of direction of arrival of sound sources for hearing aids using gyroscopic information. *POMA* **19**:030046
- Boyd, A.W., Whitmer, W. M., Brimijoin, W. O., Akeroyd, M. A. (2013) Recording and analysis of head movements, interaural level and time differences in rooms and real-world listening scenarios. *International Symposium on Room Acoustics*
- Boyd, A.W., Whitmer, W. M., Brimijoin, W. O., Akeroyd, M. A. (2013) Using head movement information to increase the accuracy of computational auditory scene analysis in a microphone array. *Proceedings of the Institute of Acoustics* **35**(1):275-282

Nomenclature

ADMA adaptive directional microphone array

AED adaptive eigenvalue decomposition

B-DOA biomimetic direction of arrival

BRIR binaural room impulse response

BTE behind the ear

CIC completely in the canal, a type of hearing aid

DMA directional microphone array

DOA direction of arrival. The (angular) direction of a sound incident on a listener or microphone array

DRR direct-to-reverberant ratio

FFT fast Fourier transform

GCC generalized cross correlation

GCC-PHAT generalized cross correlation with phase transform

GCD gyroscopically controlled directionality

GC-DOA gyroscopically compensated direction of arrival

GCS gyroscopically controlled system

GUI graphical user interface

HEBIR headphone-equalized binaural impulse response

HRTF head-related transfer function

IFFT inverse fast Fourier transform

ILD interaural level difference, the difference in sound level measured at the two ears

IMU inertial measurement unit

ITC in the canal, a type of hearing aid between ITE and CIC in size

ITD interaural time difference, the difference in arrival time of a sound between the two ears

ITE in the ear

jnd just noticeable difference. Also known as a difference limen

LMS least mean squares

MAA	minimum audible angle
MEMS	micro-electromechanical system
MIMO	multiple input multiple output
MuSHRA	multiple stimuli with hidden reference and anchor
n	discrete time index
PIQ	perception of internalization questionnaire
SIMO	single input multiple output
SNR	signal-to-noise ratio
SRP-PHAT	steered-response power with phase transform
SRT	speech reception threshold
SS	swept-sine
STFT	short-time fast Fourier transform
Wiimote	Nintendo Wii remote, primary controller for Nintendo's Wii console, including an infrared optical sensor and accelerometer
γ_2	kurtosis

Table of Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

An externalized sound is perceived by a listener as originating outside the head, while internalized sounds are those that appear inside the head. It is a perceptual continuum that has received little attention in auditory research, especially in hearing-impaired listeners. There is some clinical and survey evidence to suggest that internalization of sound by hearing-impaired listeners becomes more prevalent with the number of hearing aids used (Noble and Gatehouse, 2006). It is equally important to note that when a listener uses a hearing aid, the physical location of the sound source is internal, or at the ear, as the listener is hearing sounds produced by a receiver and emitted from an aperture at or inside the entrance to their ear canal. The observation that sounds are not continuously perceived as internalized when wearing hearing aids is testament to the brain's ability to use other auditory cues to properly externalize a sound.

Listeners' heads often move to communicate non-verbally, to improve talker audibility in noisy environments or to improve source localization, in particular to resolve sound sources into the front or rear hemifields. The

use of this head-movement information to improve short-term hearing-aid signal processing has received little attention, in contrast to long-term head orientation benefit (Lavandier and Culling, 2010; Grant, 2001). Head-movement information has been used to improve the performance of environmental classifiers for hearing-aid program selection using long-term analysis (Tessendorf et al., 2013). This new area of research has become viable due to the recently reduced size and cost of micro-electromechanical systems (MEMS) for accurately detecting movement.

1.1 Motivation

The initial motivation of this project was to determine the effect of hearing aids on the perception of internalization and externalization by hearing-impaired listeners. In addition to the survey evidence (Noble and Gatehouse, 2006), it was also known that hearing-impaired listeners were sensitive to the removal of pinna cues when rating headphone-presented speech as internalized or externalized (Ohl et al., 2009). In addition to restoring audibility and improving speech intelligibility, hearing aids should also preserve the spatial location of sound sources (Edwards, 2007). If sounds are perceived by a user to be internalized, the spatial location has not been preserved. An additional consideration for the research was the prevalence of the perception of internalization by hearing-impaired listeners, in order to determine the size of the problem in the hearing-impaired population. The types of sounds and listening environments that resulted in the perception of internalization were also important considerations.

The improvement of hearing-aid signal processing using head-movement information comprised the second main motivation of this project. Listeners move their heads, either to add additional signals to their communication with others, to improve audibility or to resolve front-back confusions when localizing sources. Adaptive techniques for noise reduction or sound source direction of arrival (DOA) estimation often perform less well in noisy environments, as auditory information becomes less robust. Head movement has previously been shown to improve environmental classifiers for hearing aids (Tessendorf et al., 2013). However, no research has been performed on the use of head movement information on short timescales. As head movement information is robust to acoustic noise, it could be utilized in hearing-aid signal processing to improve (automatic) microphone directionality selection and DOA estimation.

1.2 Novel contributions

1.2.1 The effect of pinna cues and number of talkers on the perception of externalization

A headphone-presentation method was used to investigate the effects of hearing-aid microphone position (in the ear and behind the ear), stimulus bandwidth (broadband and < 6.5 kHz) and number of sources (1 and 4) on the perception of externalization. It was found that hearing-impaired listeners did not fully externalize or internalize sounds and were not as sensitive to the alteration or removal of pinna cues in comparison with normal-hearing listeners. The effects of the hearing aid on pinna cues were not a cause of internalization

in hearing-impaired listeners and they perceived a “flattened” perception of externalization, especially with multiple talkers.

1.2.2 The effect of hearing aids, presentation angle and acclimatization on the perception of externalization

Using loudspeakers arranged from 0° to 90°, listeners were asked to rate the internalization or externalization of loud noise bursts with and without their hearing aids. Stimuli presented from 0° were internalized, while sounds presented from other angles were not. No effect of the hearing aid was observed; listeners gave similar responses with and without their hearing aids. Normal-hearing listeners tested immediately after being fitted with hearing aids and after six hours of acclimatization to the hearing aids showed an increase in externalization after acclimatization. This suggests that acclimatization to the altered spectrum and increased gain provided by a hearing aid is a factor in the perception of internalization.

1.2.3 Perception of Internalization Questionnaire

The prevalence of the perception of internalization by hearing-aid users was unknown, in addition to the situations and types of sounds that were most likely to cause it. A short questionnaire was developed as a supplement to the spatial part of the Speech, Spatial and Qualities of Hearing questionnaire (SSQ) to research the perception of internalization by hearing-impaired listeners further. The overall prevalence of internalization was 25% to 29.5% of the surveyed population. Prevalence increased with the number of hearing aids worn. Unilaterally aided listeners who experienced internalization rated themselves

as significantly less able to spatially localize using auditory signals. Impulsive noises were more frequently internalized than other sounds for unaided and unilaterally aided listeners.

1.2.4 Effect of dynamic-range compression on interaural level difference (ILD) distributions

The effect of hearing-aid compression on the distribution of ILDs was investigated, as a narrower distribution of ILDs results in the perception of sounds as closer to the listener (Catic et al., 2013). It was found that compression does not narrow the distribution of ILDs. Anechoic conditions produced narrower distributions than reverberant and a lower sensitivity to changes in ILD also produced a narrow distribution. Changes in the kurtosis (peakedness) of the ILD distributions were similar for variations in sentence used and source DOA, suggesting that large changes in kurtosis are required to produce the perceptual effects seen in Catic et al. (2013).

1.2.5 Gyroscopic control of directionality (GCD)

The output of a head-mounted, micro-electromechanical system (MEMS) gyroscope during head movements was investigated. The output of the gyroscope was used to control the directionality of the microphone output of a simulated microphone array in a novel system, assuming a stationary target location. Selecting microphone directionality based on head movement provided a gain advantage over a standard adaptive or fixed directional system. In listener localization tests, the system was found to reduce the occurrence of

reversals (where listeners turn in the opposite direction from the desired source) for the majority of listeners, in comparison to a fixed-directional output.

1.2.6 Gyroscopically compensated direction of arrival (GC-DOA) estimation

The head-mounted gyroscope output was used to compensate for head movements during estimates of source direction of arrival (DOA) in a novel system. Estimates of DOA for physically stationary sources were shifted against the head movement detected by the gyroscope. This resulted in improved DOA estimates during head movements in comparison to the same system without head movement compensation.

The gyroscope output was used to compensate for head movements during estimates of source direction of arrival (DOA) in a novel system. The system measured instantaneous head-movement velocity. Estimates of DOA for physically stationary sources were shifted against the head movement detected by the gyroscope. Using short-time cross-correlation based localization techniques with gyroscopic input produced robust in situ DOA estimates for several sources in reverberant environments, making use of the time-domain sparseness of speech sources. In addition, the gyroscope allowed the null of an adaptive beamformer to be steered to a noise source, compensating for head movements on a short timescale during DOA estimation.

1.2.7 Biomimetic direction of arrival (B-DOA) estimation

A 9-axis system (MEMS gyroscope combined with accelerometer and magnetometer) was used to biomimetically determine whether a source was

in the front or rear hemifield relative to the listener, using the apparent motion of physically stationary sources during head movements. Estimates of DOA for a physically stationary source were shifted clockwise or anti-clockwise with head movement and the DOA estimate histogram with the largest peak determined the hemifield in which the source was located.

1.3 Thesis layout

The remainder of the thesis is organized as follows. Chapter 2 presents the previous auditory research related to the internalization/externalization continuum, including research on auditory localization, distance perception and other auditory spatial cues that can affect the perception of externalization.

Chapter 3 presents all the signal processing techniques used in hearing aids that are relevant to the internalization/externalization continuum and some of the mathematical derivations of the algorithms used in the head-movement utilizing systems.

Chapter 4 describes the design and results of the novel psychoacoustic experiments on the internalization/externalization continuum and discusses the results.

Chapter 5 presents the design and results of a novel externalization survey and the analysis of the spatial SSQ results for the same survey population. The chapter also covers the design and results of the ILD distribution analysis and discusses the novel results in the context of the previous psychoacoustic research.

Chapter 6 describes three designs for the utilization of head movements for improved hearing-aid signal processing, including the novel hardware and software implementations that were developed.

Chapter 7 provides conclusions on the internalization/externalization continuum, including possible research areas for future work. It also concludes the research utilizing head movements for hearing-aid signal processing, including future extensions of the current systems and other areas of hearing-aid signal processing that could benefit from head-movement information.

CHAPTER 2

Review of spatial hearing research

2.1 Introduction

This chapter concerns the key historical and contemporary research published on spatial hearing in normal and hearing-impaired listeners for this study. It will cover the main areas of azimuthal auditory localization (section 2.2), auditory distance perception (section 2.3) and the internalization/externalization continuum (section 2.4). Each section will also address the research carried out using both normal-hearing (listeners with better-ear four frequency averages (BE4FA) < 20 dB HL) and hearing-impaired listeners (listeners with BE4FA > 20 dB HL).

2.2 Binaural hearing and auditory localization

"Binaural hearing" refers to our ability to compare the acoustic signals arriving at each ear and use these comparisons to determine the direction of a source of sound (Akeroyd, 2006). In this section the term "localization" refers to azimuth and nominally to elevation. Radial localization, or auditory distance

perception, will be covered in section 2.3. This section will cover interaural time and level differences (ITDs and ILDs), the head-related transfer function (HRTF), the relationship between head movements and localization and the effect of hearing-impairment and hearing aids on localization. Acclimatization and localization using altered spatial cues will also be discussed.

2.2.1 Interaural time differences

A sound emanating from a position to the left of a listener will arrive at the left ear before the right. The time delay between these arrivals is known as the interaural time difference (ITD); this varies approximately linearly with the angle of the sound (Mills, 1972). A simple and quite accurate model of ITDs uses a sphere for the head. The ears are two points on the surface of the sphere (Woodworth, 1938; Duda and Martens, 1998).

Figure 2.1 shows the two cases for a sound traveling from the source to each ear. On the ipsilateral side (same side as the source), a sound wave travels directly from the source to the ear, whereas on the contralateral side (opposite side from the source), the wave must travel from a point of tangency on the sphere to the contralateral ear. The speed of sound c is assumed to be constant in the air (d_1 in figure 2.1) and on the surface of the sphere (d_2 in figure 2.1). For the ipsilateral ear, $d_1 = \sqrt{r^2 + a^2 - 2ra\cos\theta}$ and for the contralateral ear, $d_1 + d_2 = \sqrt{r^2 - a^2} + a(\theta - \theta_0)$.

The normalized time difference, $\Delta\tau$ between the sound reaching either ear and the sound reaching the centre of the sphere in free field (a sound field without

obstacles affecting the sound) is given by a simple geometric equation,

$$\begin{aligned}\Delta\tau &= \frac{c\Delta t}{2\pi a} \\ &= \begin{cases} \frac{1}{2\pi}(\sqrt{\rho^2 - 2\rho\cos\theta + 1} - 1), \text{if } 0 \leq \theta \leq \theta_0 \\ \frac{1}{2\pi}(\theta - \theta_0 + \sqrt{\rho^2 - 1} - \rho), \text{if } \theta_0 \leq \theta \leq \pi \end{cases}\end{aligned} \tag{2.1}$$

ρ is the normalized distance from the source ($\cos\theta_0 = \frac{a}{r} = \frac{1}{\rho}$).

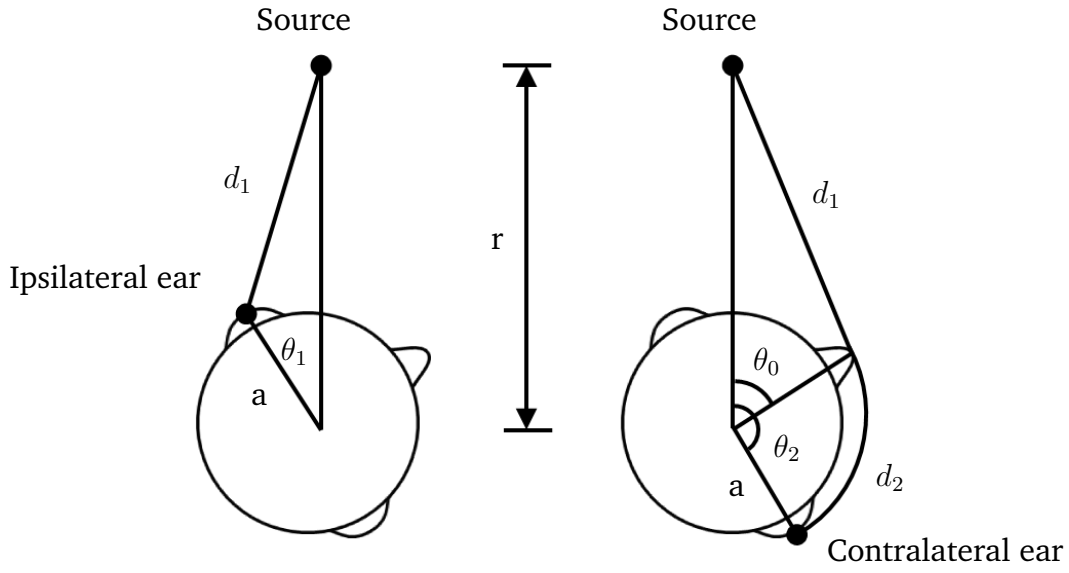

Fig. 2.1: Diagram of the Woodworth/Schlosberg model.

A sound directly in front or behind the listener gives an ITD of $0\mu s$, increasing approximately linearly to around $660\mu s$ for a sound on the interaural axis (line going through the ears to the left or right of the listener), though this maximum value can vary somewhat across individuals (Middlebrooks, 1999).

This geometric model holds for frequencies above 1500 Hz, however below 500 Hz the maximum ITD can be around $800\mu s$. This is due to the physics of how sound diffracts around a rigid sphere (Kuhn, 1977; 1987).

The Jeffress model hypothesized how the detection of these ITDs may occur (Jeffress, 1948). In this model, two arrays of neurons receive input from the ears with varying delays from each ear to individual neurons (dual-delay lines). The neurons fire only if input from both ears arrive at the neuron at same time. Each neuron corresponds to a given delay between inputs from the ears, resulting in the detection of ITDs.

The just noticeable difference (jnd) in the ITD of a pure tone is around $10\mu s$ under optimal conditions (frequencies 500-1000 Hz with a reference value of $0\mu s$) (e.g. Domnitz, 1973; Zwislocki and Feldman, 1956). This corresponds to an angular jnd of 1°. The ITDs of pure tones with frequencies above 1500 Hz cannot be detected by listeners (Klumpp and Eady, 1956). One reason for this may be the reduced accuracy of inner hair cells to phase lock to the structure of a pure tone (Pickles, 1988). More recently, it has been found that the Jeffress dual-delay line model can predict the reduction in ITD detection above 1.5 kHz, if a frequency component is added, creating a two-dimensional representation, the "centriod lateralization display" (Hartmann et al., 2013). Lateralization is the localization of sounds perceived inside the head on the interaural axis, whereas localization refers to sounds localized outside of the head (Plenge, 1974a). The ITDs present in the envelopes of high-frequency pure tones can be detected (Bernstein and Trahiotis, 2002; 2004), often with similar jnds and perceived position as low-frequency tones.

2.2.2 Interaural level differences

In addition to a time delay between the ears, the head also casts an acoustic or head shadow on the far, or contralateral, ear to the position of the sound source, reducing the sound level at this ear relative to the near, or ipsilateral, ear. This difference in sound level is known as the interaural level difference (ILD). Unlike ITDs, ILDs are strongly frequency dependent, increasing with frequency from a maximum of 3 dB at 200 Hz to 21 dB at 10 kHz (Shaw, 1974; Shaw and Vaillancourt, 1985). Jnds for ILDs are for tones are 0.5 - 1 dB (Mills, 1960; Domnitz, 1973) and resolution is approximately unchanged between 200 Hz and 10 kHz apart from a small deterioration at 1 kHz (Grantham, 1984). While in general the ILD increases as a source moves towards the interaural axis, this relationship also varies with frequency, as the maxima angles shift forward and backward of the interaural axis. At distances of less than 0.5 m, ILDs are exaggerated, producing ILDs of 20 dB for a 500 Hz signal (Brungart and Rabinowitz, 1999). Conversely, ITDs are independent of distance to a first approximation (Duda and Martens, 1998). ILDs can be estimated by the effects of diffraction around a rigid sphere (Rayleigh, 1894; Kuhn, 1987; Duda and Martens, 1998). For larger wavelengths, diffraction is large, producing a small shadowing effect (i.e., little attenuation of the sound at the contralateral ear). For shorter wavelengths, diffraction is greatly reduced, producing a large shadowing effect. Variations are caused by reflections from the torso at lower frequencies (0.1-2 kHz, 0-10 dB) and the pinna (the visible part of the ear) of the listener at higher frequencies (2-14 kHz, 0-15 dB) (Begault, 1994). The effects of these reflections will be covered in section 2.2.4.

2.2.3 Combining interaural time and level differences

In reality, listeners experience sounds with both ITDs and ILDs. ILDs are large enough to be accurately utilized above approximately 1.5 kHz, while ITDs cannot be detected (with the exception of envelope ITDs) above approximately 1.5 kHz. This frequency dependent trading of ITDs and ILDs is known as the Duplex theory of localization (Rayleigh, 1907). While directional sensitivity remains at 1° for high and low frequency tones, it is reduced to $3-4^\circ$ around 1500 Hz (Mills, 1972). This is due to both ITD and ILD cues performing poorly around this range. Low frequency ITD cues are dominant for broadband sounds in silence (Banks and Green, 1973; Wightman and Kistler, 1992), while in the presence of noise neither ITDs nor ILDs (reliably) dominate. The most reliable cue depends on the situation (Lorenzi et al., 1999b). When listening to sound sources in rooms, the interaural coherence (similarity of sounds incident at each ear) of a source becomes a useful cue for ITD localization (Rakerd and Hartmann (2010) for a single source). The stronger the coherence, the better a listener's ability to discriminate ITD values. As coherence decreases, ILDs are increasingly favoured by listeners.

Localization acuity varies with both azimuthal angle and elevation. In the azimuthal plane, localization is more accurate in front of the listener than at the left or right side (e.g. Carlile et al., 1997). Errors for listeners turning to face a sound increase from 2° for brief (150 ms) broadband sources in front to 20° for sources presented from more peripheral locations (Makous and Middlebrooks, 1990). One reason for this is the reduced sensitivity of listeners to ILDs as sources are presented closer to the interaural axis (Hafter et al., 1977). Perrott

and Saberi (1990) used a two-alternative, forced-choice, three-up one-down adaptive paradigm as a function of source array orientation, with 50 ms long, 400 Hz click trains as stimuli. Distributing sources in the horizontal plane, mean MAA threshold was $0.97°$ and in elevation this value was $3.65°$, in agreement with Mills (1960). Rotating from the horizontal to the vertical, thresholds only became appreciably different from $1°$ at an angle of $80°$, when threshold increased to $1.8°$. Evidence has been found for neurons that are more sensitive to stimuli presented at $0°$ than at the periphery, known as "midline sensitive" neurons, which gives a neurological basis for the decreased minimum audible angle for stimuli presented closer to $0°$ than the interaural axis (Briley et al., 2012).

2.2.4 Head-related transfer function

ITD and ILD cues can determine angular source location in the horizontal or azimuthal plane and they provide information on the elevation of a source (e.g. de Boer, 1946; 1947; Wallach, 1940). For this, the variations in the acoustic signal caused by the pinna are used. The small folds cause reflections and scattering of the acoustic signal above approximately 4 kHz. The interactions of the pinna, head and torso with a sound source change with source direction, distance and elevation, these interactions are collectively characterized by head-related transfer functions (HRTF), containing all ITDs and ILDs across frequency for that particular source-listener configuration (Middlebrooks and Green, 1991). HRTFs play a role in other aspects of spatial listening, such as distance and externalization, which will be covered in sections 2.3 and 2.4.

Pinna cues help to disambiguate interaural cues that lead to front-back confusions, where a sound presented > 1 m in front of the listener is perceived behind and vice versa. This occurs because the overall interaural differences for a sound presented at the same angle on a circle centred on the interaural axis are the very similar due to the symmetry of the head, the "cone of confusion" (e.g. Wallach, 1938). The cone becomes perceptible as a "tori" of confusion within 2 metres of the listener, as the changes in ILD with distance are exaggerated in the near field (< 1 m; Shinn-Cunningham et al., 1999. The pinna cues can resolve the confusion by breaking this symmetry due to spectral changes, though cues specific to front-back discrimination are not confined to any particular frequency region; this was tested in a study using loudspeakers to mimic headphone presentation (Zhang and Hartmann, 2010). The study found that spectral dips in the directional transfer function (Middlebrooks and Green, 1991), which are the changes in the HRTF with angle, were more important for front-back discrimination than peaks. Pinna cues become important for localization in the absence of head movements and have no effect on localization ability when the head is allowed to move freely (Fischer and Freedman, 1968). This suggests that the overall ITD and ILD cues are more important for localization.

2.2.5 Localization and head movements

Listeners move their heads while communicating (Morency et al., 2005). By rotating tones, music speech or clicks against head movement to varying degrees, head movement was found to provide several cues to auditory localization (Wallach, 1940). Twenty loudspeakers were set up in an arched

array around a listener, whose head was connected to a rotary switch. The ratio between head movement and the switch rotation was variable and the switch connected a signal with one loudspeaker at a time, dependent on head position. Moving the source with head movement, keeping it on axis in front of the listener, produced a percept of the source being above the listener. Moving a source in front as if it was behind the listener caused a perception of the source in the rear during head movement. This often persisted after head movement. The same effect was found for sources in the rear rotated as if in front. Varying the ratio of rotation also produced perceptions of elevation for the sources.

Thurlow et al. (1967) analyzed head movement during source localization. They reported that horizontal head rotations (yaw) were the most common type of head movement when localizing sound. This movement could be combined with a "tipping" movement (nose up or nose down, normally referred to as pitch) or a combination of pivot (more commonly referred to as roll) and tipping with the rotation. Head rotation has been shown to improve localization accuracy during binaural listening (Begault, 1994; Moore, 2003). Rotation in the horizontal plane and free movement of the head during stimulus presentation had the greatest positive effect on accuracy (Thurlow and Runge, 1967; Perrett and Noble, 1997). These rotations may occur to place the sound in front of the listener, in order to minimize the minimum audible angle (Blauert, 1997) and improve accuracy after the head movement. Listeners have been shown to move their heads more the further the sound is initially presented from 0° azimuth (Iwaya et al., 2003). Front-back confusions can also be reduced by head movements, as the binaural cues from a stationary sound

source will move in a predictable way with head movement (this is discussed further in section 2.4.8).

When listeners are instructed to keep their heads still in the absence of auditory stimuli, movement can be up to 5° over several seconds (König and Sussman, 1955). The degree to which listeners move their heads when rating different auditory features has been studied (Kim et al., 2013). It was found that listeners moved their head most when judging source width or envelopment in comparison to source direction or timbre. This was comparable to tests carried out in real-world listening scenarios, such as PC gaming and watching a musical performance. It has also been observed that listeners turn their heads less for broadband white noise than low-pass and high-pass noise (Morikawa et al., 2013).

2.2.6 Localization, hearing impairment and hearing aids

In general, hearing impairment degrades localization performance in comparison to normal-hearing listeners. This reduction can be due to sensorineural or conductive hearing loss. Sensorineural hearing loss is the result of damage to the hair cells within the cochlea or the auditory nerve. A conductive hearing loss is the result of sounds not being able to pass freely to the inner ear, due to blockages or deformities in the middle ear. Unlike sensorineural hearing loss, bone-conducted sound (as opposed to air-conducted sound through the ear canal) is not affected by a conductive hearing loss.

The decrement in localization performance of hearing-impaired listeners is only moderately predictable from sensorineural hearing loss, suggesting that audibility is only one factor in the reduction in performance. A conductive loss

introduces an additional reduction in performance, due in part to the distortion of low-frequency ITD cues due to the higher proportion of low-frequency sound transmitted by bone conduction relative to air conduction (Noble et al., 1994). Low-frequency (< 1.5 kHz) sensorineural hearing loss reduces horizontal localization ability. However, restoring audibility at both ears results in very little loss up to 50 dB HL (Byrne et al., 1992). Vertical localization decreases markedly with a sensorineural hearing loss, as elevation cues are generally in the high frequencies, where hearing loss is often greatest (Byrne and Noble, 1998). If the high frequencies are made audible, the reduced frequency selectivity of listeners with a sensorineural hearing loss mean they cannot make use of the peaks and troughs in the spectrums of high-frequency pinna cues. In contrast, a substantial conductive loss causes a marked decrease in localization ability, due to the lower interaural attenuation for bone-conducted sound in comparison to air-conducted sound (Durlach et al., 1981). This means that ITDs and ILDs at the cochlea are smaller than the corresponding differences at the eardrum (Zurek, 1993). In addition, phase differences between bone and air paths can alter interaural phase differences when the two paths are combined in the cochlea (Dillon, 2001).

Assymetrically impaired listeners have been shown to move their heads to attempt to increase the level of a target source in noisy conditions when listening without hearing aids (Brimijoin et al., 2012). Brimijoin et al. (2010) found that orienting behaviour to auditory stimuli changes with hearing impairment compared to normal-hearing listeners. Hearing-impaired listeners showed a greater complexity of movement, exhibiting rapidly changing velocities, reversals of direction and many fixation corrections, whereas

normal-hearing listeners showed smooth, sigmoidal motion. There were also larger differences in auditory versus visual fixation and greater latency between initial movement and fixation on a target among the hearing-impaired listeners.

There is evidence for adaptation to the increased gain provided by a unilateral hearing-aid fit (Robinson and Gatehouse, 1995). Bilateral fitting is beneficial to localization of broadband noise for moderate to severely impaired listeners (Byrne et al., 1992). However, in the same study, listeners with mild hearing impairment performed similarly whether bilaterally or unilaterally fitted. When listeners with mild to moderate losses are acclimatized to a mainly bilateral fitting, horizontal localization of speech is better when bilaterally aided (two hearing aids) than unilaterally aided (one hearing aid) listeners (Köbler and Rosenhall, 2002). The difference between these studies for mild losses may be due to the difference in acclimatization time to unilateral and bilateral cues.

In normal-hearing listeners, a decrease in signal to noise ratio also decreases localization accuracy and precision. Azimuthal accuracy is less affected by noise than elevation localization and front-back confusions (Good and Gilkey, 1996). The addition of noise produces a greater detrimental effect on hearing-impaired listeners' ability to localize than normal-hearing listeners (Lorenzi et al., 1999a). Performance on a click-train localization task was only slightly poorer than normal-hearing listeners when a white noise was presented from directly in front. However, their performance decreased at higher signal to noise ratios when the noise was presented from $\pm 90^{\circ}$, though the effect was more variable across listeners. The decrease in performance was related to the inaudibility of high frequencies and additional factors of hearing impairment. Similar results were found for hearing-impaired listeners when localizing speech in

multi-talker environments (Best et al., 2011). The reduced audibility of high-frequency cues accounted for some, though not all, of the reduction in performance. Results for localizing targets without noise were similar between hearing-impaired and normal-hearing groups, suggesting that the addition of noise had a greater detrimental effect on hearing-impaired listeners. It has been suggested that the reduced spectrotemporal sensitivity exhibited by hearing-impaired listeners increases "mutual masking" of speech sounds, resulting in the (partial) loss of available spatial cues (Moore, 1985; Dillon, 2001). A priori knowledge of speaker configuration can provide substantial benefit to normal-hearing listeners (G. Kidd et al., 2005; Kopčo, 2010), however it is not known whether hearing-impaired listeners would experience the same benefit.

Hearing aids affect localization cues in a number of ways, depending on their design and how localization performance is measured. Hearing-aid users report improved localization when wearing them (Noble et al., 1995). Under laboratory conditions, listeners with sensorineural hearing losses perform better without their hearing aids when stimuli are presented at clearly audible levels (Markides, 1977; Byrne et al., 1992; Orton and Preves, 1979). The findings both inside and outside the laboratory would suggest that audibility is important when localizing sounds. Vertical localization is particularly degraded when using bilateral BTE-type hearing aids with occluding earmolds. The reasons for this are threefold. Firstly, the microphones are placed outside the pinna, reducing the spectral effects they introduce that are so important for elevation localization. Secondly, the ear-molds attenuate and distort the pinna cues available from the direct acoustic signal. Finally, standard digital hearing aids

operate up to a maximum audio frequency of 8 kHz, though the usable limit is 5 kHz (Füllgrabe et al., 2010), meaning that even if the cues were preserved at the microphone, they would be removed in the hearing-aid output. The reduced pinna cues from BTE microphones also increase front-back confusions (see section 2.2.4) in comparison to microphones positioned inside the ear (found in ITE, ITC and CIC hearing aids) (Orton and Preves, 1979; Westerman and Toepholm, 1985; Turk, 1986). Removing front-back confusions, there is no significant difference between localization with BTE and CIC hearing aids (Byrne and Noble, 1998). The interference between the residual acoustic signal and the delayed hearing-aid signal could cause distortion of the ITD cues in listeners with mild low-frequency hearing loss (Noble et al., 1998). Multiband, adaptive processing algorithms in modern hearing aids could also distort the cues (Byrne and Noble, 1998; den Bogaert et al., 2006). Adaptive directional systems can continuously change the phase relationships between signals at each ear if constantly adapting (Keidser et al., 2006).

The previous review addresses localization by bilaterally aided listeners with symmetric hearing losses. Bilateral amplification has produced more accurate localization results than unilateral amplification in a number of studies (DiCarlo and Brown, 1960; Heyes and Ferris, 1975; Noble and Byrne, 1991). However, these early studies did not control for acclimatization effects so that localization was potentially better with their own fitting, which was generally bilateral. In a later study listeners were tested using their own hearing aid(s), whether bilaterally or unilaterally aided (Byrne et al., 1992). In that study, while the aided results were generally worse than unaided, the bilateral and unilateral results were not significantly different. After a period of acclimatization to

bilateral aids, they have been found to provide benefit to hearing-impaired listeners in comparison to their performance when unilaterally aided (Boymans et al., 2008). The advantage was both objective (speech perception and localization) and subjective (questionnaire responses). A large variation in benefit was found among the users and not all differences between the unilateral and bilateral performance were clinically relevant. In addition, pre-trial diagnostic tests were unable to predict the benefit that a user would obtain from a bilateral fitting.

2.2.7 Acclimatization to altered cues

The use of hearing aids and to some extent the development of a hearing impairment alter the cues available to the auditory system and requires adaptation to them. Listeners have the ability to adapt to altered pinna cues and switch between the original and altered cues, though altering pinna cues has a greater effect on elevation localization than angular (Hofman et al., 1998). By compressing spatial vision, Zwiers et al. (2003) showed that auditory localization had been compressed by a similar amount in azimuth. There have been experiments with "super-normal localization," where interaural differences are enhanced by essentially simulating a larger head (Durlach et al., 1993). These have shown that with training, listeners can adapt to the new cues, though not completely. The adaptation is never better than when using their original cues (Shinn-Cunningham et al., 1998a). Non-linear transformations of cues result in a best-fit linear adaptation (Shinn-Cunningham et al., 1998b). Since interaural cues vary approximately linearly with angle, this finding suggests that localization by the auditory system is inherently a linear process.

The factors discussed above determine mainly angular localization. Listeners localize sounds using a variety of cues. Listeners are able to detect time differences between signals arriving each ear of 10 μs and level differences due to head shadowing (sound diffraction due to the head) effects of 1 dB. They combine these binaural signals with variable weighting depending on the listening situation. Pinna filtering provides additional spectral cues for localization, particularly in the vertical plane, and head movements reduce front-back confusions. Both sensorineural and conductive hearing losses can be detrimental to a listener's localization ability, although increasing the level of stimuli can enable those with a mild to moderate loss to achieve similar performance to normal-hearing listeners in localization tasks. In general hearing aids have a detrimental affect on localization when audibility is accounted for. Accounting for acclimatization, unilaterally aided listeners do not perform significantly different from bilaterally aided listeners. The next section will detail the perception and estimation of radial localization, or auditory distance.

2.3 Auditory distance perception

This section concerns radial localization, or auditory distance perception. It will cover the research performed on the main cues relating to the determining the distance to a sound source. These are the ability to see the source (vision), level and loudness of the source at the listener position, the direct to reverberant energy ratio, binaural and spectral cues and the weighting of these cues in different listening environments. The perception of auditory distance by hearing-impaired listeners will also be discussed.

2.3.1 Vision

A number of cues may be utilized by a listener to determine the distance to a sound source. Using localization cues as discussed in section 2.2, the listener can orient towards the sound source and identify the distance to it visually. A visible target can, under certain conditions, attract the location of the sound perceived by the listener. This phenomenon, known as "the ventriloquism effect," can cause a sound source and visual target to be perceived in the same position for angular separations between them of up to 30° (Jack and Thurlow, 1973). This effect has also been observed for auditory distance estimation (Gardner, 1968; Mershon et al., 1980). Improved accuracy and reduced variability also occur with the use of visual targets (Zahorik, 2001). This may be due to the large difference in distance over which two sounds will be judged spatially coincident with a given a single visual target, especially when both sounds are distant (Zahorik, 2001).

2.3.2 Level and loudness

The absolute level of a sound source is one of the earliest studied cues to auditory distance (Thompson, 1892). Sound level decreases as distance increases by an inverse square law in the free field. A reduction of 6 dB represents a doubling in distance from the source. Early work using level as the primary cue for auditory distance found that a change in source distance of 20% is just noticeable (Gamble, 1909; Edwards, 1955). Lower thresholds of 13% have been found, though this increased to 48% for nearby sources (Simpson and Stanton, 1973) and thresholds for sources moving towards the listener where generally lower than those for sources moving away. Conversely, in an

outdoor environment, thresholds as low as 3% have been reported for sources between 6 and 49 m (Strybel and Perrott, 1984). From 0.5 to 3 m in this study, thresholds again increased, ranging from 9% to 20% . For sources at 1 and 2 m in an anechoic space, a 6% threshold has been reported (Ashmead et al., 1990). Level jnds of 1 dB have been reported (Jesteadt et al., 1977; Miller, 1947; Riesz, 1933). If the inverse square law between distance and level is assumed, this gives a distance jnd of 10%, which fits the research on far-field (> 1 m) distance estimation (Strybel and Perrott, 1984). Additional cues for distance may have been utilized to produce the $< 10\%$ thresholds observed, as listeners could still discriminate changes in distance when the level at the position of listener was held constant (Ashmead et al., 1990). A number of studies have determined the psychometric function for relationship between perceived and physical distance (von Békésy, 1949; Haustein, 1969; Simpson and Stanton, 1973; Cochran et al., 1968). These studies have found that the function is compressed beyond 1 metre, meaning that distance is generally underestimated beyond this distance. Other studies have asked listeners to change the level of a stimuli until its apparent distance has increased by a factor of two (Mershon and King, 1975; Petersen, 1990; Stevens and Guirao, 1962). A change of > 6 dB is required, agreeing with the compressed psychometric functions.

The physical quantity of level is closely related to the perceptual quantity of loudness, though under natural conditions, level may be varied by changing the distance, without necessarily changing perceived loudness (Mohrmann, 1939; Zahorik and Wightmann, 2001). Studies have shown that a doubling in perceived distance and a halving of loudness requires a 6 to 10 dB decrease in level (Warren, 1958). Given the similarity of these results to the inverse square

law, it has been postulated that loudness may be developed from experience with changes in source distance (Warren, 1981). However, using level alone as a cue, perceived distance and physical distance may bear no relationship to one another (Gardner, 1969). Changes in the acoustic power of or distance to an unfamiliar source can be differentiated by listeners, suggesting a more complex process for auditory distance estimation in the auditory periphery (Zahorik and Wightmann, 2001). That is, there are more cues to distance than the fundamental cue of level. For example, the effect of visual feedback, suggested as a possible cue in section 2.3.1, was not addressed in these studies.

2.3.3 Direct to reverberant energy ratio

The direct to reverberant energy ratio (DRR) refers to the ratio between the sound power following a direct path to the listener (the direct energy) and the sound power reaching the listener as delayed copies of the direct sound from reflecting surfaces (the reverberant energy). The DRR cue provides auditory distance information to the listener, as the DRR of a nearby source will be greater than that of a far away one. Reverberation is created when a sound is produced causing a number of echoes to build up due to the sound reflecting from nearby surfaces. The higher-order reflections decay as the sound is absorbed by those surfaces and the air (Lloyd, 1970). A diagram of a reverberant impulse response is shown in figure 2.2. The “early” part of the reverberation is defined as the reflections reaching the listener up to 80 ms after the direct sound, not including the direct sound. The “late” part of the reverberation is defined as the reflections reaching the listener after the early part, up to the point where attenuation causes the reflections to fall below the

noise floor, set at 60 dB below the level of the direct sound if possible (Kinsler et al., 2000).

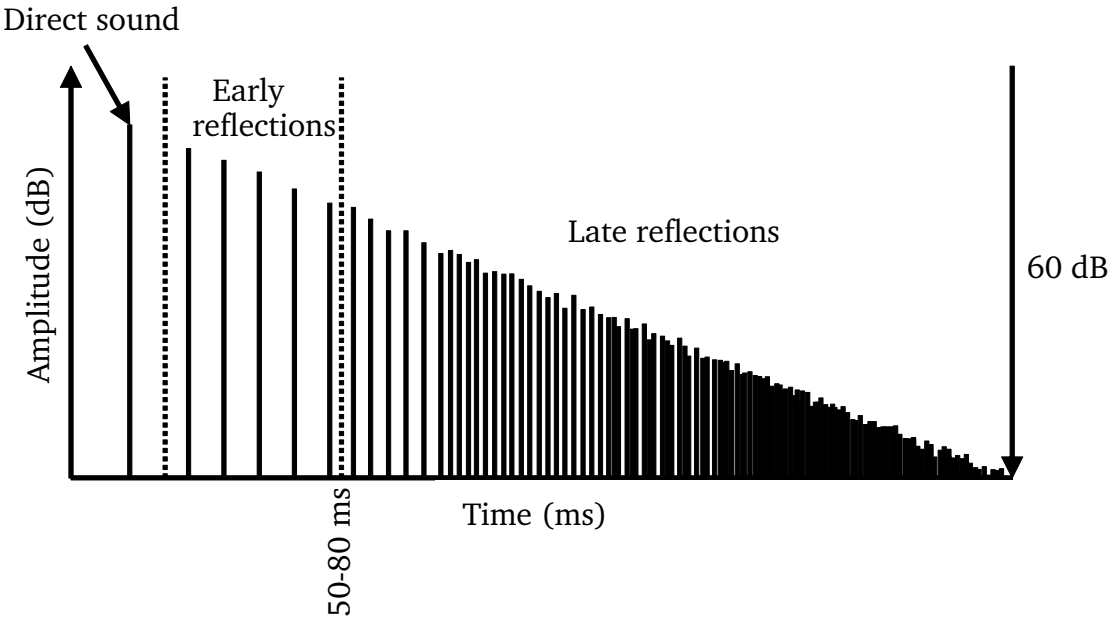

Fig. 2.2: Diagram of reverberant impulse response.

The inverse square law results in increased attenuation of the reflections, due to increased path length to the listener in comparison to the direct sound. As the reflective surfaces are absorptive, attenuation will also occur. The exact nature of the reverberant energy decay depends on the size, shape and acoustic properties of a room and the positions of the source and receiver. Any outdoor environment with partially reflective objects – even trees – will also produce reverberation (Richards and Wiley, 1980).

The reverberant energy of a sound may be approximated by a diffuse sound field, having uniform energy over varying observer positions (Zahorik et al., 2005). There is a particular exception to this. First arriving (often the largest)

reflections are generally floor or ceiling reflections. These reflections do vary with source distance, changing the reverberation level. This effect is removed when the head and torso absorb these reflections, as is the case for sources moving along the interaural axis (Shinn-Cunningham, 2000). Increasing the level of a broadband noise masker has the same effect as increasing the audible DRR, reducing the perceived source distance (Mershon et al., 1989). Increasing the number of reflections in simulated reverberation also decreases the DRR, resulting in the source being perceptually further away.

The earliest work on DRR and distance was performed by von Bekesy (von Békésy, 1938). The DRR was changed by mixing between signals in damped and reverberant environments. Distance perception was most salient in the damped condition, which did not fit the theory at the time. It was concluded that this mixing technique did not capture the distance information in the DRR cue. An updated version of this work has recently been successfully used in externalization research (Ohl et al., 2009; Boyd et al., 2012). Later work has shown that DRR improves the accuracy of apparent source distance judgments in comparison to judgments made in anechoic conditions (Mershon and King, 1975). This improvement holds over a wide range of source distances (Mershon and Bowers, 1979) and acoustic environments (Wettschureck et al., 1973; Nielsen, 1993). Little effect of learning has been found in the use of the DRR cue, whereas initially poor estimates made in anechoic environments improve over time (Mershon and Bowers, 1979). These results suggest that DRR is an absolute distance cue, in comparison to level, which requires relative comparisons of sources at different distances to be useful.

Familiarity with a reverberant listening environment can also affect listener performance. This can improve localization accuracy (in both distance and angle) of nearby sources, generalizing to other source positions in the same environment (Kopčo et al., 2004). DRR has been found to have the greatest impact on listeners' abilities to judge source distance using the level cue where the DRR is 15 to 9 dB, although large differences were found among listeners (Sheeline, 1983). Other research suggests that manipulation of the ratio of early to late reflections alters the perception of distance more effectively than the more traditional DRR (Albrecht and Lokki, 2013). The jnd for DRR corresponds to more than a doubling of the source distance, suggesting that DRR alone is a poor cue for relative distance (Zahorik, 2002a). It has been suggested that the DRR provides absolute distance information for use with other more sensitive, but less reliable distance cues, such as level (Zahorik et al., 2005).

2.3.4 Binaural cues

Auditory distance cues have been considered monaurally in previous sections. Binaural cues are important for angular localization as previously discussed in section 2.2. The use of direct comparisons in changes of ILD and ITD cues for distance estimation has been postulated in a number of papers (Coleman, 1968; Hirsch, 1968; Molino, 1973), as the range of a sound source can theoretically be calculated from the ITD and ILDs between the ears, without prior knowledge of the level of the source or moving ones head. However the psychophysical results vary for far-field sources (distances > 1 m). Some studies suggest that binaural cues are useful for distance estimation of lateral

sources (Gardner, 1969; Holt and Thurlow, 1969), while others show no effect of listener orientation (Cochran et al., 1968; Koehnke et al., 2000).

The utility of binaural cues for determining the distance to near-field sources is less in doubt. ILDs change rapidly for sources within 1 m, whereas there is little change in near-field ITDs (Brungart and Rabinowitz, 1999). For near-field sources, a listener may determine angular location using ITDs and source distance using ILDs. Blindfolded listeners have been shown to be accurate in their near-field distance estimates for sources close to the interaural axis, where ILDs are largest and extremely poor when sources are directly in front, providing evidence for the use of ILDs in near-field distance estimation (Brungart et al., 1999).

The distribution of ILDs over time due to reverberation also has an effect on auditory distance perception. Increasing the kurtosis of ILD distributions for speech stimuli recorded binaurally in a reverberant room ($RT_{60} = 0.5s$) resulted in sounds being perceived closer to or "inside the head" (see section 2.4). Binaural recordings taken at different distances showed a similar change in kurtosis, suggesting that this is another distance cue related to reverberation (Catic et al., 2013).

2.3.5 Spectral cues

The removal of low frequencies up to 3 kHz severely degrades distance estimation performance, whereas a 500 Hz low pass filtered source produces similar results to the broadband source for near-field sources (Brungart, 1999). This introduces the importance of spectral cues in auditory distance perception. For distances greater than 15 m (Blauert, 1997), air absorption significantly

alters the spectrum of the sound. High frequencies are most attenuated, with an approximate loss of 3 to 4 dB per 100 metres at 4 kHz (Ingard, 1953). The proportion of energy reflected and absorbed by the surfaces in a room usually changes with frequency. This will change the frequency spectrum of the reverberant portion of the sound. As distance increases and the DRR changes, the spectrum of the direct and reverberant portions of the sound combined will change, potentially providing a further spectral distance cue. Simply changing the frequency content of a stimulus has been shown to change its perceived distance from the listener. It has been reported that a decrease in high-frequency content can increase perceived auditory distance, when compared with similar sounds with different frequency content (Little et al., 1992). In an earlier study where no direct comparisons were made, greater high-frequency content was perceived as a closer sound for far-field sources, whereas it was reported as further away for a near-field source (Coleman, 1968).

The opposing high-frequency results of the Coleman and Little et al. studies could be partially explained by the accuracy of auditory distance estimation. During distance localization tasks listeners tend towards underestimation of far-away sources and overestimation of those nearby. The pattern of under and overestimation of distance is present across different response procedures, from magnitude estimation and paired comparison (Zahorik, 1997) to actively moving to the perceived location of the sound (Loomis et al., 1998). Moreover, listeners' variability may be between 5 and 25% of the source distance in terms of under or overestimation (Haustein, 1969). For individual listeners, standard deviations can be as large as 60% of the source distance when represented

in logarithmic coordinates (Zahorik, 2002b). A meta-analysis of 84 data sets across 21 articles revealed that this compressed perception of auditory distance can be modeled as a power function of the form $r' = kr^a$, where r' is an estimate of perceived distance, r is the physical source distance, and k and a are the fit parameters (Zahorik et al., 2005). k values were consistent, being slightly greater than 1, while a values varied from 0.15 to 0.7, with an average value of 0.4. The underestimation of distant sources may be indicative of a maximum perceived distance or "auditory horizon", though this has yet to be proven (von Békésy, 1949). It could also provide a safety margin for avoiding objects in the real world (Ghazanfar et al., 2002). The compression of auditory distance suggests a perceptual organization factor known as "specific distance tendency" (Mershon and King, 1975): listeners may tend towards a baseline distance value in the absence of other distance cues. This value can be estimated, using the power law above, as 1.6 m (Zahorik et al., 2005). Research using sources of theoretically indeterminate distance (fixed-level noise bursts in an anechoic chamber) support the result of 1.6 m (Wightman and Kistler, 1997; Zahorik et al., 1994).

2.3.6 Cue weighting

The relative weighting and processing given to each of the auditory distance cues described depends on the type of source, its position (angularly and radially) from the listener and the listening environment (Zahorik, 2002a). In a room, level and DRR are the primary distance cues. Reverberation and level have different effects on perceived loudness and distance for speech and non-speech sounds (Warren, 1973). DRR becomes less important than level for

speech signals than it does for noise (Zahorik, 2002a). Background noise makes speech sound further away (Cabrera and Gilfillan, 2002) and non-speech seem closer (Mershon et al., 1989). These results would suggest prior knowledge of the source or environment characteristics allow the reliability of different cues for each particular source and environment to be determined and the cues weighted accordingly. It also suggests that speech and non-speech signals' distances are perceived differently.

2.3.7 Distance perception and hearing impairment

Self-report data have shown a link between hearing impairment and difficulties in auditory distance perception (Gatehouse and Noble, 2004). Using an image-source method (Allen and Berkeley, 1979) and a circular 24-speaker horizontal loudspeaker array, the distance judgments of hearing-impaired listeners has been compared with normal-hearing listeners (Akeroyd et al., 2007). Hearing-impaired listeners performed as well as normal-hearing listeners when judging distance if the level cue was available. However, when DRR was the only cue available, they performed significantly less well. The same experimental apparatus was used to test the effect of hearing-aid dynamic-range compression on distance perception (Akeroyd, 2010; compression is explained in detail in section 3.3). Compression affects the variation in the level of sounds reaching the ear. No effect of compression on distance perception was found, suggesting that the hearing-aid users had become acclimatized to the altered cues provided, or the compression of the level cues did not affect the relative variations in level, which was the factor under test in this study.

The cues for both angular localization and auditory distance perception combine to allow listeners to place sounds with their most likely sources. When these cues are absent, contradictory or difficult to determine, sounds can appear "inside the head" instead of outside it. The next section describes the continuum between these percepts.

2.4 Internalization and externalization

An externalized sound is one that is perceived to originate outside the head, in the world, whereas an internalized sound appears to originate inside the head (Durlach et al., 1992). The continuum is shown in figure 2.3. This section discusses the research on the perception of internalization and externalization of sound. Early research up to the mid-1960's is initially discussed, followed by Plenge's research, the effect of reverberation and stimulus on externalization and the use of individualized and non-individualized HRTFs. The effect of hearing impairment and hearing aids, head movements and visual cues will also be covered. The neurological research on the perception on internalization and externalization will be discussed. Table 2.1 and 2.2 summarize the important research on internalization and externalization.

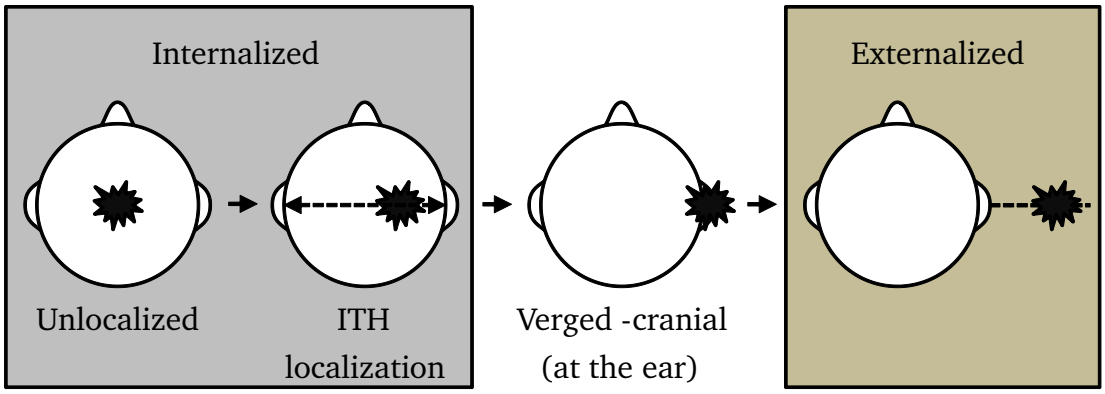

Fig. 2.3: Internalization/externalization continuum.

Author(s)	Research	Results
Laws (1973)	First simulation of loudspeaker listening over headphones to include headphone equalization	Increasing the level of headphone stimuli decreased externalization.
Plenge (1974a)	Externalization of speech by listeners presented with dummy head recordings in anechoic and reverberant rooms	Internalization/ externalization not dependent on electroacoustic transmission type
Wightman and Kistler (1989a)	Production of individualized simulations of a loudspeaker for headphone presentation	Acoustic conditions at the ear when listening to a loudspeaker successfully recreated up to 14 kHz using headphones
Wightman and Kistler (1989b)	Validation of individualized simulation of a loudspeaker for presentation over headphones	Externalization achieved by listeners. Front-back confusions increased
Loomis (1990)	Investigated the effect of head movement	Head movements important for externalization
Durlach et al. (1992)	Overview & discussion of internalization/ externalization research	HRTFs, reverberation, head movements and visual effects important for externalization

Table 2.1: Summary table of important internalization/externalization research up to 1992.

Author(s)	Research	Results
Hartmann (1996)	The perception of changes in the ITD, ILD & IPD changes in individual harmonic components of a synthetic /a/ vowel	Externalization dependent on low-frequency IPDs, ILDs at all frequencies and spectral profile at each ear
Kulkarni and Colburn (1998)	The effect of HRTF "smoothing" (reducing filter coefficients) investigated	Externalization only significantly reduced at min. number of coefficients
Noble and Gatehouse (2006)	One question on internalization/ externalization by hearing-impaired hearing-aid users	Internalization increases with the number of hearing aids used
Begault and Wenzel (2001)	Effects of head tracking, reverberation, and individualized head-related transfer functions directly compared with respect to the degree of externalization by listeners	Inclusion of the first 80 ms of reverberation sufficient for externalization. The effect of small head movements not significant
Ohl et al. (2009)	First experimental research on internalization/ externalization by hearing-impaired listeners	Hearing-impaired listeners are sensitive to externalization, but less so than normal-hearing listeners

Table 2.2: Summary table of important internalization/externalization research after 1992

2.4.1 Early research

Early theories of binaural hearing suggested that gross interaural differences were sufficient to localize a sound, whereas headphone signals containing only these cues could only be lateralized by listeners (Jeffress and Taylor, 1961). Mills found that listeners' jnds were similar for localization and lateralization (Mills, 1960). Therefore, internalization and externalization can be considered a separate and closely related field to both localization (section 2.2) and auditory distance perception (section 2.3). Internally produced sounds such

as tinnitus (Jastreboff, 1990) and one's own voice are correctly localized by the auditory system as being inside the head and will not be mentioned further in this review.

The processes and cues that produced stable, consistent "outside world" visual and auditory images were considered as early as 1848 (Weber, 1848). The first report referring explicitly to sounds appearing inside the head was in 1859 (Purkyne, 1859). It described sounds being heard inside the head if identical or highly similar signals were presented to both ears by means of listening tubes.

Early binaural research made reference to the phenomena of internalized sound, though not by name, with binaural "beats" being "most distinctly heard (and) taking place within the cerebellum" (Thompson, 1877). This internalized percept was produced using telephone receivers placed at each ear. Replacing the receivers with headphones and reversing the input signal phase in one headphone shifted the internalized auditory event towards the rear of the head (Thompson, 1878). Similar research was also carried out by Urbanschitsch (1889). This early work was reviewed by Schaefer (1890) and it was established that an auditory "image will even appear inside the head when each source, considered alone, is heard as being directly before the ear on its side." It is interesting to note that the physical source of sound when listening through a hearing aid *is* directly before the ear.

Early externalization experiments with headphones used microphones placed on a sphere or manikin, with the distance between them being similar to the distance between a person's ears. Feeding the output of one microphone to the left ear and the other output to the right, listeners reported a sensation of

externalization and auditory distance (de Boer and Vermuelen, 1939; Koenig, 1950; Kock, 1950). Initial explanations of inside the head localization listed by Blauert (Blauert, 1997) include; natural resonances of the microphones and headphones (Kietz, 1953); "overmodulation" of the nervous system (Franssen, 1960) and a mental "switch" used by listeners to internalize sound whenever headphones are detected. Schirmer (1966) also considered and ultimately rejected explanations such as; invariability of the ear input signals when the head is moved; different ear drum impedance loading from the free field; mechanical pressure of the headphone against the head; the absence of sound presented to the rest of the body. Schirmer's hypothesis suggested that the transfer characteristics of the electroacoustic transmission channels were the source of inside the head localization. Later work has shown some of the discounted theories (impedance loading and head movements in particular) to have an effect on the perception of externalization.

The effect of the electroacoustic transmission channels as the source of inside the head localization was partially disproved by Reichardt and Haustein (1968). Listeners in their study could not perceive the difference between a sound presented through listening tubes of equal length and the same signal presented through two microphones in the same position as the tube entrances and a pair of high-quality headphones. In both cases, inside the head localization was perceived by listeners. They showed that internalization was caused by the removal of pinna cues. They also found that sound presented from two loudspeakers, each placed close to an ear, could produce internalization. This effect was strongest when the signals at each loudspeaker were similar or highly coherent. There is some evidence to show that sounds presented over one

loudspeaker in an anechoic chamber may be internalized if the loudspeaker is directly in front of the listener (Hanson and Kock, 1957; Schirmer, 1966; Krumbacher, 1969). The signals reaching the ears are also highly coherent (coherence = 1) in this situation due to a lack of reverberation and the similar paths for the sound to the ears.

Bone-conducted sound was also considered as a factor in externalization. In a paper by Sone et al. (1968), listeners heard a tape recording played from a loudspeaker placed in front of them, with the sound relayed to them from microphones placed on either side of their heads through headphones. In the second experiment, only the headphone signal was played, resulting in the removal of the bone-conducted portion of the sound that had come from the loudspeaker. In the first experiment, listeners externalized the sound, whereas without the bone-conducted portion from the loudspeaker, the sound was internalized. From this it was concluded that bone-conducted sound leads to externalization. However, when the bone-conducted portion was substituted by an exciter placed on the subjects' foreheads, the sound was not externalized, in opposition to their previous conclusion.

2.4.2 Research by Plenge

Plenge researched the perception of internalization in several papers (1974b; 1974a). Pulsed noises with 300 Hz bandwidth and 400 Hz centre frequency were presented over two loudspeakers in an anechoic room (Plenge, 1974b). The noises were either in phase or 180° out of phase. Listeners were asked to judge where they heard the stimuli in the room. In-phase sources were perceived to be out of the head and out-of-phase sources as inside the head

or behind the head. The more audible the phase difference between the two loudspeakers, the greater the variation in responses. The same study found that listeners could not perceive a difference between the same signals whether they were produced by loudspeakers in the same room as the listener or played over headphones from recordings made by a dummy head in an identical room. Using concert hall recordings of speech and music made with a dummy head and presented over headphones, listeners were given time to acclimatize to the sound field. After this, an additional signal was presented from a loudspeaker presented two metres from the listener. This signal commonly appeared inside or close to the head. It was postulated that inside the head localization occurs when a subject has little or no information about the sound source and the spatial environment.

A second study researched the differences between lateralization and localization (Plenge, 1974a). Using similar apparatus to the previous study and speech as the stimuli, it was found that internalization of one speaker and externalization of another could occur at the same time, a finding similar to that of Toole (1969), but for different sources, not the same source. A second experiment in a concert hall transformed signals recorded close to a source with a single microphone into “ear-adequate” stimuli that were indistinguishable by listeners from a more distant dummy head recording. This was achieved using artificial reverberation, delays and filters to mimic the directional response of the pinna. This showed that the perception of in the head or out of the head localization did not depend on the type of electroacoustic transmission, similar to the Reichardt and Haustein (1968) findings.

2.4.3 Reverberation and externalization

The effect of reverberation on externalization was first explored by Sakamoto (1976). The type of stimuli used was not stated in the paper. In the first experiment, stimuli were presented from a loudspeaker in a reverberant room at four different direct to reverberant ratios (DRR). It was found that the percentage of outside the head judgments made by listeners increased with decreasing DRR. Sounds presented at 90° were externalized with a lower direct to reverberant ratio than sounds presented on axis. A second experiment mixed direct and reverberant stimuli artificially, using dummy heads placed in anechoic and reverberant rooms ($RT_{60} = 0.1, 0.25s$). Similar results to the first experiment were found. Shorter reverberation times required a lower DRR to produce the same percentage of externalization ratings. A third experiment used electronic delays to simulate reverberation. Externalization was achieved in this condition, even for monophonic sources. A later study by Begault (1992) showed a similar effect of DRR for externalization over headphones, using non-individualized HRTFs, speech stimuli and synthetic reverberation. The study found that inclusion of the first 80 ms of early reflections was sufficient to increase the externalization of the sound.

2.4.4 Stimulus type and externalization

The degree of in the head localization experienced has been found to depend on the bandwidth of a signal and its source position (Toole, 1969). White noise was presented over a visually hidden, four-loudspeaker arrangement at 0° (front), 90° (right), 180° (rear) and 270° (left) with radius 2 metres. The source was presented in five spatial conditions: rear loudspeaker only; front

loudspeaker only; left and right loudspeakers simultaneously; front and rear loudspeakers simultaneously; all four loudspeakers simultaneously. The noise source was also octave filtered at 0.25, 0.5, 1, 2, 4 and 8 kHz and "wideband" bandpass filtered (75-10 kHz). All sources were turned on and off with 100 ms rise/decay times.

Inside the head localization was reported in all conditions, with the strongest internalization in the "all loudspeakers" condition and the weakest in the "rear only" condition. During small (within a 5-10 centimetre diameter circle) translational movements, internalization was occasionally lost. It was restored after the movement had ceased. After the movement, inside the head localization perception was often increased. This could be related to the effect of envelope on localization as it has been shown that sound sources are more difficult to localize and therefore possibly more readily internalized if they have no audible change in envelope (Rakerd and Hartmann, 1986).

The effect of frequency was variable among listeners. Some reported that low-frequency sounds were perceived to be too large to be contained within the head. This is the only study where both internalized and externalized percepts of one stimuli have been simultaneously experienced by listeners. A second experiment mounted loudspeakers on the head. In the same configuration as the loudspeakers, the perception of internalization was found to be similar using the transducers. Finally a comparison of internalization ratings with or without the head clamped was made using speech, music and noise signals. High inter-subject variability was observed and the effect of head-clamping was found to be small and inconsistent. The effect of bandwidth was also variable

between subjects, while responses often did not vary between types of signal used.

Krumbacher (1969) also researched signal distortions and the location of headphone-presented signals in the head. It was found that symmetric configurations of transducers on either side of the head with identical or $180°$ out of phase signals produced inside the head localization, similar to that of Toole (1969).

In a study by Levy and Butler (1978), the relationship between ITDs, ILDs, pinna cues and stimulus frequency content was explored in relation to externalization over headphones. 30 ms noise bursts with 10 ms onset/offsets were presented from a loudspeaker 1.5 m to the right ($270°$) of a listener and the sound recorded with microphones placed at the entrance of the listener's ear canal. The noise bursts were broadband (up to 9.5 kHz due to the response of the microphones), high-pass (2, 4 and 6 kHz cutoff frequencies) and low-pass (0.5, 1 and 2 kHz) filtered. The same recording was used for all participants (i.e., the recordings were not individualized) and presented over headphones. The recorded stimuli were presented in three conditions: with ITDs, ILDs and pinna cues; without ILDs; ITDs only. Listeners were asked to indicate on a visual scale from loudspeaker to ear where they heard the sound. No significant effect of ILD or pinna cue and ILD removal was found. High-frequency stimuli were heard significantly closer to the head than low frequency stimuli, with broadband responses placed perceptually between the high and low frequency stimuli. In a second experiment, the effect of arrival time ITDs and ongoing, fine structure ITDs was investigated, using 4 kHz high-pass, 1 kHz low-pass and broadband filtered noise stimuli. A significant effect of frequency content

was found, similar to the first experiment and the removal of arrival time ITDs (fine structure ITDs only) resulted in sounds appearing closer to the head across all frequency conditions.

2.4.5 Individualized and non-individualized HRTFs and externalization

Inside the head localization was shown to principally depend on the content of the ear input signals alone by Laws (1973). The study used probe microphones in the ears of 12 subjects to measure binaural signals of white noise being presented from a loudspeaker 3 metres from the listener and over headphones. The inputs to the loudspeaker and headphones were identical. Difference curves between the spectra of the two presentation modes were plotted and an electrical circuit was built to reproduce the average of curves. By placing this into the headphone circuit, Laws created the first headphone presentation of a simulated loudspeaker that at least partially compensated for the response of the headphones. It was found that, without the difference circuit, white noise signals were localized inside the head, regardless of changes in loudness. Using the difference circuit to simulate the loudspeaker, the signal was localized outside of the head and moved closer to the head with increasing level. Stimuli were not placed as far from the head as the original loudspeaker, possibly due to the non-individualized nature of the simulation circuit.

Wightman and Kistler (1989a) developed a method of extracting head-related impulse responses (HRIR), the time domain representation of the HRTF, and "pre-equalizing" them for playback. Pre-equalization involves constructing a filter that is the inverse of the headphone-to-ear transform, effectively

removing the spectral colouration caused by headphone playback. Using in-ear microphones, virtual stimuli were directly compared with the real stimuli and were shown to have equivalent spectra up to 14 kHz. Psychophysical validation of the technique reported similar localization results in the horizontal plane and good externalization (Wightman and Kistler, 1989b), using broadband, 250 ms noise bursts. However, more front-back confusions were reported and elevated positions were less well defined. Variability in HRTF characterization has also been reported by Kulkarni and Colburn (2000).

Hartmann (1996) used a subtly different method to study externalization. In their study, the signal used was a synthesized vowel, /a/, using 38 harmonic amplitudes taken from Klatt (1980). Using a synthetic vowel has two main advantages over noise bursts, as they have fewer and more predictable frequency components. Firstly, it made it easier for listeners to detect very small differences between the synthetic vowel presented over loudspeakers and over headphones. Secondly, it allowed complete control over the stimulus content, as individual harmonics could be changed in order to produce a virtual signal that was as close to the real signal as possible, using an iterative "baseline synthesis" technique. Listener responses were recorded using the scale shown in table 2.3. Control of individual harmonics also enabled the relative contributions of interaural phase and level differences over different frequencies to be measured. It was found that externalization depended on the interaural phase differences of low-frequency ($<$1 kHz) harmonics, but not at high frequencies. ILDs were equally important across frequencies. It was also found that a realistic spectral profile at each ear was required for externalization. Maintaining only the difference in spectra between the ears

and not the spectrums themselves was not sufficient for externalization. Small phase changes in the ITD with frequency did not play a role in externalization, as a single ITD was sufficient for externalization.

Response	Perception
0	The source is in my head
1	The source is not well externalized. It is at my ear, or in my skull, or very diffuse
2	The source is externalized but it is diffuse or else in the wrong place
3	The source is externalized, compact, and located in the right direction and at the right distance

Table 2.3: Hartmann and Wittenberg's (1996) 4-point scale.

Other studies involving externalization used non-individualized HRTFs (i.e mannequin recordings). In a study by Begault and Wenzel (1993), anechoic, non-individualized HRTFs were convolved with speech signals and listeners were asked to plot on a top-down view of the virtual auditory space with their head at the centre of it, where they heard the stimuli angularly and radially. The radial scale was plotted in inches, with 4 inches being the verged-cranial (on the skull) position. Results were variable across listeners. One possibility is that the closer one's own HRTFs are to the HRTF used, the more externalized a sound will be. Sounds presented from directly in front or directly behind the listener were internalized 50% of the time, which matches earlier results from Laws (1973). At other positions, internalized sounds were heard 15 to 46% of the time by listeners.

Kulkarni and Colburn (1998) researched the role of the fine detail in the spectral response of the HRTF. Listeners' own individual HRTFs were measured

from 350 to 15000 Hz. A "smoothing" factor was then applied, reducing the Fourier coefficients used to describe the HRTF spectrum from 256 to 8. Listeners were then asked to localize sounds presented over headphones using the "smoothed" HRTFs. Though localization performance was degraded only for the lowest (eight coefficient) condition, the sound was reported as externalized by all listeners for all conditions. This would suggest that the fine detail in the HRTF is not necessary to produce a fully externalized effect.

Kim and Choi (2010) used a Wiener filter approach to optimize externalized headphone presentations for KEMAR and individualized HRTFs. The technique used to obtain the HRTFs was similar to Wightman and Kistler (1989a). Headphone equalization was performed by building an optimal Weiner filter model of the real headphone to in-ear microphone system. The degree of externalization was measured for all angles across all combinations of unequalized and headphone equalized, KEMAR and individual HRTFs. The strongest externalization was found with the equalized responses and the individualized responses produced the most stable externalization. Across all conditions, internalization occurred at 0°, suggesting that other cues must be included (such as reverberation or head movement) to successfully externalize headphone stimuli presented with 0 dB ILD and 0 μs ITD.

2.4.6 ILD distribution and externalization

The effect of ILD distributions on auditory depth perception and externalization has recently been investigated (Catic et al., 2013). ILD distributions have previously been used in research on modeling human detection of interaural incoherence using binaural models (Goupell and Hartmann, 2007). Speech

was presented at different distances in a room and recorded using a dummy head. The ILD distributions were analyzed and it was found that closer sources (higher DRR) produce narrower distributions. The ILD distributions of speech convolved with impulse responses from the same room were artificially narrowed using the system shown in figure 2.4. This system was based on frequency analysis and resynthesis using a complex gammatone filterbank (Hohmann, 2002).

ILD modification is integrated into the frequency analysis and resynthesis. In each frequency channel, left and right envelopes were converted to the logarithmic domain and the instantaneous ILDs, $ILD_i(n)$ (n is the discrete time index) calculated by subtracting the left and right envelopes. Subtracting the mean ILDs, $\overline{ILD_i(n)}$, the ILDs can be compressed,

$$ILD_{c,i}(n) = \alpha(ILD_i(n) - \overline{ILD_i(n)}) \tag{2.2}$$

where α is the compression parameter. An average envelope, $env_{i,avg}(n)$ was derived from the left and right envelopes,

$$env_{i,avg}(n) = \frac{env_{i,L}(n) + env_{i,R}(n)}{2} \tag{2.3}$$

The modified left and right envelopes, $env_{i,L,mod}(n)$ and $env_{i,R,mod}(n)$ were produced by adding the compressed ($ILD_{c,i}(n)$) and mean ILD values to the

average envelope,

$$\begin{aligned} env_{i,L,mod}(n) &= env_{i,avg}(n) + \frac{ILD_{i,c} + \overline{ILD_i(n)}}{2} \\ env_{i,R,mod}(n) &= env_{i,avg}(n) + \frac{ILD_{i,c} + \overline{ILD_i(n)}}{2} \end{aligned} \tag{2.4}$$

Using the modified envelopes, a gain g_i was calculated and applied to the real part of the filter outputs to obtain the compressed ILD. For example,

$$y_{i,L}(n) = g_{i,L}(n) Re(g_{E,i,L} \times w_{i,L}(n)) \tag{2.5}$$

The gain was lowpass filtered at 500 Hz to remove the audible artifacts of the compression and because it is unlikely the auditory system can detect faster fluctuations than this.

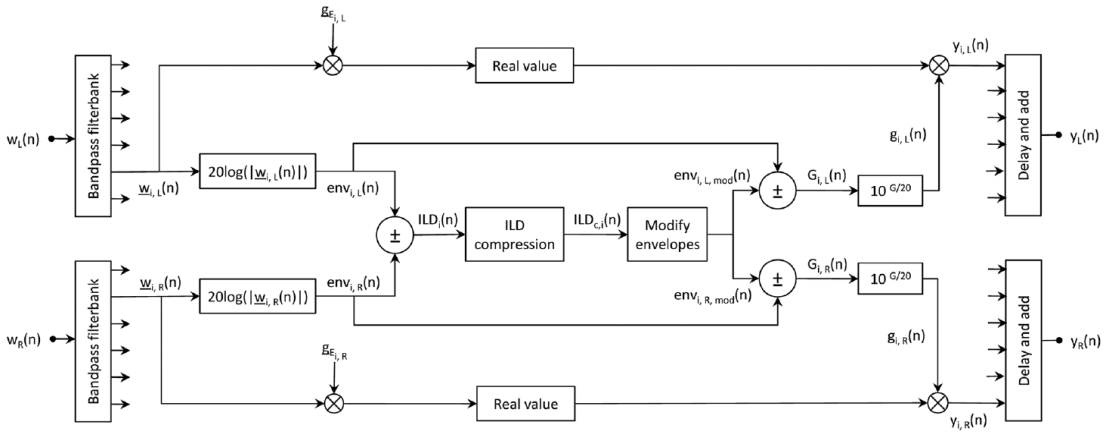

Fig. 2.4: ILD distribution compression system, reproduced from Catic et al. (2013).

It was found that narrowing the distribution of ILDs caused the headphone presented speech to be perceived as closer to and in some cases, inside the

head. This result agreed with the effect of moving a source closer on recorded ILD distributions.

2.4.7 Externalization, hearing impairment and hearing aids

The speech, spatial and quality of hearing questionnaire (SSQ) includes one question on externalization, "Do the sounds of things you are able to hear seem to be inside your head rather than out there in the world?" (Noble and Gatehouse, 2006) Responses to this question could range on an 11-point scale from 0 ("Inside my head") to 10 ("Out there"). It was recorded that hearing-aid users experienced internalized sounds more often than hearing-impaired listeners without hearing aids. Bilaterally aided listeners experienced more internalized sounds than unilaterally aided. This survey represented the only evidence that hearing-aid users experienced internalization of sound before the current research was undertaken.

Recently, a factor analysis of the SSQ has been undertaken (Akeroyd et al., 2013). Three clear factors were identified, corresponding to the three main sections of the SSQ: speech understanding, spatial perception, and clarity, separation and identification. Very little of the variance in the responses to the internalization question could be accounted for by these factors. A question on externalization was also uncorrelated with other questions in a study (Wiggins and Seeber, 2012) of dynamic-range compression with normal-hearing listeners.

One previous study has researched externalization in hearing-impaired listeners. Ohl et al. (2009) investigated whether hearing-impaired listeners were sensitive to changes in HRTF cues and hence changes in externalization. Speech stimuli

were convolved with headphone-equalized head-related impulse responses (HRIRs, head-present) in a room ($RT_{60} = 0.5s$). The speech stimuli were also convolved with headphone-equalized impulse responses taken by microphones in the same position, but without the person present (head absent). Therefore only the ITD cue was preserved, producing a theoretically more internalized percept. Mixing between these head present (theoretically most externalized) and head absent (theoretically most internalized) conditions, the sensitivity of normal-hearing and hearing-impaired listeners to the HRTF cues for externalization was tested using a 2-alternative forced choice discrimination task with a 1-up 2-down rule. The question asked was "Which interval sounded more like coming from the loudspeaker?" The experiment was run in both directions, adaptively varying the mix of HRIRs against a fixed head absent and a fixed head present stimuli. It was reported that hearing-impaired listeners were sensitive to changes in externalization, but they were less sensitive and more variable across individuals than normal-hearing listeners. Due to the variability across listeners, the results were not significant.

2.4.8 Head movements and externalization

Fixing a listener's head in place can increase the perception of inside the head localization when stimuli are presented from 0° azimuth in an anechoic chamber (Bauer, 1965). The perceptual constancy of angular localization of sources during head turns has been shown (Day, 1968), meaning that stationary sources are not perceived as moving sources during head turns. Attaching the ears to long tubes connected to fixed, spatially separated funnels, head movement has been decoupled from binaural cues (Young, 1931; Klensch,

1948). These studies found that listening through the apparatus and moving one's head caused the sound to be perceived inside or at the back of the head. Toole (1969) found no effect of small, involuntary head movements ($< 3°$) on internalization.

Begault and Wenzel (2001) found the effect of head movement on the externalization to be negligible in relation to HRTF information and reverberation. This study used headphone presented speech as the stimuli with individualized and non-individualized HRTFs. The degree of head movement made by listeners was not stated.

Loomis (1999) showed that head movements may be all that is required to externalize sounds using an ingenious apparatus. A listener wore a pair of in-ear, sound-insulating earphones, sound-attenuating hearing protectors placed over the earphones and microphones placed on top of the hearing protector earcups. A stereo amplifier then drove the earphones according to the microphone signals. This system removed the pinna cues and altered the lower frequencies of the HRTF, while preserving the relationship between broadband interaural cues and head movement. Listeners experienced a compelling sense of externalization. In more recent work, emulating head tracking effects, by randomly moving the presented sound by $2°$, has also been found to increase externalization (Wersényi, 2009). This study used a variety of stimuli, from low-pass (1.5 kHz cutoff), high-pass (7 kHz cutoff) noise to single words and sound effects. Listeners were asked in a forced-choice paradigm, "Is the sound source externalized or in-the-head?", "Where is the simulated sound source in virtual space?" and "Do you have the percept of a moving source?" Front-back errors were not reduced using head movement and the speed of the

movement was not significant to externalization. It is possible that the head movements were not large enough to resolve front-back errors and the use of non-individualized HRTFs simultaneously increased them.

Further collaborative research undertaken during this PhD shows the importance of head movements for externalization (Brimijoin et al., 2013). In this study, the previously described HRTF mixing technique was used (Ohl et al., 2009). Individualized HRTFs were dynamically updated with a listener's head movements using a motion-tracking system. Listeners were asked to judge whether a presented sound was internalized or externalized. Listeners were asked to move their heads or keep them still. HRTFs for the source were either dynamically updated, producing a relatively stationary source relative to the head or not, resulting in a source that moved with the head. It was found that in the dynamically updated, head-moving case, listeners would perceive speech stimuli as externalized with most of the HRTF removed, much more so than in the other conditions. When the source moved as expected with head movement, the HRTF cues were not important for externalization. This finding is contrary to previous similar studies, where head movement was not found to be important (Begault and Wenzel, 2001).

2.4.9 Visual cues for externalization

Though visual cues have been studied in relation to auditory localization and distance perception through the ventriloquism effect (section 2.2), there has been no research specifically conducted on the effect of visual cues and externalization. However, given the strength of the ventriloquism effect in distance perception and anecdotal evidence, one may expect the presence of

suitable visual candidates for a sound source may improve externalization. This suggestion was also made by Ohl Ohl et al. (2009).

2.4.10 Neurological research

The difference between internalized and externalized sound has been studied at the neurological level. Normal subjects' brains were scanned using fMRI while listening to internalized and externalized speech stimuli presented over headphones. Comparisons of brain activity showed that an area of association auditory cortex posterior to Herschel's gyrus (the planum temporale, PT) was activated for externalized sounds only, regardless of position (Hunter et al., 2003). A recent study also suggested the PT as an area involved in externalization in addition to other prefrontal regions (Looijestijn et al., 2013). It has been shown that auditory brainstem and middle latency responses are similar for internalized and externalized sounds (Junius et al., 2007).

CHAPTER 3

Review of signal processing research relevant to hearing aids

3.1 Introduction

This chapter concerns the key historical and contemporary research related to hearing-aid signal processing and the techniques used during this research. Hearing-aid hardware, dynamic-range compression, microphone directionality, beamforming and adaptive directional microphone arrays will be discussed. The results of listener tests and preferences for omnidirectional and directional microphone settings will be covered, in addition to bilateral directional hearing-aid systems and alternatives to directional arrays. Source direction of arrival (DOA) estimation techniques will be discussed, including the derivation of the generalized cross correlation with phase transform (GCC-PHAT) and alternatives to GCC-PHAT. The micro-electromechanical systems (MEMS) inertial measurement units (IMUs) utilized in the research will be discussed. The methods for combining the outputs of the MEMS gyroscope (measuring angular velocity), accelerometer (measuring proper acceleration or g-force) and magnetometer (measuring the strength and direction of a magnetic field)

to preserve head-tracking accuracy over time will be explained. Existing virtual auditory display systems used during the research will be described and industrial perspectives on internalization and externalization obtained at the beginning of the PhD will also be briefly summarized.

3.2 Hearing-aid hardware design

A hearing aid is a sound amplifier designed to help hearing-impaired listeners overcome hearing loss. At its most basic a hearing aid consists of a microphone, amplifier and output transducer (receiver) (Kates, 2008). Most hearing aids are now digital, incorporating algorithms to enhance desired signals such as speech, suppress unwanted noise, and compensate for the properties of the impaired auditory system. They can also be fine-tuned to an individual's particular needs with several situation-specific programs.

Hearing aids are available in several hardware designs or "form factors." Behind-the-ear (BTE) hearing aids house their electronics, including the microphone, in a case sitting on top of the ear. The output sound from the receiver is conducted to the ear canal via tubing, terminating at an earmold. This may be a custom earmold (with or without a small open-ended tube known as a "vent") or an "open-fit", which is a soft, silicone dome that sits inside the ear. They are more comfortable than a custom earmold as they remove the problems of own-voice amplification and pressure build-up in the ear canal (e.g. Lybarger, 1980). Open-fit molds allow more sound to be radiated outside the ear canal, leading to an increased chance of the sound being picked up by the microphone and transmitted as feedback. This means that open-fit hearing aids usually operate at lower gains for mild to moderate hearing losses and require adaptive

feedback algorithms to function (e.g. Siqueira and Alwan, 2000). They also have a reduced low-frequency response as it is difficult to produce the levels required in the ear canal without the seal that a custom mold would provide. Feedback also occurs with vented custom molds, as sound may be radiated through the vent (Dillon, 2001). In-the-ear (ITE) hearing aids fill the bowl of the concha in the outer ear. In-the-canal (ITC) and completely in-the-canal (CIC) hearing aids sit in the ear canal and CICs do not protrude into the concha. ITE and ITC microphones are placed in the concha of the pinna and the entrance of the ear canal respectively. CIC hearing-aid microphones reside in the canal. Placing the microphones inside the pinna or canal preserves the directional characteristics of the ear (Dillon, 2001).

ITE often fill the entire concha of the pinna, while ITC and CIC hearing aids are small and may not be visible to most observers when fitted. The position of the microphones inside the pinna can also improve localization, though this is principally due to decreased front-back confusions (section 2.2.6). Their size produces some disadvantages. It decreases the space for electronics and the complexity of the algorithms that may be used, limiting their performance in difficult listening situations. The battery size is also reduced in comparison to BTEs, leading to a shorter running time between changes and increased battery replacement difficulties for less dexterous users. The proximity of the microphone(s) to the receiver limits the output power of the receiver, due to the increased possibility of feedback (Kates, 1988).

Vents reduce the uncomfortable effects of plugging up the ear canal, such as occlusion. Occlusion is the increase in low-frequency sound pressure when the user is talking, caused by a combination of low-frequency, bone-conducted

sound and blocking the open end of the ear canal (Killion et al., 1988). Occlusion leads to sounds produced inside the head, such as the user's voice and mastication, being amplified in comparison to high-frequency sounds. An open fit causes almost no occlusion (Kiesling et al., 2005). Though occlusion is a purely acoustic effect of blocking the ear canal, recent survey work reports that hearing-aid users experience issues with their own voice quality in comparison to unaided listeners, even when using an open fitting (Laugesen et al., 2011).

3.3 Compression

Dynamic-range audio compression was introduced for hearing aids for two reasons; reduced dynamic range and recruitment. Reduced dynamic range occurs because the threshold of audibility increases in hearing-impaired listeners, while the threshold of pain does not. Recruitment refers to the phenomena of rapidly increasing loudness perception above the threshold of audibility (Buus and Florentine, 2002). If a normal-hearing listener has a threshold of audibility at 10 dB SPL and a threshold of discomfort at 120 dB SPL, their dynamic range will be 110 dB SPL. A hearing-impaired listener's audibility threshold may be 60 dB SPL resulting in a 60 dB SPL dynamic range. The aim of compression is to make sounds audible within a hearing-impaired listener's reduced dynamic range (Kates, 2010).

Most digital hearing aids available today include some kind of compression algorithm. Compressors have four main parameters which can be set to produce the desired output from a given input: threshold, compression ratio, attack and release. The threshold (also known as the knee-point) sets the level at which the compressor starts to operate. The knee-point can be either "soft," gradually

applying the compression ratio over several dB around threshold or "hard," immediately applying the chosen compression ratio at threshold. Below this an input signal will have gain applied linearly to it.The compression ratio gives the change in output gain for a given level above the threshold com (2003),

$$\text{Compression ratio} = \frac{\Delta\text{Input SPL}}{\Delta\text{Output SPL}} \tag{3.1}$$

For example, if threshold is 75 dB SPL, the compression ratio is 2:1 and the input signal is 81 dB SPL, the output will be 78 dB SPL, assuming that the gain applied to the signal below the kneepoint is 0. Attack time is the time delay that occurs between the onset of an input signal loud enough to activate the compressor (exceeding the threshold) and the resulting reduction of gain to the target value com (2003). Release time is the time delaz that occurs between the offset of a sound sufficiently loud to activate the compressor (therefore, the input falls below the threshold) and the resulting increase in gain to the target value com (2003). Compressors may be applied to a signal split into several frequency bands (i.e., multi-channel), applying different compressor settings to each band. This changes the spectral content of the output signal, attempting to fit the input to the dynamic range of the listener at different frequencies.

A low compression threshold (< 50 dB SPL) is referred to as wide-dynamic range compression (WDRC) (Dillon, 1996; Kuk, 2000). These typically have a low compression ratio ($< 5:1$, more often $< 2:1$) (Walker and Dillon, 1982). Often WDRC is combined with a high compression ratio above thresholds of > 75 dB SPL in hearing aids. Attack times for hearing aid compressors can range from 8 ms (very fast) to 500 ms (slow) and vary across frequencies

(Moore et al., 2004). A hearing-aid compressor's release time is referred to as slow-acting (release time > 200 ms) or fast-acting (release time < 200 ms) (Dreschler, 1992). A fast release time will produce a compressor more capable of reacting to changes in the auditory environment or speech quickly, though this can result in an unnatural, distorted listening experience. With a high compression threshold they can act as a limiter (see below), preventing saturation distortion from momentary loud sounds. With a low compression threshold they act as "syllabic" compressors, working on the timescale of and reducing the level differences between syllables and phonemes (Braida et al., 1982).

A very high compression ratio ($\geq 60 : 1$) is referred to as a "limiter" (Dillon, 2001). If the limiter has an instantaneous attack time, the limiter threshold is the maximum level at which the hearing aid can output sound. This is used to protect the listener's hearing from high SPL sounds in the auditory environment. Compression ratios generally increase with increasing level, meaning that a true limiter should only be employed in the most extreme cases. A gain-reducing compressor (towards lower input levels) is known as an expander (Dillon, 2001). In this case the compression ratio is $< 1 : 1$, resulting in quiet sounds being made quieter. This is useful for reducing the level of unwanted sounds such as internal hearing-aid noise.

Hearing-aid compression has possible effects on localization cues. If a listener is bilaterally aided, the compressors in each aid will typically operate independently of the other. The "head shadowing" effect that produces natural ILDs can produce distorted ILD cues when combined with compression, as the level difference between the hearing aids will result in

different degrees of compression. Multi-channel compressors also change the spectral content of the sound at each ear, again distorting localization cues. Modest effects on localization performance have been found for both WDRC and linear amplification (Keidser et al., 2006). Dynamic compression has a significant effect on high-frequency ILD discrimination in both normal-hearing and hearing-impaired listeners: ILD jnds increased with shorter attack times and unrealistically high ($8:1$) compression ratios. No effect was found for low-frequency ITD discrimination (Musa-Shufani et al., 2006). Both normal-hearing and hearing-impaired listeners performed poorly when discriminating ITDs at 4 kHz, however hearing-impaired listeners were worse, suggesting that they rely on ILD cues for high-frequency sounds because they are unable to make use of the ITD information. The effect of independently operating compressors also affects the lateral position of sounds for normal-hearing listeners (Wiggins and Seeber, 2011). Fast-attack compressors shifted the perceived position of high-frequency, abrupt onset/offset sounds to a more central position by 24 units on a 100 unit scale, "0" being centre of the head and "100" fully lateralized. Gradual onset/offset sounds, such as some portions of speech, were perceived to move or become a split sound, in addition to becoming broader in lateral extent. Inclusion of low-frequency binaural cues significantly reduced the measured effects. Compression also has a small but significant effect on externalization in normal hearing listeners using non-individualized HRTFs (Wiggins and Seeber, 2012). Fast-acting, independent compression increases the spatial separation between maskers and a target required for a normal-hearing listener to report the target message correctly by 5° on average (Schwartz and Shinn-Cunningham, 2013).

An effect was found for independent, slow-acting compression, but only for small spatial separations. Linking the compressors to restore the original ILD, by providing the same gain to each ear, restored the deficit in performance, but only for fast-acting compression.

3.4 Directional microphones

Most modern hearing aids now include a directional microphone, often utilizing adaptive acoustic beamformer techniques. In many listening scenarios, the listener is assumed to be facing a talker, while unwanted sound sources come from other directions. An omnidirectional microphone has the same response at all angles (in the free field). A directional microphone "spatially filters" (Veen and Buckley, 1988) the incoming signals, having a maximum sensitivity in one direction (generally directly in front of the listener) and reduced sensitivity to signals incident from other directions. Directionality is a useful and important form of noise-reduction signal processing for hearing aids because it is largely independent of the input signals the hearing aid is exposed to, unlike other noise-reduction techniques such as spectral subtraction, where an estimate of the noise source spectrum is required. Directionality can provide robust SNR improvements to target sources for hearing-impaired listeners in noisy situations where they are needed most (Hoffman et al., 1994).

Directional microphones in hearing aids work by comparing incoming signals sampled at two different locations, 4-15 mm apart (Ricketts and Mueller, 1999; Kayser et al., 2009). Timing differences between the signals sampled at these two points are used to attenuate sounds arriving from unwanted angles. This attenuation may be achieved in two ways. A single microphone and

an acoustic phase-shifting network may be used for acoustic cancellation (a two-port directional microphone) or two omnidirectional microphones may be used with electronic delays (a directional microphone array, DMA).

The two-port directional microphone operates as follows: A signal incident on the hearing aid from the rear will reach the rear port first and the front port and the front of the microphone diaphragm after some delay (external delay). The signal incident on the rear port is subject to an acoustic delay equal to the external delay before reaching side of the microphone diaphragm. The signal incident on the front port is guided to the other side of the diaphragm with no internal delay. This means that the signal incident from the rear reaches both sides of the diaphragm at the same time and cancellation occurs, attenuating (in theory, perfectly canceling out) the signal. A signal incident from the front is subject to no delay through the front port and an external and internal delay through the rear port, resulting in less cancellation. Varying the acoustic time delay will produce different types of directional response. Directional microphones of this type were introduced to hearing aids in the late 1960s and early 1970s (e.g. Arentsschild and Froeber, 1972).

The directionality of the two-port microphone can also be achieved electronically, by adding together two omnidirectional microphones with a fixed time delay between them. This is known as a fixed directional microphone array or a delay-and-subtract beamformer (Dillon, 2001). The spatial and spectral response of the output depends on the spatial separation of the microphones and the time delay applied. The microphone arrays in hearing aids are approximately 15 mm long (Kayser et al., 2009; Teutsch and Elko, 2001). The frequency response of a directional microphone is dependent on the array

spacing and decreases with decreasing frequency, due to the wavelength of low-frequency sounds being much greater than the array spacing and their phases being similar (Kates, 2008). Additional amplification (equalization) is applied to improve the frequency response at low frequencies. However, this also amplifies the internal noise of the microphone, which can be an additional source of unwanted noise for the hearing-aid user (see section 3.7). The position of the microphones on the head affects the directionality of the microphone, as head shadowing changes the response of the microphone to different frequencies (Kates, 2008).

3.5 Beamforming

Beamforming techniques were first proposed for hearing aids in the mid-1980s (Petersen et al., 1987). These methods were adapted from radar applications (Elliott, 1981) where they were developed as a way of improving spatial noise reduction. Constrained, adaptive beamformers such as the Frost (Frost, 1972) and Griffiths and Jim (Griffiths and Jim, 1982) designs build upon the most basic delay-and-sum technique by minimizing the output power of a microphone array, using some variation of a least mean squares (LMS) technique (see section 3.6 for a full derivation of the LMS technique), with the constraint that signals in the target direction be preserved.

The Frost beamformer utilizes a tapped-delay line for each microphone, adapting filter weights at each tap, before summing the output from the delayed, weighted samples (Frost, 1972). Using the fact that the average output power of the beamformer is quadratically related to the filter weights, an LMS technique can be used to minimize the output power. Assuming that the target

and noise are uncorrelated and the target direction is known (often assumed to be on axis with the microphone array), the noise will be reduced and the target enhanced. Constantly adapting many filter weights to minimize a chosen measure (e.g. the output power of the array) can be both time consuming and costly in terms of processing.

The Griffiths and Jim beamformer (also known as the generalized sidelobe canceler) simplifies the constant filter weight adaptation by introducing a two-stage system (Griffiths and Jim, 1982). Firstly, the microphone signals are summed and subtracted. The summed signal contains the target and interference, while the subtracted signal contains only interference, if the assumptions above hold. The second stage then becomes a noise canceler, again utilizing an LMS adaptive filter, in this case unconstrained. This simplification means it is trivial to extend the system to more microphones. A balance must always be struck between adaptation time and filter quality. A fast adaptation may produce an unstable filter, while a slow adaptation will not be sufficiently sensitive to changes in the signals and environment.

3.6 Adaptive directional microphone array

A common system used in modern hearing aids is the adaptive first-order directional microphone array (ADMA) (Teutsch and Elko, 2001; Dillon, 2001). This system removes the need for continuously variable time delays between the microphones (as required in the previous techniques) using back-to-back cardioids, one facing the front and one facing the rear. The cardioids are formed by combining two omnidirectional microphones with a fixed delay (in the same way as the fixed directional microphone) by subtracting the microphone

outputs. By assuming the target signal is on-axis (on a line directly in front of or behind the listener, orthogonal to the interaural axis), a one-tap LMS algorithm can be used to adaptively minimize the output power of the the system by varying the level of the rear cardioid. This produces a variable response that adaptively places a null at the angle of the loudest interfering source. In common with the Griffths and Jim beamformer, the first-order array can also be extended to more microphones. When extended to three microphones in two pairs, the system becomes a second-order array. The array spacings are smaller for a second-order array (Kayser et al., 2009). This increases the frequency below which equalization is required and the rate of gain reduction in the low frequencies increases from 6 dB/oct. to 12 dB/oct. (Chung, 2004).

ADMAs are commonly used in hearing-aid signal processing as they do not require constantly varying fractional-delay filters, which are computationally expensive and instead simply mix the relative gains of front and rear-facing cardioid microphones.

The derivation of an adaptive directional first-order array is as follows (Teutsch and Elko, 2001; Chatlani, 2011). For a plane-wave (sources are assumed to be in the far-field) signal $s(t)$ with spectrum $S(\omega)$ and a wavescalar k incident on a two element array with displacement vector d as shown in figure 3.1, the output can be written as,

$$Y(\omega,\theta) = S(\omega)\left(1 - e^{-j(\omega T + k\cdot d)}\right) \tag{3.2}$$

$$= S(\omega)\left(1 - e^{-j\omega[T + (d\cos\theta)/c]}\right) \tag{3.3}$$

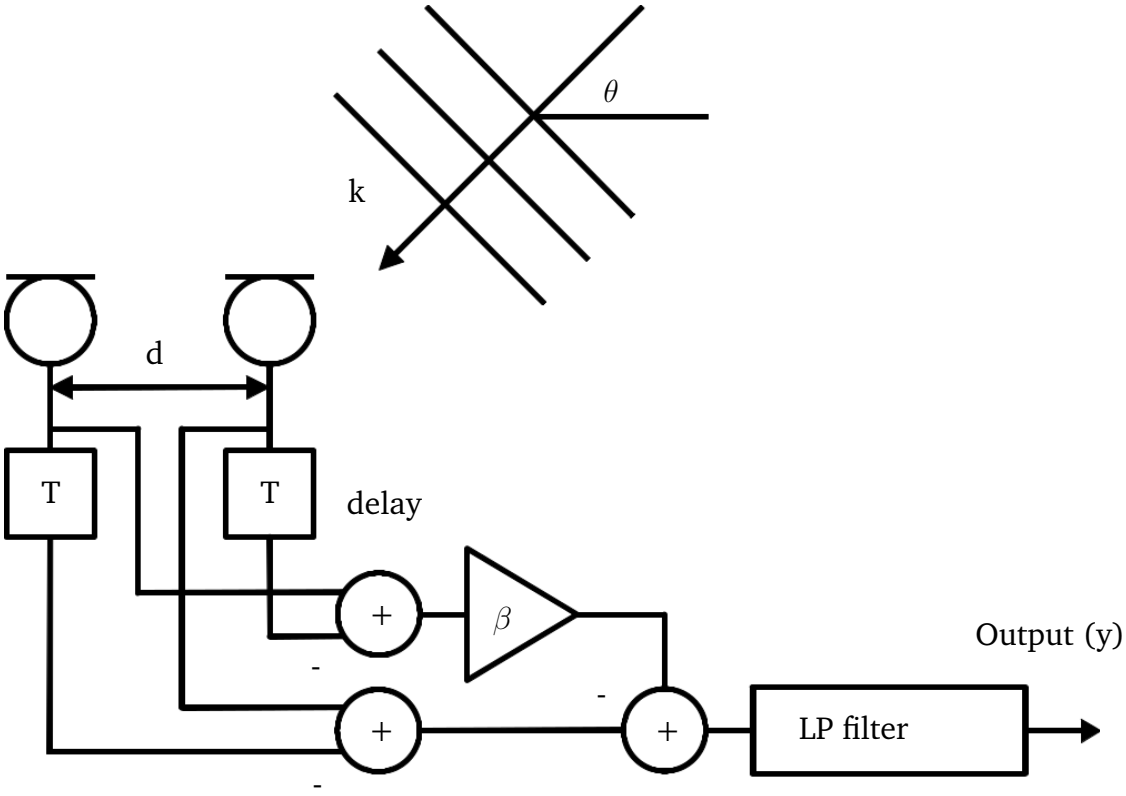

Fig. 3.1: Directional microphone array.

where d is the inter-element spacing, T is equal to the delay applied to the signal from one sensor and $|k| = k = \omega/c$ where c is the speed of sound. The magnitude of 3.3 gives,

$$|Y(\omega,\theta)| \approx 2|S(\omega)\sin\frac{\omega[T+(d\cos\theta)/c]}{2}| \tag{3.4}$$

By assuming a small inter-element spacing and delay ($kd \ll \pi$ and $\omega T \ll \pi$),

$$|Y(\omega,\theta)| \approx \omega|S(\omega)[T+(d\cos\theta)/c]| \tag{3.5}$$

The first-order differential array has a monopole term and a first-order dipole term $\cos\theta$. This resolves the acoustic particle velocity component along the sensor axis. Equation (3.5) also shows a first-order differentiator frequency dependence, which is a linear increase with frequency. A first-order low-pass filter compensates for this in practice. The directional response of the array is $[T + (d\cos\theta)/c]$.

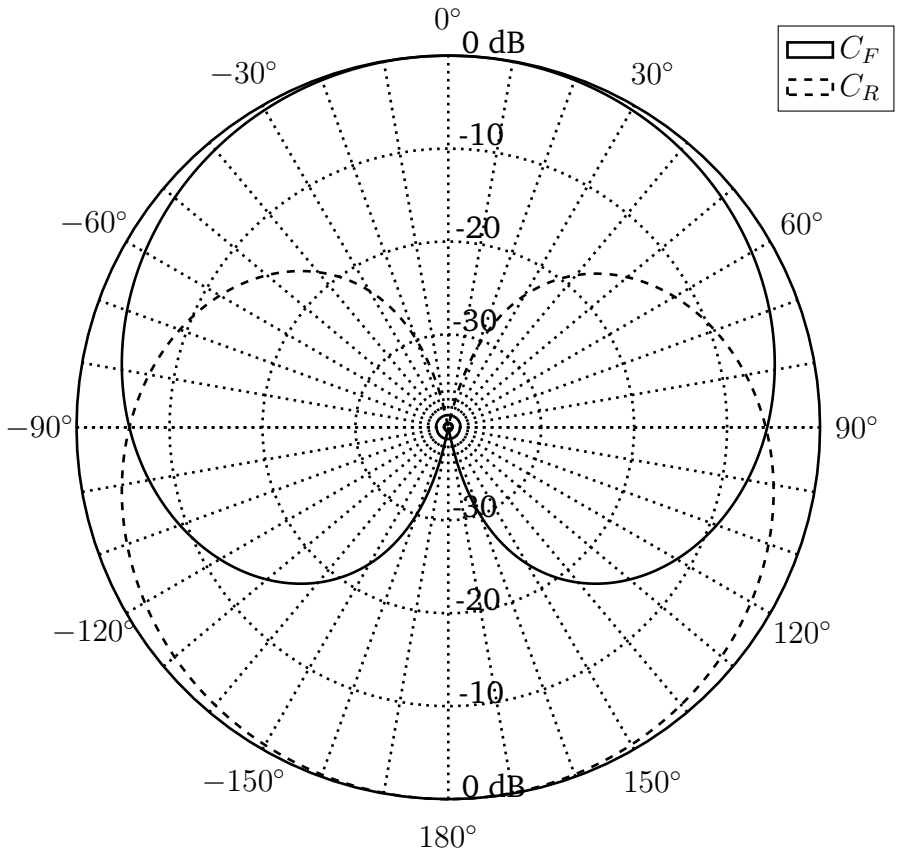

Fig. 3.2: Back-to-back cardioid patterns formed from the delayed sums and differentials of two omnidirectional microphones

The array design is implemented in hearing aids using back-to-back cardioid outputs, shown in figure 3.2. Setting the sampling period equal to d/c,

back-to-back cardioid microphone outputs can be formed directly from two omnidirectional microphones.

The spatial origin is taken as the centre of the array. Using (3.3) and using a sample period of $T = d/c$, the expressions for the forward-facing cardioid C_F and rear-facing cardioid C_R are as follows,

$$C_F(\omega, \theta) = 2jS(\omega)e^{-j\omega T/2} \sin \frac{kd(1 + \cos\theta)}{2} \tag{3.6}$$

$$C_R(\omega, \theta) = 2jS(\omega)e^{-j\omega T/2} \sin \frac{kd(1 - \cos\theta)}{2} \tag{3.7}$$

The polar pattern produced by the array is constrained for use in hearing aids. The maximum output of the array is assumed to be required in the front hemisphere. Therefore, the position of the null points in the pattern are controlled by varying the gain applied to the rear-facing cardioid. The gain factor, β is constrained to be between 0 and 1, resulting in the null points of the array output varying between $\pm 90^\circ$ in the rear hemisphere only for the following equation of the normalized output signal $Y(\omega, \theta)$,

$$|\frac{Y(\omega, \theta)}{S(\omega)}| = 2|\sin \frac{kd(1 + cos\theta)}{2} - \beta \sin \frac{kd(1 - cos\theta)}{2}| \tag{3.8}$$

Setting β to a fixed value produces the directional microphone array. A first-order directional microphone array pattern can be seen in figure 3.3.

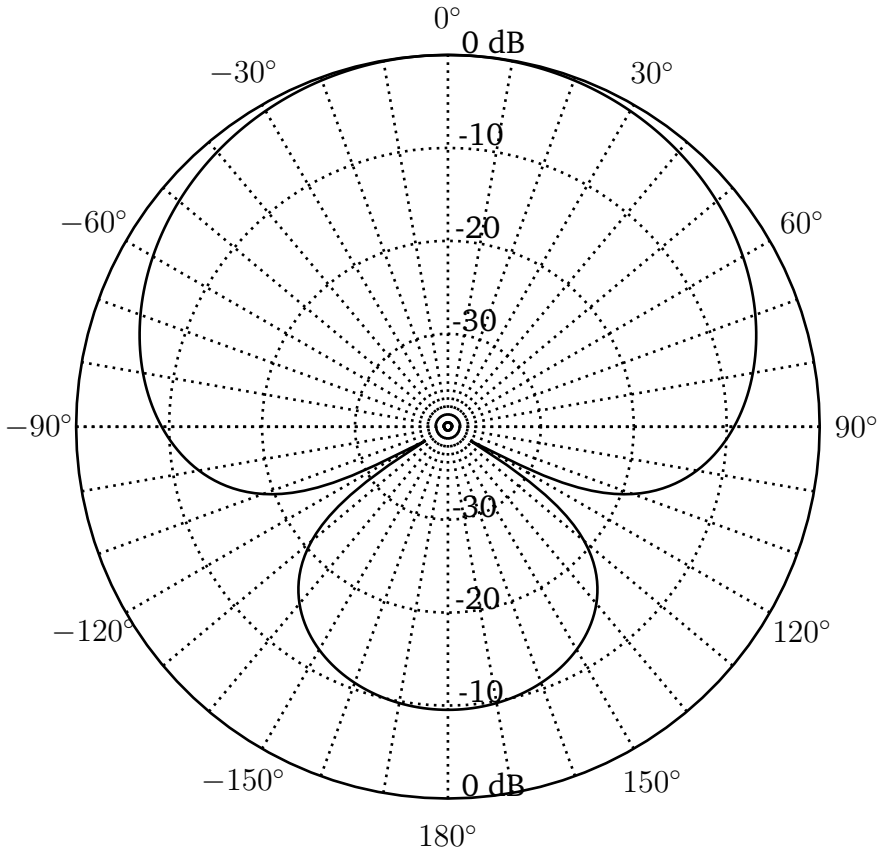

Fig. 3.3: First-order directional array patterns formed from the sum of back-to-back cardioids. $\beta = 1/3$

The adaptive directional microphone array attempts to optimize β by minimizing the mean-square value of the array output. The value of β is constrained ($0 \leq \beta \leq 1$). The time domain representation of the back-to-back cardioid output is,

$$y(t) = c_F(n) - \beta c_R(n) \tag{3.9}$$

This equation is squared and the expected value is given by,

$$E[y^2(t)] = R_{c_F} - 2\beta R_{c_F,c_R} + \beta^2 R_{c_R} \tag{3.10}$$

where R_{c_F} and R_{c_R} are the autocorrelations of the front and rear cardioid outputs and R_{c_F,c_R} is the cross-correlation between the front and rear cardioid outputs. The output power $E[y^2(t)]$ is a quadratic function of β. The derivative of 3.10 is taken for β and setting the result to 0 we obtain,

$$\beta_{opt} = \frac{R_{c_F,c_R}}{R_{c_R}} \tag{3.11}$$

Since the second derivative is positive, the value of β_{opt} is a minimum. Assuming zero lag, the second derivative is positive. The autocorrelation R_{c_R} must be positive or zero for zero lag and this produces a concave $E[y^2(t)]$, with one minimum. This results in an optimal solution for a one-tap Wiener filter.

The real-world signal processing implementation uses short-time estimates of R_{c_F,c_R} and R_{c_R}, which allows it to theoretically adapt in non-stationary environments. For this the least mean squares (LMS) or stochastic gradient algorithm is implemented. The square of (3.9) is,

$$y^2(t) = c_F^2(n) - 2\beta c_F(t) c_R(n) + \beta^2 c_R(n)^2 \tag{3.12}$$

The steepest descent algorithm finds the minimum of the error surface $E[y^2(n)]$ by stepping in the direction opposite to the surface gradient, with respect to β_n. This update equation is written as,

$$\beta_{t+1} = \beta_n - \mu \frac{dE[y^2(n)]}{d\beta_n} \tag{3.13}$$

where μ is the update step-size and the differential gives the gradient of the error surface $E[y^2(n)]$ with respect to β_n. The mean of y^2 should be minimized and the LMS algorithm uses the instantaneous estimate of the gradient. Differentiating (3.12) gives,

$$\frac{dy^2(t)}{d\beta} = -2c_F(t)c_R(n) + 2\beta c_R^2(n) \qquad = -2y(n)c_R(n) \tag{3.14}$$

The LMS algorithm is then,

$$\beta_{n+1} = \beta_n + 2\mu y(n)c_R(n) \tag{3.15}$$

μ is usually normalized to remove the dependence of convergence on the input power, resulting in the normalized LMS,

$$\beta_{n+1} = \beta_n + 2\mu y(n)\frac{c_R(n)}{< c_R^2 >} \tag{3.16}$$

where $< \ldots >$ indicates a time average.

The optimal value of β can also be calculated directly, given a desired null angle, θ_1 or vice versa. Setting (3.8) to zero and solving for $\theta = \theta_1$,

$$\theta_1 = arccos\left(\frac{2}{kd}\text{arctan}\left(\frac{\beta-1}{\beta+1}\tan\frac{kd}{2}\right)\right) \tag{3.17}$$

Assuming a small spacing and delay, this approximates to,

$$\theta_1 = \text{arccos}\frac{\beta-1}{\beta+1} \tag{3.18}$$

3.7 Listener tests with hearing-aid directional microphones

While laboratory tests show that directional settings provide benefit to listeners in terms of improved SNR and speech intelligibility (Hawkins and Yacullo, 1984; Ricketts and Dahr, 1999; Valente et al., 1995), in the real world, this is not always the case. As previously mentioned, reverberation can reduce the performance of directional settings (Ricketts and Dahr, 1999; Hawkins

and Yacullo, 1984; Ricketts and Hornsby, 2003). Directional settings work best when the desired source is directly in front, reverberation is low and other sources are spatially separated from the desired source (Walden et al., 2003; 2004). When listeners are able to switch between omnidirectional and directional settings, their ratings of their performance are improved over omnidirectional or directional only (Ricketts et al., 2003). Adaptive systems are theoretically able to provide greater noise reduction than fixed directional settings due to their ability to move their null response to the strongest noise source. In real-world scenarios these systems may only provide 2 dB more attenuation and only at the highest DRR ratios (Woods and Trine, 2004).

Laboratory tests show that directional microphones provide benefit to listeners in a number of ways. SNR increases using directional microphones for on-axis sources range from 3-4 dB when listening in reverberant conditions (Hawkins and Yacullo, 1984) or with noise sources places at 90° (Wouters et al., 1999), to 6-7 dB (Valente et al., 1995; Lurquin and Rafhay, 1996) in anechoic conditions with noise sources at 180°.

Listeners with severe hearing losses experience significant benefit from directional microphones in the laboratory when they are combined with visual information (Ricketts and Hornsby, 2003). When only audio information is provided, the benefit of a directional microphone is only significant at the least positive SNRs. Directional microphones have been shown to provide benefit for hearing-impaired children in laboratory settings (Gravel et al., 1999; Kuk et al., 1999).

A preference for directional over omnidirectional microphones has been shown in laboratory tests of sound-treated, living and classrooms (Amlani, 2001) and speech-in-noise tests (Walden, 1999). In the real world, the benefit of directional microphones is less obvious. A systematic review of the effectiveness of directional microphones for hearing-impaired listeners outside of the laboratory found nine suitable studies and found evidence for a weak, positive effectiveness (Bentler et al., 2004). Three of the studies showed no preference for directional over omnidirectional microphones amongst listeners (Palmer et al., 2006; Surr et al., 2002; Walden et al., 2004). The study by Palmer et al. (2006) found no subjective preference between omnidirectional and a second-order directional microphone. The Surr et al. (2002) study found that, while directional microphones were favoured over omnidirectional in some situations, listeners often found it difficult to perceive a benefit. The Walden et al. (2000) study found speech understanding and sound quality in directional and omnidirectional modes to be rated equal by users in everyday listening situations and directional combined with a further spectral noise reduction to be more comfortable than omnidirectional. The talker location, type of or absence of background noise and the type of listening environment were factors affecting the selection of directional over microphone settings. Another study found that a greater perceived benefit in the real world did not lead to better performance in a clinical setting (Cord et al., 2004).

Listeners report using the directional microphones of their hearing aids on average 25% of the time when allowed to switch freely between omnidirectional and directional microphones in field trials of BTE hearing aids (Cord et al., 2002). Their satisfaction with both modes was rated equally. In

another study using switchable ITE hearing aids, laboratory tests of hearing in noise showed benefits for the directional setting (Preves et al., 1999). Questionnaire, paired-comparison and interview data showed that listeners preferred the directional mode in noisier environments and omnidirectional in quiet.

The effectiveness of an adaptive directional microphone was tested in anechoic conditions and a moderately reverberant classroom (Bentler et al., 2004). The microphones were tested in a fixed (cardioid) and adaptive setting. Five speakers were arranged around the listener from $110^\circ - 250^\circ$ and speech weighted noise presented from them. Noise was increased by 8 db from one speaker at a time and the other speakers reduced in level to maintain a constant overall level of 65 db. Using speech perception tasks as outcome measures, the adaptive system was shown to be unable to follow the dominant noise source. Self-report measures agreed with the test findings, as listeners did not perceive the adaptive system to be more effective than the fixed.

Internal noise generated by the hearing aid increases when using directional microphones (Chung, 2004). Two factors contribute to this. Internal noise from modern omnidirectional microphones is about 28 dB SPL. Combining two microphones increases this noise by about 3 dB. Though this is normally masked by environmental sounds or inaudible to hearing aid users, the equalization applied due to the reduced response of the directional microphone at low frequencies increases the noise level. In addition, directional microphones are more sensitive to wind noise due to their higher sensitivity to incoherent near-field signals. Wind noise can be up to 20 or 30 dB higher than omnidirectional microphones (Kuk et al., 2000; Thompson, 1999). Laboratory

tests have shown a large variability in the objectionable level of internal noise for hearing-impaired hearing-aid users (Lee and Geddes, 1998). Users requiring higher-gain hearing aids will tolerate a higher internal noise, however the relationship between the increased tolerable level and gain is non-linear and a simple rule for the increase has not been found (Macrae and Dillion, 1996).

3.8 Bilateral microphone systems

Binaural listening provides listeners with advantages in noise, reverberation and other complex listening situations (Zurek, 1993). The types of directional microphone described previously have been *monaural* designs for single, or unlinked pairs of hearing aids. The effect of using these monaural directional microphones bilaterally has been studied (Hawkins and Yacullo, 1984). The SNRs required for a constant performance level in normal and hearing-impaired listeners, under three reverberation levels (0.3, 0.6 and 1.2 s) were compared. Listeners used monaural or independent bilateral hearing aids in omnidirectional and directional settings. A significant binaural advantage due to better-ear listening, independent of reverberation and microphone type was found, of 2-3 dB. The advantage of the directional microphone was 3-4 dB, significant and dependent on reverberation level (decreasing with increasing reverberation), but independent of monaural or bilateral listening. Additive bilateral and directional microphone advantages were found and the effect of reverberation was larger than both the binaural and directional effects. Leeuw and Dreschler (1991) also found the advantage of directional microphones to be dependent on reverberation, in addition to the position of the noise source.

Two independent, two-microphone, endfire (perpendicular to the interaural axis) fixed-directional arrays were tested against a four-microphone, broadside (parallel to the interaural axis) fixed-directional array (Desloge and Rabinowitz, 1997). The broadside array preserved localization cues in the low frequencies by sending the leftmost and rightmost microphone signals to the left and right ears respectively. At higher frequencies, all microphones were combined in a fixed-directional array. In listener tests, a 2-4 dB gain in SNR was found for the system, relative to bilateral cardioids. A two-microphone bilateral adaptive system was also developed (Welker et al., 1997). The high-frequency portion of the left and right microphone signals were combined in a Griffiths and Jim beamformer, modified to halt adaptation of filter weights during periods where the target source was active (Greenberg and Zurek, 1992). Low-frequency signals were used to provide binaural cues. A cutoff frequency of 500 Hz for low and high-frequency components provided an improvement of 40% in sentence intelligibility over unaided listening for normal-hearing listeners, while allowing adequate localization performance. Comparison of this system with omnidirectional, bilateral hearing aids found improvements in SRTs of approximately 3 dB in mildly reverberant environments and 5 dB in comparison to monaural amplification.

The effect of using one, two or three microphone signals on each hearing aid to perform adaptive, bilateral (combining signals from both ears) beamforming has been investigated (Gordy et al., 2008). Monaural and bilateral systems were evaluated with respect to directivity pattern, directivity index and noise gain. Simulations showed that for a target source in the front hemifield, the bilateral beamformer provided a 3 dB gain in directivity in comparison to a

monaural beamformer, with no increase in noise gain. This improvement was only observed if more than one omnidirectional microphone signal was taken from each hearing aid.

The head-shadowing effect can also be used to improve SNR and noise estimates (Chatlani et al., 2010; Chatlani, 2011). In Chatlani's system, the noise estimate for a bilateral adaptive (Wiener) filter is taken from the ear furthest from the target and combined with first-order unilateral and bilateral ADMAs. The increased array size and directivity makes it both feasible and useful to steer the response of the array towards target angles other than 0° (Chatlani et al., 2010; Löllmann and Vary, 2012). Combining signals from both sides of the head provides an SNR benefit over monaural arrays, however it destroys any binaural cues in the original signal. It is unknown whether the benefit of the SNR improvement would become a deficit in a reverberant environment due to the loss of the binaural listening advantage. Bilateral beamforming algorithms have been developed that either preserve the cues through the application of a cost function (Klasen et al., 2007), or re-instate them after the beamformer processing (Lotter and Vary, 2006).

Beamformers have also been mounted on eye-glasses, using a five microphone broadside or endfire delay-and-sum beamformer array (Soede et al., 1993b). The increased number of microphones and spacing of the arrays over BTE hearing aids provided an attenuation of reverberant noise of 6 dB in the free-field and an improvement of 7 dB in SNR in a diffuse noise field (Soede et al., 1993a). Hearing-impaired listener tests showed a 7 dB improvement in SRTs using the broadside array and 6.8 dB using the endfire array over monaural omnidirectional listening. Binaural listening with two endfire arrays

showed a similar improvement to the improvement obtained listening with two ears or two conventional hearing aids. However, physically connecting hearing aids in order to pass program and audio information between them has been bypassed for aesthetic reasons in favour of more computationally expensive and power consuming wireless links (Walden, 1999). Due to the constraints of wireless transmission, data-limited techniques have been developed, such as combining only the low-frequency part of the signal (Srinivasan, 2008) and the trade-off between the rate of data transfer and the improvement in noise reduction performance has been studied (Srinivasan and den Brinker, 2009; Roy and Vetterli, 2006). Commercially available aids currently transfer data such as program selection and volume (Doclo et al., 2009). They can also stream music, however in all cases the transfer cost to battery life is high.

3.9 Alternatives to directional arrays

Alternatives to beamforming include spectral noise-reduction techniques such spectral subtraction (Doerbecker and Ernst, 1996) and Wiener filters (Doclo et al., 2009). These require an estimate to be made of the spectrum of the noise source in multiple bands, before attempting to attenuate the noise source in the spectral domain. Determining what is a target and what is noise can be unreliable in the real-world, therefore Wiener filters are often combined with a beamformer, in effect assuming a location for the target and treating sources arriving from other directions as a noise estimate for the filter (Chatlani et al., 2010; Chatlani, 2011).

3.10 Direction of arrival (DOA) estimation

To improve the performance of adaptive spatial noise reduction algorithms, such as adaptive beamformers and binaural Wiener filters, it is necessary to reliably and accurately detect the direction of arrival (DOA) a signal of interest. ADMA algorithms theoretically find the DOA of the loudest off-axis source incident on the rear portion of the array by minimizing the output power of the array. After adaptation, the null position of the cardioid response should point to the DOA of the loudest noise source (Teutsch and Elko, 2001). Many factors can reduce the benefit of ADMAs, such as reverberation and the presence of multiple active sources (See section 3.4). Often, more robust DOA techniques are required to properly localize both noise sources and targets of interest. Table 3.1 summarizes the main research carried out on DOA estimation for hearing aids.

Author(s)	Research
Knapp and Carter (1976)	Derived the generalized cross-correlation method with phase transform (GCC-PHAT), used to determine direction of arrival (DOA) using the time delay between microphones
Benesty (2000)	Developed the adaptive eigenvalue decomposition (AED) method for determining source DOA in reverberant environments. The source signal impulse response is blindly determined
Liu et al. (2000)	Developed a dual-delay line DOA estimator, based on the Jeffress model of localization
DiBiase (2000)	Developed and tested the steered-response power with phase transform (SRP-PHAT), combining GCC-PHAT DOA estimates from multiple microphone pairings and beamforming
Rohdenburg et al. (2008b)	Found GCC-PHAT to be more robust in noise than the dual-delay line approach
Goetze et al. (2007)	Introduced a propagation model and interpolation for improved DOA estimation using a head-mounted array
Brutti et al. (2008)	GCC-PHAT was less computationally complex and performed better at < 15 dB SNR than AED
Usagawa et al. (2011)	Differential pinna filtering used to resolve the position of a source in the front or rear hemifield using only two microphones

Summary table of important research on DOA techniques.] Summary table of important research on DOA techniques.

Table 3.1: (

3.10.1 Generalized cross-correlation with phase transform (GCC-PHAT)

Several systems have been developed to estimate DOA from the electroacoustic inputs. One of the oldest is the generalized cross-correlation (GCC) technique (Knapp and Carter, 1976). This algorithm performs a cross-correlation in the frequency domain between two microphone signals receiving a single source, using a short-time Fourier transform (STFT). When the inverse fast Fourier transform (IFFT) of the STFT is taken, the position of the largest peak in

the output gives a time delay (or lag) in samples between the microphones corresponding to the largest cross-correlation. If the distance between the microphones and the sample rate is known, the DOA of the sound source can be obtained. The mathematics of the GCC algorithm can be found below. In noisy or reverberant conditions, the technique is less robust, producing spurious peaks in the cross-correlation (Benesty, 2000). In the original paper on GCC, a number of transforms and additions were suggested (Knapp and Carter, 1976), such as the commonly used (e.g. Löllmann and Vary, 2012) generalized cross-correlation with phase transform (GCC-PHAT). This additional phase transform whitens the microphone signals, making them more spectrally uniform. This makes the analysis more robust to reverberation. The phase transform has been derived as a special case of the theoretically optimal maximum-likelihood weighting in low noise, reverberant environments (Zhang et al., 2008), however it was originally added to GCC for a single source in a post-hoc analysis, meaning that it was not initially derived directly. The phase transform has also been derived analytically for two or more active sources, though not tested on a real electroacoustic signal path (Kwon et al., 2010).

The most simple configuration of a single active source in an anechoic space recorded by two spatially separated microphones can be described by,

$$x_1(n) = s(n - \tau_1) \tag{3.19}$$

$$x_2(n) = s(n - \tau_2) \tag{3.20}$$

$$\tau_s = \tau_2 - \tau_1 \tag{3.21}$$

where x_1 and x_2 are the source, s received at each microphone, n is the time-step and τ_s is the relative time-delay of the source signal between the microphones.

The estimation of time-delay estimation (TDE) using GCC is a frequency-domain technique for calculating τ_s, defined as,

$$\psi_{GCC}(n) = F^{-1}\{X_1^*(k) \cdot X_2(k)\} \tag{3.22}$$

for frequencies $k = 0, ..., N-1$, where N is the analysis window size and F^{-1} is the inverse discrete fast Fourier transform (IFFT). X_1 and X_2 are frequency domain representations of the microphone signals and * is the complex conjugate. To calculate the delay, τ_s, the maximum peak in the GCC function is selected,

$$\tau_s = \arg\max_n \psi_G(n) \tag{3.23}$$

The GCC can be made more robust to noise and reverberation by applying the phase transform (PHAT), setting all frequency magnitudes equal to 1, thus preserving phase,

$$\psi_{GCC-PHAT}(n) = F^{-1}\left\{\frac{X_1^*(k) \cdot X_2(k)}{|X_1^*(k) \cdot X_2(k)|}\right\} \tag{3.24}$$

Using an N-point Hanning window also improves the result.

The robustness of all GCC-based algorithms can be improved by aggregating multiple DOA estimates over time into a histogram, the largest histogram peak providing the most likely DOA (Strobel and Rabenstein, 1999). Splitting the signal up into frequency sub-bands and building up an estimate histogram can also improve estimation accuracy over the broadband implementation (Löllmann and Vary, 2012). In addition, if both the microphones and the sources are assumed physically static, longer segments of the microphone output can be used for each DOA estimate (Silverman and Kirtman, 1992). Cepstral (the inverse Fourier transform of the logarithm of the signal spectrum) prefiltering of the signals attempts to deconvolve the effects of reverberation before GCC (Stephenne and Champagne, 1995). This requires signal segments equivalent to the length of the reverberation time and a static source-sensor relationship during the segment, possibly making it impractical for real-world use unless coupled with a robust reverberation time and head-movement detector. Cepstral prefiltering is also highly susceptible to variable and non-stationary signals such as speech. Other improvements can be obtained by minimizing the weighted least-squares function of the phase data (Brandstein and Silverman, 1997). Given that the phase data is discontinuous, a complex search algorithm must be applied to find the minimum phase, resulting in a high computational cost for a marginal improvement.

3.11 Alternatives to GCC-PHAT

The steered response power with phase transform (SRP-PHAT) technique combines multiple microphone signals instead of multiple estimates of DOA,

in order to steer a beamforming array to the loudest source (DiBiase, 2000). In terms of GCC-PHAT, it is equivalent to the sum of cross-correlations over all possible pairs of microphones in an array. This means that over two microphones, SRP-PHAT and GCC-PHAT are the same. Due to the small spacings between microphones on hearing aids, extensions of bilateral GCC-PHAT to SRP-PHAT have yielded little improvement in DOA estimation over GCC-PHAT using a single bilateral microphone pair (Rohdenburg et al., 2008a).

Adaptive eigenvalue decomposition (AED) attempts to determine the direct path for the source-sensor by blindly identifying the impulse response (Benesty, 2000). The impulse response is assumed to be finite, allowing a dispersive propagation model to be used. By minimizing the variance in the output signal, a single input, multiple output (SIMO) system is created. Extending this to a multiple input, multiple output (MIMO) system results in the ability to track more sources. In electroacoustic tests AED was shown to stably localize a single speech source under reverberant and noisy conditions, whereas the GCC-PHAT was unstable (Aichner et al., 2006). The MIMO system was able to track two sources, including one moving target. The system requires shorter segments than the cepstral-filtered GCC, enabling better tracking of moving sources. It is assumed that the number of simultaneously active sources does not exceed the number of microphones and that the sources are mutually uncorrelated (Aichner et al., 2006). GCC-PHAT is less computationally complex and performs better than AED below 15 dB SNR for single sources (Brutti et al., 2008)

A dual-delay line approach has also been developed for two microphones (Liu et al., 2000). Signals from each microphone are transformed into the frequency

domain and analyzed for coincidences in their delay-line pairs across frequency, mimicking the Jeffress model of localization (Jeffress, 1948). The system also considers coincidences at phase delays greater than 2π, which are generally regarded as ambiguous. Integrating coincidences across frequency determines the DOA of up to six simultaneous sources in anechoic conditions. AED has not been shown to work with more sources active than microphones, while GCC-PHAT will detect the strongest source in each estimation window, meaning that its maximum simultaneous source detection is based on the length of the estimation window, the temporal sparsity of the signal (producing silences in which other, quieter sources can be detected), and the number of aggregated estimates.

A comparison of similar systems using either the GCC-PHAT or the dual-delay line approach revealed that GCC-PHAT is more robust to noise and reverberation (Goetze et al., 2007; Rohdenburg et al., 2008b). These studies also combined the techniques with HRTFs, parametric head models, and steerable beamformers to determine their functionality in a head-mounted system, highlighting their potential for use with hearing aids and self-steering bilateral beamformers. Goetze et al. (2007) also used interpolation of an oversampled IFFT to improve the angular resolution of the DOA estimation. DOA estimates could be placed in seven different DOA sectors in the front hemifield ($\pm 105°$, $30°$ wide bins). Test were shown only for single sources.

Another DOA technique uses both ITDs and ILDs (Raspaud et al., 2010). The algorithm uses a time-frequency representation to estimate the level difference and time difference between the ears and uses either HRTF look-up or a parametric HRTF model to estimate the DOA of the signal. Joint estimation

of ILD and ITD is used, as the level difference gives a single, wide angle for the source direction, while the time difference gives several highly precise estimates. By choosing the time difference estimate within the area defined by the level difference estimate, the correct angle of arrival can be found. Several studies have explored aspects of this technique, comparing the spectral content of microphone signals from outside the pinna with those inside the pinna (Keyrouz and Saleh, 2007), deconvolution of the HRTF from the source(s) using a library of inverse HRTFs (Rothbucher et al., 2012), and deconvolution combined with a binaural source cancellation algorithm (Usman et al., 2008; Keyrouz et al., 2006). HRTF-based techniques are also able to use the differential filtering produced by the pinna for sources in the front or rear hemifield to resolve position using only two microphones (Usagawa et al., 2011).

Line arrays parallel to the interaural axis (broadside) are unable to determine whether a source is in front or behind of the listener using two-microphone DOA techniques without HRTF information. This is due to the identical time delays produced by sources at the same angle in the front or rear hemifield (i.e. the cone of confusion, see section 2.2.4). Systems using two-dimensional array shapes with more than two microphones can resolve front-back DOA estimates. Examples of these systems include a three-microphone array that uses a summation of GCC estimates (Kwon et al., 2008), a four-microphone square array developed for use with mobile robotic systems (Li et al., 2011), and a circular, eight-microphone array (Pavlidi et al., 2012). Due to the size and form factor of most hearing aids, these techniques are not suitable for use with hearing aids.

GCC-PHAT and AED DOA techniques have been shown to track moving sources in low reverberation and background noise (Rohdenburg et al., 2008b; Benesty, 2000). Under optimal conditions, a two-microphone AED system has been shown to simultaneously track one static and one moving source at the same time, whereas the equivalent GCC-PHAT technique jumps between the two sources (Aichner et al., 2006). No research on compensating for head movements during DOA estimates has been published, though the result would be expected to be similar to tracking results for moving sources. However, DOAs would potentially move much faster with head movements. The additional information provided by head movements over time could be used to improve the performance of GCC-PHAT, avoiding the computational complexity and two-source DOA estimation limit for two-microphone AED systems.

3.12 MEMS inertial measurement systems

In order to be viable for real-world use, head-motion tracking for hearing aids cannot use external points of reference. This requires the use of inertial measurement sensors (IMUs), such as MEMS accelerometers, gyroscopes and magnetometers. The combination of these three, three-axis inertial measurement units is referred to as a "9-axis" sensor. MEMS cover the technology of very small devices, comprising components built at the 1 to 100 micrometre scale and ranging in size from 20 micrometres to 1 millimetre.

3.12.1 MEMS accelerometer

The accelerometer used in the 9-axis SparkFun sensor (SEN-10724) was the ADXL345. Accelerometers behave conceptually as damped masses on springs.

The mass is displaced when the whole system experiences an acceleration and this displacement can be measured in a number of ways.

MEMS accelerometers operate using several physical properties that change with linear acceleration and the ADXL330 uses a capacitive mechanism. The sensor itself is a polysilicon surface micro-machined structure built on top of a silicon wafer. Polysilicon springs suspend the structure over the surface of the wafer and are resistive to acceleration forces in X, Y and Z axes, resulting in highly orthogonal sense directions. A differential capacitor measures the deflection of the moving mass at the centre of the polysilicon springs using capacitive plates fixed on the wafer and moving mass. Acceleration deflects the moving mass, which in turn unbalances the differential capacitor and produces an output from the sensor that is proportional to acceleration. The ADXL330 has a measurement range of $\pm 3g$. Readings are output digitally using the I^2C protocol.

Accelerometers measure linear acceleration and translational motion and therefore measurements of head rotations, an angular acceleration, may produce noisy results. As an accelerometer requires a double integration to obtain orientation information, numerical integration errors (see appendix 3.12.4) are increased by a factor of approximately two in comparison to the gyroscope. However, in applications that do not produce large centrifugal forces, such as head movement, accelerometers provide a reliably invariant reference of acceleration due to gravity.

3.12.2 MEMS gyroscope

The gyroscope used in this research was the Invensense ITG-3200, included in the Sparkfun ITG3200 breakout board and the 9-axis SparkFun sensor. It is a single-chip, digital-output, 3-axis gyroscope. Gyroscopes measure angular velocity and utilize the Coriolis force, most commonly associated with rotating weather systems (Wallace and Hobbs, 2006). MEMS gyroscopes use vibrating instead of rotating structures, as vibrating structures of similar accuracy to rotating structures are easier and cheaper to fabricate. The physical mechanisms used for detection of the Coriolis force range from piezoelectric, resonating and "tuning fork" sensors to vibrating wheels. The mechanism used by the ITG-3200 is similar to the ADXL330/345 accelerometer. One capacitive plate of a vibrating element is deflected during rotation due to the Coriolis force. This deflection causes a change in capacitance between the deflected and fixed plate. This capacitance is converted to a voltage that is proportional to the angular velocity of the sensor. The maximum range of the gyroscope is $\pm 2000^{\circ}/s$. Readings are output digitally using the I^2C protocol.

3.12.3 MEMS magnetometer

The magnetometer used in the SparkFun 9-axis sensor is the Honeywell HMC5883L. This magnetometer operates on the principle of anisotropic magnetoresistance, where a material's electrical resistance depends on the angle between an electric current and the direction of a magnetic field. A change in resistance across the resistive elements of this type of sensor corresponds to a voltage change, which can be related to the direction of a magnetic field. Three elements are aligned along a common axis,

resulting in a positive voltage change with increasing magnetic field along that axis. Additional sensors are placed orthogonally to the first to enable three-dimensional measurements of the magnetic field. The unit has a 2 mG field resolution in $\pm 8g$ fields, corresponding to a $1-2^{\circ}$ compass accuracy after proper calibration. Readings are output digitally using the I^2C protocol.

Other MEMS magnetometers have been designed to detect displacement of a current-carrying conductor in a magnetic field due to the Lorentz force. The mechanical displacement can then be detected using Wheatstone bridges or optical measurements.

3.12.4 DCM algorithm overview

The sensors are combined using the discrete cosine matrix algorithm (DCM). The DCM algorithm was first applied to MEMS inertial sensors housed in model airplanes (Premerlani and Bizard, 2009; Mahony et al., 2008). The software used to implement the DCM algorithm and produce estimates of head movement was developed at TU Berlin. A tutorial and more information on the software used can be found online (Bartz, 2012). A diagram of the algorithm is shown in figure 3.4. The gyroscope is used as the primary source of inertial information, integrating the rotational velocity over time to obtain its change in orientation. As the measurements are samples of a continuous movement, numerical errors violate the orthogonality constraints of the DCM. This requires small adjustments to the elements of the DCM to satisfy the required orthogonality. The accumulated gyroscope drift errors are detected by using invariant reference vectors and a proportional plus integral (PI) negative feedback controller (Ang et al., 2005). This reduces the errors in the DCM

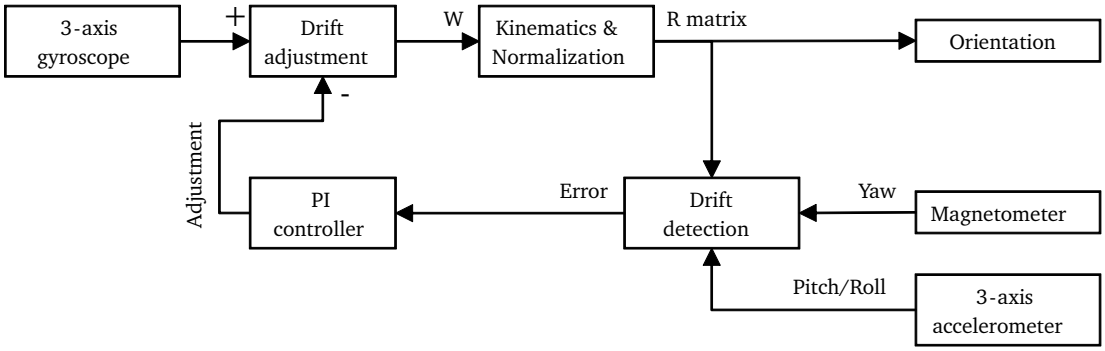

Fig. 3.4: The DCM algorithm for measurement drift cancellation

elements faster than they can accumulate. A calibrated magnetometer was used to detect yaw error, while the accelerometer was used to detect the error in pitch and roll.

3.13 Research using micro-electromechanical systems and multi-modal hearing aids

This section will focus on research into human movement using combinations of 3-axis MEMS accelerometers (measuring acceleration due to gravity), gyroscopes (measuring orientation and in-plane velocity) and magnetometers (measuring magnetic field strength and direction) IMUs. IMUs are increasingly found in consumer electronics, from game-controllers to smartphones. They have been utilized for head tracking in virtual acoustics (Bartz, 2012). IMUs have the potential to resolve DOA estimation issues for moving sensor arrays such as hearing aids. Resolving these issues could make the adaptation of directional systems such as beamformers more robust.

Much of the research undertaken with IMUs has been in the fields of robotics (e.g. Wongwirat and Chaiyarat, 2010), head tracking for mobile phones (e.g.

Pörschmann, 2007) and pedestrian tracking (e.g. Kim et al., 2012). An object can be easily tracked by a rudimentary sensor if the sensor remains in a fixed position and the object to be tracked has recognizable markings (Brimijoin and Akeroyd, 2012). For true freedom of movement however, the tracking must be performed by a system based entirely on the object itself, without an external measurement point. With a continuously measuring, calibrated analogue 3-axis accelerometer, the combination of the acceleration output and vector calculus would give the correct orientation of the object at all times. However, most MEMS devices are digitally sampled and some sampling and measurement error builds up over time.

One area of multi-modal hearing-aid research has focused on body movements and eye movements as additions to audition. These modalities were introduced for improved context recognition for hearing instruments (Tessendorf et al., 2010). Context recognition in this case means detecting the type of listening situation the listener is currently in based on sensor information from microphones, IMUs and eye sensors and a pre-trained classifier. It allows hearing aids to automatically select the programs and program parameters that are broadly correct for the listening situation, such as an appropriate directional setting for one-to-one speech in noisy environments (Lamarche et al., 2010).

Initially these acoustic, body (attaching sensors to the arms and legs), head and eye-movement modalities were used to distinguish between two situations. The first situation was named “Conversation” and included situations in which the listener was conversing while sitting or standing. In this case the hearing aid should be optimized for speech intelligibility. The second situation was named “Work” and included situations in which the listener was not talking, with no

conversation taking place near them or a conversation was happening nearby in which the listener was not involved. This could also occur while the listener was sitting or standing (Tessendorf et al., 2011a). The other modalities were useful for determining whether the person was sitting or standing and the direction of the listener's attention. Body-movement detection was performed using nine IMUs and eye movement was measured using electro-occulography. Sound was also recorded using modified hearing aids. Using the additional modalities, context recognition was improved from 77% accuracy to over 92% combining all modalities. Eye movement was the most useful part of the body to track, achieving a higher recognition rate (86%) than head movement alone (84%) and a lower standard deviation. Sensors placed on other parts of the body, such as the arms and legs, provided little benefit for context recognition.

Identification of relevant multimodal cues for context recognition (Tessendorf et al., 2011b) led to the design and testing of an ear-worn, combined IMU and sensor data recorder (Tessendorf et al., 2012). Analysis of listener head movements during different types of listening showed that acceleration data could be used to distinguish directed conversation from undirected listening and sitting or standing from walking. The device size is currently similar to a hearing aid, with a runtime of five hours (Tessendorf et al., 2013). This research has shown it is possible to include head-movement detection in hearing-aid signal processing and the sensors used can be housed in hardware similar in size to most current BTE hearing aids.

Selecting between simultaneous talkers has been shown to be faster using eye tracking than using either pointing or button selection (Hart et al., 2009). In the same study, eye tracking was rated by listeners as the easiest and most

natural way to select a talker. This has been implemented in a prototype hearing aid using a high-order, monaural beamformer whose main lobe could be steered by eye movement (Kidd and Favrot, 2013). In one part of the testing, concurrent target digits were presented from 5 spatially separated loudspeakers spanning $60°$. Listeners heard the digits remotely, through a KEMAR head or the eye-tracking microphone array mounted on the KEMAR. In a separate room, the listener wore an eye-tracker and was presented with a screen showing the loudspeakers from KEMAR's viewpoint. The target digit to be selected by the listener was preceded by a visual cue on the screeen at the correct loudspeaker, one second before presentation. Eye movement was similar in both the KEMAR and eye-tracking beamformer conditions and the lag in steering the main lobe of the beamformer using eye tracking was negligible. This produced similar target digit identification results in both conditions, meaning that the monaural eye-tracking beamformer was at least as good as the binaural-listening condition for this test. How this would translate to real-world benefit is unclear.

A patent exists for the use of a gyroscope for calibrating the three-dimensional orientation of hearing aids on the ear before use (Ho et al., 2008). Another patent proposes an accelerometer to detect head nods for switching between programs (Samuels, 2011).

3.14 Industrial perspectives on the internalization/externalization continuum

A research trip to four leading hearing-aid manufacturers (Siemens, Phonak, GN ReSound, and Oticon) in Europe was undertaken early in the project to gain industrial perspectives on internalization/externalization and hearing aids.

Internalization was believed by Siemens to be a result of inadequate modeling and implementation of pinna effects, relegating the other research areas to the control of secondary effects on internalization. Pinna effects were also thought important at Oticon. Acclimatization to new cues was not considered to be important, whereas at Phonak acclimatization was suggested as a possible factor in externalization.

At Phonak and Oticon, externalization was thought of as a continuum, rather than two separate phenomena separated by a verged-cranial condition. GN ReSound suggested that externalization be treated as a cost function to be minimized. Phonak also suggested that any attempt to reduce internalization would mean a trade-off between "naturalness" and intelligiblity. At Phonak, Oticon and GN Resound, occlusion and acoustic coupling in the ear canal were thought to be a major factors in the internalization of sounds

Testing normal-hearing listeners, or at least more tightly defining the type of hearing-impaired listener in the research was suggested at Oticon and Phonak. This would make it more likely that usable results be obtained from the research and the effect of only the hearing aid on the internalization/externalization continuum be found.

A strong case for the preservation of ITD cues was made during a hearing-aid demonstration at Oticon. Removing ITD cues resulted in complete internalization, whereas preserving them while listening through hearing aids preserved externalization, though sound sources were perceived as closer.

Several issues were clearly identified from the industrial perspectives. Firstly, helpful input on focussing the research was received, such as defining the type of hearing-impaired listener who could benefit from the research The reinstatement of HRTFs and the effects of occlusion were highlighted as viable areas of research.

CHAPTER 4

Psychoacoustic experiments on internalization/externalization

4.1 Introduction

This chapter describes the psychoacoustic experimental research carried out on the internalization/externalization continuum, including initial experimental designs, the effect of microphone position, bandwidth and number of talkers, source angle and acclimatization on externalization in normal-hearing and hearing-impaired listeners. The experiment described in section 4.3 was published in: Boyd, A.W., Whitmer, W. M., Soraghan, J. J., Akeroyd, M. A. (2012) Auditory externalization in hearing-impaired listeners: The effect of pinna cues and number of talkers. *J. Acoustic Soc. Am.* **131**(3):EL268-EL274.

Two initial experiments on auditory distance perception (determining the distance to the origin of a sound) were designed and piloted (section 4.2). The first (initial) experiment treated reduction in speech reception thresholds as virtual listener to talker distance was increased as analogous to a reduction in the perception of externalization experienced. The second (initial) experiment

allowed listeners to use a visual response to estimate the perceived auditory distance to virtual talkers.

The results of the initial experiments (section 4.2) and industrial perspectives (section 3.14) highlighted that the ability to reliably produce an internalized percept (such as headphone presentation) should be considered in a new experimental design. The experiment described in section 4.3 investigated the importance of high-frequency pinna cues for externalization in normal-hearing and hearing-impaired listeners. Participants rated the degree of externalization using a multiple-stimulus test in increasingly complex listening situations and using progressively reduced pinna cues. Pinna cues were reduced by recording impulse responses in and above the ear, mimicking ITE and BTE microphone placement. Stimuli were also lowpass filtered to the bandwidth of a hearing aid. It was found that hearing-impaired listeners experience a compressed or "flattened" perception of externalization in relation to pinna cues.

Informal tests by the author while wearing hearing aids and preliminary results of the Perception of Internalization Questionnaire (PIQ, section 5.2) suggested that short, loud, impulsive and high-frequency sounds may be most readily perceived as internalized. Therefore a novel experiment (section 4.4) was developed to investigate the role of changes in the stimulus and source DOA on the perception of externalization by hearing-impaired listeners with and without their hearing aids. The effect of short-term acclimatization to hearing aids by normal-hearing listeners performing the same task was also investigated. An effect of angle and no effect of the hearing aid on externalization was found for hearing-impaired listeners. An effect of acclimatization was found for normal-hearing listeners.

The remainder of this chapter is organized as follows: Section 4.2 describes the initial psychoacoustic experiments on auditory distance perception. Section 4.3 shows the design, results and discussion of the headphone-based externalization experiment. Section 4.4 gives the design, results and discussion of the loudspeaker-based externalization experiment. Section 4.5 summarizes the findings.

4.2 Initial experiments

Two experiments were initially designed and piloted on auditory distance perception. Table 4.1 outlines their design and results.

Title	Method	Results
The effect of source to listener distance on speech intelligibility	Speech reception thresholds (SRTs) measured for variable talker-listener distance in the presence of two maskers	No significant differences between conditions found
Perception of auditory distance using a visual response paradigm	Auditory distance perception measured using virtual audio and novel visual response paradigm	Responses showed over-estimation of distance in the near field and under-estimation in the far field. This was in agreement with previous studies (Zahorik, 2002a).

Table 4.1: Summary table of initial experiments.

4.2.1 The effect of source to listener distance on speech intelligibility

The experiment was conducted using a 0.9 m radius, 24-loudspeaker (Phonic SEp 207 loudspeakers) ring. If the experiment had continued beyond the pilot stage, the use of loudspeakers would have enabled hearing-impaired listeners to be tested with their hearing aids, avoiding the task of modelling the listener's hearing aid over headphones. The image-source artificial reverberation technique was used to model a virtual room (Allen and Berkeley, 1979). The technique is described in more detail in appendix A.

Five normal-hearing listeners were tested. The stimuli used were sentences from the IEEE York corpus (Stacey and Summerfield, 2007). Listeners were tested using a one-up, one-down adaptive track to find their threshold of intelligibility as a function of the virtual distance to a target talker. Listeners responded by repeating the sentence heard. Three step sizes were used for the adaptive track: 4 (reducing step size after 1 reversal), 2 (2 reversals) and 1 dB (4 reversals). An incorrect repetition of the sentence would result in a reversal. The talker and maskers had a level of 68 dB when presented from a virtual distance of 2 m from the listener. The starting position of the talker was randomly varied between 2.8 and 3.5 m (corresponding to a roving level of 5 dB). Four conditions were used, comprising two maskers and a target, either spatially collocated directly in front of the listener or with distracters separated by $\pm 30^{\circ}$ from the target (figure 4.1). Both spatial conditions were presented in a virtual acoustically dampened room or RT_{60} = 0.4 s.

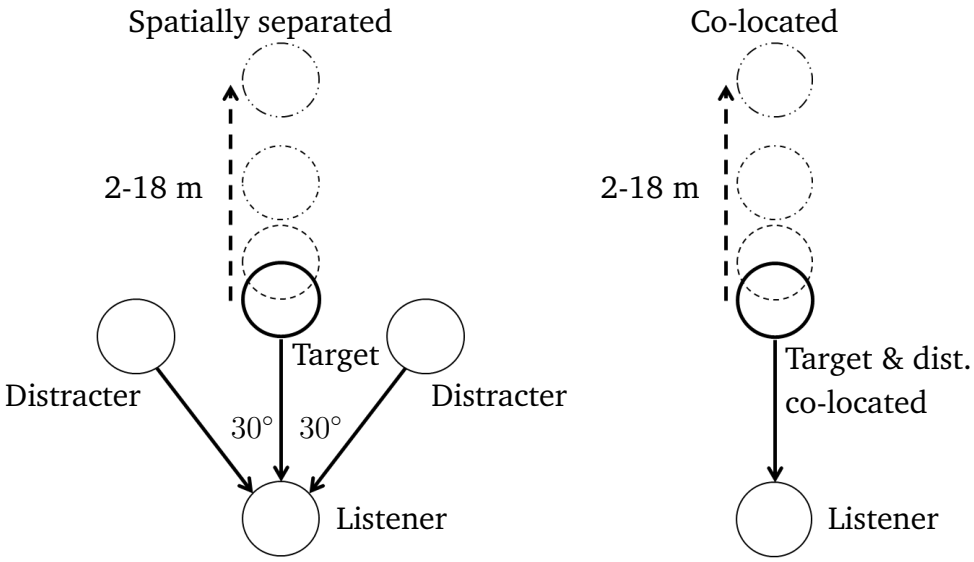

Fig. 4.1: Positions of the talker and maskers in each condition

Figure 4.2 shows the results. Larger distances mean the listeners are performing better at the task (i.e. achieving a lower SRT score). The thresholds were not significantly different from one another, therefore little comment can be made about the effect of the conditions on SRT scores. It is possible that reverberation may be helping some listeners achieve better scores in the spatially separated condition.

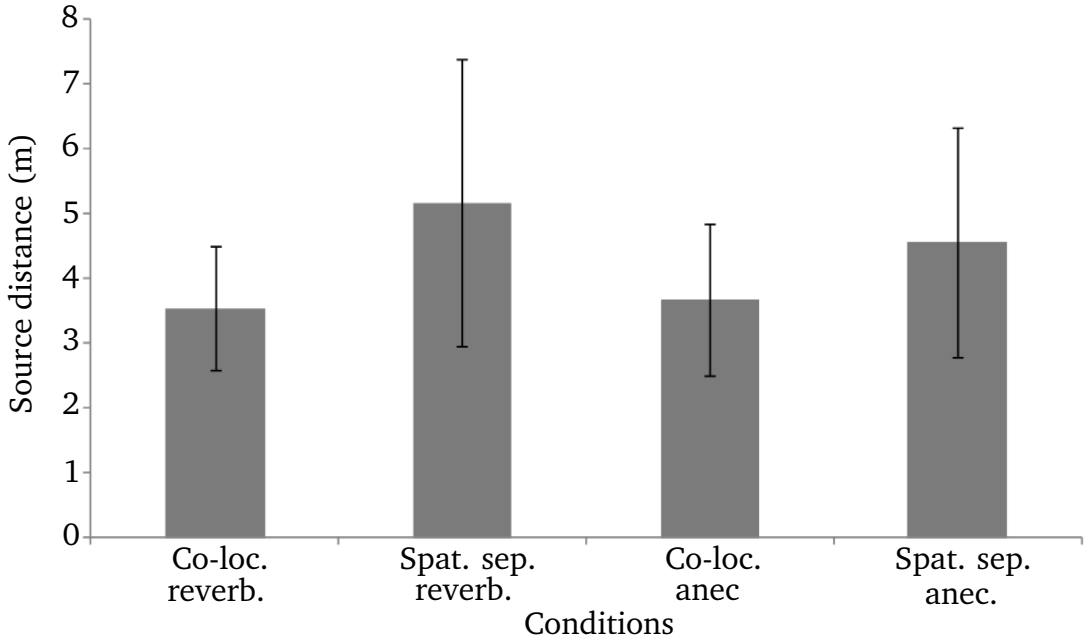

Fig. 4.2: Speech reception thresholds for five normal-hearing listeners for spatially co-located and separated talker and maskers, in reverberation and anechoic conditions. Plotted with one standard deviation of the mean.

It is unclear from the initial data whether any significant differences would result from further testing. Speech reception thresholds as a function of distance may relate to externalization for near-field sources (Brungart and Simpson, 2002), but there is no evidence to suggest this for speech presented in the virtual far-field.

4.2.2 Perception of auditory distance using a visual response paradigm

In the next experiment, listeners were required to estimate auditory distance using a visual response. An improved image-source method and 0.9 m radius,

24-loudspeaker ring was used. See appendix A for a description of the image-source method.

Three talkers were placed at $-30°$, $0°$ and $+30°$ around the listener, who faced $0°$. A model of the room was built using Google SketchUp, displaying the talkers at the positions modeled by the image-source technique. The visual images of the talkers were moved incrementally and combined as a stop-frame animation. Listeners could move the talkers smoothly by scrolling through the stop-frame movie to the distance at which they perceived the talkers to be. Distances varied, both aurally and visually, between 0.25 and 4 metres. Figure 4.3 shows the two visual extremes seen by the listener. There were five talker positions; 0.25, 0.5, 1, 2 and 4 metres. Listeners could move the visual representations of the talkers to 16 possible positions that increased logarithmically from 0.25 to 4 metres.

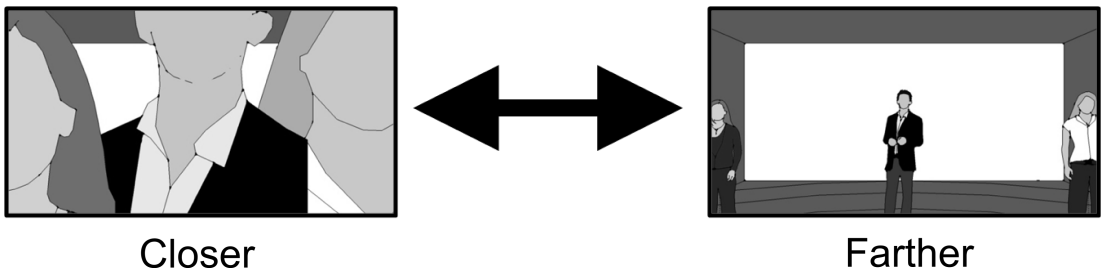

Fig. 4.3: GUI for the visual estimate of auditory distance, showing the possible extremes (0.25 and 4 m).

Six normal-hearing listeners participated, with better-ear four-frequency averages < 20 dB HL. Listeners were given a training period, where the talkers were presented aurally and visually at each modelled distance. During the experiment, listeners heard the talkers first and were then asked to move the visual representations of the talkers to the visual distance they perceived them.

Sounds could not be replayed. The visual start point was at 1 metre after every presentation to prevent order bias. Each condition was presented twice.

Figure 4.4 shows the results plotted with 95% confidence intervals computed from the standard error of the mean. In the near-field, most participants over-estimated the visual distance from the auditory stimulus, while in the far-field, all participants under-estimated the distance. There is a clear knee-point in the results around 1 metre: visual estimates increase significantly (tested using a two-tailed t-test) with simulated auditory distance up to 1 metre (0.25-0.5m, $p < 0.001$; 0.5-1m, $p < 0.01$) and then do not increase significantly after 1 metre.

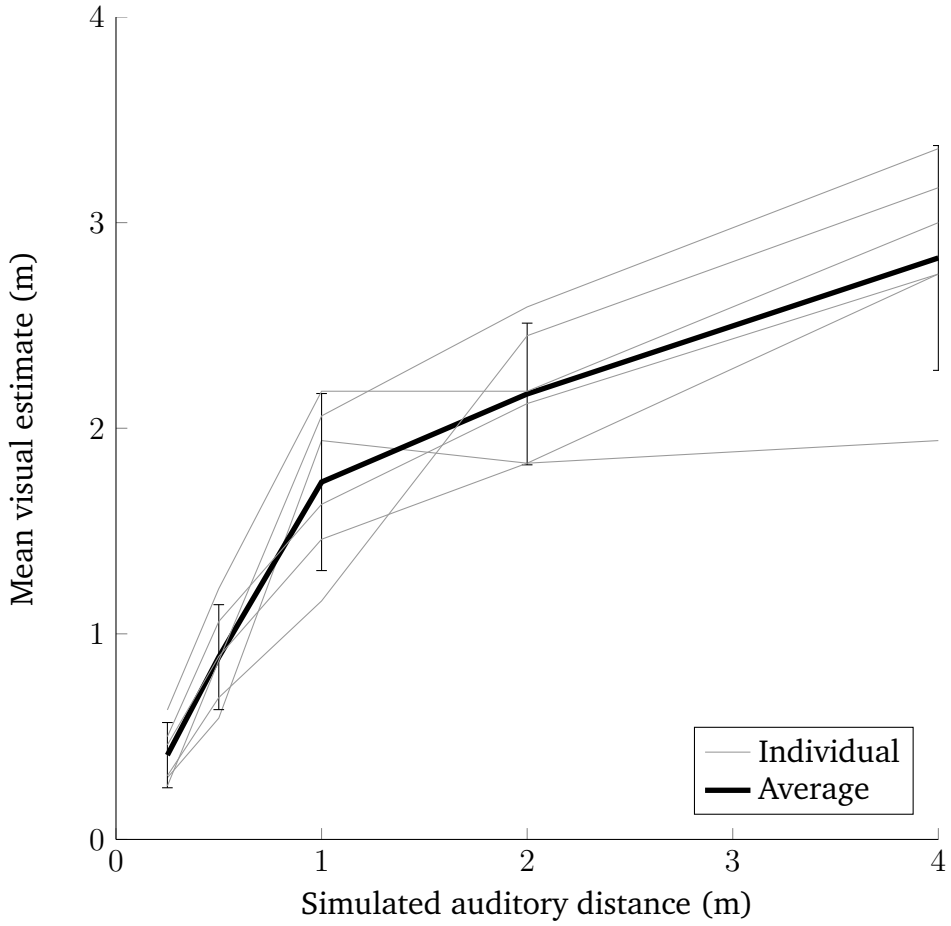

Fig. 4.4: Mean source-listener distance estimates vs. modeled source-listener distance plotted with 95% confidence intervals computed from the standard error of the mean.

These results agree with work by Zahorik (2002a) on auditory distance perception, who also found that the distance to near-field sources is over-estimated and far-field sources are under-estimated. The experiment was a proof of concept for a novel way of measuring auditory distance perception. However, the experiment did not allow listeners to respond to a sound perceived as internalized, as the allowed response space did not extend into the head.

4.2.3 Discussion of initial work

The internalization/externalization continuum may, based on previous literature (see chapter 2) to be simply a sub-section of auditory distance perception. This assumption formed the basis for both initial experiments. With the benefit of subsequent research, the continuum appears to be as related to angular localization as it is to auditory depth perception for hearing-impaired listeners (see section 4.4). The first experiment was not designed to measure the continuum, only a factor (SRTs as a function of simulated distance) that may have been analogous to it. The second experiment measured distance perception, but it had no definitively "internalized" condition and no way for the listener to respond to a perceptually internalized sound, as all possible responses were external. While both offered novel methods to examine auditory distance, neither produced novel results, nor offered an insight into the internalization of sounds. After these initial experiments, the experimental focus shifted towards stimuli that could be internalized and the cues that lead to externalization.

4.3 The effect of pinna cues and number of talkers

The experiment described in this section compared the sensitivity of hearing-impaired and normal-hearing listeners to externalization cues when listening with their own ears and simulated behind-the-ear (BTE) hearing aids. This was carried out in increasingly complex listening situations and using progressively reduced pinna cues. Participants rated the degree of externalization using a multiple-stimulus listening test for mixes of internalized

and externalized speech stimuli presented over headphones. Previous research has shown that hearing-impaired listeners are sensitive to changes in the HRTF and that this relates to a change in externalization rating (Ohl et al., 2009).

4.3.1 Simulation of open-ear listening over headphones

The experiment simulated open-ear listening over headphones using a modification of a method used by Ohl et al. (2009), after Wightman and Kistler (1989a;b). Figure 4.5 shows the signal path used to produce open-ear transfer functions. All impulse responses (time domain) and hence transfer functions (spectral domain), were extracted using the swept-sine technique given by Berdahl and Smith (2008) and described in more detail in appendix B.

For the in-the-ear microphone position (ITE condition) the in-ear microphones (Sound Professional MSTFB-2) were placed at the entrance to the ear canals. For the behind-the-ear microphone position (BTE condition), ear-hooks with integrated cable-guides (supplied with the Sound Magic PL30 in-ear headphones) were used to place the microphones above the front portion of the pinna, simulating the microphone position of a BTE hearing aid. These positions also produced the (theoretically maximally externalized) head-present impulse responses. For the (theoretically maximally internalized) head-absent conditions the in-ear microphones were placed on a horizontal bar, 18 cm apart and at the same height as the loudspeakers (1.2 m). Eight concatenated swept-sine signals (X in figure 4.5) were played from each loudspeaker in succession at 75 dBA and simultaneously recorded by the microphones. The transfer functions of the extracted impulse responses contained the spectral information from the loudspeaker (($\mathrm{H}_{LS}(\omega)$) and

binaural room (BRTF$_{L,R}(\omega)$), as shown in figure 4.5. The BRTF$_{L,R}(\omega)$ could refer to the the ITE, BTE (head present), or head absent conditions.

The stimuli were spectrally equalized for headphone playback by presenting the swept-sine signals over headphones and recording them through the microphones in the ITE position (H$_{HP,L/R}(\omega)$ in figure 4.5). In the frequency domain, using the inverse of the H$_{HP,L/R}(\omega)$ for equalization could result in large peaks in the filter and small variations in the position of the headphones in relation to the ITE microphones could vary the filter shape (Kulkarni and Colburn, 2000). To reduce these effects, the headphones were removed by the participant after presentation of two swept-sine signals and replaced before recording again (for a total of eight signal presentations). Applying the average of the H$_{HP,L/R}(\omega)$ to the BRTF$_{L,R}(\omega)$ as an inverse filter created the headphone-equalized binaural-impulse responses (HEBIR$_{L,R}(\omega)$ in figure 4.5) for headphone playback. As all signals were recorded using the same microphones, the microphone transfer function (H$_{mic,L/R}$ in figure 4.5) was also removed.

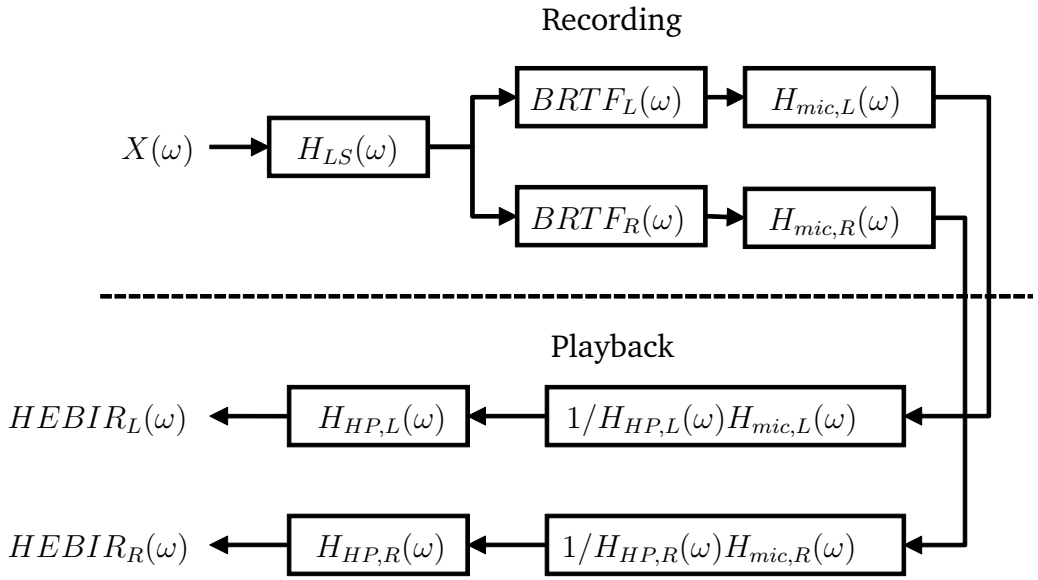

Fig. 4.5: The signal path used to create open-ear transfer functions.

4.3.2 Methods for the investigation of the effect of pinna cues and number of talkers

The listener's head was not fixed; however, a fixation point on the wall facing the listener was provided which helped the listeners obey instructions to maintain a fixed head. Binaural room impulse responses (BRIRs) were recorded with microphones at ITE and BTE positions, with the listener head present (theoretical maximally externalized condition) and head absent, the microphones being placed on a horizontal bar in the same position as the listener's ears (theoretical maximally internalized condition). The recording positions are shown in figure 4.6. Participants rated the degree of externalization using a multiple-stimulus listening test for five mixes of the head-absent and head-present HEBIRs convolved with speech stimuli presented over headphones. The mix parameter was the fraction (expressed in

percent) of head-present recording power compared to the sum of head-present and head-absent recording power. Eight stimulus conditions were created, comprising all combinations of microphone position (ITE and BTE), frequency response (broadband and lowpass filtered), and number of talkers (1 and 4).

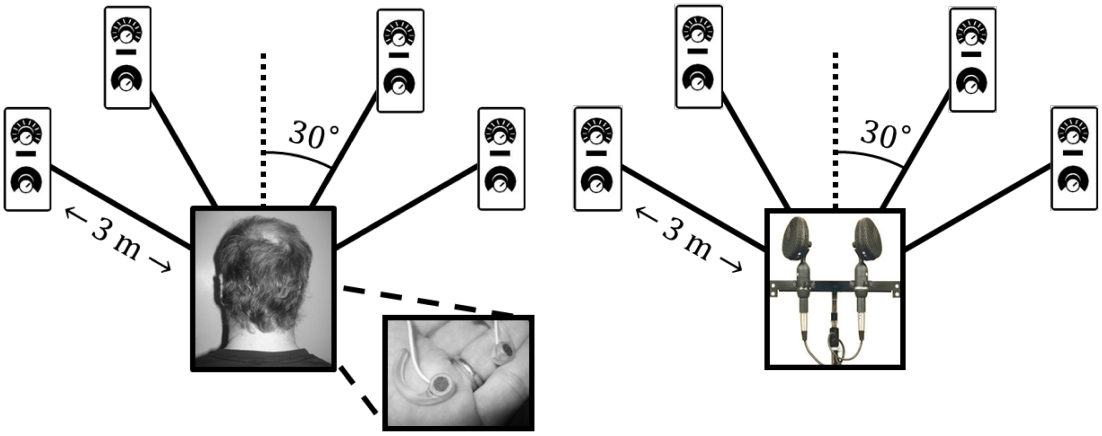

Fig. 4.6: Diagrams of the experimental apparatus used to record the head-present (left panel) and head-absent (right panel) conditions.

Participants. Seven normal-hearing (1 female) and 14 hearing-impaired (6 female) listeners participated. Normal-hearing listeners had a better-ear four-frequency (0.5, 1, 2, and 4 kHz) average (BE4FA) of less than 20 dB HL, and hearing-impaired listeners had a BE4FA of 34 dB HL and range of 21-51 dB HL. Average asymmetry for hearing-impaired listeners was 6 dB. Seven hearing-impaired listeners wore one hearing aid and two wore two at the time of testing; all who wore hearing aids wore BTE-type aids. The normal-hearing listeners had a median age of 27 years (range of 22-46 years) and the hearing-impaired listeners had a median age of 61 years (range of 48-72). Figure 4.7 shows the audiograms for the hearing-impaired listeners. Most had mild to moderate sloping losses and two listeners displayed a notch in

their audiogram around 4-6 kHz, indicative of a possible noise-induced hearing loss.

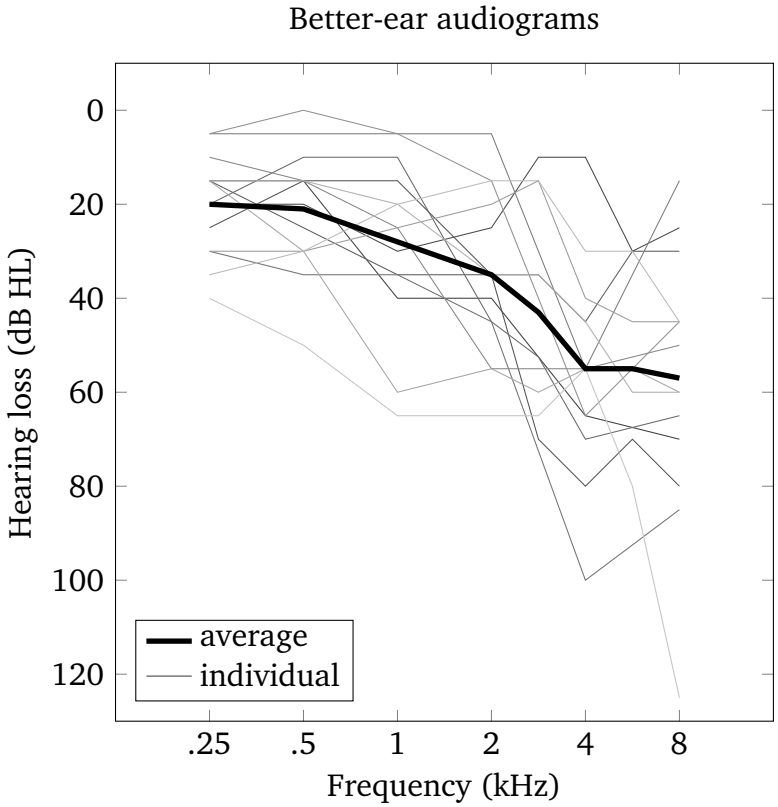

Fig. 4.7: Hearing-impaired participant better-ear audiograms. Individual audiograms are shown in greyscale. Average better-ear audiogram is shown in solid black.

Apparatus. Listeners were seated along the central short axis of a room measuring 6.5 $\times$ 5 $\times$ 3 m that was acoustically treated to perform as a hearing-impaired classroom under BB93 regulations 2003. The reverberation time (RT_{30}), measured at the same position as the listener's head, was 0.35 s. Four loudspeakers (JBL Control 1) where placed at a height of 1.2 m, at $\pm 30^{\circ}$ and $\pm 60^{\circ}$, as previous work has shown a similar degree of externalization across this range of angles (Kim and Choi, 2010). This configuration also provided a

large spatial separation while keeping all loudspeakers in the visual field of the listener. The loudspeakers were placed at a distance of 3 m from the listener.

Stimulus. The signals were 3 s of concatenated (or truncated) random sentences from the same talker of the IEEE York corpus (Stacey and Summerfield, 2007). The corpus was recorded at 16-bit, 44.1 kHz sampling rate. Spectrograms of the corpus recordings displayed sufficient speech energy above 6.5 kHz up to 15 kHz for high-frequency HRTF cues to be present in the convolved stimuli. For the one-talker condition, a male talker was convolved with HEBIRs at $\pm 30^{\circ}$. For the four-talker condition, female talkers were convolved with HEBIRs for the $\pm 30^{\circ}$ and $\pm 60^{\circ}$ positions, and male talkers were convolved with HEBIRs for the $\pm 60^{\circ}$ and $\pm 60^{\circ}$ positions. The broadband condition used the full bandwidth of the convolved stimulus. The lowpass condition was created by applying a tenth-order Butterworth filter with a cutoff at 6.5 kHz, approximating the bandwidth of a standard digital hearing aid. The convolved ITE/BTE, broadband/lowpass (BB/LP) sentences were mixed in the time domain with the same sentences convolved with the head-absent responses. The amount of ITE/BTE signal mixed in amplitude with the head-absent signal (the "mix point") varied from 0% (i.e., hypothetical maximum internalization) to 100% (i.e., hypothetical maximum externalization) in 25% increments. The head-absent responses were time-shifted if necessary before mixing, both within pairs and as pairs, to produce identical ITDs between the ITE/BTE and the head-absent convolved sentences. The root-mean-square (rms) values were standardized before and after each stage of the mixing process for both the one- and four-talker conditions. The playback level was 70 dBA, ensuring greater

than 15 dB sensation level for all listeners, using their worse-ear 4FA as a reference.

Test procedure. There were eight test conditions (a $2 \times 2 \times 2$ design) consisting of all combinations of the chosen parameters. There were 8 blocks of trials, each block consisting of 5 head-present to head-absent mix points and (if not already included in the mixes) the hidden reference. Each block was a test condition, presented in a randomized order during the experiment. The parameter combinations were: ITE/BB/1 talker; ITE/BB/4 talkers; ITE/LP/1 talker; ITE/LP/4 talkers; BTE/BB/1 talker; BTE/BB/4 talkers; BTE/LP/1 talker; and BTE/LP/4 talkers. They are also shown in table 4.2.

ITE/BB/1 talkers	ITE/ LP/1 talker	BTE/BB/1 talker	BTE/LP/1 talker
ITE/BB/4 talkers	ITE/ LP/4 talkers	BTE/BB/4 talkers	BTE/LP/4 talkers

Table 4.2: Parameter combinations for headphone experiment. ITE = in the ear; BTE = behind the ear; BB = broadband; LP = lowpass.

In training the listener was played the same sentence in the ITE/BB condition at five mix points in order from 100% to 0%. The participant was asked if the stimuli appeared to move toward them over successive plays, beginning at the loudspeaker. All participants reported this effect, indicating successful creation of an externalized sound for the ITE/BB condition with a mix point of 100% and the viability of the mixing technique.

The participant was trained in the use of the modified multiple stimulus with hidden reference and anchor (MuSHRA) test to rate the stimuli (ITU-R BS.1534-1,2003). The reference (and hidden reference) for all conditions was the ITE/BB stimulus with a mix point of 100%. The response screen, a diagram

of which is shown in figure 4.8, consisted of a row of five or six "mix" buttons (five in the ITE/BB conditions, as the reference was the same as the 100% mix point) and a slider corresponding to each button. Mix points were randomly assigned to each button and slider. Upon pressing a "mix" button, the reference stimulus was played followed by a 1 s pause for the target stimulus to be rated. The participant was instructed to rate the second target stimulus against the first, using the slider and a 0 - 100 point scale with five referents: "At the loudspeaker" (100); "In the room" (75); "At the ear" (50); "In the head" (25); "Center of head" (0). The referents for the scale were modified from Hartmann and Wittenberg's four-point scale (Hartmann, 1996; see figure 2.3). To prevent listeners responding after insufficient listening, each mix had to be played at least twice to enable progression to the next condition. The training was repeated for the four-talker, ITE/BB condition.

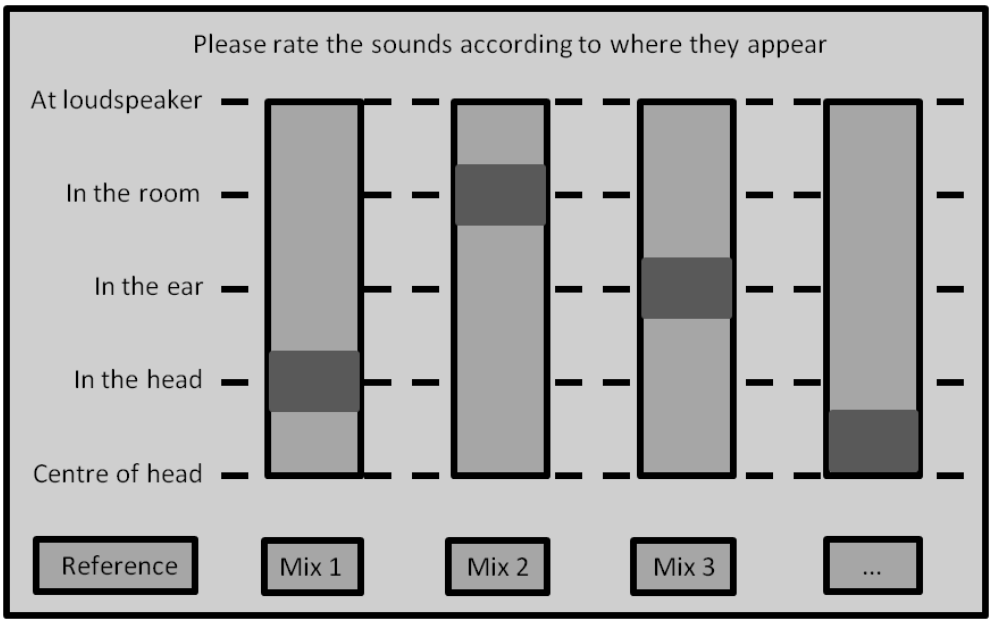

Fig. 4.8: A diagram of the modified MuSHRA interface used by the listener during the experiment. The "reference" uses the ITE, broadband impulse response for convolution, assumed to be the condition that would be most externalized.

4.3.3 Results of the investigation of the effect of pinna cues and number of talkers

Externalization ratings, with 100% at loudspeaker and 0% center of head, were computed from the average response for each mix point in each condition across normal-hearing and hearing-impaired listeners. Each point is based on 14 hearing-impaired listeners or 7 normal-hearing listeners. Table 4.3 displays the results of a between-subjects analysis of variance on the responses, showing statistically significant main effects and interactions. The main effect of the number of talkers was not significant. Interaction effects of mix and bandwidth, talkers and microphone position and talkers and bandwidth were not significant. Three-way interactions were tested, but were not shown in Table 4.3 as they were not significant. A large effect of mix was found ($\eta^2 = 0.21$).

Figure 4.9 shows the results for the one-talker conditions for both normal-hearing and hearing-impaired listeners plotted with 95% confidence intervals computed from the standard error of the mean. In the ITE/BB condition, normal-hearing listeners fully externalized ("at the loudspeaker") the 100% ITE mix and perceived the stimuli to move toward the head as the ITE mix decreased, with a minimum externalization between "at the ear" and "in the head." In contrast, the hearing-impaired listeners did not fully externalize the 100% ITE/BB stimuli. Hearing-impaired listeners also show greater variability in their placement of the reference condition. Normal-hearing listeners experienced a reduction in maximum externalization – rating for the 100% mix – across the other conditions. This reduction is similar in these

Source	Effect size (η^2)	*df*	F	*p*
HI/NH	**0.024**	**1**	**26.0**	**< 0.001**
Mix	**0.211**	**4**	**57.0**	**< 0.001**
Talkers	0.001	1	1.0	0.3
MicPos	**0.013**	**1**	**13.9**	**< 0.001**
Bandwidth	**0.019**	**1**	**20.2**	**< 0.001**
HI/NH $\times$ Mix	**0.067**	**4**	**18.0**	**< 0.001**
HI/NH $\times$ Talkers	**0.005**	**1**	**5.3**	**< 0.05**
HI/NH $\times$ MicPos	**0.020**	**1**	**22.1**	**< 0.001**
HI/NH $\times$ Bandwidth	**0.006**	**1**	**6.0**	**0.015**
Mix $\times$ Talkers	**0.023**	**4**	**6.1**	**< 0.001**
Mix $\times$ MicPos	**0.019**	**4**	**5.0**	**< 0.001**
Mix $\times$ Bandwidth	0.003	4	0.9	0.48
Talkers $\times$ MicPos	0.001	1	1.2	0.28
Talkers $\times$ Bandwidth	< 0.001	1	0.0	0.86
MicPos $\times$ Bandwidth	**0.012**	**1**	**13.4**	**< 0.001**
Error	0.585	631		

Table 4.3: Analysis of variance of responses. Statistically significant effects are in bold. The first column is the source of the effect, column 2 is the effect size (η^2), column 3 the degrees of freedom (*df*, followed by the F ratio and probability (*p*) value for the effect.

conditions, with responses becoming more variable in the BTE/LP condition. The hearing-impaired listeners appear relatively unaffected by microphone position and frequency response, placing maximum externalization "in the room" and minimum externalization "at the ear." A two-tailed t-test revealed that hearing-impaired listeners also rate the 100% head-absent condition as significantly more externalized than normal-hearing listeners in both lowpass conditions and the BTE broadband condition ($p < 0.05$).

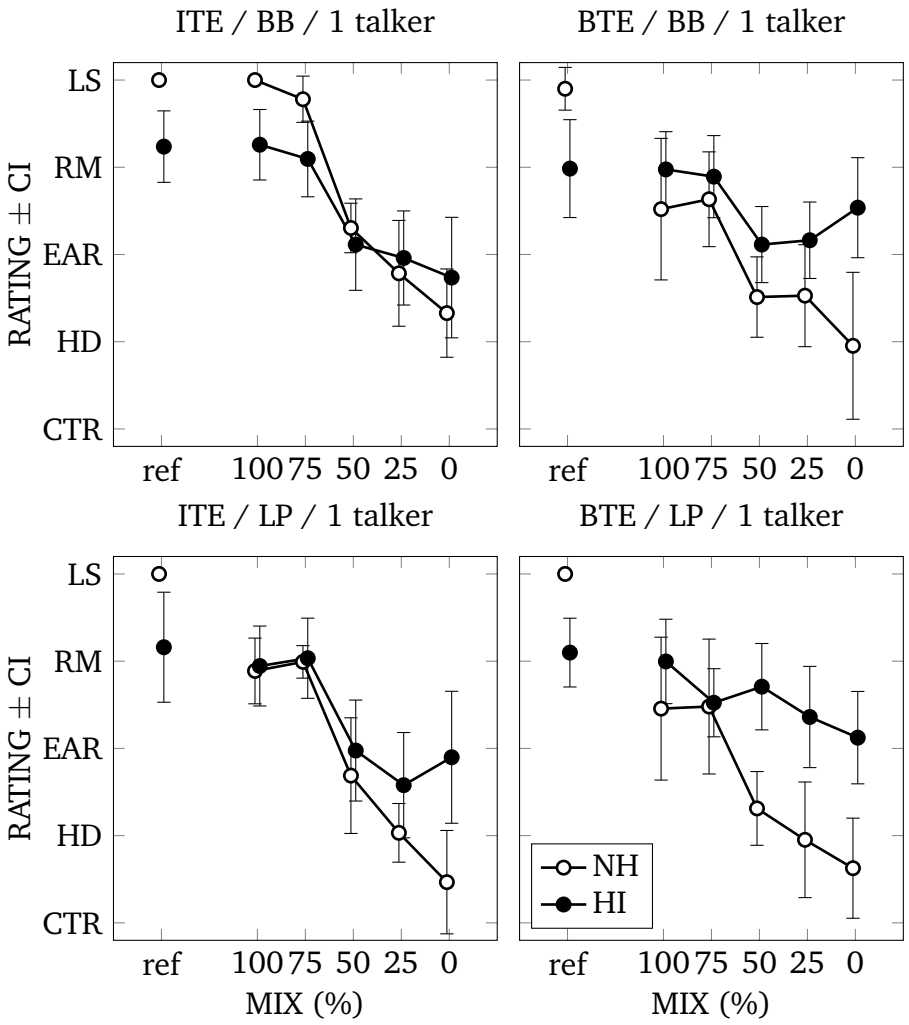

Fig. 4.9: Average externalization ratings of 1 talker for normal-hearing (open circles) and hearing-impaired (closed circles) participants against mix point as a function of microphone position (ITE/BTE) and frequency response (BB/LP). The reference condition (ref) is the same as ITE/BB. Error bars show 95% confidence intervals.

Figure 4.10 shows the average normal-hearing and hearing-impaired listener responses for the four-talker conditions plotted with 95% confidence intervals computed from the standard error of the mean. Normal-hearing listeners display similar responses in the four-talker conditions to their one-talker equivalent conditions. Overall, the hearing-impaired listener results display

both a reduced maximum externalization and internalization in comparison to the one-talker conditions, with all results lying between “in the room” and “at the ear”. Maximum externalization occurs for 50% mix in the ITE/LP condition and 0% mix (100% head absent) in the BTE/LP condition, perhaps indicating more variability in responses and a general lack of difference across the mix variable. Hearing-impaired listeners rate the 100% head-absent condition as significantly more externalized than normal-hearing listeners in both BTE conditions ($p < 0.05$).

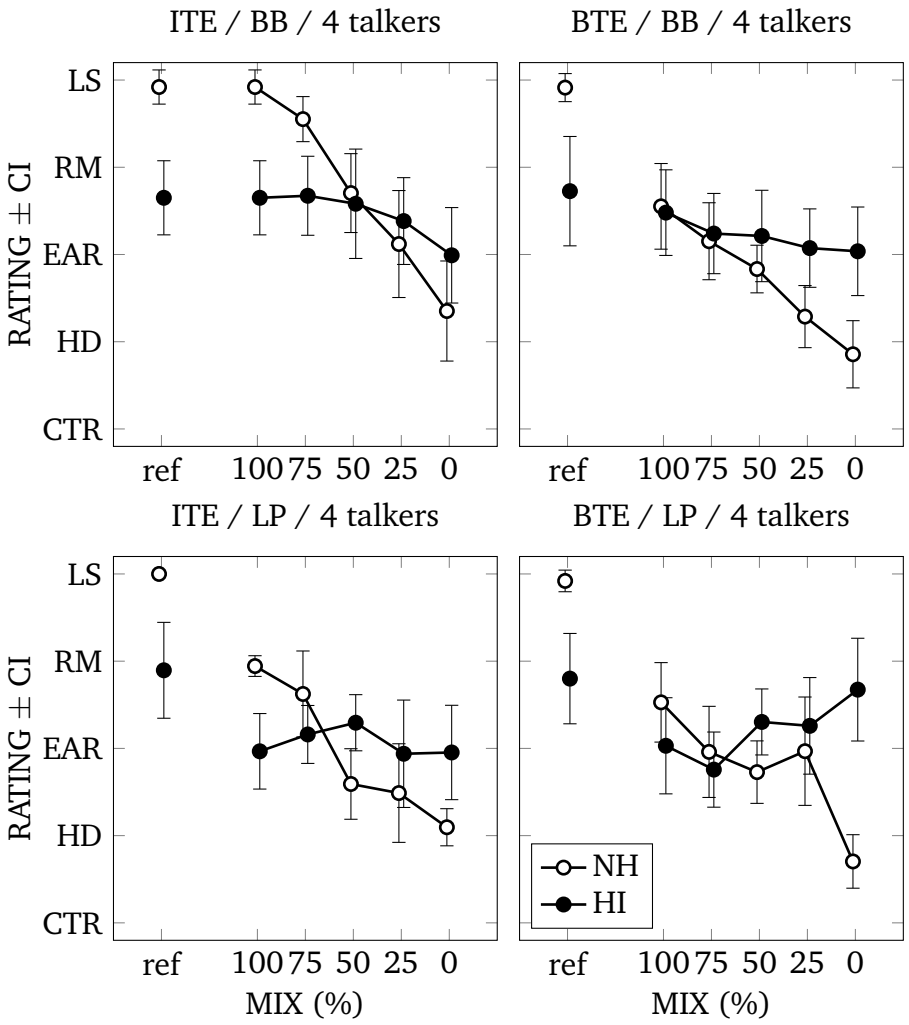

Fig. 4.10: Average externalization ratings of 4 talkers for normal-hearing (open circles) and hearing-impaired (closed circles) participants against mix point as a function of microphone position (ITE/BTE) and frequency response (BB/LP). The reference condition (ref) is the same as ITE/BB. Error bars show 95% confidence intervals.

Figure 4.11 shows the individual mean ratings across mixes as a function of high-frequency hearing loss, calculated as the average of the 6 and 8 kHz pure-tone thresholds in the better ear. The correlation of high-frequency hearing-loss with externalization of the reference condition was statistically significant ($r = -0.61$, $p < 0.05$), but for externalization of the head-absent

condition it was not statistically significant. A best-fit linear regression line for the externalization of the reference condition as a function of high-frequency hearing loss gives a reduction in externalization rating of 7% per 10 dB HL high-frequency hearing loss. The correlation of age with reference externalization rating was also not significant.

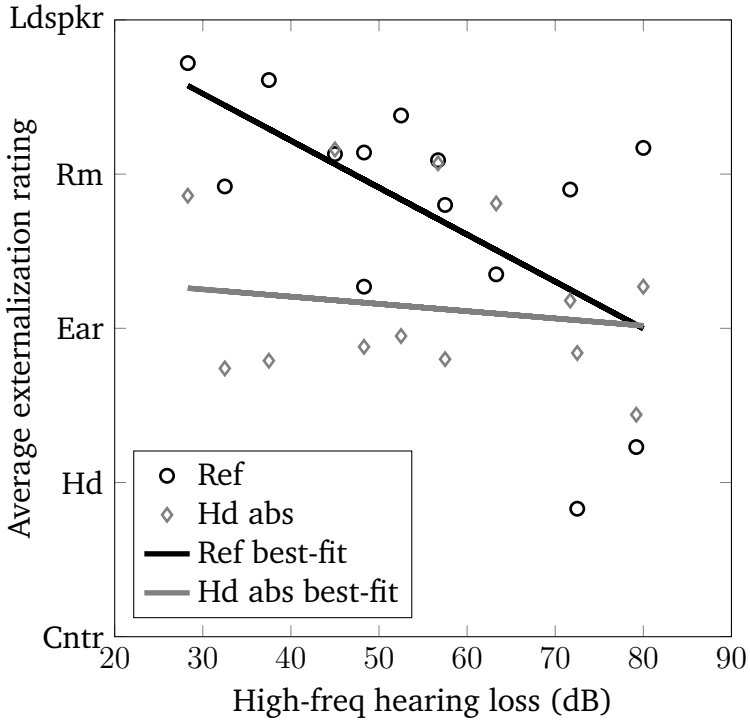

Fig. 4.11: Scatter-plot of the average externalization rating of the reference condition and the head-absent condition against high-frequency hearing loss (average of better ear 6 and 8kHz values). Lines of best-fit are also plotted for each condition.

4.3.4 Discussion of the effect of pinna cues and number of talkers

The normal-hearing listener results show that a continuum can be produced between full externalization and internalization using a mixing technique that

varies the strength of the HRTF filtering present, holding all other cues constant. Due to the preservation of the ITD cues, the head-absent (0% mix) stimuli could still be lateralized in the head, hence the absence of ratings for "center of head," which would have required a diotic stimulus. The results across conditions show the importance of high-frequency pinna cues for the externalization of static acoustic scenes. The removal of these cues resulted in an immediate reduction in maximum externalization. Therefore, in normal-hearing participants, the design and frequency response of a BTE hearing aid produces a reduction in externalization. This may be reduced with acclimatization.

The results for the one-talker conditions show an effect of high-frequency inaudibility of high-frequency HRTF (i.e., pinna) cues for hearing-impaired listeners with respect to externalization. However, they also show that this insensitivity produces a contracted perception of externalization. A simulated representation of sound in a room at an egocentrically fixed position is not fully externalized (i.e., rated directly at the loudspeaker) for hearing-impaired listeners, whereas the preservation of the ITD and DRR is enough to partially externalize sound in many cases, resulting in a reduced perceptual range. The correlation between high-frequency hearing-loss and reference externalization suggests audibility of pinna cues in addition to sensitivity may be a factor in loss of externalization. The reduction in internalization compared to normal-hearing listeners when only the static ITD and DRR cues were present suggests that hearing-impaired listeners place a greater perceptual emphasis on these cues than pinna cues. The inclusion of dynamic binaural cues (not present in this study) may have resulted in a greater maximum externalization in the

HI listeners. The inaudibility of high-frequency HRTF cues results in a smaller variation in maximum externalization across conditions.

The aforementioned insensitivity and reduced perceptual range are compounded by a more complex acoustic scene of four spatially separated talkers. In these conditions, hearing-impaired listeners often reported perceiving no change or movement in the stimuli presented and an inability to perceive four distinct talkers. Since the normal-hearing listeners did not report this problem in the four-talker conditions, this may be due to the hearing-impaired listeners' reduced spectrotemporal resolution, reducing their ability to detect changes in HRTF cues in talker mixtures and hence changes in externalization. An alternative explanation is that the combination of talkers presented at the same level as the one-talker stimuli resulted in each individual talker being 6 dB lower, further reducing the audibility for hearing-impaired listeners The results in the lowpass, four talker conditions show that responses become decoupled from the mix played, suggesting that they are more insensitive to the remaining low-frequency cues in these complex listening conditions.

A number of participants found the task very difficult to perform, either due to the complexity of the graphical user interface used or an inability to hear any difference between the reference and head-absent conditions during training on the four-talker condition. This produced a large variation across HI listeners, resulting in the large confidence intervals shown in figure 4.10. The acoustic scene delivered over the headphones, though acoustically identical to externalized listening, within the documented limitations (Wightman and Kistler, 1989a), was fixed in space. Therefore, movements of the head resulted

in the room moving with it, which could break the externalization illusion. To mitigate this, participants were asked to listen to the stimuli while looking at the focus point. Combined with the visual localization of the loudspeakers, the externalization was reported to have remained stable for the duration of the task.

The production of the lowpass stimuli could have been improved by using a different type of filter. The Butterworth filter was chosen for its smooth response in amplitude, though the phase is distorted above the cutoff frequency. Modern hearing aids often use biquadratic filters (Kates, 2008) and similar research has used filter designs that minimize phase distortions (Wiggins and Seeber, 2011). However, at the frequencies involved, there is no phase-locking in the auditory system, making the phase distortions inconsequential to the experiment. In addition, identical filters were used on both ears. The similarity of the responses between the unfiltered BTE condition and the filtered ITE condition suggest that amplitude and not phase differences where important, as these conditions produce similar reductions in high-frequency pinna cues. Hearing aids also produce some phase distortions in everyday use, due to processing delays and microphone combinations.

4.4 The effect of hearing aids, presentation angle and acclimatization

The experiment described in this section is an abstract version of a situation that appeared to cause internalization in the hearing-impaired listeners consulted. If cutlery was dropped in a quiet restaurant, the sound of the cutlery would

often be internalized. This situation was broadly simulated with impulsive noise bursts representing the cutlery falling and a single distracter representing a low-level ongoing conversation nearby.

These high-gain, short-duration noise stimuli were presented over loudspeakers from several angular positions to the participants. Participants rated the degree of externalization of these stimuli with a male continuous talker as a distracter. Sixteen stimulus conditions were created, comprising all combinations of presentation angle ($0°$, $30°$, $60°$, $90°$), bandwidth (1-8 and 3-8 kHz), and linear amplitude ramp time (10 and 100 ms).

The normal-hearing participants were also tested for the effect of short-term acclimatization. Participants wore flat-gain hearing aids for 6-7 hours and were then asked to rate the degree of externalization for the same stimuli as before.

4.4.1 Methods for the investigation of the effect of hearing aids, presentation angle and acclimatization

Participants: Five normal-hearing (one female) and five hearing-impaired (one female) listeners participated. All normal-hearing listeners had a better-ear four-frequency (0.5, 1, 2, and 4 kHz) average (4FA) of less than 20 dB HL. Hearing-impaired listeners had a median 4FA of 52.5 dB HL and a range of 48.75-60 dB. For one listener the distracting speech was at threshold (60 dB) when listening without their hearing aids. However, the low frequency threshold (250 and 500 Hz average) was less than 60 dB HL and their responses were not different from the other participants. The asymmetry for hearing-impaired listeners ranged from 0-1.25 dB. All hearing-impaired

listeners wore two behind-the-ear (BTE) hearing aids, which were their own (1+ year experience with those hearing aids). The hearing aids worn were: Danalogic 6071, Oticon Spirit Zest, Oticon Spirit 2P, Siemens Reflex DP and Siemens Reflex L. Normal-hearing listeners used Danalogic 6071 BTE hearing aids set to a gain target of 14 dB for 50 dB input sounds and 3 dB for 80 dB input sounds across frequencies, and a target compression ratio of 1.6. These settings were chosen to ensure a realistic compression ratio without exposing the listener to dangerously high sound levels. Table 4.4 shows the measured compression ratio, attack and release times for the hearing aid across frequencies, measured using a 40 to 90 dB and back to 40 dB pure tones. The effective bandwidth of the hearing aid was 0.2-5 kHz.

Freq. (kHz)	Comp. ratio	Att. time (ms)	Rel. time (ms)
1	1.67	12	42
2	1.54	10	33
4	1.54	9	46

Table 4.4: Information on the Danalogic 6071 hearing aid used for the normal-hearin listeners' hearing-aid trials. Compression ratio was measured using the input/output values for a narrowband (third octave) noise signal.

All normal-hearing participants performed the acclimatization (HA+6) task. The normal-hearing listeners had a median age of 28 (24-44) and the hearing-impaired listeners had a median age of 65 (57-73) years. Individual and average audiograms of the hearing-impaired participants are shown in figure 4.12.

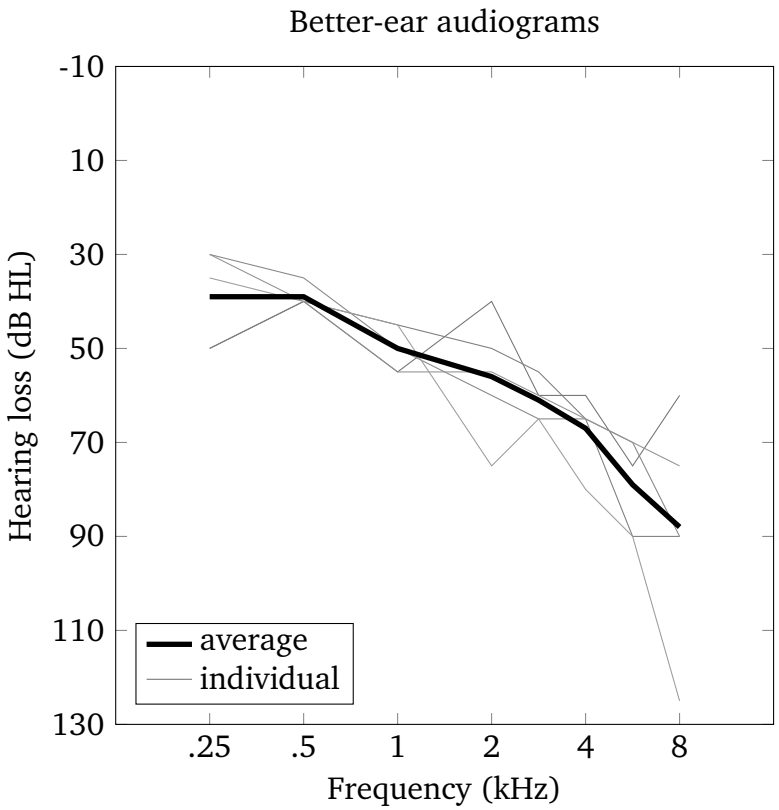

Fig. 4.12: Hearing-impaired participant better-ear audiograms. Individual audiograms are shown in greyscale. Average better-ear audiogram is shown in solid black.

Apparatus. The test room was the same 6.5 $\times$ 5 $\times$ 3 m room used in the previous experiment (see section 4.3 for further details). Listeners were seated along the long axis of the room. Five loudspeakers (JBL Control 1) where placed at a height of 1.2 m and a distance of 2 m from the listener at angles of $0°$, $30°$, $60°$, $90°$ and $180°$. The experimental apparatus can be seen in figure 4.13.

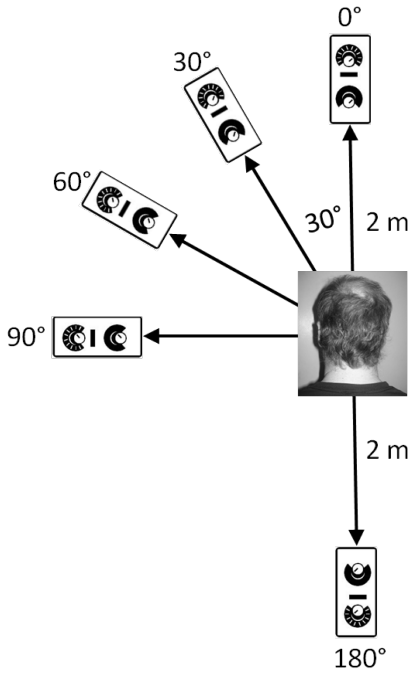

Fig. 4.13: Experimental apparatus for experiment 2. The loudspeaker at $180°$ presents a continuous male talker. All others present the noise stimuli.

Stimuli. The signals were 200 ms noise bursts played at 16-bit, 44.1 kHz sampling rate. They were band-pass filtered by applying a tenth-order Butterworth filter with a lower cutoff of 1 or 3 kHz and a high cutoff of 8 kHz. Linear amplitude ramps were applied to the beginning and end of the signals of either 10 or 100 ms duration. The playback level for these impulsive signals was 80 dBA.

The distracter signal was a continuous concatenation of sentences by the same male talker from the IEEE York corpus (Stacey and Summerfield, 2007). The corpus was recorded at 16-bit, 44.1 kHz sampling rate. The playback level for the distracter signal was 60 dB. The distracter was presented from the loudspeaker at $180°$.

Anecdotal reports suggested that unexpectedness appeared to be a common factor in the internalization of sounds. To reduce the effect of expectation on the responses, the interval between listener response and presentation of the next stimulus was varied randomly between two and five seconds.

Procedure. There were ten blocks of trials, each block consisting of one presentation of each of the 16 combinations of stimulus position ($0°$, $30°$, $60°$, $90°$), low-frequency limit (1-8 or 3-8 kHz) and onset/offset ramp (10 or 100 ms). In training, the participant was informed that they would hear a constant male talker behind them and short noises that could come from anywhere in the room. The participant was instructed to rate the noise stimulus using a discrete, ordinal, five-point scale: (5) "At the loudspeaker," (4) "In the room," (3) "At the ear," (2) "In the head," and (1) "Center of the head" (i.e., the same referents for the continuous five-point scale used in our previous study). After explanation of the scale and one training block, the participants performed the same task unsupervised for ten blocks with their hearing aids and ten blocks without their hearing aids. This was not randomized across listeners to avoid variable effects of acclimatization by hearing-impaired listeners. The (16) conditions were presented randomly within each block, resulting in ten presentations of each test condition.

Normal-hearing listeners performed the same task three times: without hearing aids (No HA), non-acclimatized (within ten minutes of being fitted, HA) with hearing aids and after 6-7 hours of acclimatization to wearing the hearing aids (acclimatized, HA+6). During the acclimatization period, participants were instructed to perform their daily tasks as normal, which for most involved office, laboratory and outdoor environments. The trials without hearing aids were

randomized to be before or after those performed with hearing aids, with at least a day separating trials performed with hearing aids and without.

4.4.2 Results of the investigation on the effect of hearing aids, presentation angle and acclimatization

Table 4.5 displays the results of a between-subjects analysis of variance on the responses for hearing-impaired listeners, showing statistically significant main effects and interactions. The main effect of angle was significant. Interaction effects of hearing impairment and hearing aid, hearing impairment and angle and hearing aid and angle were not significant. Three-way interactions were tested, but were not shown in table 4.5 as they were not significant.

Source	Effect size (η^2)	*df*	F	*p*
HA/NHA	0.0042	1	0.87	0.3517
Angle	**0.35**	**3**	**24.12**	**< 0.001**
Bandwidth	0.00057	1	0.12	0.7295
Envelope	0.0096	1	2.01	0.1587
HA/NHA*Angle	0.0084	3	0.59	0.6258
HA/NHA*Bandwidth	0.00012	1	0.00141	0.8751
HA/NHA*Envelope	0.0013	1	0.26	0.6077
Angle*Bandwidth	0.0046	3	0.32	0.8076
Angle*Envelope	0.0050	3	0.35	0.7906
Freq*Envelope	0.0036	1	0.76	0.3850
Error	0.61	159		

Table 4.5: Analysis of variance of responses for hearing-impaired group (HA and no HA). Statistically significant effects are in bold. The first column is the source of the effect, column 2 is the effect size (η^2), column 3 the degrees of freedom (*df*, followed by the F ratio and probability (*p*) value for the effect.

Table 4.6 displays the results of a between-subjects analysis of variance on the responses of the normal-hearing listeners, showing statistically significant

main effects and interactions. The main effect of envelope was not significant. Interaction effects of acclimatization (NHA/HA/HA+6) with angle and bandwidth were significant. Three-way interactions were tested, but were not shown in table 4.6 as they were not significant.

Source	Effect size (η^2)	*df*	F	*p*
NHA/HA/HA+6	**0.57**	**2**	**308.98**	**< 0.001**
Angle	**0.045**	**3**	**16.32**	**< 0.001**
Bandwidth	**0.10**	**1**	**110.84**	**< 0.001**
Envelope	< 0.001	1	0.22	0.64
NHA/HA/HA+6*Angle	**0.054**	**6**	**9.67**	**< 0.001**
NHA/HA/HA+6*Bandwidth	**0.040**	**2**	**21.59**	**< 0.001**
NHA/HA/HA+6*Envelope	< 0.001	2	0.23	0.80
Angle*Bandwidth	0.0012	3	0.43	0.73
Angle*Envelope	< 0.001	3	0.2	0.90
Freq*Envelope	< 0.001	1	0.01	0.92
Error	0.18	239		

Table 4.6: Analysis of variance of responses for normal-hearing group (no HA, HA and HA + 6).

Figure 4.14 shows the mean ratings as a function of angle obtained from normal-hearing listeners plotted with 95% confidence intervals computed from the standard error of the mean. As no effect of envelope was observed, the envelope responses have been combined. Effective stimulus bandwidth when wearing the hearing aids was 4 kHz (1 kHz cutoff) and 2 kHz (3-8 kHz bandwidth). Responses without hearing aids (No HA), immediately after fitting hearing aids (HA) and after 6-7 hours of acclimatization to the hearing aids (HA+6) are plotted. The effect of testing immediately after hearing-aid fitting (HA) and six hours after fitting (HA+6) was significant. Post-hoc, two-tailed t-tests (using a Bonferroni correction for 6 comparisons) revealed a significant effect of angle immediately after fitting (HA) for the 3-8 kHz bandwidth

($0° - 30°$, $p < 0.01$), but no effect of angle for the 1-8 kHz bandwidth over the same angle. There was also a significant effect of bandwidth at all angles after acclimatization ($0°$, $p < 0.01$, $30°$, $p < 0.01$, $60°$, $p < 0.01$, $90°$, $p < 0.01$). After acclimatization and without hearing aids there was no significant effect of angle. Without hearing aids (NHA), all sounds were externalized.

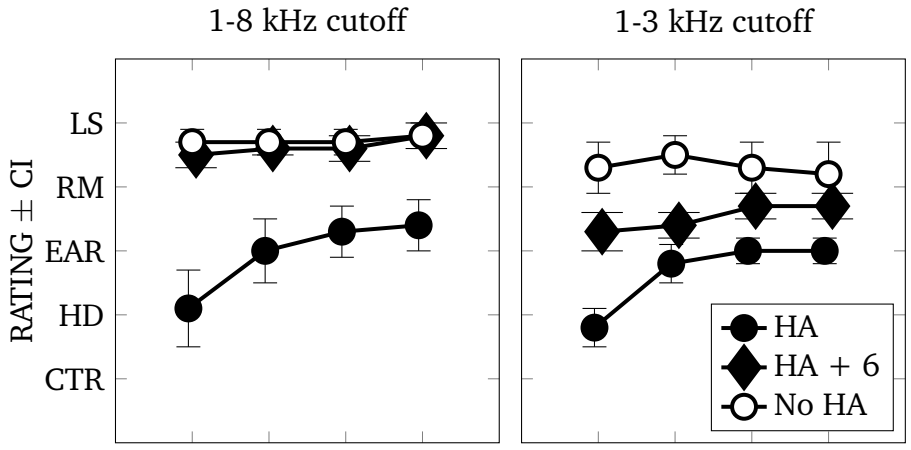

Fig. 4.14: Average externalization ratings for normal-hearing with hearing aids (closed circles, immediately after fitting), without hearing aids (open circles) and after wearing hearing aids for six hours (closed diamonds) against presentation angle as a function bandwidth (1-8 and 3-8 kHz).

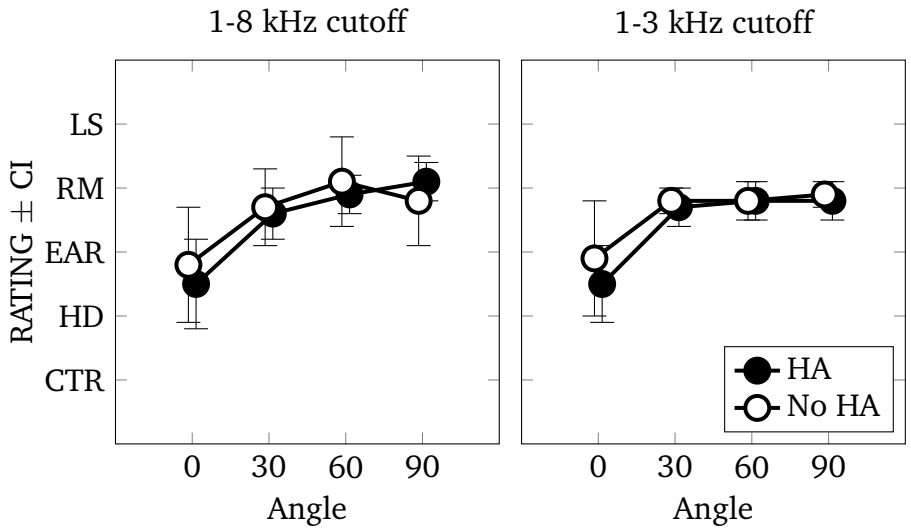

Fig. 4.15: Average externalization ratings for hearing-impaired listeners with hearing aids (closed circles) and without hearing aids (open circles) against presentation angle as a function of bandwidth (1-8 and 3-8 kHz).

Figure 4.15 shows hearing-impaired listener results. As no effect of envelope was observed, the envelope responses have been combined. Responses when wearing hearing aids were not significantly different from those without hearing aids. A significant effect of angle (using Bonferroni-corrected, post-hoc t-tests) was observed for hearing-aid responses for the 1-8 kHz condition ($0° - 30°$, $p < 0.05$) and the 3-8 kHz condition ($0° - 30°$, $p < 0.01$). No significant effect of angle was produced for the no hearing-aid responses, due to the greater variability in responses among listeners. In general, the most internalized sounds were presented at $0°$ and sounds were placed around "in the room" from $30°$ to $90°$. Hearing-impaired listeners' ratings were most variable at $0°$, indicating that some hearing-impaired listeners reported the $0°$ stimulus as far more internalized than others for whom responses did not vary (greatly) as a function of angle. Listeners rarely reported sounds as being "at the loudspeaker". No effect of stimulus type or bandwidth was observed.

4.4.3 Discussion of the effect of hearing aids, presentation angle and acclimatization

A summary of the main experimental findings from section 4.4 is shown in table 4.7.

Presentation angle has a significant effect on perception of internalization
Hearing aids, stimulus spectrum and envelope do not have a significant effect on internalization in hearing-impaired listeners in this particular experiment
Aided normal-hearing listeners initially perceive internalized sounds
Short acclimatization reduces internalization for aided normal-hearing listeners

Table 4.7: Summary of section 4.4 experimental findings

Normal-hearing responses without hearing aids showed responses between "in the room" and "at the loudspeakers" for all stimuli. No effect of presentation angle could be seen. The non-acclimatized responses showed a strong effect of presentation angle and greatest internalization at 0°. After wearing the hearing aids for six hours, the responses show increased externalization in relation to the non-acclimatized responses. The greatest change can be seen for the 0° presentation of the 1-8 kHz bandwidth stimuli, shifting from "in the head" to between "in the room" and "at the loudspeaker" because angle had no significant effect on the acclimatized, HA+6 condition. For the 1-8 kHz stimuli, the effect of presentation angle is no longer significant and the responses are not significantly different from the no hearing-aid trials. A significant effect of stimulus spectrum is observed, as the 3-8 kHz stimuli are not as externalized as the 1-8 kHz responses (or the no hearing-aid condition).

Previous research has shown that acclimatization to altered pinna cues requires days and weeks to show an effect in localization tasks (Wanrooij and Opstal, 2005). The effect of presentation angle in the current experiment was reduced after just 6-7 hours of acclimatization to hearing aids. Given this rapidity of acclimatization, it is possible that other factors than altered pinna cues are involved in externalization acclimatization.

Acclimatization to the increased gain provided by the hearing aid may explain the shift in normal-hearing responses. This effect is similar to the effect observed by Laws using externalized headphone presentations 1973: stimuli presented at a higher gain were more likely to be perceived by listeners as internalized. The effect of increased level reduced as the normal-hearing listeners adapted to the increased gain. The hearing-aid gain also introduced a spectral tilt, increasing the level of low frequencies more than high frequencies relative to listening without hearing aids. Normal-hearing listeners also acclimatized to this change after just a few hours. It is possible that internalization in hearing-impaired listeners is caused by an inability to acclimatize to the increased audibility of sounds and the changes to their spectrum when wearing hearing aids.

Why does the increased loudness of the sound initially induce a perception of internalization? At its simplest, auditory distance can be determined by the loudness of a sound source, so louder sounds will be placed closer to the listener than quieter sounds (see section 2.3). If all sounds are perceived as loud, they will all be placed close to the listener. In the absence of other strong spatial cues, the auditory system places the sound source from where it originates when wearing a hearing aid, at the ear. If the signal at each ear is very similar, as is

the case in the 0° condition, the sound is similar to a diotic presentation, which results in an "inside the head" perception.

Hearing-impaired listeners showed an effect of angle on their perception of internalization (section 4.4.2. No other factors produced a significant effect. The reason for this is unclear and could be related to an insensitivity of the listeners, or the experimental paradigm, to the effect of bandwidth.

The lack of any difference between the hearing-aid and no hearing-aid responses in hearing-impaired listeners is surprising, especially given the normal-hearing results. The pinna cues are compromised in the no hearing-aid condition due to audibility and in the hearing-aid condition due to the position of the microphones behind the ear. Hearing-impaired listeners do not appear to be using pinna cues to determine how externalized a sound is. This lack of difference with changing pinna cues is similar to the findings in section 4.3. It appears that for the types of sound presented in both experiments, hearing-impaired listeners are externalizing sounds solely on interaural differences. In the 3-8 kHz bandwidth case, listeners are using ILD and interaural envelope cues only. In addition, audibility is reduced in the no hearing-aid condition in comparison to the hearing-aid condition. A sound that is perceived to be quieter is often perceived as further away from the listener (see section 2.3).

The range of responses for hearing-impaired listeners is similar to those reported in section 4.3. Sounds are rarely placed "at the loudspeaker", with responses between 30° and 90° being between "at the ear" and "in the room". For short, loud, high-frequency noises, the results show that hearing-impaired

listeners do not perceive the stimuli as fully externalized. This may also be due to the sensitivity of hearing-impaired listeners to pinna cues, similar to the results of section 4.3.

Both hearing-impaired and normal-hearing listeners showed no significant effect of envelope on internalization ratings. It is possible that an effect would be seen using longer stimuli and more slowly changing envelopes (see section 5.3 for an alternative explanation based on ILD distribution). A higher compression ratio may have produced a greater interaction with envelope for the normal-hearing listeners wearing hearing aids.

One hearing-impaired listener showed no effect of presentation angle or any other factor on their externalization ratings and was not included in the results. This listener's responses were between "at the ear" and "in the room" across all conditions. The reason for this lack of variation in responses may relate to insensitivity to the changes in the stimuli in relation to externalization, or to externalization in general. This highlights a major issue with performing externalization psychoacoustic experiments on untrained listeners.

One normal-hearing participant did not acclimatize significantly to the hearing aids, giving similar responses after initial hearing-aid fitting and six hours of acclimatization. This participant's responses were not included in the results. This was possibly due to the types of sound environment the participant was exposed to during acclimatization. Participants were instructed to perform their daily tasks as normal, which for most involved office, laboratory and outdoor environments. However, the participant who did not acclimatize spent the six hours in a small, quiet office environment. The reduced exposure to

different sound environments may have led to the reduced acclimatization effects observed.

4.5 Conclusions

The initial experiments in section 4.2 attempted to research externalization through auditory distance perception. The first experiment (section 4.2.1) provided inconclusive results and the second initial experiment (section 4.2.2) developed a novel response method for auditory distance perception research. Responses broadly agreed with previous research in this area. An inability to produce stimuli that could be perceived as internalized meant the experiment did not proceed beyond the pilot stage.

Section 4.3 examined the effects of static HRTF cues, hearing-aid microphone placement, bandwidth, and number of talkers on participants' ability to externalize speech. Using a headphone simulation of open-ear listening and a modified MuSHRA testing paradigm, the study demonstrated:

1. Externalization can be perceived as a continuum. This continuum was shown, albeit incompletely, using a mixing technique varying the strength of HRTF filtering present and keeping all other cues constant, for hearing-impaired listeners (e.g., ITE/BB/1 talker).

2. The normal-hearing results showed that the microphone placement and frequency response of a BTE hearing aid adversely affected the perception of externalized sounds, due to the removal of high-frequency pinna cues.

3. The hearing-impaired results for both the one- and four-talker conditions displayed insensitivity to HRTF cues – a lack of differences based on microphone placement – in both full and limited bandwidth conditions.

4. Hearing-impaired listeners experienced a contracted sense of externalization, as stimuli where neither fully externalized nor internalized to the same degree as normal-hearing listeners.

These last two findings suggest that binaural cues, such as ITDs and ILDs and, by extension, dynamic binaural cues are of greater importance for externalization than monaural cues such as the HRTF in hearing-impaired listeners. The last two findings could also be explained by a lack of audibility of the high-frequency pinna cues for hearing-impaired listeners.

In section 4.4 the effect of hearing aids, presentation angle, stimulus spectrum, stimulus attack and release and acclimatization on internalization in hearing-impaired and normal-hearing listeners was investigated. Using impulsive noises of varying bandwidth, envelope and presentation angle, the study demonstrated:

1. Presentation angle has a significant effect on perception of internalization for hearing-impaired listeners.

2. The use of hearing aids, stimulus spectrum and envelope do not have a significant effect on internalization in hearing-impaired listeners in this particular experiment.

3. Normal-hearing listeners wearing hearing aids initially internalize the stimuli, with a significant effect of presentation angle.

4. After a short acclimatization period, normal-hearing listeners wearing hearing aids externalize the presented stimuli. Narrower bandwidth stimuli are less well externalized than broader bandwidth stimuli after acclimatization.

These findings suggest that hearing-impaired listeners weigh interaural cues strongly when externalizing sounds and that an inability to acclimatize to the increased audibility provided by hearing aids may be a contributing factor to internalization in hearing-impaired listeners. The normal-hearing hearing-aid results in section 4.4 could be partly due to the change in spectrum, most notably the gain to 5 kHz only and the relative attenuation from 5-8 kHz (via the earmould).

The results of both experiments taken together raise the possibility that a lack of full externalization could be due more to high-frequency (pinna) cues, while internalization is due more to ITD/ILD cues related to angle.

CHAPTER 5

Survey and modeling research on externalization

5.1 Introduction

This chapter describes survey and modelling research carried out on the internalization/externalization continuum: the development of an externalization questionnaire, including the analysis of responses (section 5.2), and a study of simulated ILD distributions and their possible effect on the perception of externalization (section 5.3).

The development of the novel Perception of Internalization Questionnaire (shown in section 5.2) was motivated by several factors. First, the prevalence of the perception of internalization in the hearing-impaired population was not known. Besides informal clinical reports of hearing-impaired listeners suffering from internalization, the only known subjective inquiry of internalization is a question included in the Speech, Spatial and Qualities of Hearing Scale (SSQ). Responses to this SSQ question previously showed that the perception of internalization increased with the number of hearing aids worn (Noble and Gatehouse, 2006). This question was uncorrelated with the other questions in

the SSQ (Gatehouse and Noble, 2004; Akeroyd et al., 2013). A similar question on externalization was also uncorrelated with other questions in a study of dynamic range compression with normal-hearing listeners (Wiggins and Seeber, 2012; see section 2.4.7). The PIQ examines the types of sound and listening situations where hearing-impaired listeners experienced internalization and how often this happened. The prevalence in different hearing-impaired groups and the situations in which it occurred would provide information for and guide the focus of future research on internalization.

While subjective responses of prevalence provide a foundation for externalization research, objective measures of aided localization cues provide an explanation of behavioural tests of externalization (section 5.3). Previous research has suggested that unlinked, bilateral hearing-aid compressors increase the kurtosis of the distribution of ILDs in speech (Catic et al., 2013; summarized in section 2.4.6). They showed that the peak in the distribution of ILDs moved towards 0 dB with decreasing source-listener distance. The shift was largest at high frequencies, which was also reported by Wiggins and Seeber (2011). Wiggins and Seeber (2011) analyzed the high-frequency (>2 kHz) ILD distributions for speech among other stimuli for a fixed compression ratio of 3:1 and "syllabic" (fast) compression (5 ms attack time and 60 ms release) was studied by Wiggins and Seeber (2011). The ILD distribution of speech was most effected by compression.

The analysis in section 5.3 attempted to reproduce the narrowed ILD distribution suggested by Catic et al. (2013) using a hearing-aid compressor (Gilbert et al., 2008). The output of the simulated hearing aid only showed a narrower, sharper ILD distribution under particular constraints related to the

input signal, source DOA and the listening environment. Along with the limited prevalence also shown in this chapter, this analysis shows that internalization with hearing aids – when it occurs – is an issue that involves interactions of factors within and without the hearing aid.

The remainder of the chapter is organized as follows. The development and design of the PIQ is described (section 5.2.1), followed by the procedure for administering the survey and the participants involved. The results of the PIQ and spatial SSQ survey are given in section 5.2.2 and discussed in section 5.2.3. After this, the design of the ILD distribution model is described (section 5.3.1). The effects on ILD distribution of compression (section 5.3.4), type of speech (section 5.3.3), reverberation (section 5.3.2), angle (section 5.3.5) and temporal resolution (section 5.3.7) are shown and discussed in section 5.3.8. Section 5.4 concludes the chapter.

5.2 Perception of Internalization Questionnaire (PIQ)

The PIQ was developed to investigate internalization and externalization in listening situations with background speech babble, general background noise and short-duration, loud noises. Questions were also asked of the respondent's overall experience of internalization and the related perception of a "flattening" of auditory space, where all sounds appear to come from the same distance. Given the lack of knowledge in the domain of internalization, a final question asked for examples of sounds or situations where the listener had experienced

internalization in relation to impulsive sounds, as these were found to be the most internalized sounds in informal listening tests by the author.

5.2.1 Methods for the PIQ

The PIQ is an interpolation of a single question on the SSQ, question 14 on the spatial section ("Do the sounds of things you are able to hear seem to be inside your head rather than out there in the world?"). Like the SSQ, the PIQ asks the participant to consider a particular situation and then rate the prevalence of that situation. Table 5.1 shows the final questions in full. The response method was the same as the SSQ (Gatehouse and Noble, 2004): "Never" to "always" on an 11-point scale.

1. When listening in a room with many people speaking at different distances, such as a restaurant, do any of the talkers appear inside your head as if wearing headphones?
2. When listening in an area with a lot of background noise, such as a supermarket, do any of the sounds appear inside your head, as if wearing headphones?
3. When listening in a situation where there are many people speaking and other sounds at different distances, such as a pub or cafe, do all the sounds appear to come from the same distance?
4. How often do sounds in the outside world appear to come from inside your head, as if wearing headphones, with no obvious direction or distance?
5. When listening to short, sharp or unexpected sounds (doors closing, plates hitting together, industrial noise), do the sounds appear inside your head, as if wearing headphones, with no obvious direction or distance?
6. If the answer to question 5 was affirmative, please give examples of the types of sounds you hear inside your head?

Table 5.1: The questions used in the Perception of Internalization Questionnaire.

Question 1 was motivated by a need to discover whether internalization occurred in complex environments comprising of mainly speech signals, as the original SSQ question asked only about the perception of "sounds" in general. The question provides an example of the type of situation where this scene could occur (a restaurant) and also clarifies what is meant by internalization using a commonly experienced situation in which sounds are internalized ("as if wearing headphones"). Question 2 was similarly motivated, though in this

case the sounds in the environment were left open to interpretation. The example of the supermarket was given because 1) the author had experienced internalization while wearing hearing aids in a supermarket, and 2) it is a complex sound environment that is known to most respondents.

Question 3 was motivated by reports of some respondents to the SSQ that though sounds may not be internalized, they are perceived to be without a discernible distance, and as a result all sounds appear to come from the same distance. In addition, the "flattened" externalization responses from hearing-impaired listeners in section 4.3 suggested that this may be a common perception and one related to internalization.

Question 4 was similar to the original SSQ question on externalization, asked in the same style as the other PIQ questions. It was designed to provide a comparison between specific situations (questions 1 and 2) and the general perception of internalization in everyday listening.

Question 5 was motivated by the experiences of the author that loud, impulsive sounds were most likely to be internalized. Question 6 was intended to gain more specific information on the types of sounds that were internalized by respondents. Loud impulsive sounds were thought to be the most likely to be internalized and be most easily identified by listeners. The open-ended style of question 6 was also inspired by the Glasgow Hearing Aid Benefit Profile (GHABP, Gatehouse (1999)), as it also asked open-ended questions in order to gain more information from participants.

The questions focusing on speech babble and background noise were initially administered in two forms, once as a question on internalization, with a

response of 10 corresponding to "inside the head" and once as a question on externalization, where a response of 10 corresponded to "outside the head". For example, question 1, "When listening in a room with many people speaking at different distances, such as a restaurant, do any of the talkers appear inside your head as if wearing headphones?" was followed several questions later by the reverse, "When listening in a room with many people speaking at different distances, such as a bar or cafe, are all the sounds out there in the world?" The motivation for this reversal was to check the reliability of participants' responses, as the sum of the two responses of a consistent participant would be approximately 10. Participants often reported the scale reversal to be confusing and responses to question 1 and question 2 and their reversals were weakly correlated ($r = -0.44$, $p < 0.01$ and $r = -0.35$, $p < 0.01$ respectively). As a result of these findings, only the questions on internalization were used in the analysis.

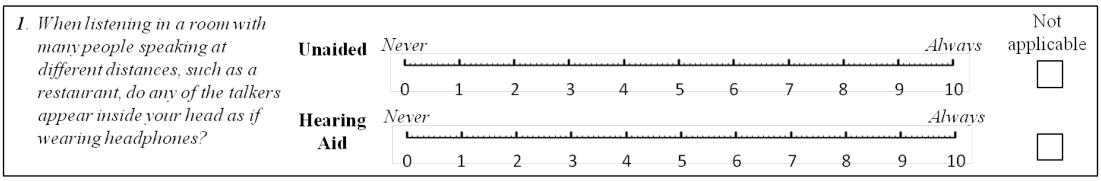

Fig. 5.1: An example of a PIQ question showing the 11-point response scale and option to ask for responses with and without hearing aids.

The response system chosen was the 11-point visual scale used in the SSQ, shown in figure 5.1. The questions related to how often sounds were internalized in a given situation, ranging from 0 (never) to 10 (always). The anchoring terms "never" and "always" were the same for each question. The questionnaire initially used paper copies to record responses; the layout, an

example in figure 5.1, was based on the SSQ. After piloting, responses to the PIQ were recorded using an HTML version of it.

The questions were given in the order given in table 5.1. Unilaterally aided (right or left ear) participants were asked for responses with and without their hearing aid if possible. Bilateral participants were not asked for responses without their hearing aids, as most wore their hearing aids for the majority of their active hours.

The participants in the study were recruited via post from other experiments or first-time visitors taking part in a number of questionnaires as part of a participant screening process. 257 participants were given the PIQ. 131 participants were male and 126 female. 72 wore no hearing aids (HA0), 118 were unilaterally aided (HA1) and 67 were bilaterally aided (HA2). Mean age was 61 years old. Ages did not vary significantly between groups or categories

If the participants wore hearing aids, these were all behind the ear (BTE) type, digital and programmable. The participants had been wearing their hearing aids for 6 months or more.

The questionnaire was given by two research assistants and an audiologist, all experienced in administering questionnaires. The questionnaire was often administered with a number of others and always in conjunction with the 17 questions comprising the spatial subsection of the SSQ. If the participant responded less than 8 to the question on internalization in the SSQ, the PIQ was given to the participant, otherwise it was not given. Those who responded ≥ 8 were categorized as non-internalizing and those who responded < 8 were categorized as internalizing. There are several exceptions to this due to the

difficulty in defining a true internalization point (without over-inflating the results). A minority of participants (HA0 = 3, HA1 = 3 and HA2 = 1) that responded ≥ 8, were given the PIQ and were responsive to it (answering > 0 to any PIQ question), providing valid results. In the non-internalizing category, only 7 (of 59) HA0 , 9 (of 82) HA1 and 0 HA2 participants responded with an 8 or 9 and were not given the PIQ. Defining the perception of internalization as a response < 10 to question 14 could be considered a valid approach. However, given the number of HA0 participants (1 of 13 respondents to the PIQ) and HA1 participants (11 of 48 respondents to the PIQ) who responded < 8 but were unresponsive to the PIQ (see section 5.2.2), this approach could be considered an over-inflation of the prevalence of internalization. Hence we have provided two values for prevalence of internalization, based on the SSQ and the PIQ.

All participants underwent standard clinical air and bone conduction pure-tone audiometry. They all had sensorineural hearing losses. Table 5.2 shows the average BE4FA (0.5, 1, 2 and 4 kHz) for the non-internalizing and internalizing categories across groups. Post-hoc t-tests (Bonferroni corrections for 9 comparisons) revealed significant differences between the non-internalizing categories: HA0 to HA1, $p < 0.001$; HA0 to HA2, $p < 0.001$. Post-hoc t-tests also revealed significant differences between some of the internalizing categories: HA0 to HA1, $p < 0.01$; HA0 to HA2, $p < 0.01$. A significant difference was not found between the internalizing HA1 and HA2 groups. No significant differences where found between the non-internalizing and internalizing categories in each group (HA0, HA1, HA2). There were no significant differences between groups, and no significant differences between categories within groups for asymmetry of hearing loss.

Group	N	Category	BE4FA/Aided-ear 4FA (dB HL)
HA0	59	non-internalizing	25.5
HA0	13	internalizing	33.4
HA1	82	non-internalizing	50.4
HA1	36	internalizing	52.9
HA2	40	non-internalizing	57.1
HA2	27	internalizing	52.7

Table 5.2: BE4FA/Aided-ear 4 FA for questionnaire participants, separated by group (HA0, HA1 and HA2) and category (non-internalizing and internalizing).

5.2.2 Survey results

The prevalence of internalization was calculated as the percentage of respondents who answered 8 or less to the question on internalization in the SSQ relative to the total number of respondents in each category: no hearing aids (HA0), unilaterally aided (left or right, HA1); bilaterally aided (HA2). Those who responded > 8 were categorized as non-internalizing and those who responded ≤ 8 were categorized as internalizing.

Table 5.3 shows the numbers of male, female and total participants in each hearing-aid group and category. Overall prevalence among all three groups was 29.5%. Prevalence of internalization among HA0 respondents was 18% (13 out of a total of 72 respondents). 30.5% (36 of 118) of HA1 respondents and 40.3% (27 of 67) of HA2 respondents reported internalization.

A chi-squared test using table 5.3 produced a value of $\chi^2 = 8.36$. As there were two degrees of freedom, this gave a significant p-value of $p < 0.05$. The observed (o) number of internalizing HA0 respondents was below the expected (e) number (o=13, e=21) and the observed number of internalizing HA2 respondents was greater than the expected value (o=27, e=20). Due partly to the larger size of the HA1 group in comparison to the HA0 and HA2 groups, the expected and observed values were the same (o=36,e=35) for HA1 internalizing respondents.

Group	Category	Male	Female	Total
HA0	non-internalizing	37	22	59
HA0	internalizing	7	6	13
HA1	non-internalizing	46	36	82
HA1	internalizing	17	19	36
HA2	non-internalizing	13	27	40
HA2	internalizing	11	16	27
	Total non-internalizing	96	85	181
	Total internalizing	35	41	76

Table 5.3: The gender balance of the questionnaire respondents, separated by group (HA0, HA1 and HA2) and category (non-internalizing and internalizing).

Figure 5.2 shows the average overall responses to the spatial SSQ section (excluding the internalization question) for the internalizing and non-internalizing categories. Higher scores indicate better self-rated spatial abilities.

An ANOVA revealed significant differences between the non-internalizing HA1 and internalizing HA1 group responses to individual questions in the spatial SSQ section (non-int/int group, η^2=0.0251, *df* = 1, F = 46.52, *p*=¡0.001). Post-hoc t-tests (Bonferroni corrections for 16 comparisons) showed significant differences for questions 2, 3, 4 and 17 ($p < 0.05$). A post-hoc statistical t-tests (Bonferroni corrections) revealed a significant difference in the average response between non-internalizing HA0 and non-internalizing HA2 ($p < 0.01$) groups. Overall responses to the spatial SSQ (excluding the response to the internalization question) were negatively correlated with BE4FA hearing loss ($r = -0.2420, p < 0.01$).

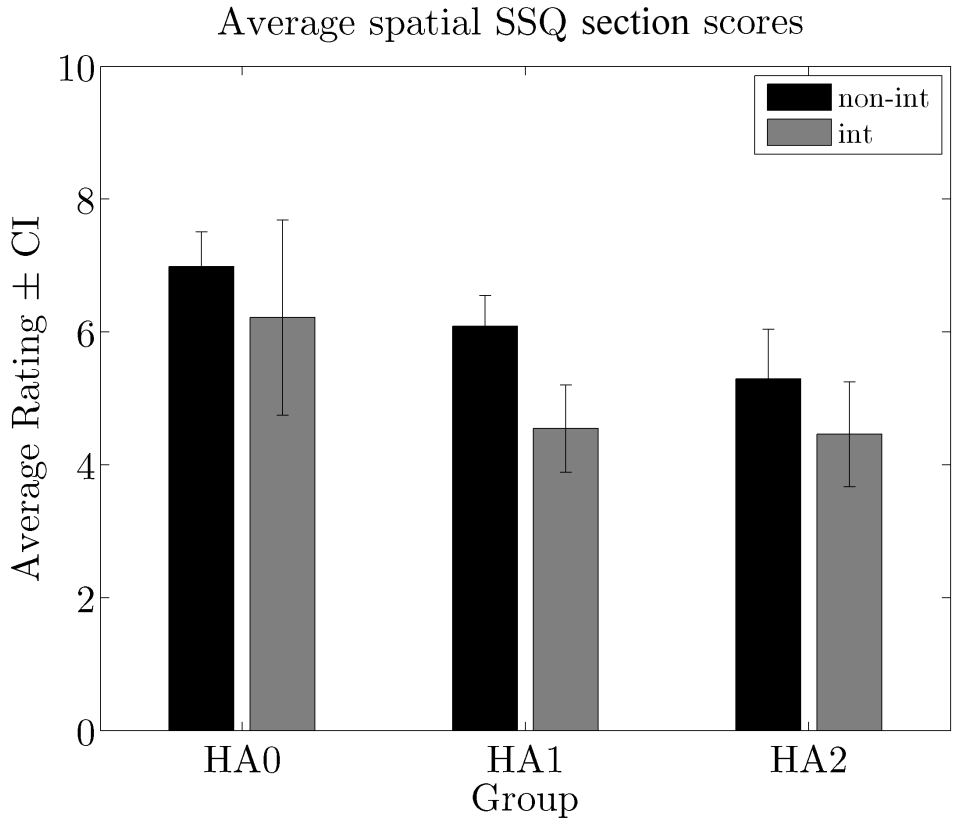

Fig. 5.2: Average responses to the spatial section of the SSQ (excluding question 14). Black bars represent the non-internalizing group and grey bars the internalizing group, for HA0, HA1 and HA2 categories. Bars are plotted with 95% confidence intervals.

Sixty-four participants were responsive to the PIQ (i.e. answered any question on the PIQ > 0). Seventy-six participants were given the PIQ and there were 11 unresponsive participants in the HA1 group and 1 unresponsive participant in the HA0 group. Due to this difference in numbers, the true proportion of those who experience internalization in the hearing-impaired population is considered to be between 25% (based on PIQ respondents) and 29.5% (based on the SSQ internalization question). Scores on the PIQ questions were not normally distributed and non-parametric statistics were used to correlate outcomes across the various measurements.

For the PIQ, the HA1 group was further split into responses when wearing their hearing aids (HA1) and without (HA1n).

Figures 5.3 to 5.7 shows the average responses to questions 1, 2, 3, 4 and 5, plotted with ± 1 standard deviation. The questions are shown in table 5.1.

A significant difference was found – using the Wilcoxon-Signed test for two related groups, Bonferroni corrected for 30 comparisons – between the HA1 and HA1n groups for question 1 ($Z = -3.19$, $p < 0.01$). No other significant differences were found between groups.

Within groups, no statistically significant differences were found between questions.

No correlations were found between BE4FA hearing loss, high-frequency hearing loss, age or durations of hearing-aid use.

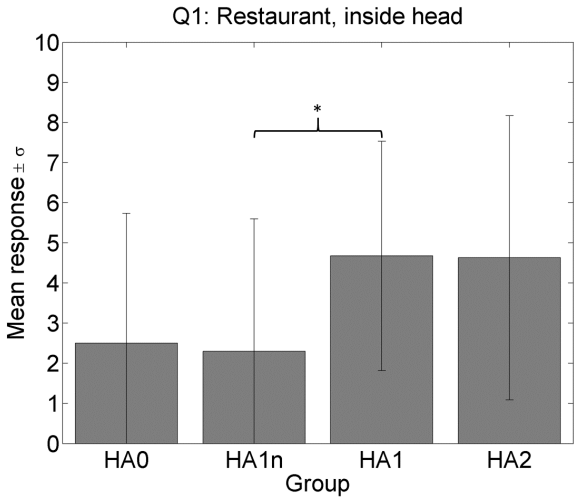

Fig. 5.3: The average responses to PIQ question 1. HA0 are the unaided group, HA1 the unilaterally aided (HA1n = HA1 response with no hearing aid) and HA2 the bilaterally aided group. * = Wilcoxon-signed, $p < 0.05$

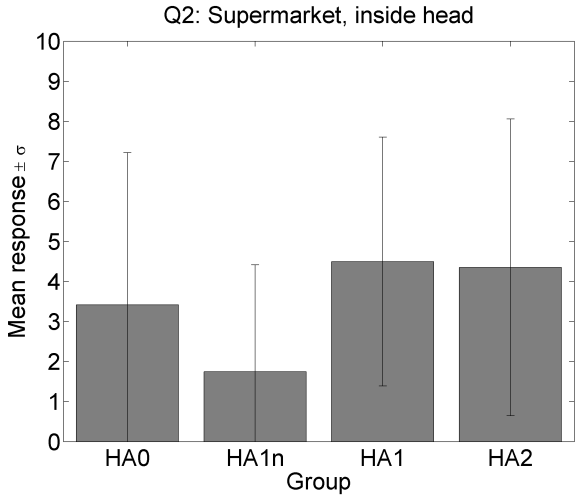

Fig. 5.4: The average responses to PIQ question 2.

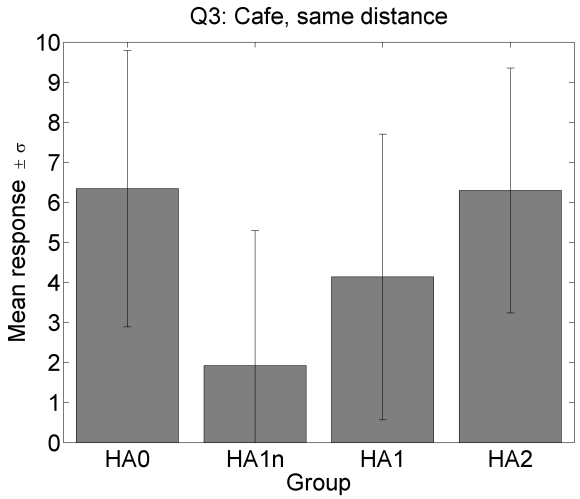

Fig. 5.5: The average responses to PIQ question 3.

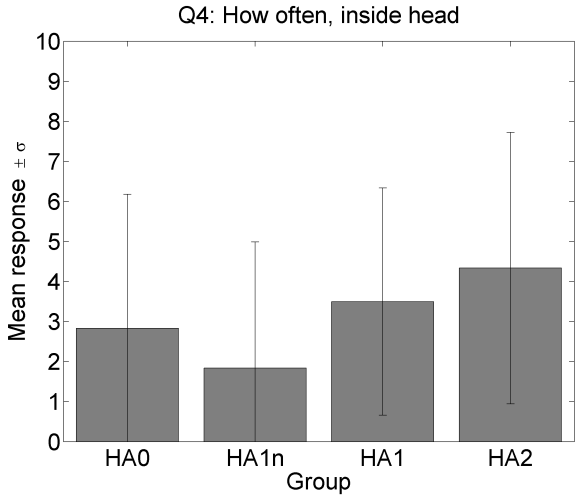

Fig. 5.6: The average responses to PIQ question 4.

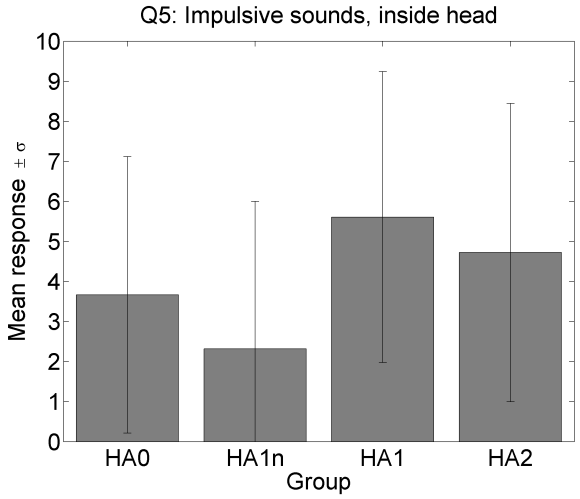

Fig. 5.7: The average responses to PIQ question 5.

Question 6 allowed respondents to report what types of sounds they perceived as internalized. Table 5.4 provides a list of all responses, grouped by number of hearing aids worn. The most common type of sounds listed as internalized were doors closing or slamming (8 instances). The related description of "banging" had 3 instances and there was one instance of "door banging." These sounds could be categorized as loud, broadband and impulsive and totalled 12 instances. 5 instances of "high-pitched" sounds were specifically referred

to. Related sounds such as “alarms” (3 instances), “sirens” (2 instances), ringing telephone (“ringtone”/“telephone ringing”, 2 instances), and door bell (1 instance) were also reported. These sounds could be categorized as high-frequency and narrow bandwidth and totalled 13 instances. Sounds with no obvious source were also reported, such as “chatter”/“voices” (3 instances) and “traffic” (2 instances) and “sound far away in a stairwell” (1 instance).

HA0	HA1	HA2
dull things banging or slamming	chatter	loud music; high-pitched noise
high-pitched sounds; lightbulb buzzing; television	door closing	traffic; chatter; background noise
doors banging	high-pitched sounds	doors slamming; floorboards (creaking)
plates; mobile phones (ringtones)	voices	traffic sirens; loud banging
	music	door bell
	plates in the kitchen	dog tags banging together; strong bass in music
	traffic	door (closing); plates
	high-pitched whistle; banging	doors (closing); alarms, cutlery
	doors closing; doors closing far away in a stairwell	Children screaming
	90% of all sounds	House alarms
	telephone ringing	fire alarms; all high frequency sounds (dog barks and crying babies)
	baby crying; dropping cutlery	sirens
		door slamming; loud shout; loud bark

Table 5.4: Responses to question 6 of PIQ questionnaire, “If the answer to question 5 was yes, please give examples of the types of sounds you hear inside your head?”

5.2.3 Discussion of survey results

The results of the questionnaires show that a minority of the hearing-impaired population experience the perception of sounds as internalized. The degree of internalization experienced increases with the number of hearing aids worn. This agrees with previous findings (Noble and Gatehouse, 2006).

5.2.3.1 SSQ spatial subscale

Internalizing HA1 participants perceived their spatial listening abilities to be significantly lower overall on the SSQ spatial subscale than non-internalizing HA1 participants. The individual questions that were not significantly different between these HA1 categories (non-internalizing and internalizing) were question 6 (a question relating to localizing a dog barking without visual cues), questions 10 and 11 (questions on determining the lateral movement of a bus or someone's footsteps) and question 14 (a question relating to auditory depth perception). These questions cover a wide range of auditory localization scenarios. The results from questions 10 and 11 suggest that internalizing HA1 listeners are equally good at determining the lateral movement of sounds as non-internalizing listeners. Further psychoacoustic testing is required to test this hypothesis.

5.2.3.2 PIQ between-group comparisons

HA2 participants had the highest prevalence of internalization. However, for those who experience internalization, the frequency of the perception of internalization was not significantly different between groups for PIQ questions 1, 2, 4 and 5, which specified different potential situations for internalizing

sounds. Therefore, increasing the number of hearing aids increases the likelihood of experiencing any instances of internalization, but does not increase the frequency with which it occurs if experienced.

Unilaterally aided participants were asked for their perception of internalization with (HA1) and without (HA1n) their hearing aids. Sounds were significantly more frequently internalized for these participants with their hearing aid than without for PIQ question 1. This result weakly suggests that the hearing aid may be exacerbating the perception of internalization in those unilaterally aided respondents who experience internalization.

A minority of HA0 participants reported internalization. Unaided listeners experiencing internalization of sounds were also reported in Noble and Gatehouse (2006). This result suggests that internalization can occur as a result of hearing impairment, regardless of the use of hearing aids. A caveat to this is the lack of correlation of the HA1n participants with high-frequency hearing loss, as these participants would be expected to respond similarly to the HA0 internalizing group. In addition, the responses of the HA0 and HA1n groups to PIQ questions 1, 2, 4, and 5 were not significantly different.

There was a trend towards lower responses by the HA1n group in comparison to the HA0 group. The reasons for this are not clear, as both groups represent the same situation - a respondent who experiences the internalization of sounds responding for periods where they do not wear their hearing aid. However, the HA0 group experience internalization without hearing aids, whereas many of the HA1n group may have only experienced internalization when wearing their hearing aids (group HA1). Therefore, the groups may not represent the

same situation. In the HA1n group, the internalization may have been caused or exacerbated by the hearing aid, while in the HA0 group another, so far unknown factor was causing the perception of internalization.

PIQ question 6 allowed the participants to report what types of impulsive sounds they perceived as internalized. The types of impulsive sounds that were most likely to be internalized could be categorized as either high-frequency and narrow bandwidth or broadband, loud, and impulsive. The source could also be visually hidden or indeterminate. These reports suggest that sounds without clear spatial cues (temporally, spectrally or visually) are most likely to be internalized by participants. Perceived loudness may also have an effect, with sounds perceived as louder being internalized more often.

5.2.3.3 PIQ within-group comparisons

The lack of significant differences between question 1, 2, 3, 4 and 5 responses within groups suggests that the perception of internalization of sounds does not depend on the type of sound or listening environment.

5.3 The effect of hearing-aid compression combined with multiple source types on ILD distribution

This study, inspired by the research of Catic et al. (2013), investigated ILD distribution using a hearing-aid compressor simulation (Gilbert et al., 2008). A model was designed to test the effect of different sentences, reverberation, source angle and temporal resolution on ILD distribution with and without hearing-aid compression. The effect on the ILD distributions was then

compared to the psychoacoustic results of the study by Catic et al. (2013) and section 4.4.

5.3.1 Model design

Figure 5.8 shows the model used to produce the ILD distributions. It was based on the model described in section 2.4.6 by Catic et al. (2013). The signals [x(n)] consisted of sentences from the IEEE York corpus (Stacey and Summerfield, 2007) and speech-shaped noise (ICRA noise number 5; Dreschler et al., 2001). Input signals were convolved with impulse responses from a hearing-aid impulse response database [HRTF$_{L,R}$], using the responses from dummy head-mounted Siemens Acuris BTEs (Kayser et al., 2009): These impulse responses were recorded at a distance of 1 m in an office ($RT_{60} = 0.3$) or 0.8 m in anechoic chamber. The impulse response from the front, omnidirectional hearing-aid microphone was used in all cases. The convolved signals were compressed using a multichannel compressor simulated in MATLAB (Gilbert et al., 2008). The compressor comprised 3 channels and 10 frequency bands, set according to CAMFIT guidelines for fitting compressive hearing aids (Moore, 2000). For the purposes of the simulation, no gain or equalization was applied to the input signals, in order to determine the effect of compression alone. Attack time and analysis window size was 4 ms and release time was 60 ms. Unless otherwise stated, the same sentence and talker was used as the input for each simulation, the compression ratio used was 3:1, and the ILD fluctuations were lowpass filtered [LPF] with a cutoff frequency of 500 Hz.

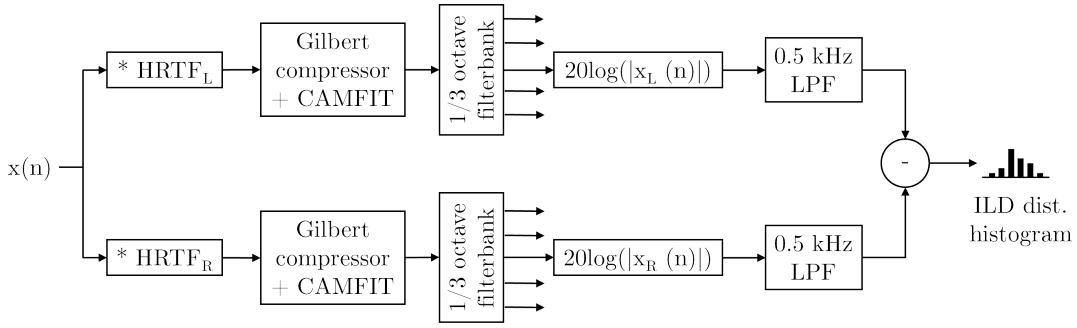

Fig. 5.8: ILD distribution calculation model, based on a model by Catic et al. (2013)

The uncompressed and compressed binaural signals were input to a fourth-order, Butterworth third-octave band filter bank designed to operate to ANSI 2004 and IEC 1995 standards. The output from each filterband was converted to decibels and the instantaneous ILD for each sample was calculated by subtracting the left and right channels. For the sake of brevity and to make some comparisons with Catic et al. (2013), the analyses focused on the 2 kHz third-octave band. Catic et al. (2013) also showed results for the 0.3 kHz band, though acoustic ILDs at 0.3 kHz are minimal unless the source is actually very near the head. In addition, ILD distributions > 2 kHz have been researched by Wiggins and Seeber (2011). A 500 Hz lowpass filter (50 Hz in section 5.3.7) was applied to the instantaneous ILD to remove any fluctuations in ILD deemed to fast to be perceived. 500 Hz was considered in the research by Catic et al. (2013) to be the upper limit of fluctuation rate that could be perceived, though there is no definitive answer in the literature. The instantaneous ILDs for 1 s of each input signal (taken from 0.5-1.5 s to avoid starting silences) were computed and then a histogram was generated to obtain the ILD distribution. Mean, standard deviation, peak ILD (ILD value with greatest relative occurrence in the distribution), and kurtosis were calculated for each distribution, as

these changes in ILD distribution have been shown to correspond to perceptual changes in psychoacoustic tests (Catic et al., 2013). Histograms in the following sections are plotted with ILD in 1 dB histogram bins on the x-axis and relative occurrence ($\frac{\text{number of samples in each bin}}{\text{total number of samples}}$) on the y-axis. Kurtosis (γ_2) is a measure of the "peakedness" of the distribution, given by 5.3.1.

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3$$

where,

$\mu_4 =$ fourth moment around the mean

$\sigma_4 =$ square of the variance of the probability distribution

$3 =$ correction factor applied to make the normal distribution kurtosis = 0

(5.1)

None of the distributions (means) shown in sections 5.3.4 to 5.3.7 are significantly different from one another. However, this does not preclude them from being perceptually relevant. The changes in ILD distribution that produced perceptual effects in Catic et al. (2013) were also not significantly different. These changes are quantified as differences in mean, standard deviation, peak ILD, and kurtosis. Changes in mean ILD – ILD jnds – for tones are approximately 0.5 - 1 dB for normal-hearing listeners (Mills, 1960; Domnitz, 1973; see section 2.2.2). Peak ILD offers an alternative to the mean ILD as an indicator of the perceptual locus of the sound source. Changes in standard deviation give a measure of the change in the width of the distribution, which has been shown to become narrower with decreasing distance by Catic et al. (2013). Kurtosis

gives a measure of the "peakedness" or gaussianity of a distribution. Increasing the peakedness of an ILD distribution caused sounds to be perceived closer to or inside the head of listeners (Catic et al., 2013). To provide a sense of scale for possible kurtosis values for a distribution, a rectangular, uniform distribution has a kurtosis of < -1 and a double exponential distribution has a kurtosis of ≥ 3. Kurtosis values for the uncompressed and compressed ILD distributions in Catic et al. (2013) were unavailable and could not be accurately calculated from the figures in the paper. However, given the similarity of the initial source type, reverberation and source DOA used, the uncompressed distribution in Catic et al. (2013) would be expected to be close to Gaussian (kurtosis ≈ 0), as measured in the current study. From other example distribution kurtosis values, the maximum kurtosis of the distributions in the Catic et al. (2013) study can be estimated as ≈ 2.

5.3.2 Effect of reverberation

Figure 5.9 shows the effect of changing the acoustic environment on ILD distributions for two scenarios: a source in an office at $30°$ to the right of and 1 metre away from the listener and a source in an anechoic chamber at $30°$ to the right of and 0.8 metres away from the listener (the closest anechoic distance available in the impulse response database used). Kurtosis was greater in the anechoic, uncompressed condition (2.93) than the reverberant, uncompressed condition (-0.06). Peak ILD is shifted by 6 dB from anechoic to reverberant distributions (8 and 2 dB peak ILD respectively). The effect of adding low threshold (10 dB below the speech rms) compression to the signal path on ILD distributions is shown in figure 5.10. Kurtosis in the anechoic condition was

reduced by compression from 2.93 to 0.73. Standard deviation was increased by compression in the anechoic condition from 1.81 to 2.84 dB. Peak ILD was increased from 2 to 6 dB by compression in the reverberant condition. Kurtosis was decreased from -0.06 to -0.19 by compression in the reverberant condition

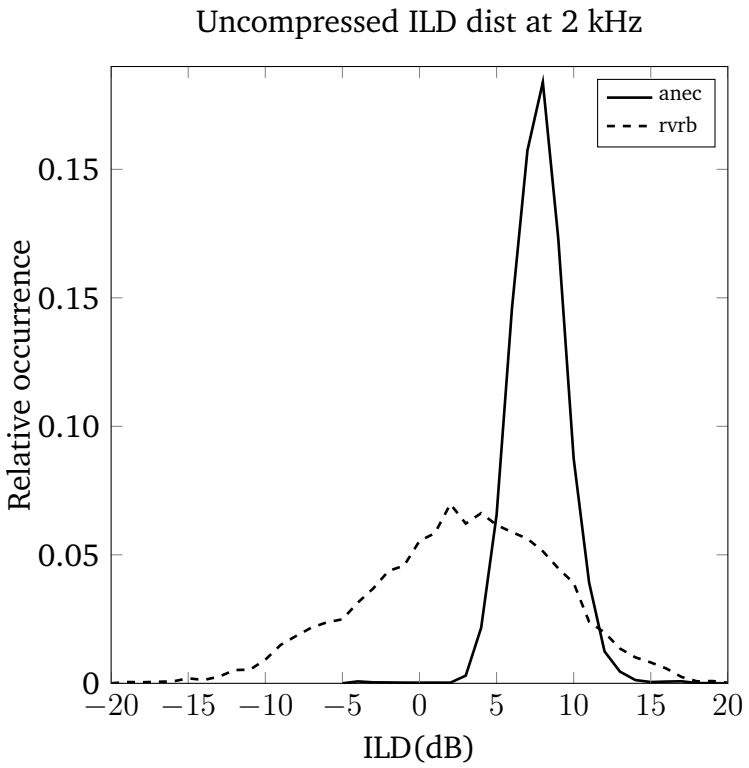

Fig. 5.9: The effect of reverberation on uncompressed ILD distributions. Source angle was $30°$.

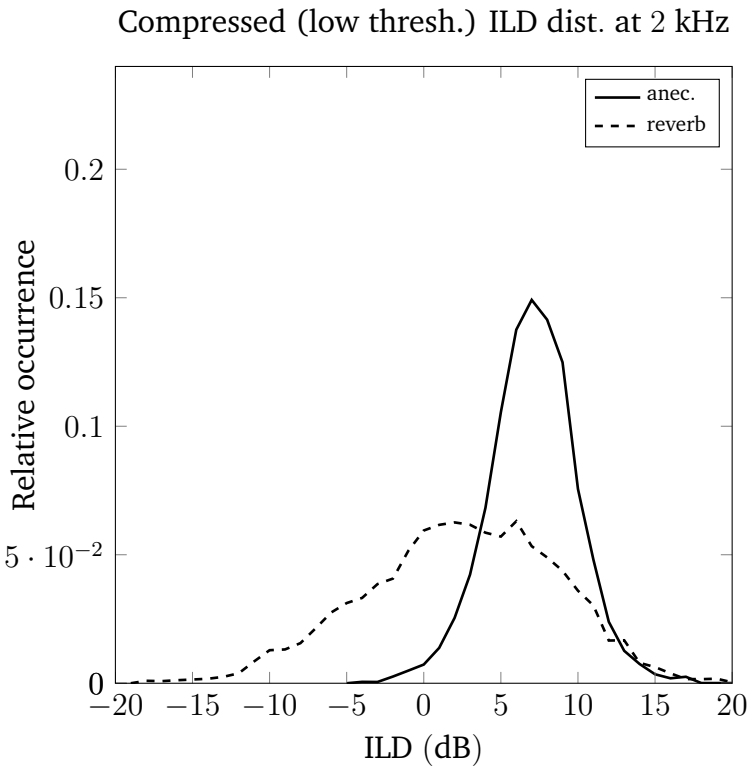

Fig. 5.10: The effect of reverberation on compressed ILD distributions. Source angle was 30°.

5.3.3 Effect of signal type used

Figure 5.12 shows the effect of changing the signal on ILD distributions in a single scenario: a source in an office at 30° to the right of and 1 metre away from the listener. Three randomly selected IEEE York corpus sentences (Stacey and Summerfield, 2007) from the same speaker and ICRA noise (track number 5 / single male talker in modulated noise; Dreschler et al., 2001) were used as input signals. Peak ILD varies between 4 and 5 dB for the signals used. Kurtosis varies between 0.13 and 1.03 (range = 0.9).

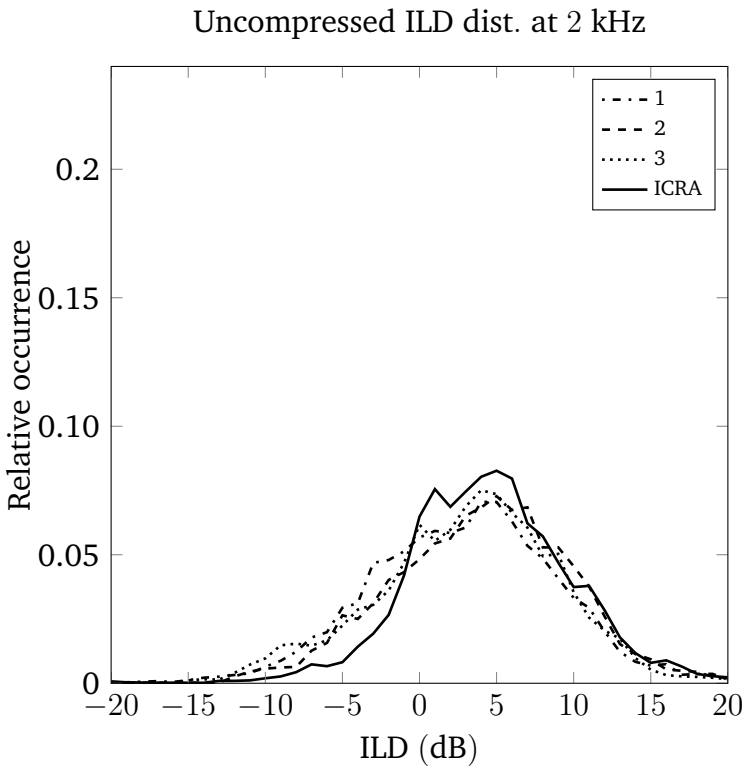

Fig. 5.11: The effect of sentence used on uncompressed ILD distributions. Source angle was 30°.

5.3.4 Effect of compression on speech

Figure 5.12 shows the effect of varying compression threshold on ILD distributions in a single scenario: a male talker in an office at 30° to the right of and 1 metre away from the listener. Compression thresholds were set to none (speech passed through compressor with 1:1 compression ratio), high (threshold set 10 dB above the average rms of the speech) and low (threshold set 10 dB below the rms of speech). Applying compression decreases the kurtosis from -0.06, close to Gaussian (uncompressed) to -0.21 (high threshold) and -0.19 (low threshold). Lowering the compression threshold shifts the peak ILD from 3 to 6 dB.

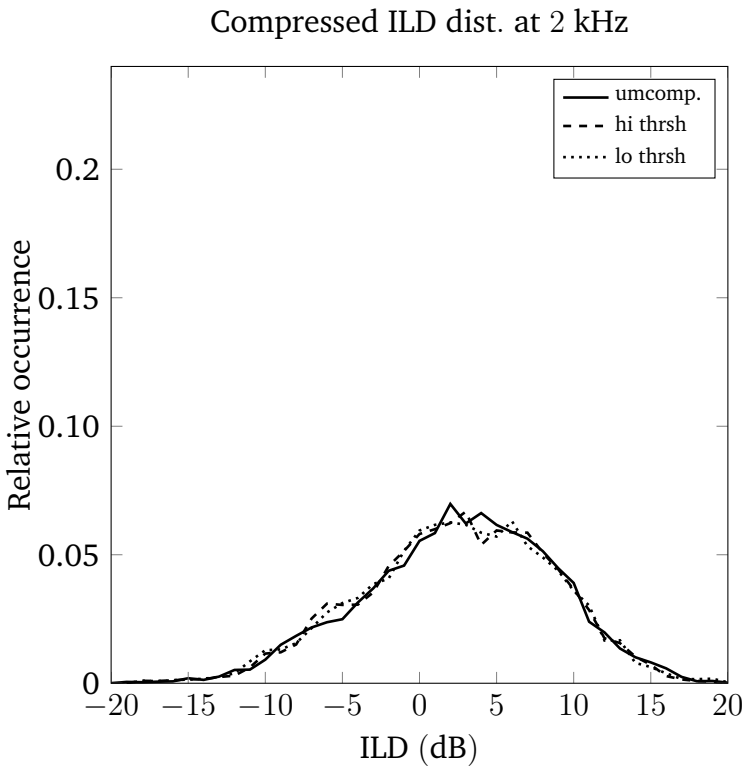

Fig. 5.12: The effect of compression on ILD distributions. The compression ratio was 3:1. uncomp. = uncompressed, hi thrsh = high threshold, lo thrsh = low threshold

5.3.5 Effect of angle

Figure 5.13 shows the effect of angle on ILD distributions for four scenarios: a source in an office at $0°$, $30°$, $60°$, and $90°$ to the right of and 1 metre away from the listener. Peak ILD is -1 dB at $0°$, 2 dB at $30°$, 4 dB at $60°$ and 3 dB at $90°$. The change in kurtosis is -0.93 between $0°$ (0.87) and $30°$ (-0.06), seven times greater than the next largest change between $30°$ (-0.06) and $60°$ (0.07), a change of 0.13. Kurtosis was 0.11 for $90°$, relatively similar to the distribution at $30°$ and $60°$.

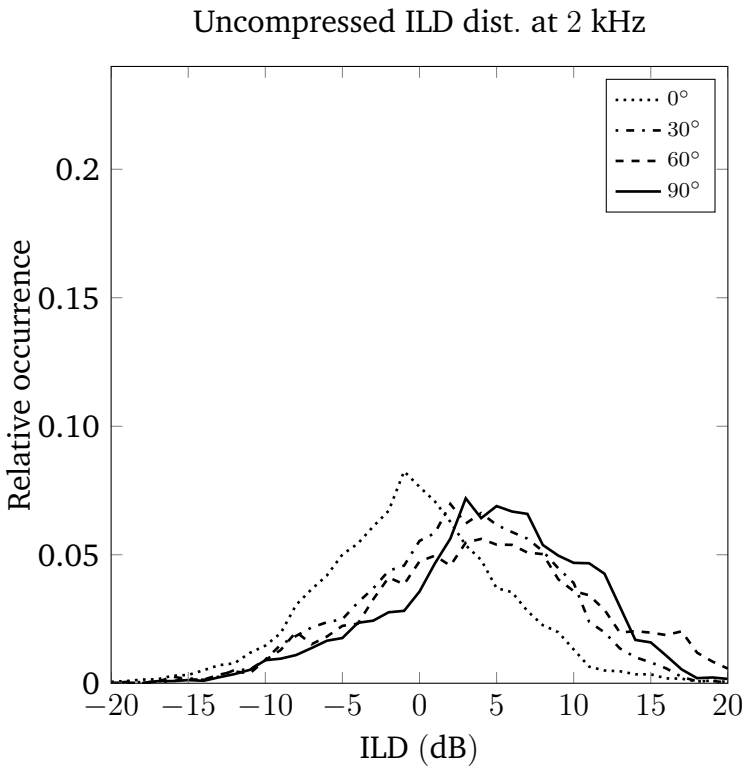

Fig. 5.13: The effect of angle on uncompressed ILD distributions in reverberant conditions.

The effect of angle on ILD distributions with compression added to the signal path is shown in figure 5.14. In comparison to the uncompressed distributions (figure 5.13), compression distorts the relationship between the changing angle and the ILD distribution statistics observed in the uncompressed distributions. Peak ILD is -1 dB at $0°$, 6 dB at $30°$, 2 dB at $60°$ and 6 dB at $90°$. The change in kurtosis from $0°$ (1.23) to $30°$ (-0.19) is 1.42, larger than the change in kurtosis between $0°$ and $30°$ in the uncompressed condition.

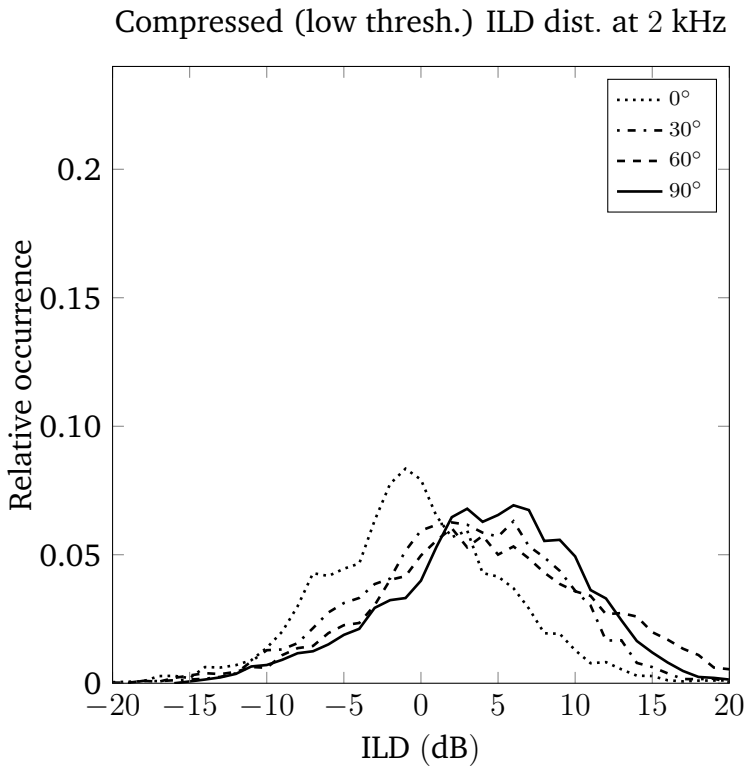

Fig. 5.14: The effect of angle on compressed ILD distributions in reverberant conditions. Compression ratio was 3:1.

5.3.6 Effect on kurtosis of angle, input signal and duration

In order to determine the variation in kurtosis and peak ILD due to the interaction of angle and input signal, the three sentences and ICRA noise number 5 used in section 5.3.3 were combined with the angle and room impulses used in section 5.3.5 ($0° - 90°$). The results are shown in table 5.5. Kurtosis does not vary linearly as a function of angle over the 4 input signals used. Peak ILD shows that the largest shift is generally between $0°$ and $30°$, with the exception of sentence 3 at $90°$. In this case the mean ILD was calculated as 5.2 dB, suggesting a larger skew in this distribution than others. ICRA noise produced the largest variation in peak ILD (11 dB).

Angle (°)	Sentence 1		Sentence 2		Sentence 3		ICRA noise		
	γ_2	peak ILD	γ_2	peak ILD	γ_2	peak ILD	γ_2	peak ILD	γ_2 range
0	0.024	0	0.52	-1	0.80	1	0.83	-3	0.8
30	0.13	4	0.50	5	0.23	5	1.03	4	0.9
60	0.23	4	0.078	6	-0.07	6	-0.01	6	0.24
90	0.46	7	0.35	6	0.46	6	0.38	3	0.11
γ_2 range	0.43	7	0.42	6	0.73	5	1.04	11	

Table 5.5: ILD distribution histogram kurtosis (γ_2) and peak ILD for multiple angle and input signal combinations.

The previous distributions were produced from 1 second samples of each input signal. Using a concatenation of sentences 1, 2 and 3 (8.23 seconds), the results are shown in table 5.6. Peak ILD varies in a similar way to the shorter samples, however kurtosis does not vary linearly as a function of angle.

Angle (°)	γ_2	peak ILD
0	0.47	0
30	0.19	4
60	-0.19	6
90	0.32	7
γ_2 range	0.66	7

Table 5.6: ILD distribution histogram kurtosis (γ_2) and peak ILD for multiple angles using three concatenated sentences.

5.3.7 Effect of temporal resolution

Figure 5.15 shows the effect of temporal resolution on ILD distributions for one scenario: a source in an office at $30°$ to the right of and 1 metre away from the listener. Decreasing temporal resolution by using a 50 Hz lowpass cutoff instead of a 500 Hz cutoff has a large effect on kurtosis relative to the other changes studied, narrowing the distribution. Kurtosis using a 50 Hz cutoff is 2.42 and -0.06 for a 500 Hz cutoff. However Peak ILD is unchanged as the shape of the distribution, not its centre, is changed.

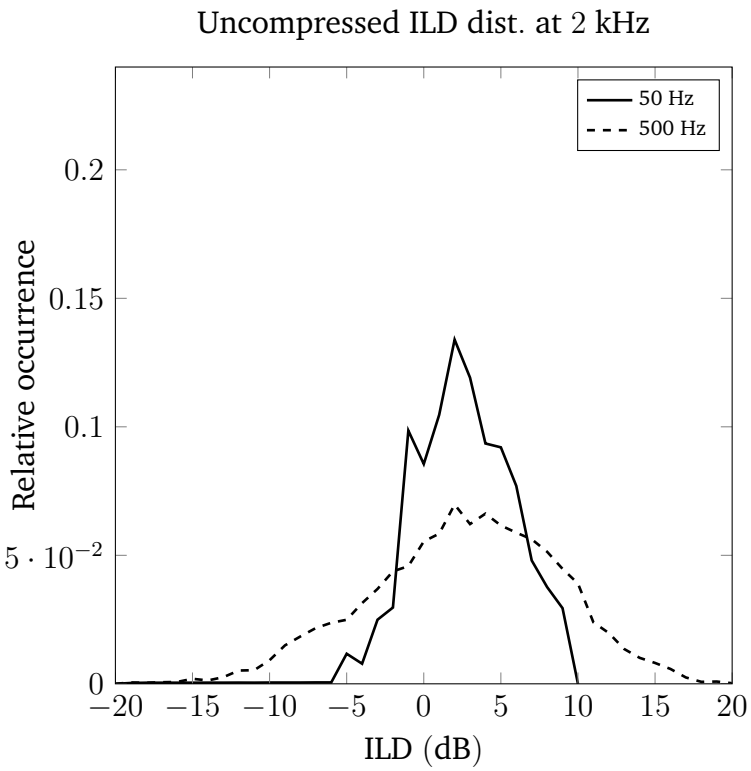

Fig. 5.15: The effect of temporal resolution on natural ILD distributions in reverberant conditions.

5.3.8 Discussion of modeled ILD distributions

The main findings of section 5.3 are summarized in table 5.7.

Hearing-aid compression does not change the shape of ILD distributions.
Small (≤ 1) changes in kurtosis occur depending on the angle of arrival or type of source signal used.
The largest change in peak ILD as a function of angle occurs between 0° and 30°, unless there is a large difference between mean ITD and peak ILD
The large ILD distribution shift toward 0 dB observed in anechoic conditions is partially smeared by reverberation.
Temporal resolution strongly affect the kurtosis of ILD distributions.

Table 5.7: Summary of observations from section 5.3.

Catic et al. (2013) suggested that a hearing-aid compressor may have the effect of reducing the width of an ILD distribution. In the results of section 5.3.4, it can be seen that a high compression ratio with a low or high compression threshold applied to a speech input signal does not affect an ILD distribution at 2 kHz. At 2 kHz, compression does not show the shift in ILD distribution in reverberant environments towards 0 dB. This shift was, however, shown by Wiggins and Seeber (2011) for frequencies $>$2 kHz in anechoic environments.

Compression does distort the relationship between source angle and ILD distribution and also increases the difference in kurtosis between on-axis and off-axis distributions (Behrens et al., 2009). The variation in kurtosis due to different input signals (different sentences from the same talker or speech-shaped noise) is the same (0.9) as the largest variation in kurtosis between on- and off-axis distributions. In addition, variations in kurtosis are not linear with angle across input signals or using concatenations of

sentences. Different sentences have not been shown to change the perception of externalization, either in experiments (section 4.3) or in every day reports, though in section 4.3 the effect of different sentences was not explicitly tested. Therefore, a change in kurtosis may need to be larger to be perceptually relevant to internalization, as more peaked distributions have been shown to increase internalization in normal-hearing listeners (Catic et al., 2013). In general, the largest shift in peak ILD due to angle is between $0°$ and $30°$. This larger shift is in agreement with the increased internalization ratings observed for stimuli presented from $0°$ in the experiment in section 4.4.

Reverberation itself widens the distribution of ILDs and decreases the distribution by a large degree relative to the other conditions tested. In anechoic conditions, the effect of compression is observed as a widening of the ILD distribution, which was also found by Wiggins and Seeber (2011). This widening of the distribution does not occur to the same degree in reverberant conditions. It is possible that the shift observed in anechoic conditions is partially smeared by the reverberant distribution. These results may explain why internalization is more easily induced in anechoic environments (Hanson and Kock, 1957; Schirmer, 1966; Krumbacher, 1969), as a narrower distribution has been shown to increase internalization (Catic et al., 2013).

The temporal resolution used to calculate ILDs has a strong effect on ILD distribution. Lower resolutions – simulated by reducing the lowpass filter cutoff – produce narrower distributions with large kurtosis values relative to the other conditions tested. There is some evidence to suggest that hearing-impaired listeners can have reduced temporal resolutions in comparison to normal-hearing listeners in relation to binaural temporal fine

structure (for a review, see Reed et al., 2009). A loss of temporal resolution may manifest perceptually as internalization, due to the narrower perceived distribution of ILDs. Further work could test the temporal resolution of internalizing and non-internalizing participants in the PIQ questionnaire.

5.4 Conclusions

The prevalence of the perception of internalized sound in unaided, unilaterally and bilaterally aided listeners was investigated using a question from the SSQ (Gatehouse and Noble, 2004) and a short additional questionnaire. Participants were categorized as internalizing and non-internalizing based on their response to the one internalization question in the SSQ. Based on 76 positive responses to the SSQ internalization question and the 64 participants who responded positively to the additional questionnaire, the proportion of our sample population that experience any internalization was 29.5% and 25%, respectively. For unilateral (HA1) participants, those who experienced internalization gave significantly lower scores on the spatial section of the SSQ. With the additional questionnaire (the PIQ), the internalization of sound in different listening conditions was also investigated. Increasing the number of hearing aids increased the likelihood of experiencing internalization at all, but did not increase the frequency with which it occurred if experienced at all.

The effect of a number of acoustic factors on the distribution of ILDs – a possible indicator of internalized sound – was also studied. Using a model of a hearing-aid compressor and hearing-aid microphone impulse responses, the study demonstrated:

1. Hearing-aid compression does not change the shape of ILD distributions in the 2 kHz band for everyday listening scenarios. This is the case at low and high compression thresholds using a high (3:1) compression ratio and very short attack and release times.

2. Small ($\leq$ 1) and inconsistent changes in kurtosis occur depending on the angle of arrival or type of source signal used. Larger changes in the kurtosis of ILD distributions may required to produce perceptual effects.

3. The largest change in peak ILD as a function of angle occurs between 0° and 30°, unless there is a large difference between mean ILD and peak ILD, as seen for sentence 3 at 90° in table 5.5.

4. The large ILD distribution shift toward 0 dB observed in anechoic conditions is partially smeared by reverberation.

5. Temporal resolution strongly affects the kurtosis of ILD distributions, providing possible analytic evidence for the internalization effects seen in previous studies for hearing-impaired listeners.

This analysis shows that ILD distribution can be affected by a number of factors, both acoustical and perceptual, and links these factors to the results of previous psychoacoustic results. The second result in particular does not support kurtosis of ILD distribution as an important predictor of percieved internalization.

CHAPTER 6

Head-motion controlled hearing-aid systems

6.1 Introduction

Human listeners do not listen passively to their auditory environments. They are often not only listening, but *communicating*, using body language such as moving their hands, heads and faces to convey additional information. In addition, listeners will move their heads to attempt to increase the level of a target source in noisy conditions (Brimijoin et al., 2012; see section 2.2.6). Until recently, these additional listening cues could not be captured or utilized by hearing aids, despite the wealth of information on both listener behaviour and the stability of the auditory scene that they could convey. Parts of section 6.4 were published in: Boyd, A.W., Whitmer, W. M., Akeroyd, M. A. (2013) Improved estimation of direction of arrival of sound sources for hearing aids using gyroscopic information. *POMA* **19**:030046. Parts of section 6.5 were published in: Boyd, A.W., Whitmer, W. M., Brimijoin, W. O., Akeroyd, M. A. (2013) Using head movement information to increase the accuracy of

computational auditory scene analysis in a microphone array. *Proceedings of the Institute of Acoustics* **35**(1):275-282.

This chapter covers novel research conducted into the use of head-mounted inertial measurement systems to improve hearing-aid signal processing. This research direction has several motivations. The original research question on the internalization/externalization continuum assumed that hearing aids were the major cause of internalization in hearing-impaired hearing-aid users and that new signal processing techniques could be developed to restore externalization. The behavioural research in chapter 4 reported some evidence for the hearing aid *exacerbating* the perception of internalization for normal-hearing listeners, but no clear direction for a signal-processing scheme to improve externalization for those listeners suffering from it. The collaborative work of Brimijoin et al. (2013) highlighted the importance of dynamic cues due to head motion for externalization (Brimijoin et al., 2013; described in section 2.4.8). Given the importance of head movements for the externalization of sounds and source localization (see sections 2.4.8 and 2.2.5), and recent advancements in inertial measurement technology, head movements could be utilized in hearing aids to improve their performance. Discussions with Stefano Cosentino (UCL Ear Institute) on the use of changes in ITD during head motion to selectively amplify and attenuate sources were indispensable in the initial stages of this research.

Previous research on and patents for the use of head movement and inertial sensors in hearing-aid signal processing is described in section 3.13. There are three main innovations covered in this chapter. First, instantaneous head-movement cues were used to auto-select optimal microphone modes

based on user behaviour. Second, algorithms were developed to compensate for head movements, stabilizing the acoustic scene and improving source direction of arrival (DOA) estimation performance during head movements. Third, head motion cues were used to extend the limits of two-microphone correlation-based DOA estimation techniques from just the front or rear hemifield to both ($360°$) using a biomimetic technique.

This chapter is organized as follows. Novel hardware and software implementations designed to simultaneously record head movements and binaural audio are presented in section 6.2. The development of offline and real-time simulations of head-movement controlled directionality, including the initial measurements of a head-mounted gyroscope and motion tracking system are presented in section 6.3. A system that compensates for head movements while estimating the DOA sources is presented in section 6.4. Section 6.5 describes a biomimetic system for estimating the DOA of sound sources in the front or rear hemifield by utilizing the relative motion of sound sources due to head movements.

6.2 Novel hardware systems and software implementations

This section describes the novel systems designed for the research in this chapter. The integration of the Wiimote (game controller with infrared camera) head-tracking system with a MEMS gyroscope is described in section 6.2.1. The synchronization of the MEMS gyroscope and 9-axis (combined accelerometer,

gyroscope and magnetometer) inertial sensor (see section 3.12) with binaural audio recording is described in section 6.2.2.

6.2.1 Integration of portable Wiimote head-tracking system and MEMS gyroscope

The portable Wiimote-based head-tracking system (previously described in detail in section C.2) was combined with the gyroscope and Arduino Uno microcontroller to measure both the listener's head movements and the output of the gyroscope during head turns. The combined system is shown in figure 6.1. Head movement was output as angular position relative to a calibrated axis. The gyroscope was connected to the Arduino Uno as seen in figure 6.1. Its output was raw voltages, linearly related to angular velocity, in three measurement axes using the I^2C protocol to the Arduino and output to MATLAB over a USB-serial connection.

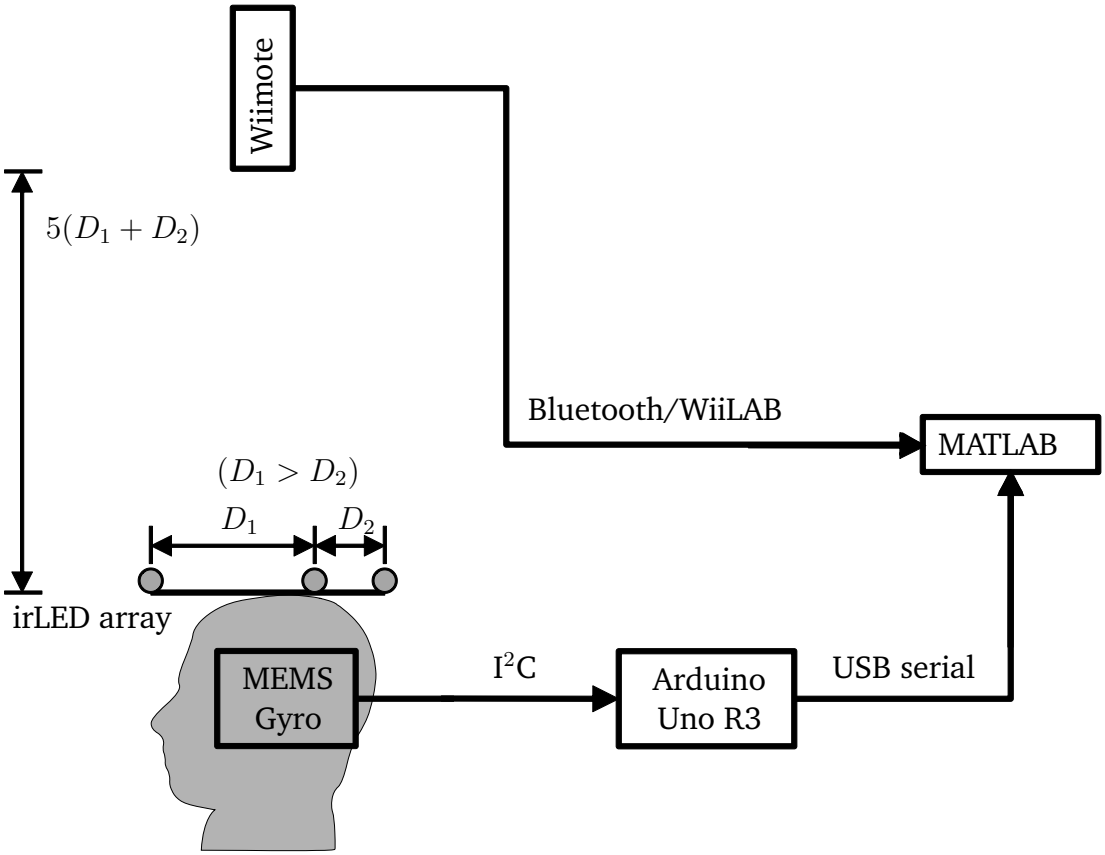

Fig. 6.1: The angular head movement and gyroscopic information measurement hardware.

The head movement and gyroscope readings were output using separate clocks. In order to maintain synchrony between the readings, the readings were recorded in MATLAB using a for-loop, calling the most recent reading from the head tracker and gyroscope each time. Both systems could individually be recorded by MATLAB at $\geq$ 100 Hz. However, in order to record from both simultaneously, MATLAB required that the serial port for the gyroscope be closed before reading from the head tracker (also output over USB serial). This opening and closing of the serial port on each loop reduced the sample rate to between 10 and 20 Hz, the variability being due to the variable time

taken to open and close the serial port on each loop. The sample rate was calculated using a timer object within the MATLAB for-loop. Using a buffering system could have improved the sample rate of the system, however it would not have solved the synchronization issue, as the exact time alignment of the gyroscope and head movement samples within the buffer would be unknown. Sampling rate was improved in later systems (section 6.2.2) as the serial port could remain open during recording.

6.2.2 Gyroscope and 9-axis integration and synchronization with audio recording in MATLAB

The gyroscope information had to be recorded in synchrony with information from the binaural microphones. The system is shown in figure 6.2. This posed a design problem, as the Arduino-based gyroscope and the microphone recordings ran on two different internal clocks, leading to eventual asynchronization even if both were started at the same time. One solution to this problem is to include a third external clock and slave both the Arduino and the audio recorder to it. However, the difficulty of implementing a third clock for the audio recording was great, as MATLAB and the driver software for the external soundcard (a Zoom H4n) did not allow low-level control of the audio sampling rate. In addition, a variable sampling rate would not be playable using standard audio codecs and would most likely result in gaps and artifacts in the audio.

A solution was found using the *playrec* function for MATLAB (www.playrec.co.uk), which uses the PortAudio audio input/output control library to control the recording buffer size directly from MATLAB. As the buffer

could be controlled in a for-loop within MATLAB, the serial connection to the Arduino could be read during each loop. By clearing the serial buffer each time, only the latest output from the Arduino would be recorded. Running the Arduino at a faster rate than the MATLAB for-loop resulted in the serial always being full when called from MATLAB. Finally, the *playrec* buffer was made large enough to run through one loop before the audio recording buffer was full, but no larger. This process allowed the recorded serial information to be synchronized to within one buffer length and the systems remained time-aligned throughout recording.

The 9-axis sensor and audio recording system operated in a similar way as the system described above. The 9-axis sensor was connected to the Arduino Uno as seen in figure 6.2. Though all three sensors had digital outputs, the output was transmitted to the analog inputs of the Arduino as a multiplexed I^2C signal, where they were decoded back into digital signals. All compensation and orientation calculations were performed using the pre-loaded Arduino firmware (see section 3.12.4). The calculated orientation information was then output over USB serial and read into MATLAB using serial commands.

In the 9-axis sensor system, the issue of synchronization with the soundcard and Matlab was solved by using a 0.04 s audio buffer and reading the last output value from the 9-axis sensor during that time. As the sample rate for the 9-axis sensor was optimized for 50 Hz, the two systems remained synchronized to within 1 sample from the 9-axis sensor, again without the need for an external clock. The use of two different clocks meant that the audio and 9 axis sensor outputs could be out of sync by one iteration. If this maximum asynchrony occurred when running at 50 Hz, the serial buffer would be empty, resulting

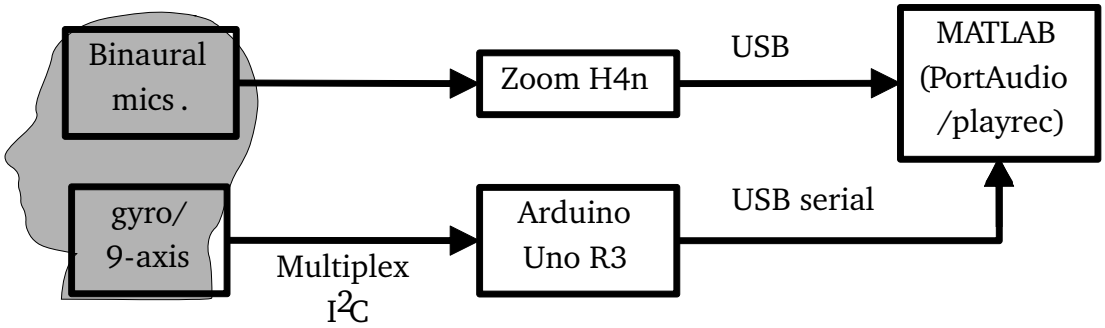

Fig. 6.2: Combined audio and gyroscope/9-axis sensor system.

in either an error, an empty reading or a gap in the audio. Running at half the sample rate of the 9-axis sensor guaranteed the serial buffer would have at least one reading at all times. The 9-axis sensor required several seconds at start-up to obtain enough information to calculate its position. When using the 9-axis sensor (section 6.5, the first five seconds of each recording were discarded.

The audio path for the system is shown in figure 6.2. Sound Professionals MS-TFB-2 binaural microphones recorded from positions inside or outside the pinna. A Zoom H4n sound recorder was used as an external soundcard and the audio signals were transferred to MATLAB via USB.

It is acknowledged that this system was not efficient, resulting in several conversions between analog and digital domains and several calculation steps. Ideally the system would output raw voltage readings to MATLAB and all calculations would be performed within MATLAB. For testing purposes, however, this amalgamation of systems could be built more quickly and was more flexible than a purpose-built system would have been, while not sacrificing on required synchronization and performance for testing. In common with the earlier system described in section 6.2.1, the use of a buffering system would have improved the sampling rate at the expense of

maintaining synchrony between the audio and gyroscope/9-axis sensor within each buffer.

6.3 Gyroscopic control of directionality (GCD)

This section describes the development, offline and real-time testing of a system to control the directionality of a hearing-aid microphone array output using gyroscopic information. The concept is described in section 6.3.1 and the implementation of the offline first-order ADMA and real-time DMA simulations are descried in section 6.3.4. The results of offline and real-time simulations are given in sections 6.3.3 and 6.3.4.2 and discussed in section 6.3.5.

6.3.1 Concept for GCD

Hearing aids with two microphones generally have two modes selectable by the listener: omnidirectional and directional (see section 3.4). In the omnidirectional mode, the hearing aid amplifies sounds coming from all directions. Attenuation is provided by the head and torso, in the same way as the listener's ears, though there is minimal pinna filtering in BTE hearing aids, due to the microphone position outside and above the ear. In the directional mode, the hearing aid amplifies sounds coming from one direction more than others. Directional modes provide improved SNRs for speakers directly in front of the listener in noisy environments. Listeners can take longer to successfully localize new talkers when using directional modes, due to the reduced audibility of off-axis sources when using a directional microphone (Brimijoin et al., submitted). This puts them at a disadvantage in group conversations, potentially leading to increased listening effort. The

omnidirectional mode improves the audibility of off-axis sources, however the SNR of a target source in front of the listener is reduced, potentially reducing intelligibility when the new talker is correctly localized.

During head turns, it is impractical to ask a listener to switch to omnidirectional mode to localize a new source then switch back to directional mode, due to the reduced dexterity of many aged hearing-impaired listeners and the time elapsed for switching modes and the distraction involved. To obtain maximum benefit, the hearing aid could rapidly and intelligently switch between modes. The directional mode provides maximum SNR benefit for a listener attending to a single, on-axis source and the omnidirectional mode provides audibility for off-axis sources when the listener is seeking another target source to attend to. Currently available classification and automatic program selection systems for hearing aids rely on audio cues alone (e.g. Xiang et al., 2010).

The aims of this section (6.3) were to research how the output of a gyroscope can determine listener behaviour and how a GCD could theoretically resolve the behaviour of a listener attending to one talker and searching for another, and set the directionality of a microphone array accordingly. The system presented here uses a head-mounted MEMS gyroscope to detect horizontal rotational head movement and select the microphone directionality that may be desired by the user based on this information.

6.3.2 Implementation of offline ADMA

This section describes the implementation of the GCD in offline simulations. Section 6.3.2.1 describes the methods for measuring the gyroscope output.

Section 6.3.2.2 describes the system implemented for the offline ADMA simulation. Section 6.3.3 presents the results of the offline ADMA simulations.

6.3.2.1 Methods for measuring gyroscopic output

The system was first designed as an offline simulation of an adaptive directional microphone array (ADMA) and omnidirectional output (see section 3.4) in a simulated hearing aid, using head-movement information recorded using the infrared head tracker described in section C.2 and gyroscope information simultaneously recorded using the system described in section 6.2.1. The head movement and gyroscope outputs were recorded over a 25-second measurement time. One listener (the author) sat in the centre of the 24-loudspeaker ring shown in figure 6.3 facing the 0° loudspeaker, with the head-mounted irLED array (infrared light-emitting diode, see section C.2) and the gyroscope/Arduino also mounted on the head. One PC was used to present 10 seconds of concatenated sentences of one talker in the IEEE York corpus (Stacey and Summerfield, 2007) from the 0° loudspeaker. A second PC was used to record the head-movements and gyroscope output. After 8 seconds, a second talker from the IEEE corpus was presented from another loudspeaker (45°, 60°, 90°, 135° and 180°) for 10 seconds. During the two second overlap, the listener moved his head to point towards the second talker. After the presentation of the second sentence was complete, the listener turned back to 0° to await the next presentation.

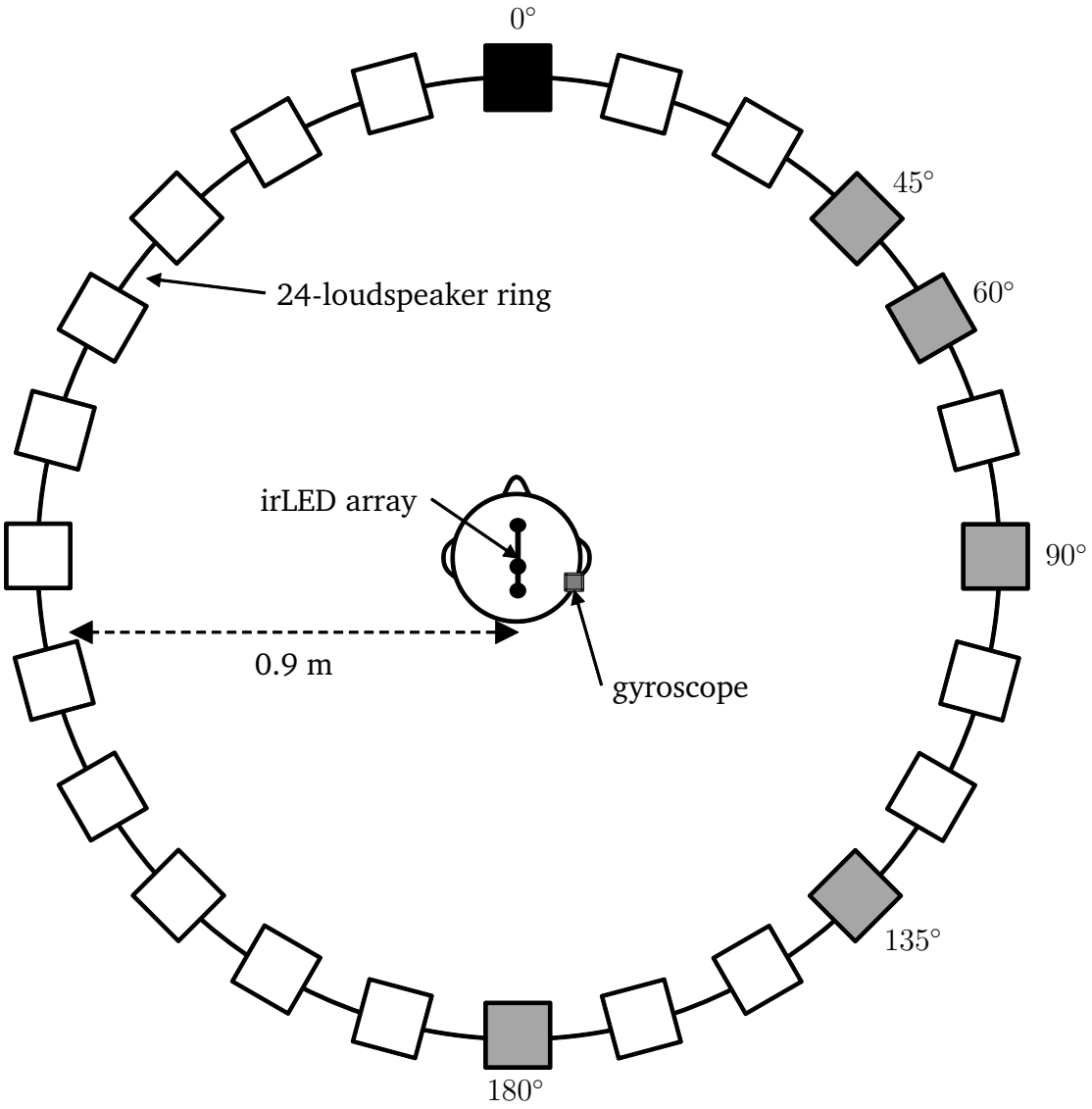

Fig. 6.3: Experimental apparatus for recording head movements and corresponding gyroscopic output.

6.3.2.2 Simulation of offline ADMA

The ADMA and omnidirectional hearing-aid outputs were calculated using a hearing-aid impulse response database, using the responses from the front and rear microphones of a dummy head-mounted Siemens Acuris hearing aid in 0.8 metre source-microphone distance anechoic conditions (Kayser et al., 2009).

Responses for each microphone were available from 0° to 180° in 5° increments for left and right hearing aids. The ADMA was modeled by combining the impulse responses using equations 3.7, 3.8 and 3.9.

Simulations were run using the modelled hearing-aid microphone responses and the recorded head movements. The gyroscope output controlled the mix between the omnidirectional and the ADMA output. A threshold of 3 mV was set for the gyroscope output. Below this threshold, the ADMA directional microphone response was selected. Above this threshold, the omnidirectional response was increased as the directional ADMA response was decreased. Once the gyroscope output dropped below threshold, the ADMA response was gradually increased and the omnidirectional response decreased. The time for mixing from one response to the other was set at five samples, which in the case of the 135° turn was 0.5 seconds.

Figure 6.4 shows the method used to produce the simulated GCD output. A front and rear sound-source direction was chosen. Fourier transforms of the front and rear microphone head-related impulse responses and the magnitude of the front and rear cardioid responses were obtained at a chosen frequency. The head angle at each timestep and the relative angle of the sound sources were calculated. The β value was calculated to attenuate the rear source and the mix between the directional $((\alpha - 1) * \text{directional})$ and omnidirectional $((\alpha * \text{omnidirectional})$ outputs was selected (α values between 0 and 1). The output of the array at a chosen angle was then plotted for a selected timestep.

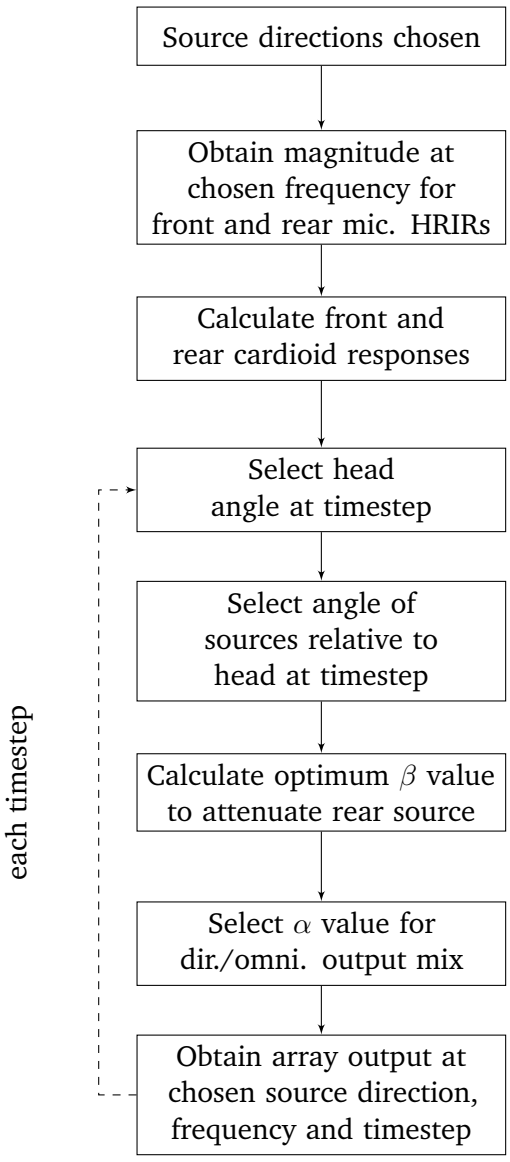

Fig. 6.4: Offline simulation of the gyroscopic control of directionality system.

6.3.3 Results of head-turn recordings and offline ADMA simulations

Figure 6.5 shows the output of the gyroscope in 3 axes during a 45° head turn. Motion occurred between 9 and 10.5 seconds. The maximum absolute output of the gyroscope was 43 mV in the yaw (x) plane, 11 mV in the pitch (y) plane and 14 mV in the roll (z) plane. The duration of the head turn was 1.2 s. Duration was measured as the time the x plane value of the gyroscope output was > 0. The gyroscope clearly captures movement in the x-axis, and there is some movement in the y and z axes. The output is relatively flat during the turn, indicating a constant angular velocity.

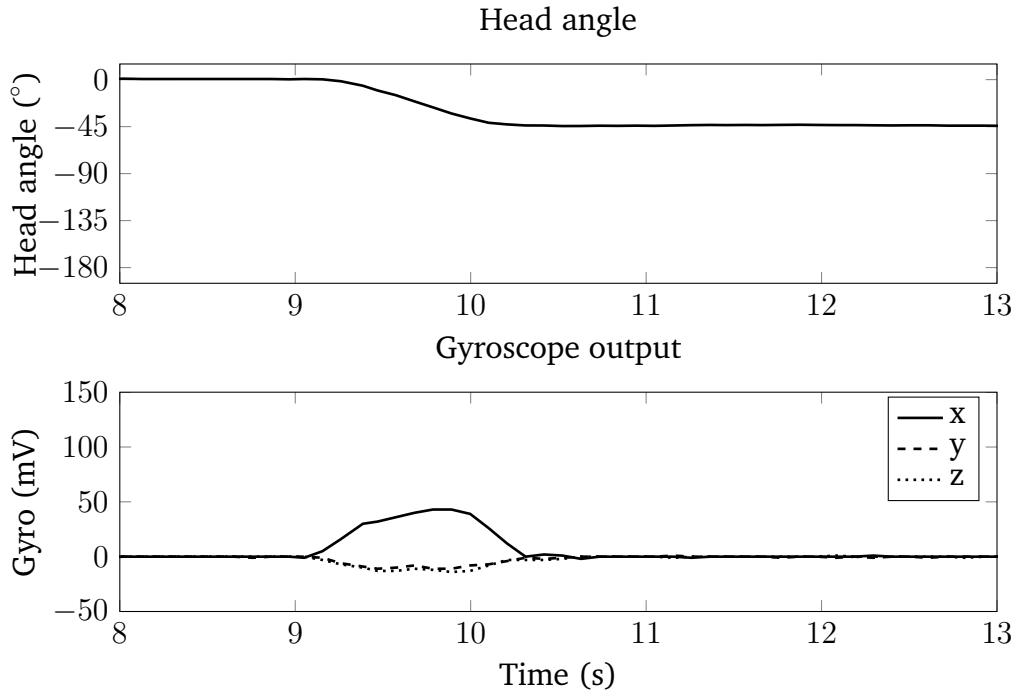

Fig. 6.5: Head angle (upper panel) and gyroscope output (lower panel) for 45° head turn.

Figure 6.6 shows the output of the gyroscope in 3 axes during a 60° head turn. Motion occurred between 10.5 and 12 seconds. The maximum absolute output of the gyroscope was 56 mV in the x plane, 15 mV in the y plane and 17 mV in the z plane. The duration of the head turn was 1.8 s. The gyroscope output differs from that in figure 6.5 in its shape, as it has more of a peak and tails.

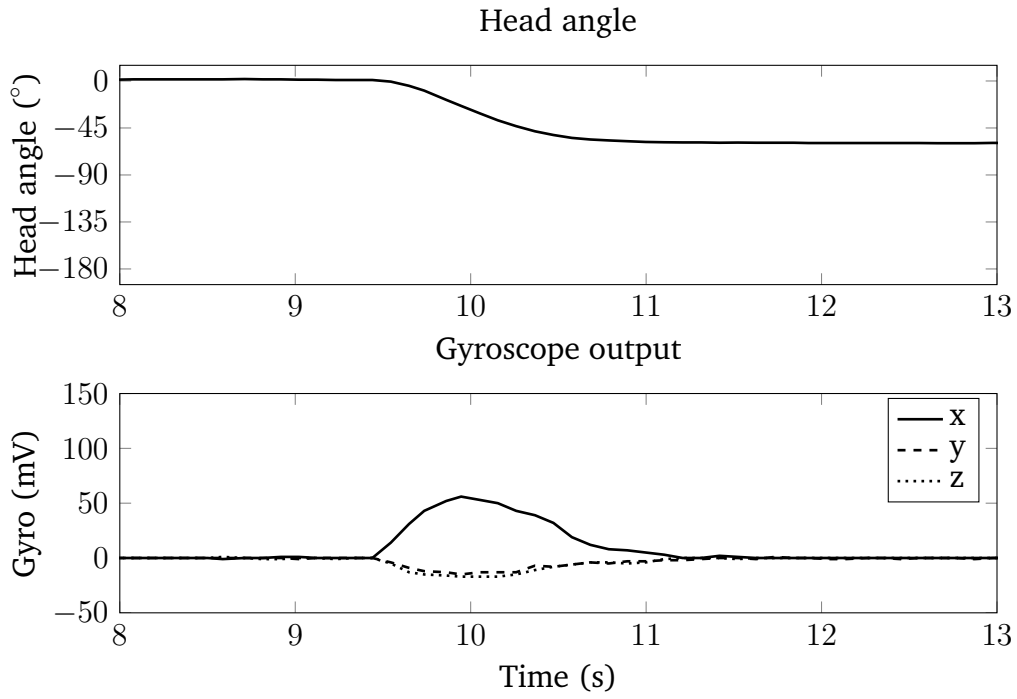

Fig. 6.6: Head angle (upper panel) and gyroscope output (lower panel) for 60° head turn.

Figure 6.7 shows the output of the gyroscope in 3 axes during a 90° head turn. The maximum absolute output of the gyroscope was 70 mV in the x plane, 36 mV in the y plane and 36 mV in the z plane. The duration of the head turn was 2.6 s. In comparison to figures 6.5 and 6.6, there is more movement of the head, manifested in the gyroscope voltage change along the y and z axes. The

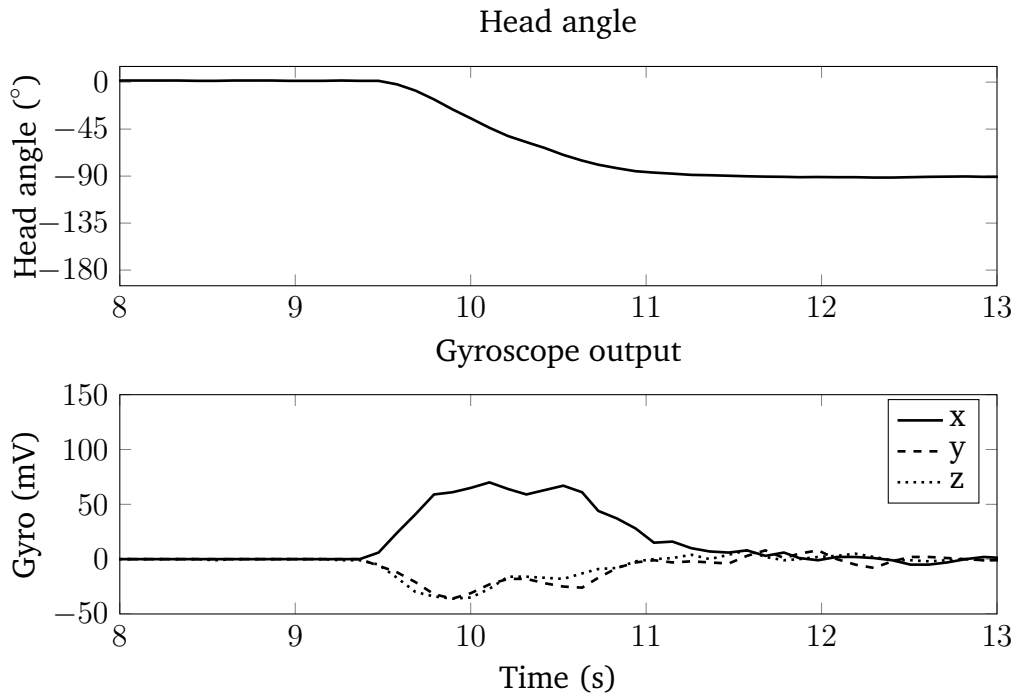

Fig. 6.7: Head angle (upper panel) and gyroscope output (lower panel) for 90° head turn.

tail of the output indicates that proper setting of the gyroscope threshold may require additional care in its calculation.

Figure 6.8 shows the output of the gyroscope in 3 axes during a 135° head turn. The maximum absolute output of the gyroscope was 80 mV in the x plane, 40 mV in the y plane and 33 mV in the z plane. The duration of the head turn was 3.1 s. Again, the increased head movement is manifest in the y and z axes outputs and movement has a period of constant velocity, similar to, but more variable than figure 6.5. Again, a long tail can be seen, indicating a slow deceleration of the head.

Figure 6.9 shows the output of the gyroscope in 3 axes during a 180° head turn. The maximum absolute output of the gyroscope was 126 mV in the x plane, 38

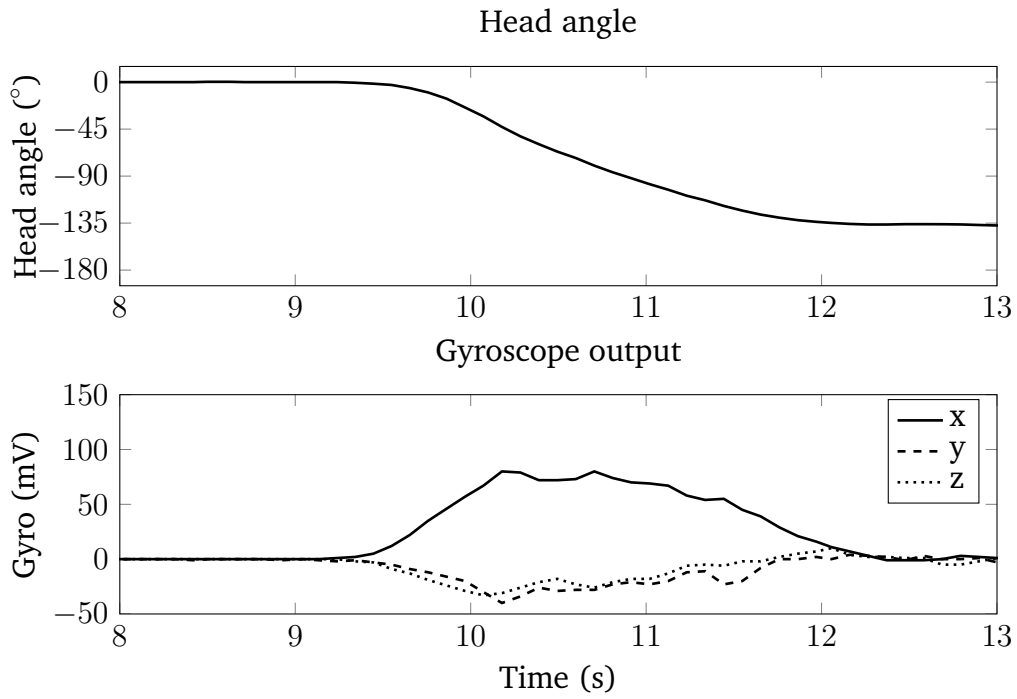

Fig. 6.8: Head angle (upper panel) and gyroscope output (lower panel) for 135° head turn.

mV in the y plane and 45 mV in the z plane. The duration of the head turn was 2.7 s. This head turn produces the largest angular velocity as shown by the peak x axis output and clear periods of acceleration and deceleration. All three figures (6.7, 6.8 and 6.9) show more movement than the smaller angle (figures 6.5 and 6.6) results.

Figure 6.10 shows the theoretical outputs of an ADMA and gyroscopically controlled mix for sources at 0° and 135° during a 135° head turn from 0° to 135°. It was modeled using the head turn shown in figure 6.8. Gyroscopic threshold was set at 3 mV and mix time from full ADMA to full omnidirectional output was 5 samples or approximately 0.5 s. At the start of the head turn (figure 6.10a), the output was full ADMA. The microphone patterns were calculated for ideal cardioid responses using equations 3.7 and 3.8. The null was calculated

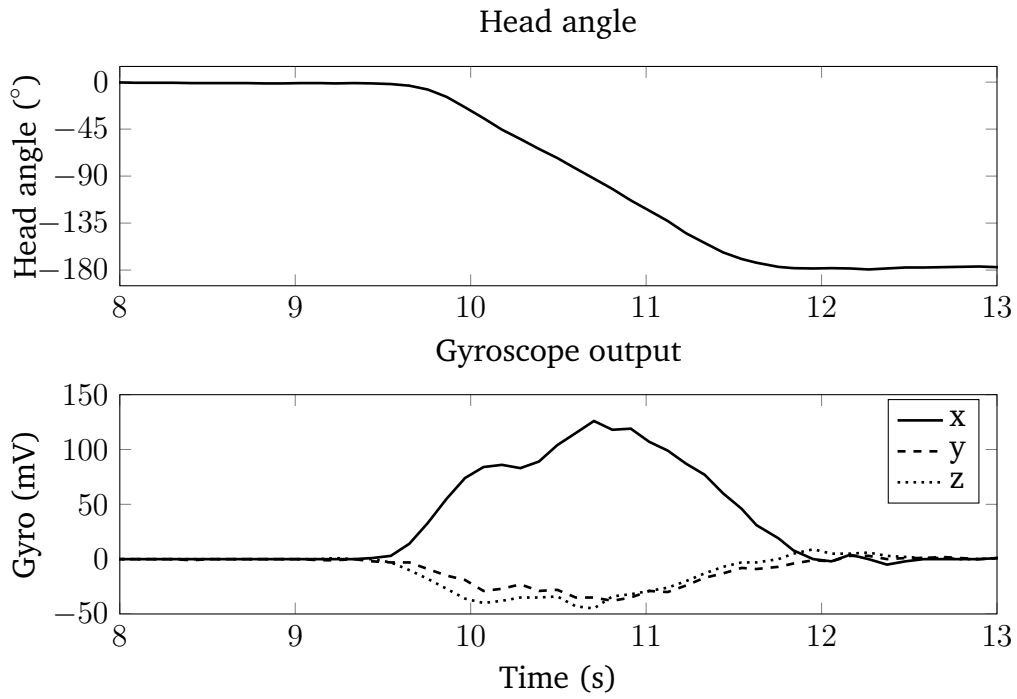

Fig. 6.9: Head angle (upper panel) and gyroscope output (lower panel) for $180°$ head turn.

using equation 3.9 in section 3.6 for an angle of $-135°$. When the head moves and the gyroscope output is above threshold, the ADMA output is decreased and the omnidirectional output increased. This has the effect of reducing the attenuation of the source at $-135°$, seen in figure 6.10b. When the head is at $90°$ (figure 6.10c), the ADMA attenuates the source at $0°$, as it is now in the rear hemisphere relative to the head. The gyroscopically controlled mix is now fully omnidirectional. In figure 6.10d, the gyroscopic output is < 3 mV, resulting in the increase of the ADMA output and decrease in the omnidirectional output. The source at $0°$ is attenuated. In figure 6.10e, the head turn is complete and the output is fully ADMA, with the null focused on the source at $0°$.

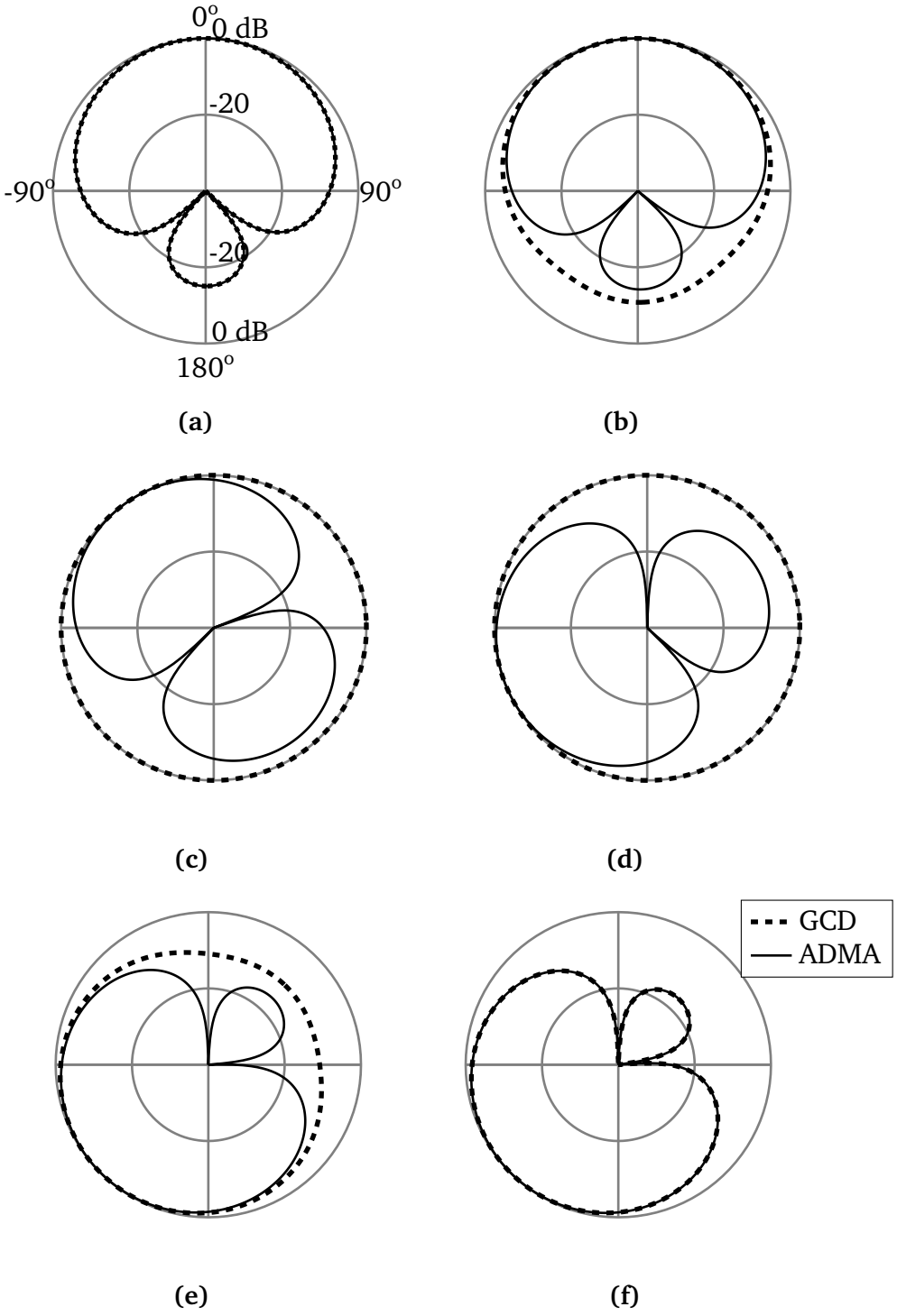

Fig. 6.10: Polar plots of theoretical ADMA and gyroscopically controlled mix outputs during 135° head turn for sources at 0° and 135°.

The top panel of figure 6.11 shows the modeled outputs at 135° of a gyroscope-controlled mix of fixed directional and omnidirectional outputs ("mix") as well as the outputs of the fixed DMA ("fixed") and ADMA ("adpt") array outputs at the same angles for comparison. The hearing-aid microphone HRIRs were used in this case to give a more accurate representation of real world results. The lower panel of figure 6.11 shows the 135° head turn used to model the outputs. The gyro-controlled mix provides a maximum gain difference over the fixed DMA and the ADMA of 18.7 and 30 dB respectively at a head angle of -24.3°. The gyro-controlled mix is within 1 dB of maximum output 0.6 s before the fixed DMA or ADMA outputs. This means that there is a large gain benefit for potential off-axis sounds during head motion with the GCD compared to the fixed DMA and ADMA. This is an extreme result, using anechoic conditions and a far-field microphone response.

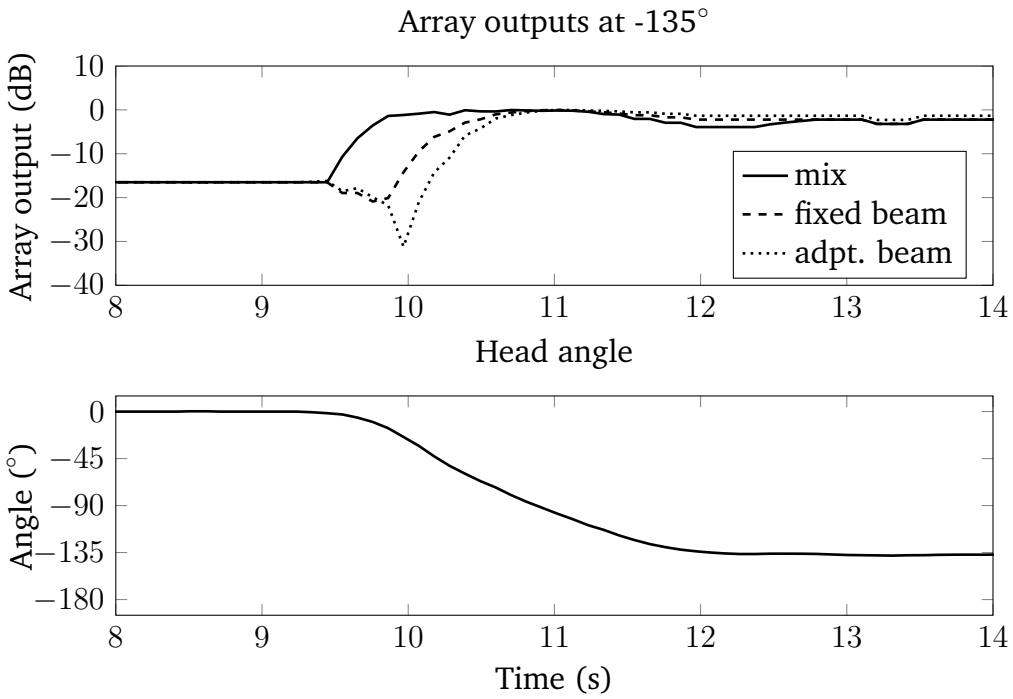

Fig. 6.11: Gyro-controlled mix, fixed directional and adaptive directional array outputs at 2 kHz for source at 135° (top panel), for a 135° head turn from 0° to 135° (bottom panel)

6.3.4 Implementation of realtime DMA simulation

This section describes the implementation of the GCD in realtime simulations. Section 6.3.4.1 describes the methods for implementing the realtime simulation. Section 6.3.4.2 presents the results of the realtime DMA simulation.

6.3.4.1 Methods for GCD simulation

The gyroscopically controlled directionality (GCD) was also implemented as a real-time simulation for determining the benefit (if any) of the system for listeners. An experiment using this simulation tested the effect of speed and accuracy of localization using a modeled fixed first-order directional array and a mixing scheme similar to that described in section 6.3.2.2.

To test the efficacy of the GCD, four normal-hearing listeners oriented to off-axis sounds which were attenuated by fixed directional and GCD patterns in a virtual auditory environment. The experimental apparatus used is shown in figure 6.12. The listeners were asked to virtually localize an off-axis source and turn their heads as quickly as possible towards the virtual source presented over headphones at $\pm 100^{\circ}$, $\pm 120^{\circ}$ and $\pm 140^{\circ}$ in the presence of another talker presented at 0° and unmodulated ICRA noise (number 5; Dreschler et al., 2001) at the same level as the talkers (without directional attenuation). Once localized, the listeners held their head in that position until the stimulus stopped, whereupon they turned their heads back to 0° to await another presentation. This was repeated five times for each position and array setting in random order. Each stimulus was presented for 8 seconds.

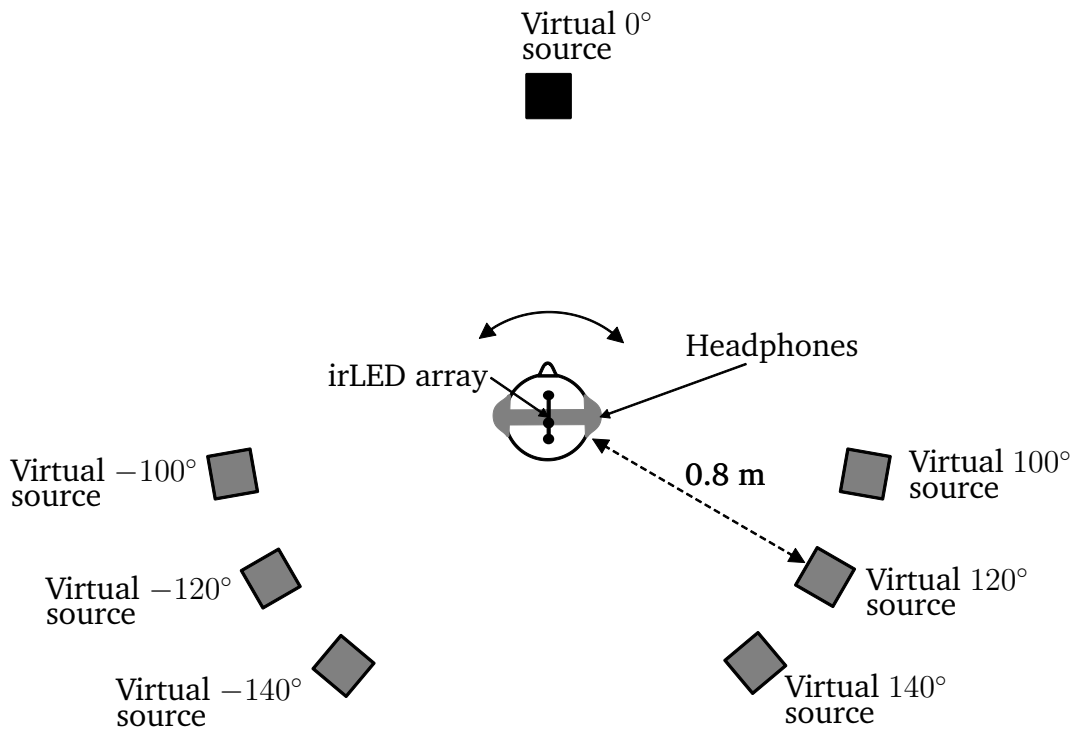

Fig. 6.12: Experimental apparatus for recording listener head movements during localization using simulated directional and gyroscopically controlled microphone array outputs.

Due to the computational complexity possible with a real-time system, the offline simulation described in section 6.3.2.2 was simplified in a number of ways. Firstly, the adaptive response was changed to a fixed directional response to remove the need for constant calculations of the null position. The null point was fixed at $120°$. Secondly, due to the number of convolutions required using real microphone responses, the theoretical response of a DMA was used (as given in equation 3.8). This meant that the MIT head-related impulse responses for each source could be simply multiplied by the theoretical gains for the DMA at that angle. Since the motion of the head was known using the head-tracking system, recorded head movement was used to control the mixing. In addition, a faster mixing speed (near instantaneous as the listener moved their head) was

used as the objective was to determine whether the additional gain provided to the $\pm 100^{\circ}$, $\pm 120^{\circ}$ and $\pm 140^{\circ}$ sources by the switch to omnidirectional output provided a measurable benefit in speed of localization.

To remove any possibility of the virtual auditory display processing being different between the fixed directional and motion-controlled conditions, the same logic decisions and threshold monitoring was performed for both, however in the fixed directional condition it had no effect on the final output.

Four normal-hearing listeners (one female) participated, aged 24-43 years old. The audio was presented over headphones at 65 dB SPL.

6.3.4.2 Results of real-time DMA simulation

The benefit of the system to listeners during localization was tested using a simplified DMA simulation (see section 6.3.4.1). As behaviour varied across the listeners tested, each listeners results are shown separately. Head position was recorded while listeners turned to a virtual off-axis source with either a theoretical fixed DMA (fixed) or GCD (gyro) applied. Successful localization is defined here as the first time the head angle was within 10° the correct source angle. The targets were not visible. THe order of presentations was randomized.

Figure 6.13 shows the real-time simulation results – angular head position as a function of time – for listener one. When the source to be localized was placed at 100° (top left panel of figure 6.13), head movement paths in the fixed and gyro conditions were similar. Three fixed trials took noticeably longer to reach their target angle than the other trials. Placing the source at 120° (top right panel of figure 6.13), head movement paths in the fixed condition were often

slower to reach target and less direct than the gyro condition. This was probably due to 120° being at the ideal null of the simulated directional microphone. This was therefore a special case. The 140° condition (bottom left panel of figure 6.13) showed head movement paths that were quicker to reach their target angle in the gyro condition than the fixed condition. No reversals, defined here as turning more than 40° in the opposite direction from the source, were observed for this listener. The average results (bottom right panel of figure 6.13) show a benefit of the gyro condition in terms of time to successful localization, henceforth referred to as localization speed. For the sources presented at 100°, 120° and 140°, the benefits were 0.1, 1 and 0.6 seconds respectively.

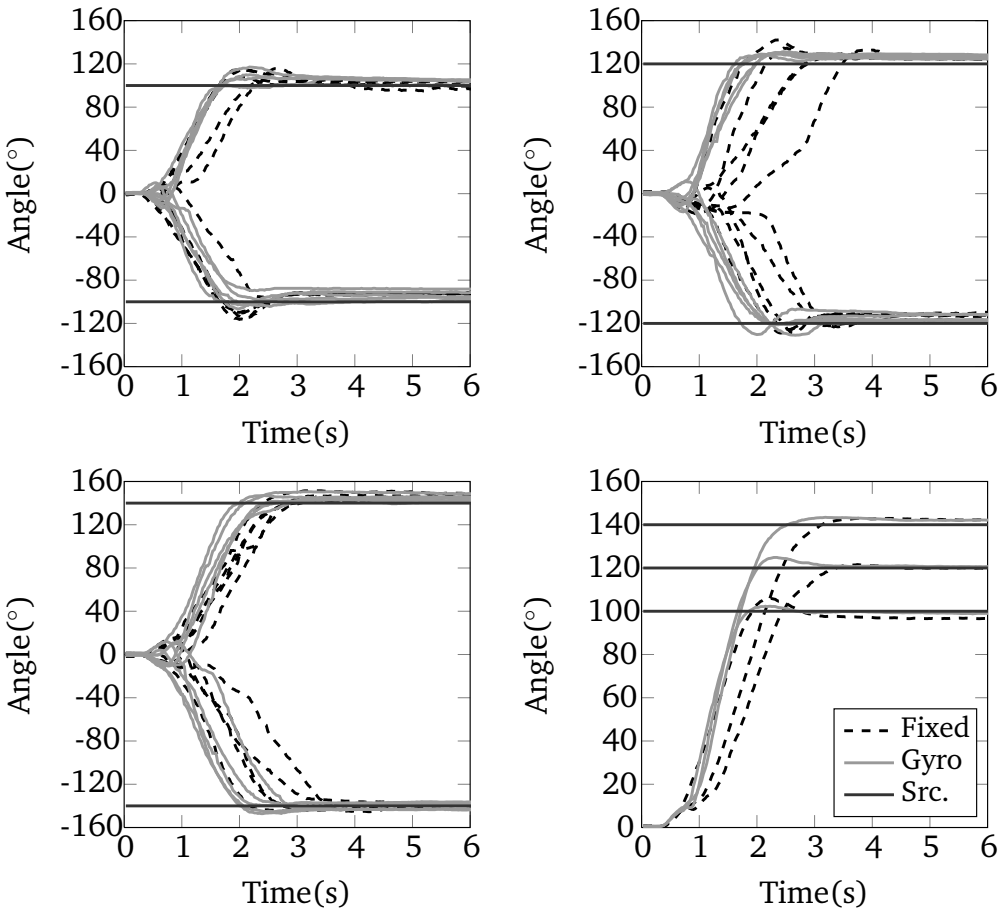

Fig. 6.13: Individual trial and average localization trajectories for listener one. The fixed DMA condition is plotted in dashed black and the gyroscopically controlled mix in solid grey. Src. shows the angular position of the sources.

Figure 6.14 shows the real-time simulation experimental results for listener two. When the source to be localized was placed at $100°$ (top left panel of figure 6.14), head movement paths in the fixed and gyro conditions were similar. Placing the source at $120°$ (top right panel of figure 6.14), head movement paths in fixed showed more reversals than those in gyro. The two slowest trials were delayed reversals in the gyro condition. The $140°$ source (bottom left panel of figure 6.14) produced a greater number of reversals in the fixed condition than the gyro condition. The slowest trial was also a delayed reversal gyro trial.

The average results (bottom right panel of figure 6.14) show a benefit of the gyro condition in terms of localization speed. For the sources presented at 100°, 120° and 140°, the benefits were 0.1 s, 1 s and 0.7 s. The dips in the trajectories observed between 3 and 4 s for sources at 120° and 140° were due to reversals.

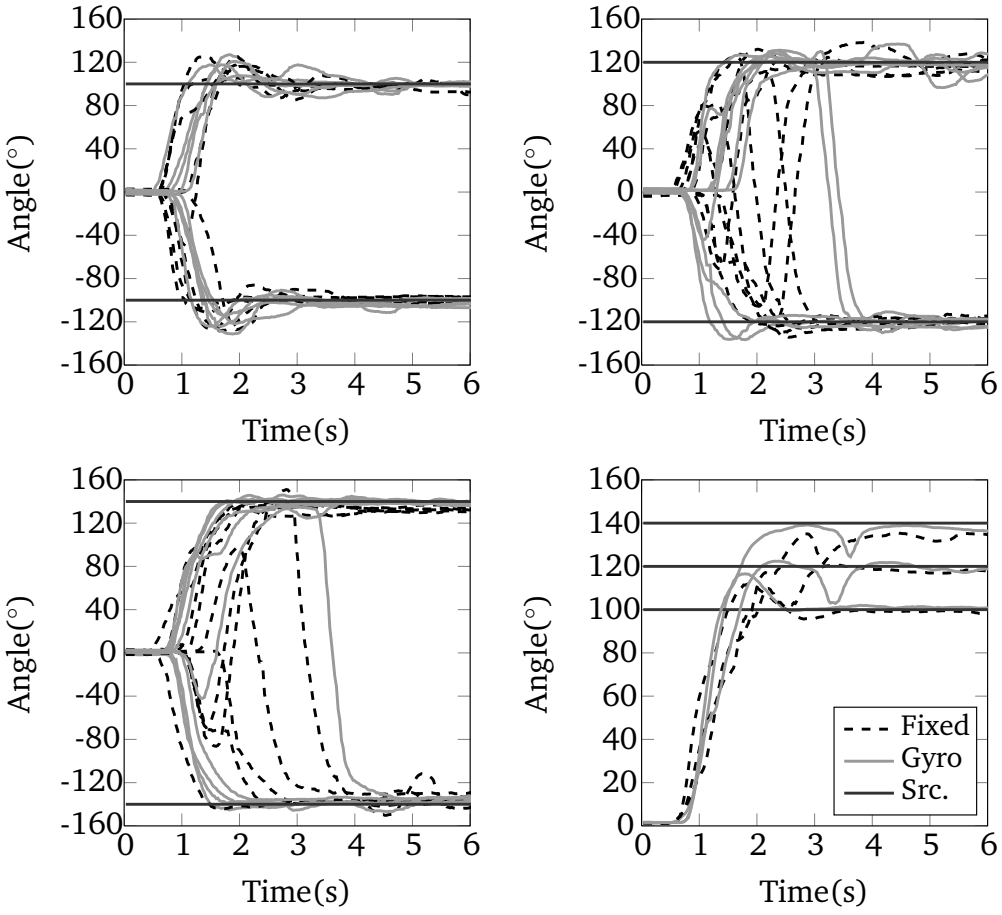

Fig. 6.14: Individual trial and average localization trajectories for listener two. The fixed DMA condition is plotted in dashed black and the gyroscopically controlled mix in solid grey. Src. shows the angular position of the sources.

Figure 6.15 shows the real-time simulation experimental results for listener three. When the source to be localized was placed at 100° (top left panel), head movement paths in the fixed and gyro conditions were similar. Placing the source at 120° (top right panel of figure 6.15), head movement paths in the fixed

condition showed more reversals than those in the gyro condition and longest trajectories were for the fixed condition. Two of the gyro trajectories started with borderline reversals, turning 40° in the wrong direction. The slowest trials were in the fixed condition. The 140° condition (bottom left panel of figure 6.15) showed similar results for most of the fixed and gyro conditions. Three fixed condition trajectories were noticeably longer than the other trials and the slowest trajectory was a reversal in the gyro condition, the only reversal observed. The average results (bottom right panel of figure 6.14) show a benefit of the gyro condition in terms of localization speed. For the sources presented at 100°, 120° and 140°, the benefits were 0.1 s, 0.5 s and 0.6 s. The dips in the average trajectories across conditions and target angles were due to reversals in specific trials.

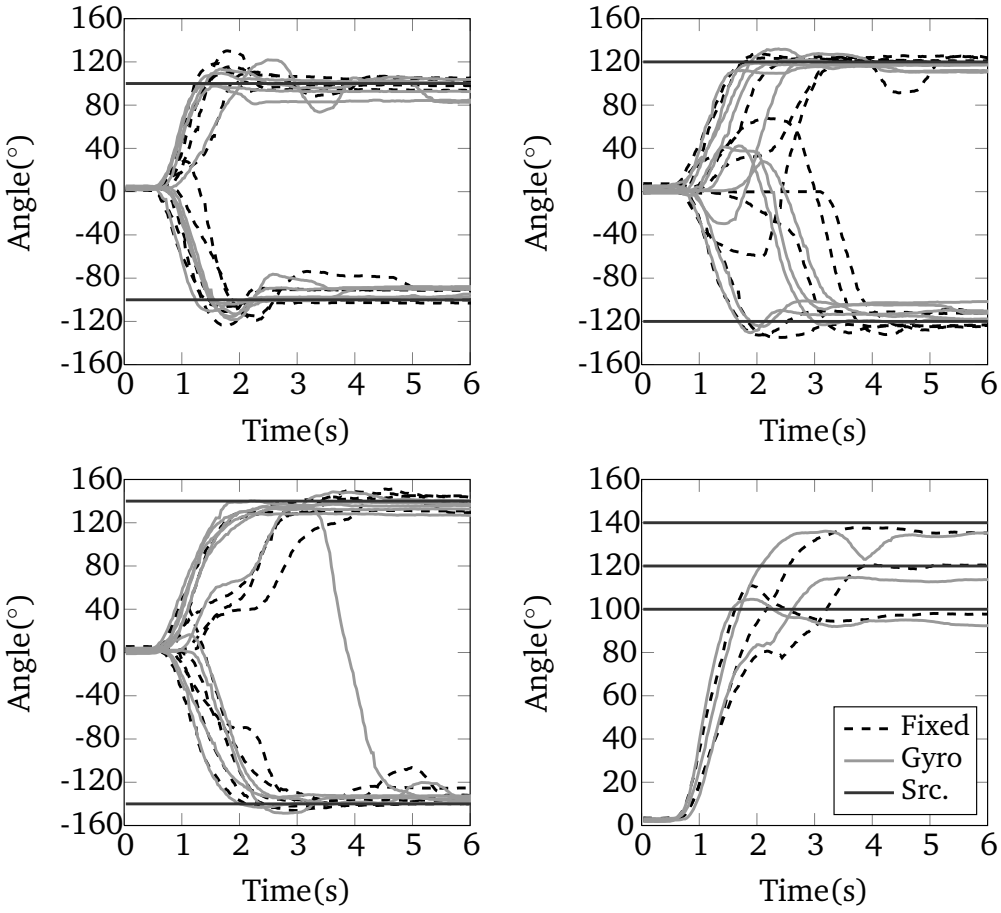

Fig. 6.15: Individual trial and average localization trajectories for listener three. The fixed DMA condition is plotted in dashed black and the gyroscopically controlled mix in solid grey. Src. shows the angular position of the sources.

Figure 6.16 shows the real-time simulation experimental results for listener four. When the source to be localized was placed at 100° (top left panel of figure 6.16), head movement trajectories in the fixed condition were faster than the gyro condition. One reversal was observed in the fixed condition and this was the slowest trajectory. Placing the source at 120° (top right panel of figure 6.16), head movement trajectories were variable and slow. The 140° source (bottom left panel of figure 6.16) showed variable results in both conditions. One reversal was observed in the fixed condition and one trial was not started

between $\pm 10°$. The average results (bottom right panel of figure 6.16) show no benefit of the gyro condition in terms of localization speed. For the sources presented at $100°$, $120°$ and $140°$, localization using the fixed condition was faster by 0.6 s, 0.6 s and 0.2 s respectively. Localization error was also reduced in the fixed condition.

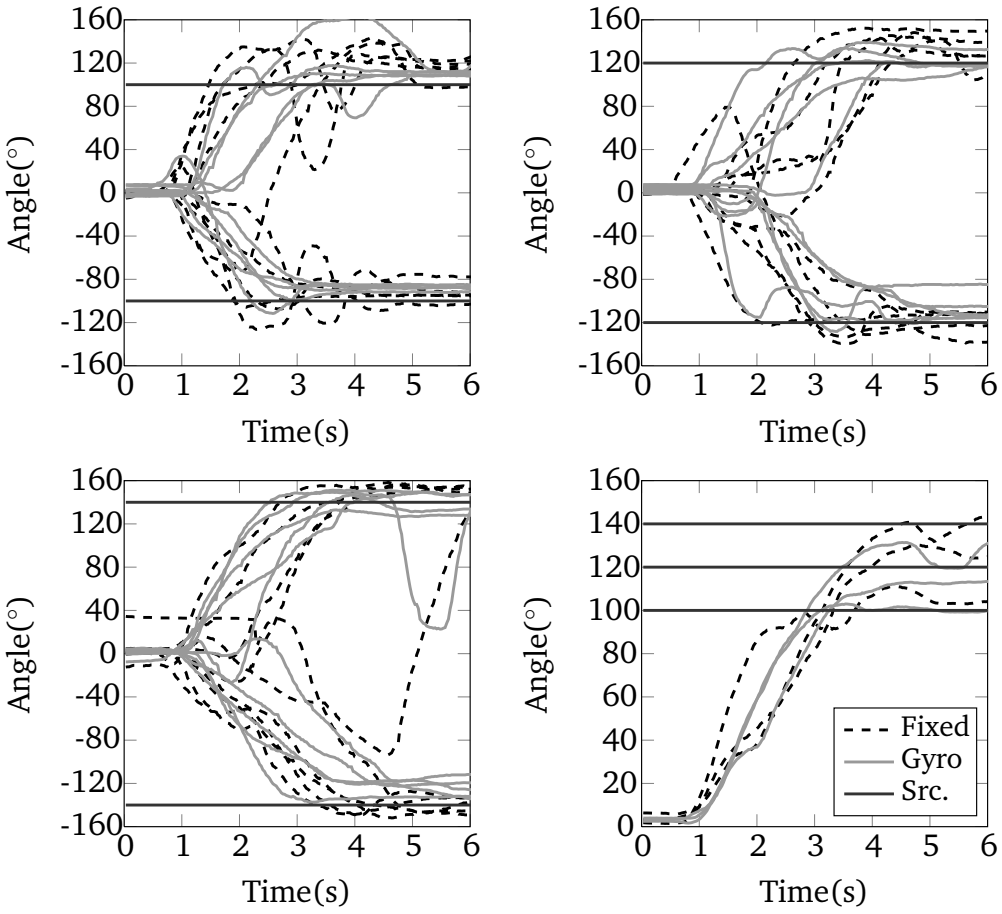

Fig. 6.16: Individual trial and average localization trajectories for listener four. The fixed DMA condition is plotted in dashed black and the gyroscopically controlled mix in solid grey. Src. shows the angular position of the sources.

6.3.5 Discussion of GCD

The initial recordings of gyroscope output with head movement shown in section 6.3.3 demonstrated that unfiltered gyroscopic output at a relatively low sample rate produces an output that is representative of head motion over a wide range of rotations. The size of the gyroscope output is proportional to the velocity of the head and it can be seen that larger outputs in these examples, appear to be related to larger head rotations. It is unclear how robust this information would be in real-world situations. The gyroscope could also detect in which plane the angular velocity of motion was largest, and showed that movements of $\geq 90^{\circ}$ produced larger y and z axis angular velocities than head movements $< 90^{\circ}$. Maximum gyroscope output in the main plane of rotation increased with the total angle of rotation. This information could be used to make more advanced predictions of listener behaviour in future research and applications.

Hearing-aid microphone array output with GCD was simulated, using pre-recorded head motion and theoretical and real hearing-aid microphone responses. Setting a mixing time (the time taken to move from full ADMA to full omnidirectional and visa versa) of ~ 0.5 s for the GCD provided additional gain for sources in the rear hemisphere earlier in a 3.1 s head movement than both ADMA and fixed DMA settings. The difference was greatest for the ADMA condition, as the null point tracks the position of the second source when it is in the rear hemisphere. These simulations show that the motion control system would provide increased audibility during head turns and, crucially, provide this audibility *earlier* than standard directional systems for sound sources at

the null-points of directional arrays. This corresponds to maximum gain being achieved 0.6 seconds earlier for a 3 second head turn (section 6.3.3). Using current first-order ADMAs, the benefit is restricted to localizing sources in the rear hemisphere due to the constraints on the β value. However, second-order and bilateral ADMAs could benefit more from the GCD, as they produce a narrower main lobe.

There are a number of unresolved issues with this system that would require further tests. The speed of mixing and the gyroscope output threshold setting could be dependent on the listener's behaviour and preferences. One could assume that a faster mixing time would provide the maximum benefit in terms of increased gain. However, the timbral change of switching from a directional array output to an omnidirectional output may cause annoyance if performed too quickly.

The localization experiment on normal-hearing listeners using a real-time simulation of the GCD (section 6.3.4.2) showed some benefit of the new system for most listeners. The average time to successful localization was faster for most listeners using the GCD over the fixed DMA. The difference in localization speed was largest for sources at $120°$, the initial position of the null. The nature of the benefit to listeners manifested itself using the GCD varied. Listener one showed similarly shaped trajectories whether the condition was the GCD or the fixed DMA, but the trajectory for the gyroscope condition was shorter (i.e. faster). Listeners two and three showed smoother trajectories in the gyroscopic condition and more reversals in the fixed DMA condition. These reversals contributed to the average difference in localization speed between the two conditions across source positions and it is known that reversals are

increased for hearing-impaired listeners using directional settings (Brimijoin et al., submitted). The GCD may improve orientations for hearing-impaired listeners too.

Listener four showed a detrimental effect on speed of localization when using the GCD across all source positions. More participants would be needed to determine the possible cause of this. However, listener four took the longest time to successfully localize the sources out of all the participants. It is possible that the GCD is only of benefit for quickly localizing off-axis sources.

The real-time simulation was different from the off-line simulation in a number of ways. The theoretical response of the fixed DMA was used, producing a greater attenuation at the null than the system modeled from head-mounted impulse responses. In addition, the mixing time was shorter for the real-time simulation. Therefore, the results of the computationally limited real-time simulation show the maximum advantage that the gyroscopically controlled system could produce over the standard fixed, first-order DMA for normal-hearing listeners.

The question of what gyroscope threshold to use is a more complex issue, requiring a trade-off between unnecessary mixing and the assumed benefit of increasing gain for attenuated sources early during head movements. Particularly emphatic communicators (large head movements) may require higher thresholds than more passive listeners, thus requiring a shorter mixing time. Given the success of using head movements for environmental classification (e.g. Tessendorf et al., 2013), however, some global thresholds for gyroscope output may be possible.

6.4 Gyroscopically compensated direction of arrival (GC-DOA) estimation

The acoustic scene experienced by a hearing aid can be dynamic if: the sources move relative to the hearing aid, the hearing aid moves relative to the sources (listener movement) or a combination of both source and listener movement. To computationally analyze these dynamic scenes, this necessitates the use of techniques that use limited information over short-time periods, such as the LMS adaptation described in section 3.6. It is a reasonable assumption that much of the auditory motion listeners experience is self-generated (König and Sussman, 1955). Therefore, knowledge of the listener's motion would allow more information to be reliably gathered by an algorithm and this increased information could improve performance. In the system described below, source direction estimates are aggregated over time and compensated for head movement, resulting in an improved source localization during head movements. This could be used for improved null-steering in adaptive beamformer systems.

The system presented in this section assumes the use of bilaterally connected hearing-aid microphones transferring audio. Previous research has also assumed the ability to transfer audio between hearing aids (e.g Chatlani et al., 2010; Chatlani, 2011) and bilateral audio transfer between hearing aids has recently become commercially available.

In this section, gyroscopically compensated GCC-PHAT will be described (section 6.4.1), followed by comparisons of standard GCC-PHAT and

gyroscopically compensated GCC-PHAT performance during head turns for single source, multi-source and noisy environments (section 6.4.2). An application of gyroscopically compensated GCC-PHAT to ADMA null-steering will also be presented (section 6.4.3).

6.4.1 Gyroscopically compensated GCC-PHAT

Figure 6.17 shows a diagram of the head-mounted, gyroscopically compensated GCC-PHAT for GC-DOA estimation. The system uses two microphones mounted on either side of the head and a single head-mounted gyroscope as inputs. The system records at 44.1 kHz, which is then downsampled to 16 kHz to approximate the sample rate of a modern hearing aid. Short-time Fourier transforms (STFT) are taken of the audio every 40 ms. Each x_L (left side) and x_R (right side) segment has a Hanning window applied to it. No overlap between windowed segments is used, as each sample from the gyroscope is synchronized with a single segment of audio and during a head turn, DOA estimations from previous audio segments would become increasingly less accurate and asynchronous with the gyroscope input. The DOA sample rate is 25 Hz. The GCC-PHAT algorithm (section 3.10.1) is then applied to each pair of audio segments. Assuming the size of the head to be roughly 16 cm diameter and the sampling frequency used to be 16 kHz, the maximum achievable delay (output of the GCC-PHAT) at 90° (the maximum distance between the microphones) is 10 samples, resulting in a range of -10 to +10 samples and a resolution of 9°. The output of the GCC-PHAT is interpolated by a factor of 8:1 using the *resample* command in MATLAB. The *resample* command interpolates using a polyphasic implementation and also applies a

linear phase, anti-aliasing (lowpass) finite impulse response filter to the original input, compensating for the delay introduced by the filter. The interpolation is performed using a weighted sum of 160 samples of the original input signal to calculate each value of the output signal. Experimental results showed that angles between $80 - 90^{\circ}$ resulted in no change in time-delay estimation (TDE) after interpolation. The interpolation resulted in a resolution of (1°) from $0^{\circ} - 80^{\circ}$ (1 sample = 1°). It is acknowledged that implementing this system in a hearing aid may necessitate a lower interpolation ratio, due to computation time. After interpolation, maximum time delay (τ_s) is selected as the peak in the interpolated IFFT, and converted to angle in degrees to produce a DOA estimate. This estimate is then placed in the corresponding histogram bin and the process repeats for the next analysis window.

During each analysis window, the rotational velocity information measured by the gyroscope is used to determine the angle through which the the head has rotated since the previous analysis window. The histogram of DOA estimates is rotated *against* the rotated angle, while the current DOA estimate is added to the histogram unchanged. At the end of a measurement frame (100 analysis windows in the current study), the peaks in the histogram correspond to the strongest sources active during the measurement frame. 100 analysis windows per measurement frame was experimentally determined as sufficient to produce robust estimates for up to four active speech sources, the maximum number used in this study. The rotated angle is reset to 0° to prevent long-term inaccuracies in the gyroscope affecting the accuracy of the gyroscopic output and the histogram is also reset.

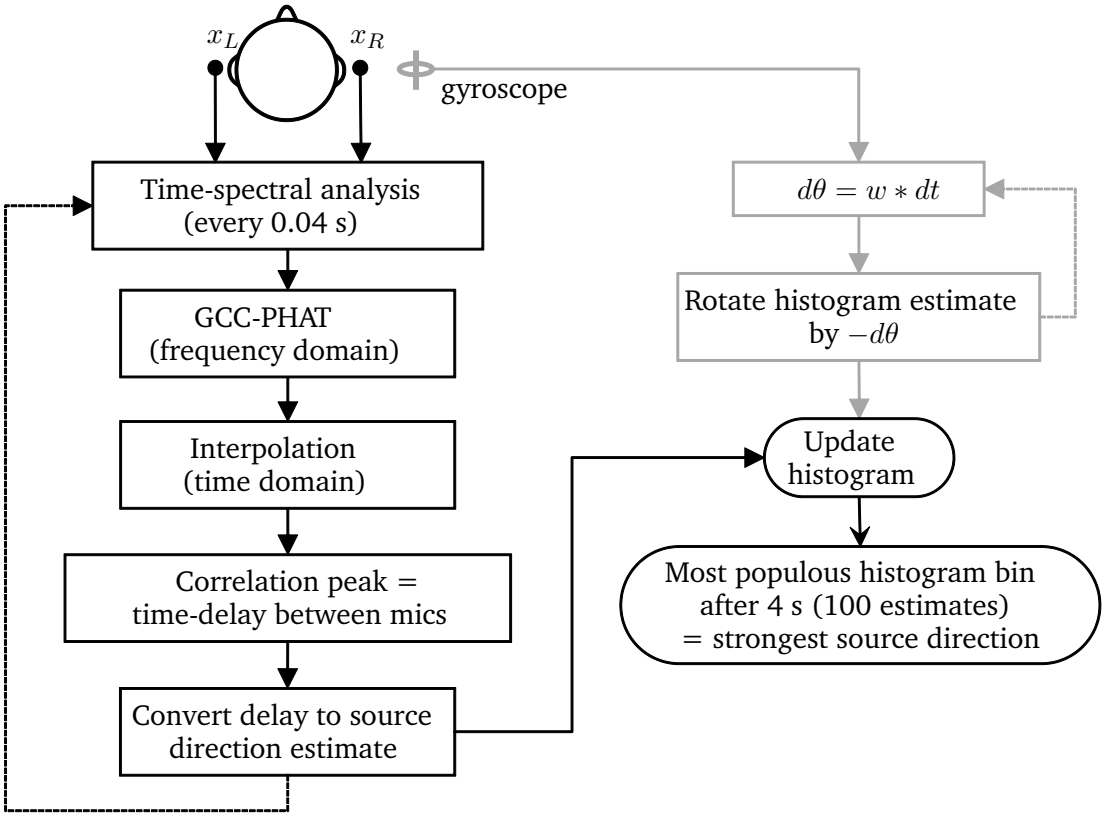

Fig. 6.17: A diagram showing the steps involved in the gyroscopically compensated GCC-PHAT.

6.4.2 Comparison of DOA estimation using GCC-PHAT and GC-DOA estimation

The performance of GCC-PHAT and GC-DOA estimation during movement of the head was compared for single sources in an acoustically deadened room, multiple sources in a reverberant space and single sources in diffuse noise. Table 6.1 gives a summary of the results, with details given below.

6.4.2.1 Experimental systems and stimuli for GC-DOA estimation

The GCS comprised a pair of in-ear microphones (Sound Professional MS-TFB-2), a 2-channel external soundcard (Zoom H4N) and a laptop

Test	Result
Single-source localization during a 30° head turn	Narrow, single peak of aggregated DOA estimates histogram produced using gyroscopic compensation. Without compensation, DOA estimates are spread across 30°
Multiple-source localization in a low-reverberation time room during a 30° head turn	Four distinct peaks correspond to four active sources in the aggregated DOA estimates histogram produced using gyroscopic compensation. Accurate to 5°. Without compensation, three peaks are detectable, one of which is correct.
Single-source localization in diffuse noise during a 30° head turn	Aggregating DOA estimates over a long timescale and compensating for head movement improves the chances of determining an accurate peak at various SNRs using gyroscopic compensation. At 0 dB SNR, robust knowledge of head movements and prior DOA estimate histograms can improve performance
Adaptive null-steering using GC-DOA estimation	Head-movement information allows nulls to be steered instantly against head movement, using previous DOA estimate histograms as a guide to noise source position

Table 6.1: Summary of section 6.4.2 experimental measurement findings

running Matlab 2012a for recording audio. To record head rotation, a triple-axis, digital-output gyroscope (InvenSense ITG-3200) was connected to a microcontroller (Arduino Uno R3) and Matlab via USB-serial.

For calibration of the gyroscope, the gyroscope was attached to the head of a Kemar manikin, mounted on a high-precision turntable (LinearX LT360EX). Knowledge of the rotational speed of the head-mounted gyroscope enabled the output of the gyroscope to be accurately converted to deg/s from the raw output of the gyroscope, using rotational equations of motion, for a 30° head turn.

This conversion was tested by tracking the output of the calibrated gyroscope for other ranges of head turn. All subsequent tests were performed using the author's head movements.

The single source and diffuse noise tests were performed using a 0.9 m radius 24-loudspeaker ring (see figure 6.3 for a schematic) in a sound-dampened room, 3.4 $\times$ 2.5 $\times$ 2.4 m. During the diffuse noise tests, 23 loudspeakers were used to present uncorrelated white-noise signals and one to present the signal. The tests in reverberant space were performed in a room with dimensions of 6.5 $\times$ 5 $\times$ 3 m, reverberation time (RT_{30}) of 0.35s, using JBL Control 1 loudspeakers, placed at a distance of 2 m from the listener's head. Randomly selected sentences from the IEEE York corpus were concatenated for the test speech signals, of duration 12 s. Single-source tests were conducted using a male talker, four-source tests using two males and two females. Tests in quiet were performed at 70 dBA. The diffuse noise was output at a combined level of 60 dBA.

6.4.2.2 Single source in quiet results

Figure 6.18 shows the output of the rotational plane of the gyroscope during a head movement of $30°$. While the head is static, small perturbations result in positional "drift". This issue is currently resolved by resetting the gyroscope output to $0°$ at the beginning of each measurement frame, as the gyroscopically compensated GCC-PHAT requires only movement *relative* to the position at the end of each frame.

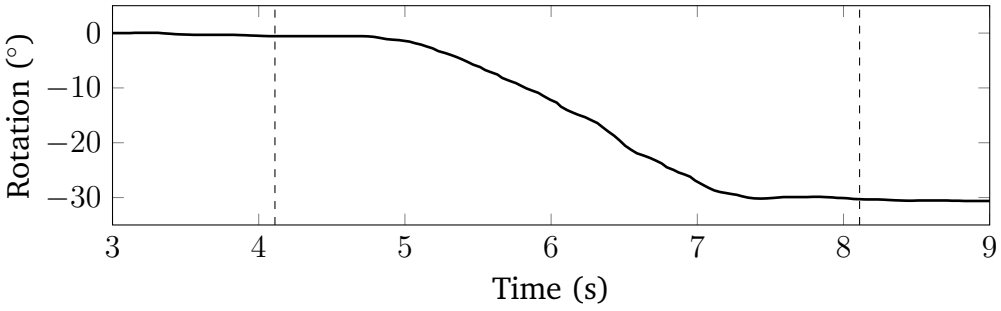

Fig. 6.18: A typical $30°$ head movement estimated by the gyroscope data during one measurement frame. The dashed lines mark the beginning and end of the measurement frame.

Figure 6.19 shows the DOA histogram output for one source in quiet, during a $-30°$ head movement, beginning with the source at $30°$ and ending at $60°$. It can be seen in the top panel of figure 6.19 that the GCC-PHAT produces a widened output, resulting in an inaccurate DOA estimation at the end of the measurement frame. However, using the GC-DOA estimation, the head-movement during the measurement frame can be compensated, resulting in an aggregation of correct DOA estimates during the head-movement and a clear peak in the DOA histogram at $58°$ in the bottom panel of figure 6.19, similar to the result that would be obtained with a static head and source at $60°$. That is, the top panel of figure 6.19 shows a source or sources between $30°$ and $60°$, whereas the bottom panel – with GC – shows a single source over the excursion. The histogram is compensated for head movement over the measurement period, in order to produce a robust DOA estimate for the current position of the source relative to the head at the end of the measurement period. This knowledge can be used for further processing of the signal.

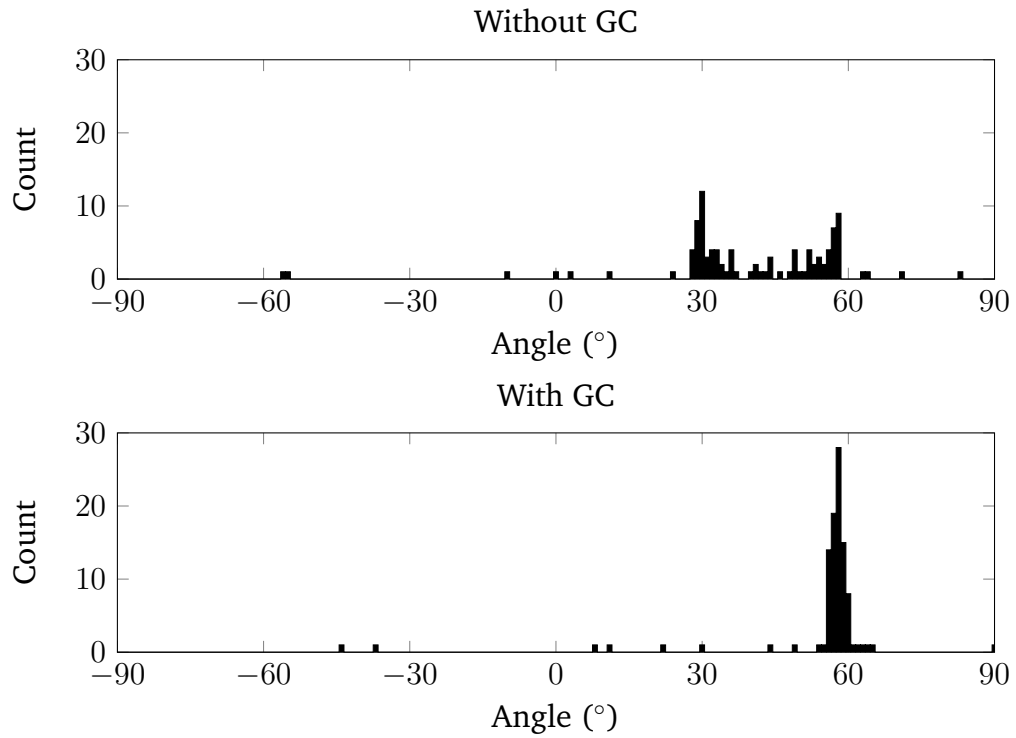

Fig. 6.19: Single source DOA estimate histograms in an acoustically deadened room.

6.4.2.3 Multiple sources in reverberant room results

Figure 6.20 shows the DOA histogram output for four sources in a reverberant room during a $30°$ head movement. The sources begin at $-30°$, $0°$, $30°$ and $60°$, and finishing at $-60°$, $-30°$, $0°$ and $30°$. It can be seen in the top panel of figure 6.20 that the GCC-PHAT once again produces wide peaks, resembling a single source at $0°$. This results in an inaccurate DOA estimation for the sources at the end of the measurement frame, except for the $0°$ source, but this is only coincidental. Using the GC-DOA estimation, the head-movement during the measurement frame was compensated, resulting in an aggregation of correct DOA estimates during the head-movement (bottom panel of figure 6.20). All identified peaks are within $5°$ of the true value. In addition, the temporally

sparse nature of the source signals (speech) and the short estimate windows used (40 ms) resulted in four peaks being identifiable, as the GCC-PHAT produces a single estimate for the strongest source in each analysis window. Again, the histogram is compensated for head movement over the measurement period, in order to produce a robust DOA estimate for the current position of the head at the end of the measurement period.

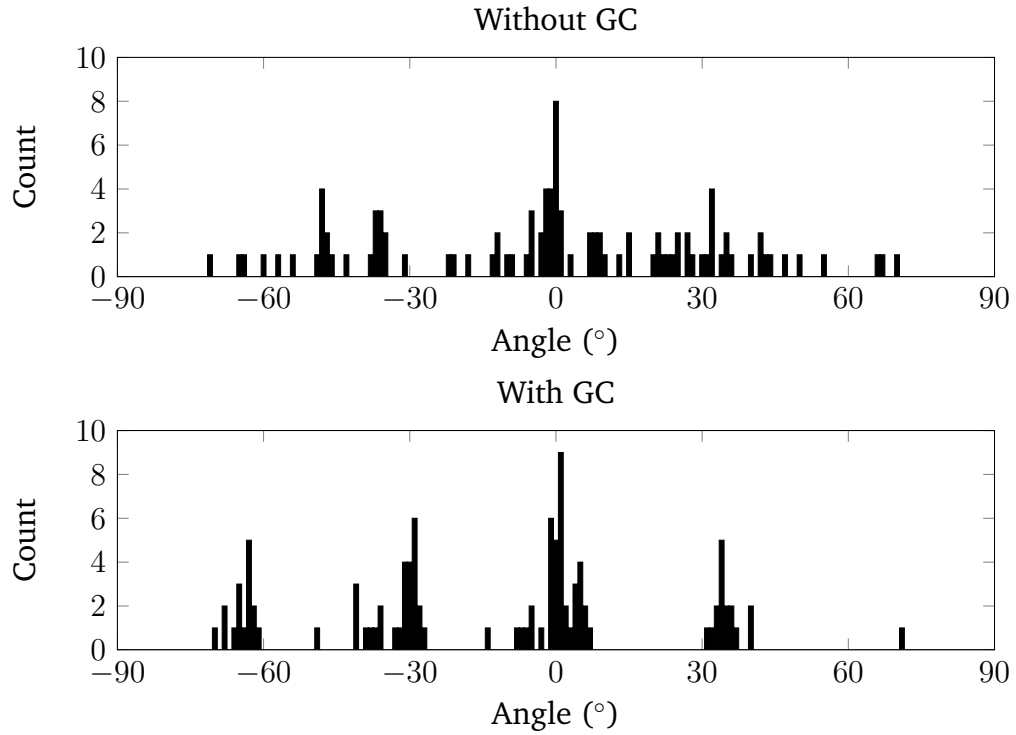

Fig. 6.20: Four-source DOA histograms in a reverberant room.

6.4.2.4 Single source in diffuse noise results

Figure 6.21 shows the DOA histogram output for one source at 12 dB SNR, during a $-30°$ head movement, beginning with the source at $30°$ and ending at $60°$. The performance differences were similar to that observed with a single source in quiet (figure 6.19). It can be seen in figure 6.21 that the GC system

produces a higher aggregated DOA histogram peak at the source position ($60°$) than the equivalent GCC-PHAT alone. The standard GCC-PHAT histogram also produces a peak at the correct angle, due to the head being at rest at the end of measurement period. Without compensation, previous correct estimates during the head turn are not shifted to the correct position, producing a lower peak.

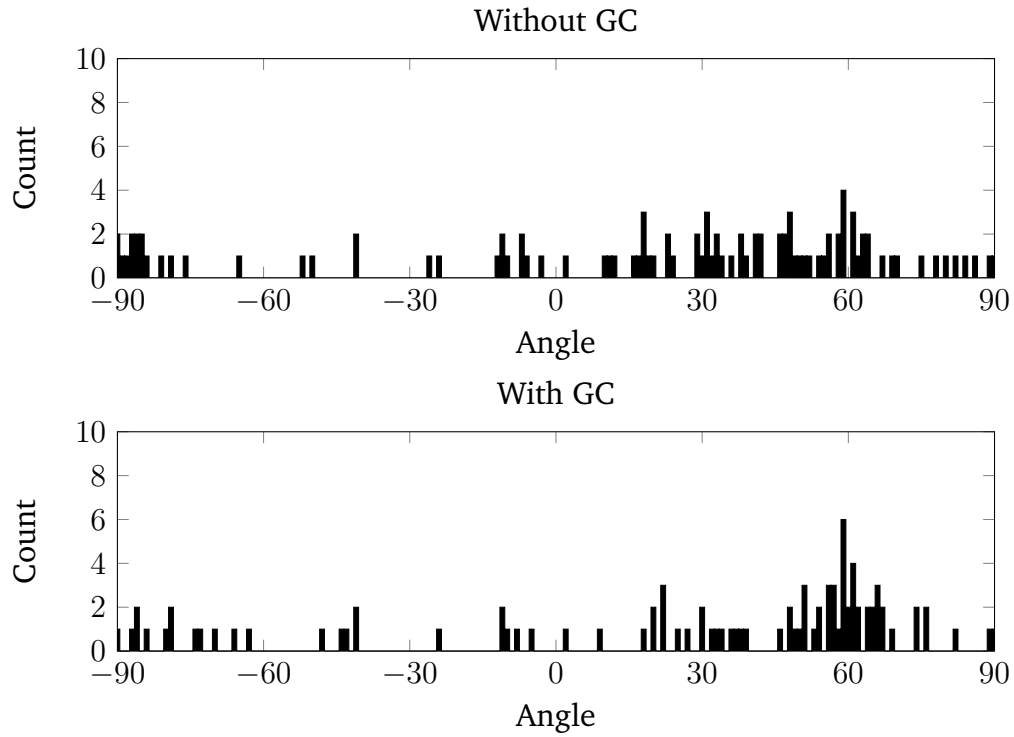

Fig. 6.21: Single source DOA estimate histograms in 12 dB SNR

Figure 6.22 shows the DOA histogram output for one source at 0 dB SNR, during a $-30°$ movement, beginning with the source at $30°$ and ending at $60°$. It can be seen that the GC system fails at this SNR, giving 3 possible DOAs (2 of which are within $5°$ of the true DOA) where only one exists. However, the equivalent GCC-PHAT histogram gives one DOA close to the starting-point of

the head turn ($-30°$). Therefore, at failure, the GC-DOA estimation provides more information on the DOA than the equivalent GCC-PHAT.

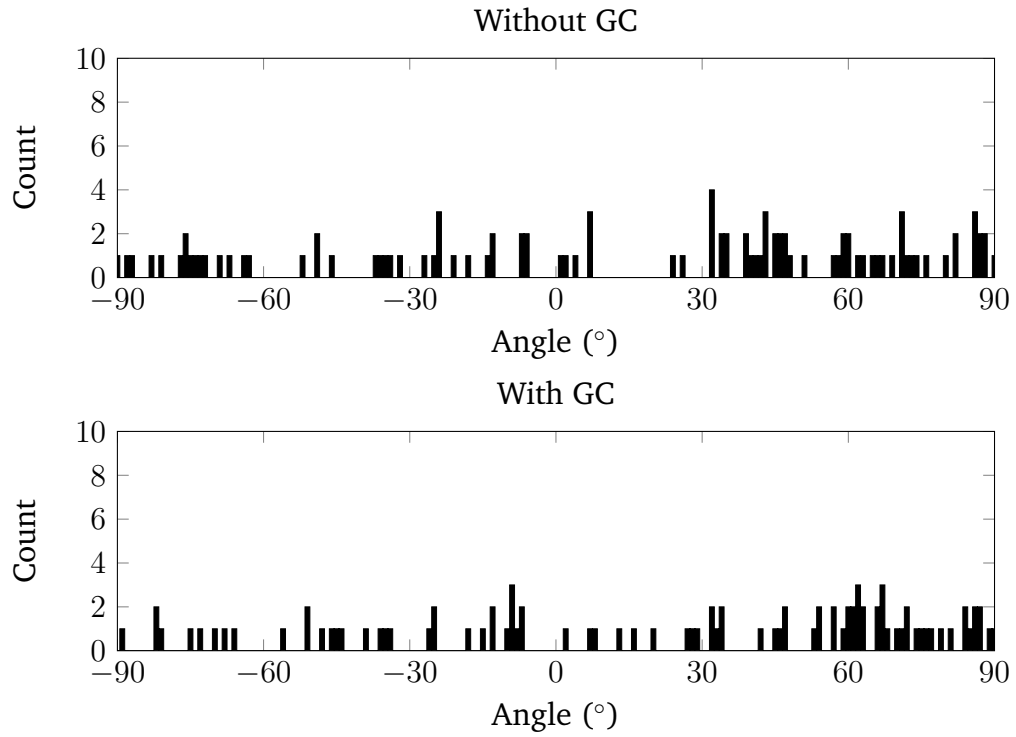

Fig. 6.22: Single source DOA estimate histograms in 0 dB SNR

6.4.3 Adaptive null-steering using GC-DOA estimation

Knowledge of head movements in relation to sound sources enables the null of an ADMA (see section 3.6) to be steered to the direction of a noise source and accurately attenuate that noise source, while compensating for head movements. Figure 6.23 shows the simulated output of the system. Using the GC-DOA estimation with the ADMA, the direction of the strongest noise source is identified by the GC-DOA estimation. The null of the ADMA can be instantaneously steered to the noise source. During the next GC-DOA estimation measurement frame, the null remains fixed on the previously

identified noise source using gyroscopic information on head movement. The lag in this system is due to the time taken to estimate DOA robustly. However, if the head is moving to a greater degree than the sources, the GC-DOA will successfully track the relative position of the noise source during DOA estimate frames. This technique could be extended to bilateral beamformers, to steer the largest "lobe" (highest output) towards a target signal, assuming DOA peaks can be accurately labelled as targets and noise sources and known listener intent.

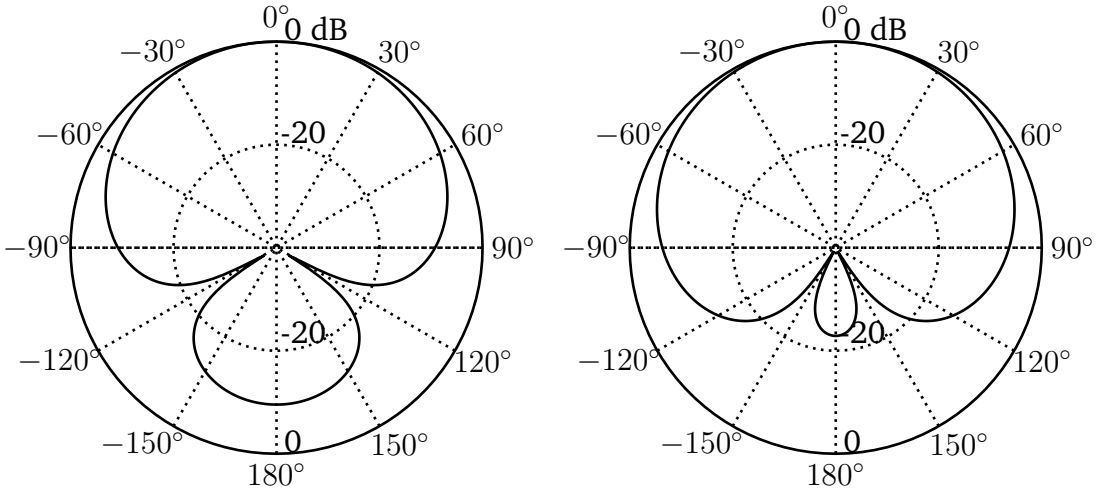

Fig. 6.23: Output of a theoretical ADMA using GC-DOA estimation at the beginning and end of a 30° head rotation

6.5 New biomimetic direction of arrival (B-DOA) estimation

6.5.1 Concept

The system developed and tested in section 6.4 improved source localization during head motion, but it has the same limitations as other two microphone

cross-correlation based systems: it must assume sources are either in the front or rear hemifield. This is due to the fundamental limitations of using time differences alone to determine location.

Sound sources at points anywhere on a circle centred on the interaural axis produce the same time difference between the microphones positioned on each ear. This is analogous to the "cone" or "tori of confusion" experienced by human listeners (see section 2.2.4). Human listeners use the differential filtering of the pinna or relative movement of the source due to head motion to resolve the position of sources on this "cone of confusion". In the horizontal plane along the azimuth only, the cone of confusion becomes two points, one in the front hemifield and one in the rear hemifield.

See section 3.10 for a review of how other localization systems expand the resolvable measurement space from $\pm 90^{\circ}$ to $\pm 180^{\circ}$, resolving this cone of confusion. Specifically, hearing aids can obtain a large array spacing in one direction by combining microphones at each ear in a bilateral array. A large spacing in another orthogonal or sufficiently different direction from the interaural axis is not possible, as individual hearing aids can only accommodate small array spacings.

The system presented utilizes head movement to resolve the cone of confusion, mimicking human listeners. Head movements and basic comparisons of aggregated localization estimate histograms are shown to robustly determine the position of a source in the front or rear of the array.

6.5.2 Implementation of B-DOA estimation

The hardware implementation of this system is similar to that used in section 6.4. The gyroscope is replaced by the 9-axis sensor stick and head movement is calculated using the DCM algorithm, described in section 3.12.4. The main advantages of this are reduced measurement drift over time and the ability to track head movement accurately in three dimensions.

The algorithm is similar to that shown in section 6.4, figure 6.17 with several differences, as shown in figure 6.24. For the B-DOA system, DOA estimation robustness is reduced to obtain a faster estimate of position. This is done by shortening the time frame over which the DOA estimates are aggregated from four seconds to one second. In addition, two DOA histograms are created for each measurement frame. One histogram is rotated clockwise by the amount of head-movement detected during each estimate, while the other histogram is rotated anticlockwise by the same amount. If head movement occurs during a measurement frame where a source is active in the front hemifield, the histogram that has been rotated anticlockwise (as in section 6.4) will produce a larger aggregated peak at one angle than the clockwise rotated histogram. If the source lies in the rear hemifield, the opposite will occur, with the clockwise-rotated histogram producing the larger peak. By choosing the histogram with the largest peak, the hemifield in which a source lies can be determined: anticlockwise indicates front and clockwise indicates rear. The hemifield of a source is determined in addition to its angular position between $\pm 90^{\circ}$ without the addition of third microphone or the use of an HRTF library (see section 3.10).

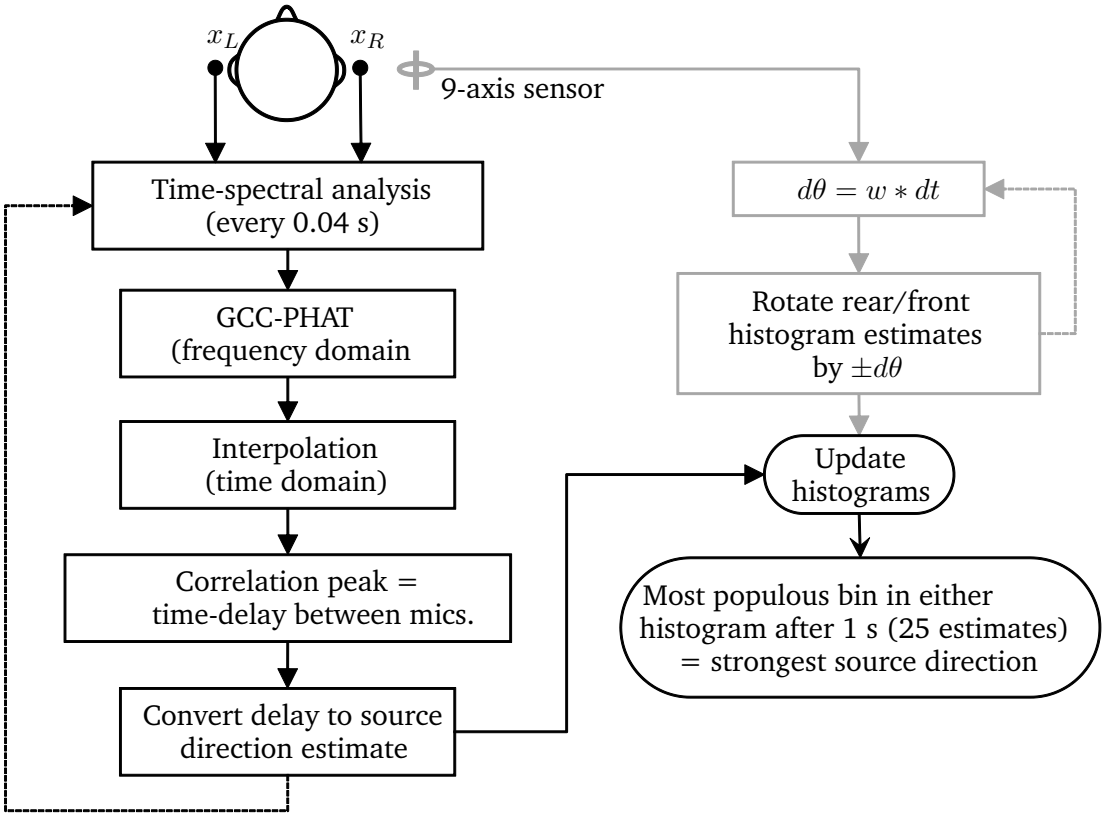

Fig. 6.24: A diagram showing the steps involved in the B-DOA system.

6.5.3 Experimental systems and stimuli for B-DOA estimation

Figure 6.25 shows the experimental apparatus for testing B-DOA estimation. A single loudspeaker (JBL Control 1 Pro) was placed 1.5 m directly in front or behind the listener at a height of 1.2 m. The listener's head moved through approximately $60°$, from $-30°$ to $+30°$, relative to the loudspeaker at $0°$ (in front) or $180°$ (behind). The speech samples used were from a male talker in the IEEE York corpus (Stacey and Summerfield, 2007), presented at 65 dB in a 6.5 $\times$ 5 $\times$ 3 m and and RT_{30} of 0.35 s.

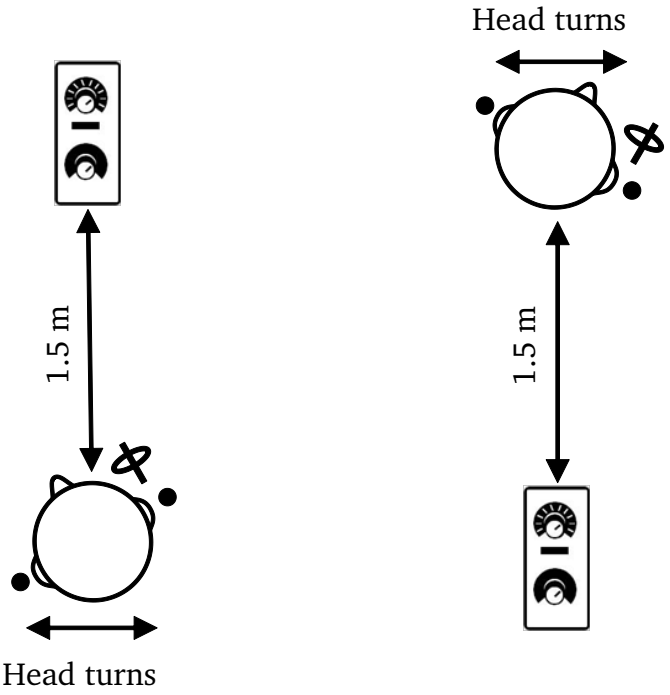

Fig. 6.25: Experimental apparatus for testing B-DOA estimation.

The in-ear microphones used were Sound Professionals MS-TFB-2, placed on top of the pinna to simulate the position of a behind-the-ear hearing-aid microphone. A Zoom H4n was used as the recording soundcard. The 9-axis sensor was calibrated for the local hard iron magnetic sources using software written by Bartz (2012) and head movement data was collected from it using an Arduino Uno and a USB-serial connection.

The 9-axis sensor required several seconds at start-up to obtain enough information to calculate its position, therefore the first five seconds of each recording were discarded (see section 6.2.2. Recordings were made for 15 seconds, resulting in 10 seconds of data and 10 measurement frames from each recording.

6.5.4 Extended empirical head model

Calibration of the previous, gyroscope-only system (see section 6.4) for the front hemifield produced an approximate 1:1 relationship between the interpolated maximum time delay using GCC-PHAT and the angle of a sound source relative to a KEMAR head up to $80°$. This was successful for the front hemifield when using the current 9-axis system. Tests using the 1:1 relationship for sources in the rear hemifield produced wider histogram distributions, suggesting that the 1:1 relationship did not hold in the rear hemifield.

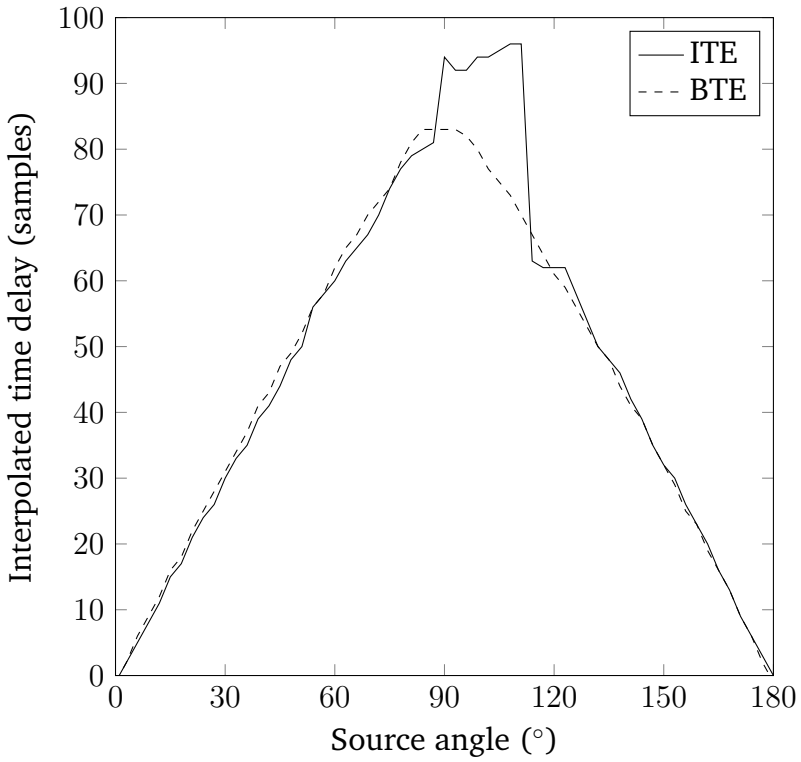

Fig. 6.26: Source angle vs. time delay for ITE and BTE microphone positions

Testing was conducted using the same apparatus used to calibrate the gyroscope in section 6.4.2.1, but with microphones placed in the ear (ITE) and at the top

front of the pinna (BTE). A single loudspeaker presenting a single male talker (Stacey and Summerfield, 2007) was used. Peak time-delays were measured in 3° increments from 0° to 357° by rotating the turntable. The peak in the DOA estimate histogram was used as the time-delay estimate for each angle measured.

Figure 6.26 shows the interpolated time delays plotted against sound source angle for front ($0^\circ - 90^\circ$) and rear hemifields ($90^\circ - 180^\circ$). Fitting a line of best fit using angles from $0^\circ - 80^\circ$ revealed that the relationship between interpolated time delay and source angle is ~ 1 for sources in the front hemisphere and > 1 for sources in the rear hemisphere for the BTE microphone position. The ears do not lie exactly on the maximum diameter of the head, but towards the rear of the head. This difference could produce the wider histogram distribution observed (see figure 6.29). Correcting for this, the rear hemifield histograms became narrower. Figure 6.26 also shows a non-linear shift in the estimated time delay between 90° and 110° for the ITE microphones. This could be due to the filtering effects of the pinna and measurement errors during the GCC-PHAT estimation.

The correction factors of 1 (front hemifield) and 1.1 (rear hemifield) were determined using KEMAR. The author's head was different in size from KEMAR. Applying the KEMAR-derived correction factor to the uncorrected DOA estimate histograms produced using the author's head (figures 6.27 and 6.29 in section 6.5.5) narrowed the distributions. By empirically varying the size of the correction applied, the distributions could be narrowed further to produce a larger peak in the DOA estimate histogram. The final ratio used was 1.09 (front hemifield) and 1.235 (rear hemifield) and the results are shown in figures 6.28

and 6.30. The correction was applied by multiplying each time-delay estimate by the correction factor (in this case 1.09 for the front hemifield and 1.235 for the rear), before placing the corrected estimate in the front and rear DOA estimate histogram. At the end of a measurement frame, the corrected time delay estimates are converted back to an angle estimate by dividing by the correction factor.

6.5.5 Experimental results for B-DOA estimation

Figure 6.27 shows two seconds of head-yaw recording and the corresponding analysis frames for a source in the front hemifield (0°) without an empirical head model. In frame one it can be seen that rotating the histogram anticlockwise by the head yaw produces a large peak in the histogram. While rotating the histogram clockwise also produces a peak, it is half the magnitude of the anticlockwise peak.

In frame two, the anticlockwise rotation produces a peak, while the clockwise rotation produces a spread of position estimates whose range is approximately twice the measured head yaw for that frame. In both frames, for clockwise and anticlockwise head rotation, the anticlockwise histogram had the larger peak and was selected as the correct histogram. This corresponded to the correct identification of a source in the front hemifield.

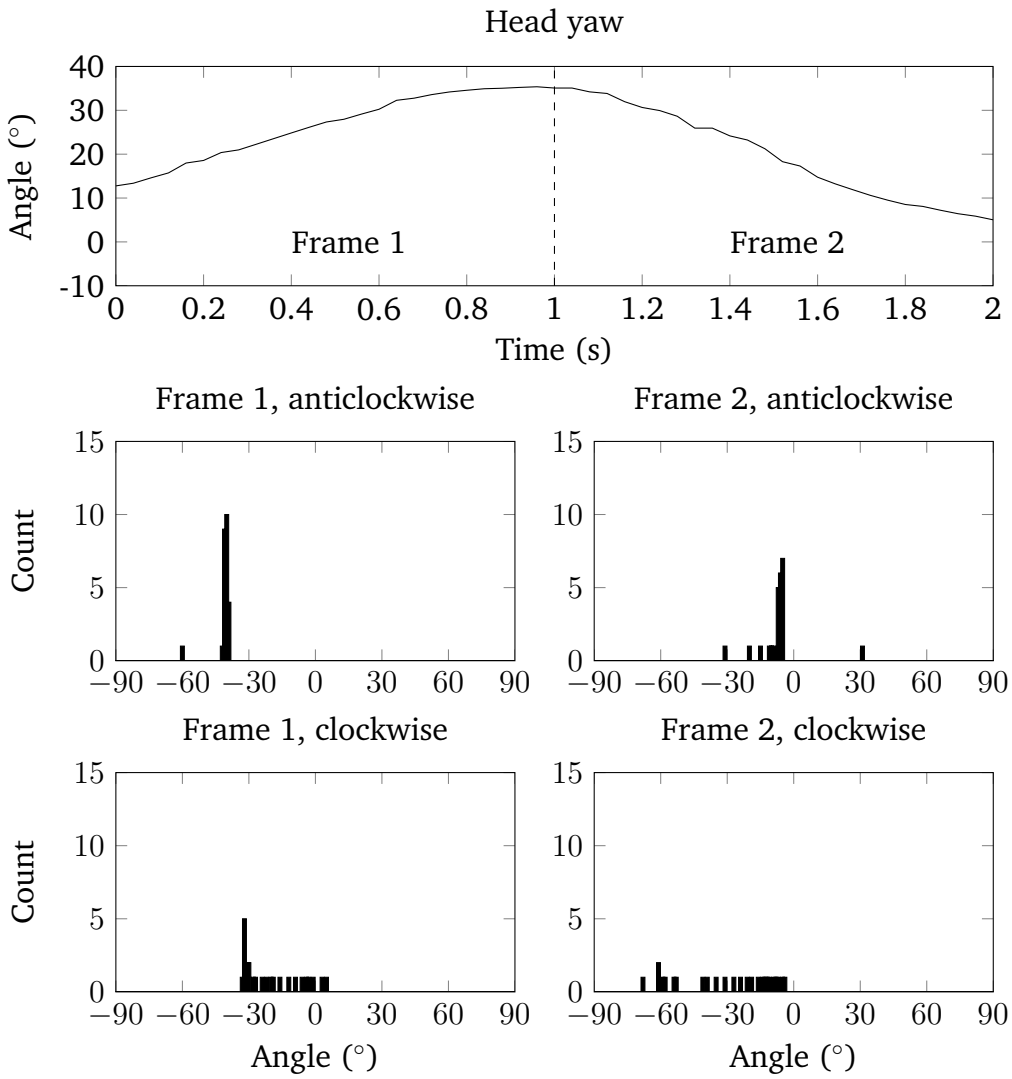

Fig. 6.27: Biomimetic source localization for source in front hemifield without empirical head model correction. Source at $0°$.

Figure 6.28 shows the same frames as figure 6.27 with the empirical head model correction applied. The histogram peak is higher and narrower with the correction factor. A comparison of localization estimates between frame one and frame two shows a movement of $33°$ ($-38°$ to $-5°$) while the reported head movement is $-31°$ ($36°$ to $5°$), giving an error of $2°$.

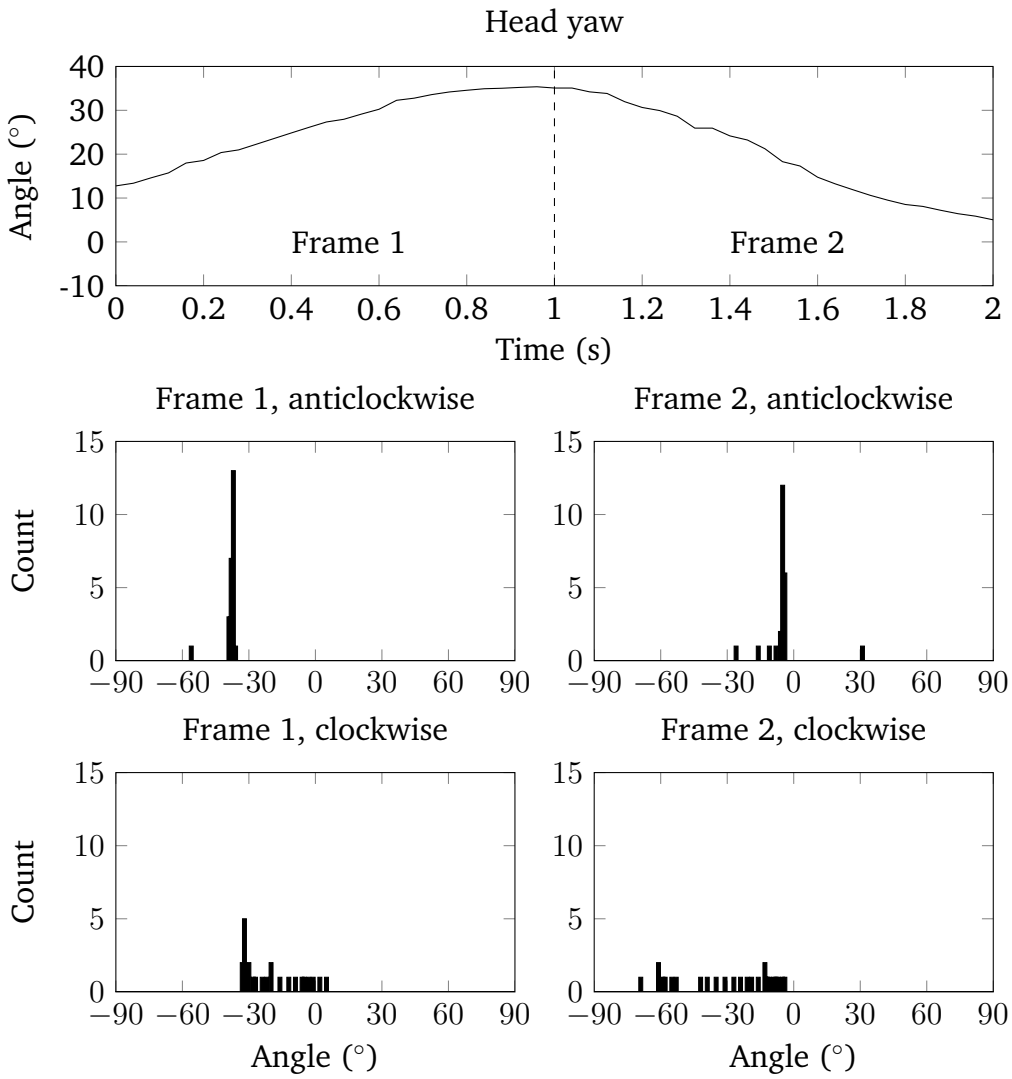

Fig. 6.28: B-DOA estimation for source in front hemifield with empirical head model correction. Source at 0°.

Figure 6.29 shows 2 seconds of head-yaw recording and the corresponding analysis frames for a source in the rear hemifield (180°), without an empirical head model and similar head motion to that shown in figure 6.27. A clockwise rotation of the histogram (bottom row of figure 6.29) produces the histograms with the largest peak in both frames. This result corresponded to the correct identification of a source in the rear hemifield.

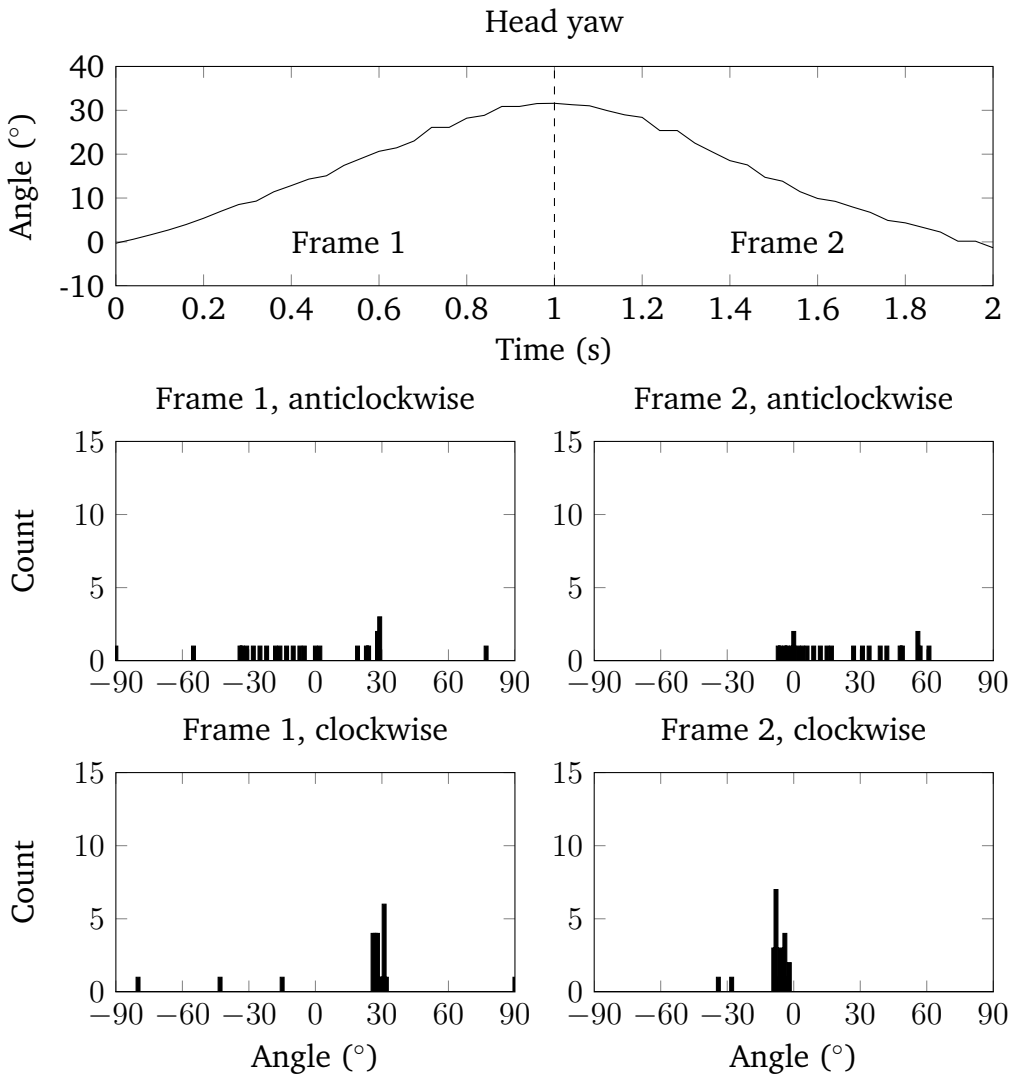

Fig. 6.29: B-DOA estimation for source in rear hemifield without empirical head model correction. Source at $180°$.

Figure 6.30 shows the same frames as figure 6.29 with the empirical head model correction applied. Similar to the front hemifield example, the histogram peak is higher and narrower with the correction. A comparison of localization estimates between frame one and frame two shows a movement of $-34°$ ($26°$ to $-8°$) while the reported head movement is $-33°$ ($32°$ to $-1°$), giving an error of $1°$.

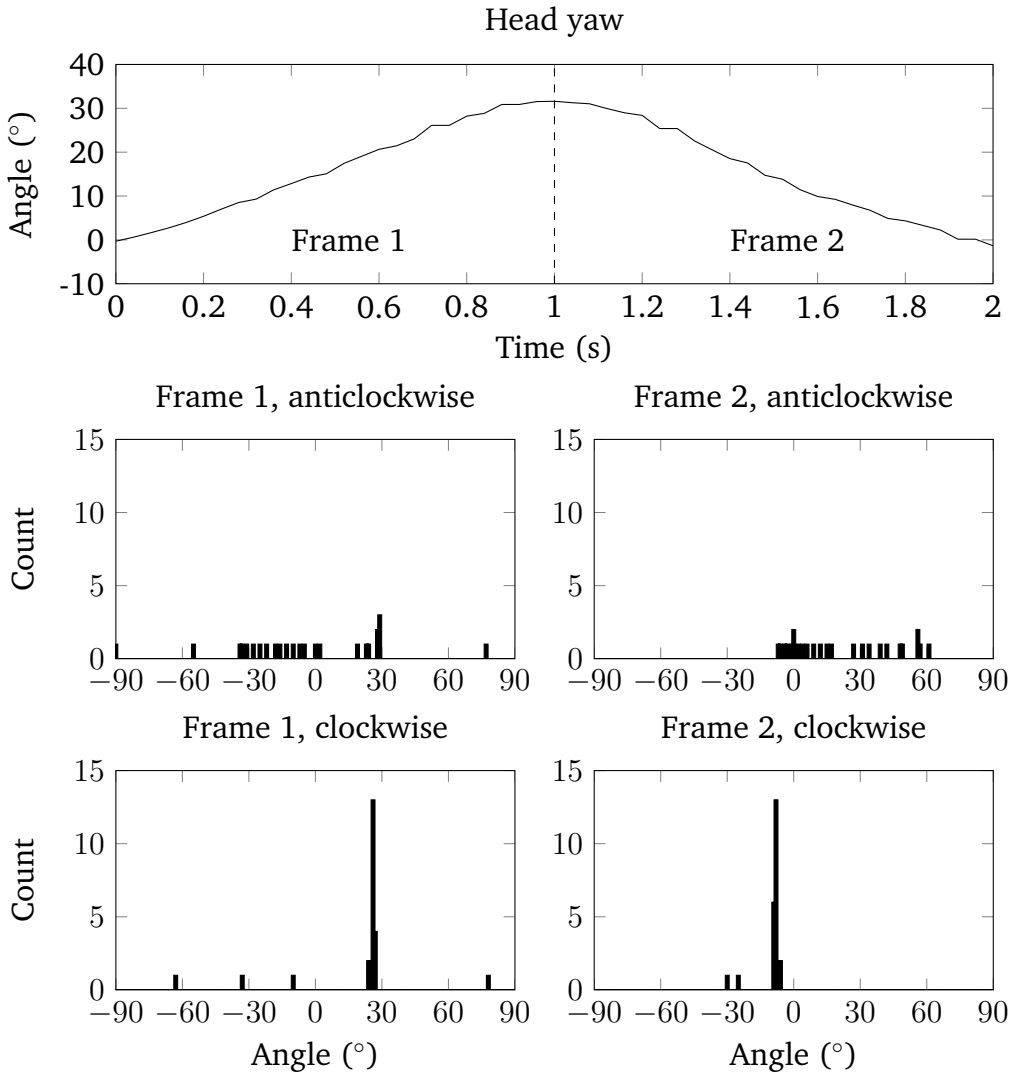

Fig. 6.30: B-DOA estimation for source in rear hemifield with empirical head model correction. Source at 180°.

6.5.6 Discussion of B-DOA estimation

The results show that a sound source can be robustly identified to be in the front or rear hemifield relative to a listener by combining a two-microphone bilateral array, the head movements of the listener, and selecting the histogram with the largest peak. Showing both clockwise and anticlockwise frames demonstrates

that the system can determine the position of a source for both rotational directions without additional a priori knowledge.

The relatively small peak observed in the clockwise rotated histogram in frame one of figures 6.27 and 6.28 is due to the period of reduced motion at the end of frame one, between 0.8 and 1 s. This shows that without head movement, the system performs in the same way as a standard GCC-PHAT and requires some head movement to determine whether a source is in the front or rear hemifield.

The use of an empirical head model makes a small improvement to the front hemifield estimates, narrowing the histogram distributions. The improvement is much greater in the rear hemifield, due to the divergence of the relationship between the interpolated time delays and source angle from the 1:1 ratio used in section 6.4. Applying the correct empirical correction narrows the distribution greatly. However, even without the correction, sources are correctly localized to the front or rear hemifield.

The shorter timescale used for each frame in comparison to the system used in section 6.4 means that a decision on the position of a source can be more quickly determined: This decision speed, though, is achieved at the expense of the robustness of the system to noise and its ability to track multiple sound sources. This system would be of use for quickly determining the position and hemifield of the strongest source in an auditory environment. The system could be extended to track the strongest source in both the front and rear by selecting the largest peak in the anticlockwise histogram as the strongest front hemifield source and the largest peak in the clockwise histogram as the strongest rear hemifield.

6.6 Conclusions

Section 6.3 described a system that mixed between directional and omnidirectional outputs based on thresholded head-mounted gyroscope output. The system was designed and simulated using a 135° head movement, the recorded gyroscope output, and real hearing-aid microphone impulse responses. The results of the simulation showed an earlier increase in the gain of a source in the rear hemisphere for the GCD relative to the output of fixed DMA and ADMA outputs during a 135° head turn. Using a dynamic virtual auditory display system, the GCD was simulated in real time using the theoretical output of a fixed DMA and HRIRs. The localization speed to sources in the rear hemisphere was tested by normal-hearing listeners. The GCD was compared to a fixed DMA. Most listeners localized sources more quickly using the GCD.

The gyroscopically compensated direction of arrival (GC-DOA) estimation used a head-mounted gyroscope to provide information on the rotational head-movements of a user was designed and tested in section 6.4. These head movements were used to compensate the DOA estimates of a GCC-PHAT system. In single and multiple source tests, with and without reverberation, the GC-DOA estimation was shown to provide an improved DOA estimate over a standard GCC-PHAT DOA estimation during head movements. Though it suffers from the same inaccuracies in increasing noise as the standard GCC-PHAT, the ability to aggregate more estimates over time partially mitigates this. This longer measurement time also allows shorter, more unstable analysis windows to be used, which allows more than one source to be identified in the case of speech signals. The system was shown to out-perform the equivalent GCC-PHAT system

during head movements for single sources, multiple sources in a reverberant space and for single sources in diffuse noise.

A biomimetic system for resolving front-back confusions in spatial signal processing was designed and tested in section 6.5. The system utilized head motion to differentially rotate GCC-PHAT estimates of DOA against head motion over time. Using simple peak size comparisons between congruent histograms, it was shown that the correct hemifield was selected for a single source. Applying an empirical correction to the localization estimates narrowed the histogram distributions, especially in the rear hemifield. This system extends the robust measurement space of a two-microphone, time-delay estimation technique from 180° to 360° without the need for a larger array. It achieves this by utilizing head movements.

CHAPTER 7

Conclusions

7.1 Discussion of internalization/externalization research (Chapters 4 and 5)

Prior to the research undertaken here, it was known that hearing-impaired listeners experienced an increased perception of internalization with the number of hearing aids worn and that they were less sensitive to the high-frequency pinna cues associated with static externalization by normal-hearing listeners.

In chapter 4, section 4.3, hearing-impaired listeners were found to have a reduced perception of externalization compared to normal-hearing listeners due to a reduced audibility of high-frequency pinna cues. This also reduced their perception of internalization compared to normal-hearing listeners when pinna cues were entirely removed and only ITD and DRR cues remained. It effectively "flattened" their perception of externalization. This suggests that pinna cues are weighted less than other cues for hearing-impaired listeners

and they are less able to extract pinna cue information for individual sources in multi-source environments. The reduced audibility of high-frequency cues for hearing-impaired listeners is however a major caveat to this finding and requires further research.

In chapter 4, section 4.4, presenting loud and impulsive noise bursts to hearing-impaired listeners showed no effect of the hearing aid, spectral content or envelope on externalization. The presentation angle of a noise burst had an effect on externalization ratings, with internalization being perceived by hearing-impaired listeners at 0° (directly in front). In the same experiment, normal-hearing listeners experienced internalization immediately after being fitted with flat-gain hearing aids, with the strongest effect at 0°. After six hours of acclimatization to the altered cues of a BTE hearing aid, the normal hearing ratings significantly increased to fully externalized for the 1-8 kHz bandwidth stimuli. In comparison, after acclimatization the effect of angle was reduced. The 3-8 kHz noise bursts were significantly less externalized after the acclimatization period than the 1-8 kHz noise bursts. The effect of angle recreates the results of externalization experiments conducted over headphones (Kim and Choi, 2010) (see section 2.4.5) and suggests that the presence of binaural cues, particularly ILDs in this case, are important for hearing-impaired listeners to externalize.

The acclimatization effect seen in normal-hearing listeners suggests that acclimatization to the altered cues provided by a hearing aid play an important role in externalization in three ways. First, an inability to acclimatize to altered cues may increase the perception of internalization, similar to the normal-hearing responses immediately after hearing-aid fitting. Second,

the speed of the acclimatization by normal-hearing listeners to the hearing aids in relation to externalization suggests that the acclimatizing factor for externalization is not the pinna cues, which previous research has shown (see section 2.2.7) take weeks to acclimatize to fully, but the increased gain and altered spectrum provided by the hearing aid. Combined with the early findings of Laws (1973) that externalized stimuli presented over headphones at higher amplitudes are internalized more, this could partially explain the cause of the internalized percept with amplification. Finally, acclimatization to hearing aids may result in a reduced weighting of pinna cues by aided hearing-impaired participants. The hearing-impaired responses show no difference between hearing-aid and no-hearing-aid conditions, suggesting an insensitivity to, or inaudibility of, changes in pinna cues in relation to externalization for the stimuli used. An increased reliance on ILD and ITD cues over pinna cues for externalization could explain the increase in internalization observed at $0°$. That is, while previous research has mostly focused on high-frequency pinna cues (see section 2.4) and reverberation as the fundamental cause of internalization, for the hearing-impaired listeners in these experiments, the fundamental factor was the sound source location.

In chapter 5, section 5.2, unilaterally aided hearing-impaired listeners who experience internalization rated their spatial listening ability on the SSQ to be lower than those who did not experiences internalization. In addition, unilaterally aided listener responses to the PIQ with and without their hearing aid showed that the perception occurred more often when wearing a hearing aid than without. The types of sounds that were reported as internalized were high-pitched, short in duration, loud and unexpected, with the exception of

slamming doors, which contain more low-frequency content. Theses types of sounds generally contain less salient spatial cues, either spectrally, temporally or audio-visually.

A recent study has focused on the effect of dynamic-range compression on the distribution of ILDs and externalization for normal-hearing listeners (Catic et al., 2013). The simulations presented in chapter 5, section 5.3 showed that a relatively high (for hearing aids) compression ratio (3:1) does not affect the shape of ILD distributions in the way suggested by that study. Anechoic conditions produce a narrower distribution than reverberant conditions. Similar changes in kurtosis (peakedness) of the ILD distributions were observed for different input signals and source DOAs. This suggests that large changes in kurtosis are required to produce perceptual effects, as there is no clear evidence that different sentences from the same talker can vary the perception of internalization. Compression can, in some cases, shift the ILD distribution towards 0 dB and has variable effects on distributions depending on the angle of the source, distorting the relationship between the source angles and their distributions. The main effect of compression on externalization then, may be due to the potential shift of the ILD distribution toward 0 dB (midline), which, as seen in section 4.4, produces the most internalized percept in hearing-impaired listeners.

From the assembled evidence, it appears that the perception of internalization by aided hearing-impaired participants is caused by (1) an inability to acclimatize to the altered cues provided by a hearing aid, in particular the increased loudness and altered spectrum of sounds, (2) an insensitivity to cues

for spatial hearing, combined with sounds and environments from which it is difficult to extract spatial cues.

7.2 Discussion of the use of head movements in hearing-aid signal processing (Chapter 6)

The research presented in chapter 6 represents the first use of head movements to improve hearing-aid signal processing techniques on *short* timescales.

A head-mounted gyroscope was used to control the directionality of a dual-microphone array in section 6.3. The directionality was directional below a set gyroscope output threshold and progressively omnidirectional above threshold. The system was designed to use head movements to determine whether a listener was listening to one source and would therefore prefer a directional response or was seeking/searching for another source, in which case an omnidirectional response would provide increased audibility for off-axis sources. This system represents a novel and first dynamic use of a gyroscope information to change hearing-aid directionality on short timescales. The system was shown in simulation to provide a benefit for orienting to sources in the rear hemifield that would normally be attenuated by an adaptive or fixed directional system during head movements. Three of four listeners showed benefit in a localization task using the system. This benefit was not due to moving their heads more quickly, but because the number of reversals that occurred were reduced. In practice, this could result in reduced listener effort in group-listening situations.

In section 6.4 a novel sound source localization system that compensates for head movements using head-mounted gyroscopes was developed. Using GCC-PHAT to make multiple estimates of source angle and aggregating them over time, this gyroscopically compensated source localization has been shown to outperform an uncompensated GCC-PHAT during head movements. Knowledge of head movements could allow hearing-aid localization algorithms to work on two timescales: (1) On short timescales ($< 1s$), head movement could be used in conjunction with previous estimates of source position to instantly update the position of a null or the main lobe of a beamformer, (2) on long timescales ($> 1s$), the algorithms could collect more information, for example to robustly detect up to four active sources.

In section 6.5 head movements were utilized biomimetically to extend the viable measurement space of a two-microphone array from one to both hemifields (front and rear). It was designed as an extension to the gyroscopically compensated source localization system (section 6.4), replacing the gyroscope with a 9-axis inertial sensor for improved head tracking. The system compensated for head movements by rotating the same DOA estimate histogram clockwise and anticlockwise by the measured head movement after each DOA estimate. The correct rotation produced the largest peak in the two histograms at the source DOA, indicating the hemifield the source originated from. This system mimics human resolution of front-back confusions by using head movements and the resultant relative movement of the apparent source position. The information provided by this system would allow future algorithms to robustly determine the position of sources of interest in the front

hemifield and noise sources in the rear hemifield without the need for larger microphone arrays or HRTF-based DOA estimation algorithms.

7.3 Future work in internalization/externalization perception

Future behavioural research on the internalization/externalization continuum and hearing aid use could focus on two main areas. The first is adaptation to level and spectrum. The aided normal-hearing results shown in section 4.4 suggest that an inability to acclimatize fully to the increased loudness and altered spectrum of sounds could result in internalization. An extension to the experiment in section 4.4 could investigate the effect of different hearing-aid gain profiles on acclimatization and internalization. In addition, the frequency content and timbre of sounds on internalization could be experimentally investigated, based on the PIQ survey evidence (section 5.2). From the prevalence data, "candidacy" – individual susceptibility to internalization – has yet to be (fully) determined.

Hearing-aid distortion artifacts is another area for future research. These are sounds that originate within the hearing aid and can also be caused by loud sounds that the hearing aid is not able to attenuate quickly enough or particularly complex auditory environments. As these sounds have no physical location in the outside world and do in fact originate "at the ear" (verged-cranial), they would be expected to be internalized by listeners. How often these sounds occur and in what situations could be important questions in internalization research for hearing-aid users.

The experiment undertaken in section 4.4 should also be revisited. As maximum externalization ratings were shown to correlate negatively with high-frequency hearing loss, the experiment should be performed while compensating for individual hearing loss.

7.4 Future work on utilizing head movements in hearing-aid signal processing

Future electroacoustic work could test the speech intelligibility benefit of gyroscopically controlled directionality (GCD). Further testing with hearing-impaired listeners is required to determine the benefit of GCD to them, as it appears from one instance that the GCD may not benefit listeners exhibiting slower dynamic localization of off-axis signals. In addition, performance may be limited by the listener's unaided ability to localize suprathreshold signals.

Determining the DOA estimate histogram produced by a moving source during head movements could also be developed in the future. Based on previous movement, the position of a moving sound source could be predicted, while compensating for head movement. Head movement information could also be applied to other DOA techniques, such as adaptive eigenvalue decomposition.

Hearing-aid algorithms that utilize head movements might be improved in the future by the incorporation of predictive models of head movement. By analyzing head movements during different types of listening, algorithms could be designed to predict the intent of the listener and select hearing-aid programmes or steer beamformers based on this prediction. However, these predictions would be required to be highly robust, with a low likelihood of

selecting the wrong setting based on the head movement and models used. In conclusion, the systems described in chapter 6 provide both implementable designs for next-generation hearing aids (algorithms), as well as a basis for future multi-modal algorithm research.

APPENDIX A

Image-source reverberation method

The image-source artificial reverberation technique was used to model a virtual room (Allen and Berkeley, 1979) in section 4.2.1. A diagram of the technique can be seen in figure A.1. Room reverberation arises from a combination of the direct path of the sound from source to listener and the multiple reflected paths arriving afterwards. Using the image-source technique, each sound reflection generates a mirror image of the original room.

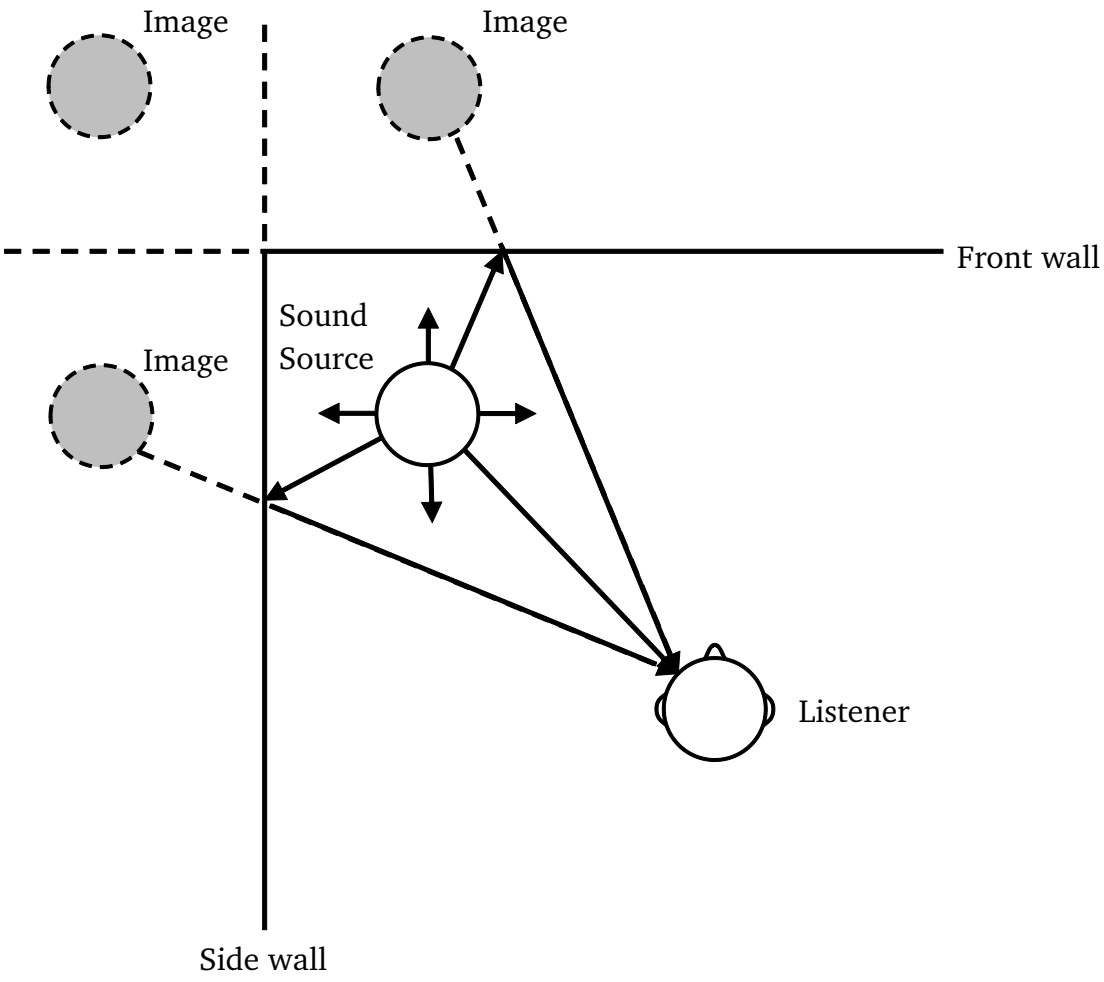

Fig. A.1: Diagram of the image-source reverberation method

The reverberant effect is created by delaying each reflection by the time it would take the sound to travel in a straight line from the "image" source to the listener. The sound is attenuated by the absorption coefficient of each "wall" it passes through, equivalent to each reflection in a real room. Each reflection is presented with the correct time delay and attenuation from the loudspeaker closest to the calculated direction of arrival, to produce an auditory perception of a room at the centre of the ring.

The method can be improved by modeling the late reflections as an exponentially decaying Gaussian noise, not as discrete reflections, producing a more natural reverberation (Lehman and Johansson, 2010). This improvement was used in section 4.2.2.

APPENDIX B

Swept-sine technique for extracting impulse responses

The swept-sine (SS) technique uses an exponential sweep which is then deconvolved with the recorded response to leave only the room impulse. This method is taken from Berdahl and Smith (2008).

Figure B.1 shows a linear system that is characterized by an impulse response $h(n)$, driven by an input signal $s(n)$ and producing the output signal $r(n)$. The swept sine measurement technique can be used to identify $h(n)$, by using the input and output signals.

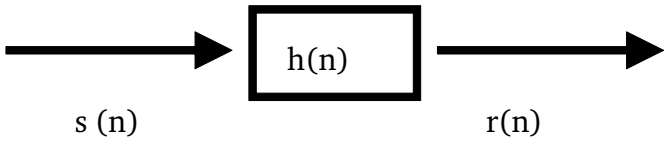

Fig. B.1: Linear system to be measured

The loudspeaker producing $s(n)$ is assumed to behave weakly non-linearly (non-linearity represented by $f(...)$). The Hammerstein model on which this system is based is shown in figure B.2 (Abel and Berners, 2005).

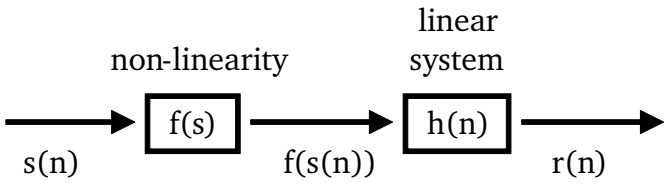

Fig. B.2: Hammerstein model

This system behaves as,

$$r(n) = f(s(n)) * h(n) \tag{B.1}$$

The input signal used to measure $h(n)$ independently of $f(...)$ is an exponential sine sweep from frequency ω_1 to ω_2 over T seconds (Farina, 2000),

$$s(n) = \sin\left[K\left(e^{\frac{-n}{Lf(...)}} - 1\right)\right] \tag{B.2}$$

where $K = \frac{\omega_1 T}{ln\frac{\omega_2}{\omega_1}}$ and $L = \frac{T}{ln\frac{\omega_2}{\omega_1}}$

The linearized impulse response is extracted by inverse filtering the recorded signal by the measurement signal. The signal $s(n)$ is created such that the time delay, Δt_n is constant between any sample n_0 and a later point with instantaneous frequency is N times larger than the instantaneous frequency at $s(n_0)$,

$$\Delta t_N = T\frac{ln(N)}{ln\frac{\omega_2}{\omega_1}} \tag{B.3}$$

By inverse filtering the response, the nonlinear terms in $f(...)$ are produced at specific places in the final response signal and the desired impulse response $h(n)$ can be separated from the nonlinear terms,.

Using simultaneous playback and recording, $h(n)$ can be obtained by,

$$h(n) = IDFT(\frac{DFT(r(n)}{DFT(s(n)}) \tag{B.4}$$

where IDFT is the inverse discrete Fourier transform and DFT the discrete Fourier transform. By recording a sweep that is longer than the reverberation time of the room under test, nonlinear components can be removed by truncating $h(n)$ without truncating the useful part of the impulse response.

APPENDIX C

Combined 9-axis and Wiimote-based head-tracking systems

This appendix describes the pre-existing systems used for head tracking.

C.1 Combined 9-axis head-tracker

The combined 9-axis sensor stick used in section 6.5 included the Analog Devices ADXL345 accelerometer, the Invensense ITG-3200 gyroscope and the Honeywell HMC5883L on one breakout board. The stick outputs digital readings using the I^2C protocol, multiplexing the outputs of the three sensors.

The combination of the accelerometer, gyroscope and magnetometer produced improved head tracking accuracy over time, as the drift in the sensor readings could be reduced. Accelerometers, unless exposed to large centripetal forces (such as those experienced by a banking aircraft) can provide a robust, invariant reference for acceleration due to gravity and thus the vertical plane. Three-axis magnetometers, once calibrated for the current magnetic field conditions, provide estimates of three-dimensional orientation. However, they are susceptible to changing external magnetic field conditions after

calibration. Gyroscopes provide fast updates on three-dimensional rotational velocity. However, they are susceptible to sensor drift due to sampling errors over time. By combining the three sensor outputs, reliable estimates of head position can be obtained.

C.1.1 MEMS magnetometer calibration

The magnetometer used required extensive calibration for the magnetic fields present, compensating for both hard iron and soft iron distortions, though other MEMS magnetometers are able to self-calibrate. Hard-iron distortions are caused by constant magnetic fields, such as the magnets of nearby loudspeakers or headphones. If the orientation of these distortions to the sensor are constant, they can be removed by applying a simple offset to the magnetometer output. Soft iron distortions occur when the orientation of the distortion to the sensor is not constant. The simplest way (implemented in this system by Bartz, 2012) is to remove these distortions is to use extensive initial measurements of the offsets in all directions and dimensions. Figure C.1 shows the output from the HMC5883L magnetometer before calibration and the ellipsoid fitted to it.

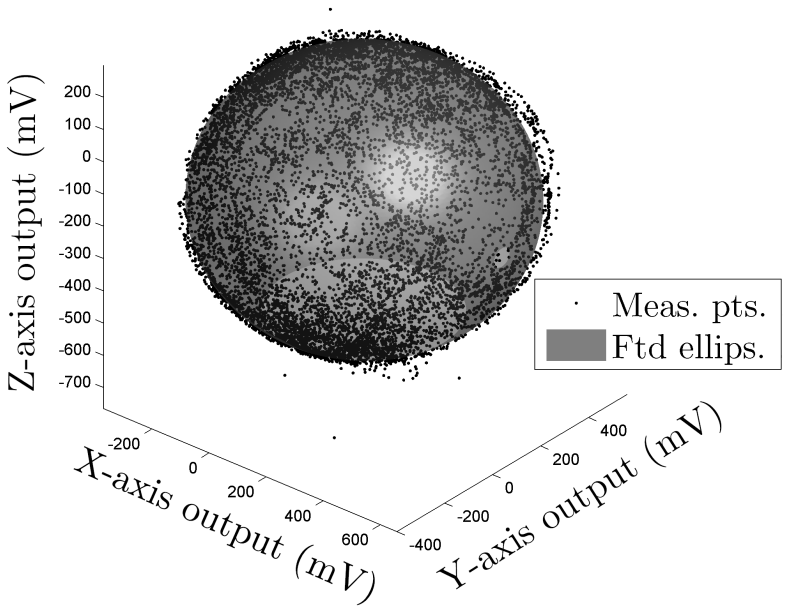

Fig. C.1: Magnetometer ellipsoid fitting. Meas. pts. = points measured by the magnetometer, Ftd ellips. = the fitted ellipsoid to the measured points

Figure C.2 shows the uncorrected and corrected output after calibration, for hard and soft iron distortions. If neither of these distortions is present, the output of the magnetometer should be a sphere centered on the origin. Hard-iron distortions result in an offset of the centre of the sphere from the origin. The soft iron distorts the shape of the sphere into an ellipsoid. By obtaining the equation of the ellipsoid, the output of the magnetometer can be calibrated.

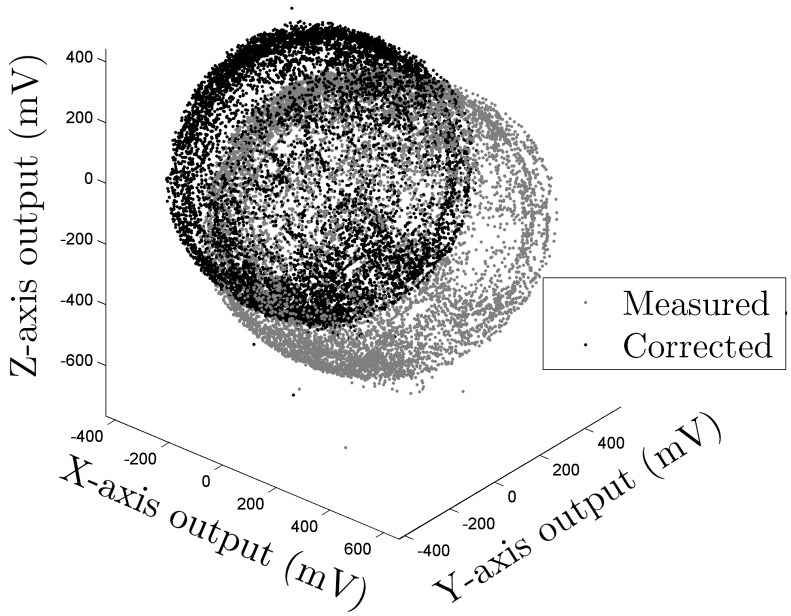

Fig. C.2: Measured output of the magnetometer and the hard and soft iron corrected magnetometer output.

C.2 Portable Wiimote-based head-tracking system and virtual auditory display

A portable Wiimote-based head-tracking system was used for the initial testing and real-time simulations in section 6.3.4.1. The system is shown in figure C.3. The system used the infrared camera and light-tracking capabilities of the Nintendo Wiimote (the controller for the Nintendo Wii games console) and a custom infrared LED (irLED) array. The Wiimote was placed 1.5 metres above the listener, producing the best compromise between the viewing angle of the infrared camera and irLED tracking accuracy. A 20 centimetre long array was mounted on the listener's head, powered by a 9V battery. The arrangement

of the irLEDs is shown in figure C.3. The central LED was offset towards the rear of the array and this asymmetric arrangement allowed determination of the $360°$ orientation of the irLED array and therefore the listener's head in the horizontal plane. Unambiguous measurement of orientation was dependent on detecting all three LEDs.

The Wiimote was connected to the host PC via Bluetooth and the XY position of the LEDs could be measured at 100 Hz within MATLAB. Communication of the Wiimote with MATLAB used the dynamic link library in the WiiLAB toolbox (Brindza et al., 2009). Measuring the relative Euclidean distance between the detected LEDs enabled the detection of the front and rear of the array. The listener's head angle was measured using the arctangent transform of the XY positions of the front and rear LEDs.

The Wiimote head-tracking system was combined with the MIT head-related impulse response (HRIR) database (Gardner and Martin, 2000) to produce a dynamic virtual auditory display system. An overlap-add technique was used to update the output of the virtual auditory display. The two closest impulse responses to the listener's current head angle (taken from the Wiimote head-tracking system) were chosen and interpolated linearly. The interpolation was required due to the $5°$ spatial resolution of the MIT impulse responses. This resulted in intermediate source directions being perceptually approximated. 3584 zeros preceded signals consisting of monaurally recorded IEEE York sentences (Stacey and Summerfield, 2007). These zero-padded signals were segmented into frames of 4096 samples overlapped by 3584 samples (a 7/8 overlap). These frames were convolved with the interpolated binaural impulse response, producing a two-channel signal of 4096 samples duration, of which

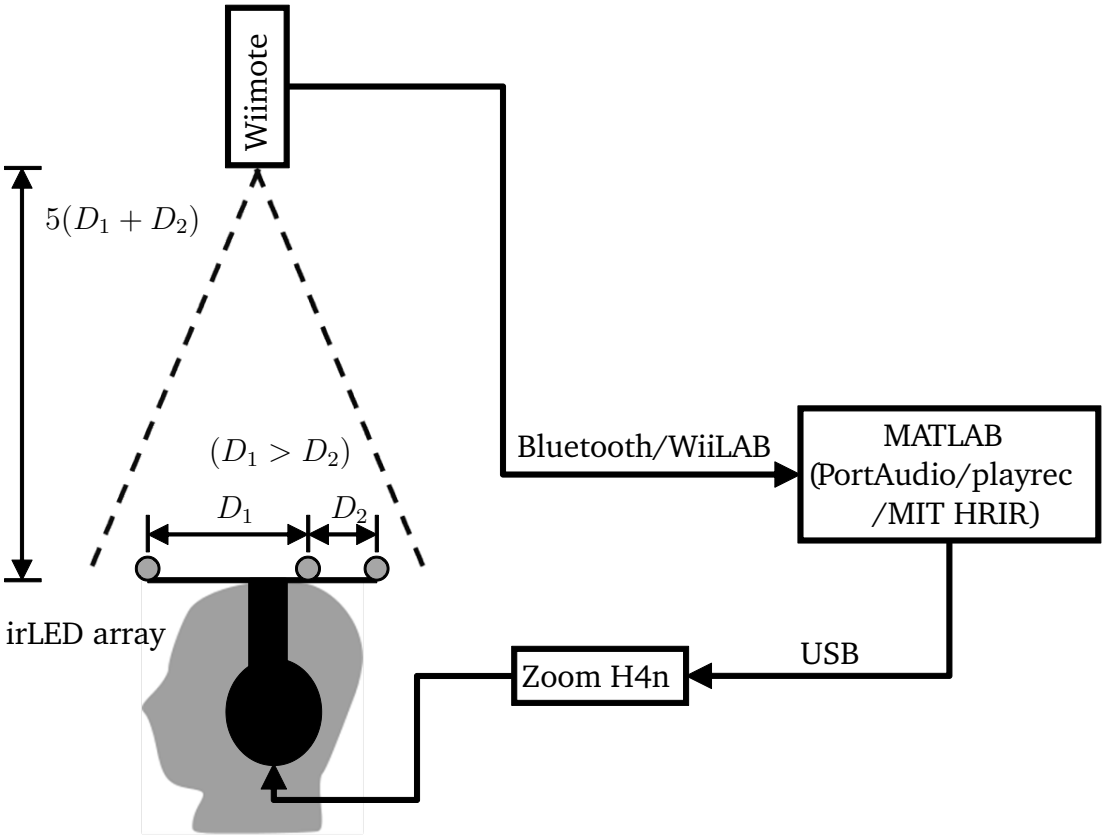

Fig. C.3: Wiimote head-tracking and virtual auditory display system.

the last 512 samples were selected for playback. 480 samples were transferred to the audio buffer and the remaining 32 where placed in an array, to be linearly cross-faded with the first 32 samples of the next buffer. This method allowed the reverberant tail of previous signal segments to be updated with the current impulse response. Head turn to a change in apparent source location latency ranged between 22 and 33 ms.

References

(1995). Iec 61260, 1st ed. 1995-07: Octave-band and fractional octave-band filters. 165

(2003). Ansi s3.22-2003: Specification of hearing aid characteristics. 60

(2004). Ansi s1.11-2004: Specification for octave-band and fractional octave-band analog and digital filters. 165

Abel, J. and Berners, D. (2005). Signal processing for digital audio effects. 252

Aichner, R., Buchner, H., Wehr, S., and Kellermann, W. (2006). Robustness of acoustic multiple-source localization in adverse environments. In *Proceedings of Sprachkommunikation*. 87, 90

Akeroyd, M. A. (2006). The psychoacoustics of binaural hearing. *International Journal of Audiology*, **45**, S25–S33. 9

Akeroyd, M. A. (2010). The effect of hearing-aid compression on judgments of relative distance. *J. Acoust. Soc. Am.*, **127**, 9–12. 34

Akeroyd, M. A., Gatehouse, S., and Blaschke, J. (2007). The detection of differences in the cues to distance by elederly hearing-impaired listeners. *J. Acoust. Soc. Am.*, **121**, 1077–1089. 34

Akeroyd, M. A., Guy, F. H., Harrison, D. L., and Suller, S. L. (2013). A factor analysis of the ssq (speech, spatial, and qualities of hearing scale). *Int. J. Audiol.*, **tba**, tba. 51, 144

Albrecht, R. and Lokki, T. (2013). Adjusting the perceived distance of virtual speech sources by modifying binaural room impulse responses. In *International Conference on Auditory Display*. 30

Allen, J. B. and Berkeley, D. A. (1979). Image method for efficiently simulating small-room acoustics. *J. Acoust. Soc. Am.*, **65**, 943–950. 34, 103, 249

Amlani, A. M. (2001). Efficacy of directional microphone hearing aids: a meta-analytic approach. *J. Am. Acad. Audiol.*, **12**, 202–214. 76

Ang, K., Chong, G., and Li, Y. (2005). Pid control system analysis, design, and technology. *IEEE Transactions on Control Systems Technology*, **13**, 559–576. 93

Arentsschild, O. and Froeber, B. (1972). Comparative measurements of the effect of a directional microphone in the hearing aid. *J Audio Tech*, **2**, 27–36. 64

Ashmead, D. H., LeRoy, D., and Odom, R. D. (1990). Perceptionof the relative distances of nearby sound sources. *Percept. Psychophys.*, **47**, 326–331. 26

Banks, M. S. and Green, D. M. (1973). Localization of high- and low-frequency transients. *J. Acoust. Soc. Am.*, **53**, 1432–1433. 14

Bartz, P. (2012). Razor attitude and head rotation sensor. 93, 94, 229, 256

Bauer, B. B. (1965). Improving headphone listening comfort. *Journal of the Audio Engineering Society*, **13**, 300–302. 52

Begault, D. R. (1992). Perceptual effects of synthetic reverberation on three-dimensional audio systems. *Journal of the Audio Engineering Society*, **40**, 895–904. 42

Begault, D. R. (1994). *3D Sound for Virtual Reality and Multimedia*. Academic Press. 13, 17

Begault, D. R. and Wenzel, E. M. (1993). Headphone localization of speech. *Human Factors*, **35**(2), 361–376. 47

Begault, D. R. and Wenzel, E. M. (2001). Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual source. *Journal of the Audio Engineering Society*, **49**(10), 904–915. 37, 53, 54

Behrens, T., Maas, P., and Neher, T. (2009). A method for quantifying the effects of non-linear hearing aid signal-processing on interaural level difference cues in conditions with multiple sound sources. In *Proceedings of the International Symposium on Auditory and Audiological Research*. 176

Benesty, J. (2000). Adaptive eigenvalue decomposition for passive acoustic source localization. *J. Acoust. Soc. Am.*, **107**, 384–391. 83, 84, 87, 90

Bentler, R. A., Tubbs, J. L., Egge, J. L., Flamme, G. A., and Dittberner, A. B. (2004). Evaluation of an adaptive directional system in a dsp hearing aid. *American Journal of Audiology*, **13**, 73–79. 76, 77

Berdahl, E. and Smith, J. O. (2008). Impulse response measurement toolbox: Real simple project. 110, 252

Bernstein, L. R. and Trahiotis, C. (2002). Enhancing sensitivity to interaural delays at high frequencies using "transposed stimuli". *J. Acoust. Soc. Am.*, **112**, 1026–1036. 12

Bernstein, L. R. and Trahiotis, C. (2004). The apparenty immunity of high-frequency "transposed" stimuli to low-frequency binaural interference. *J. Acoust. Soc. Am.*, **116**, 3062–3069. 12

Best, V., Carlile, S., Kopčo, N., and van Schaik, A. (2011). Localization in speech mixtures by listeners with hearing loss. *J. Acoust. Soc. Am.*, **129**, EL210–EL215. 21

Blauert, J. (1997). *Spatial Hearing*. The MIT Press. 17, 31, 39

Boyd, A. W., Whitmer, W. M., Soraghan, J. J., and Akeroyd, M. A. (2012). Auditory externalization in hearing-impaired listeners: The effect of pinna cues and number of talkers. *J. Acoust. Soc. Am.*, **131**, 268–274. 29

Boymans, M., Goverts, S. T., Kramer, S. E., Festen, J. M., and Dreschler, W. A. (2008). A prospective multi-centre study of the benefits of bilateral hearing aids. *Ear Hear*, **29**, 930–941. 23

Braida, L. D., Durlach, N. I., DeGennaro, S. V., and Peterson, P. M. (1982). *The Vanderbilt Hearing Aid Report: State of the Art Research Needs. Monographs in Contemporary Audiology*, chapter Review of recent research on multiband amplitude compression for the hearing impaired, pages 1–1. Upper Darby, Pennsylvania. 61

Brandstein, M. S. and Silverman, H. F. (1997). A robust method for speech signal time-delay estimation in reverberant rooms. In *Proceedings of the IEEE Internation Conference on Acoustics, Speech and Signal Processing*. 86

Briley, P. M., Kitterick, P. T., and Summerfield, A. Q. (2012). Evidence for opponent process analysis of sound source location in humans. *J. Assoc. Res. Otolaryngol.*, **14**, 83–101. 15

Brimijoin, W. O. (2013). The effect of hearing aid microphone mode on performance in an auditory orienting task (submitted). *Ear Hear.*, **N/A**, N/A. 188, 213

Brimijoin, W. O. and Akeroyd, M. A. (2012). The role of head movements and signal spectrum in an auditory front/back illusion. *iPerception*, **3**, 179–182. 95

Brimijoin, W. O., Boyd, A. W., and Akeroyd, M. A. (2013). The contribution of head movement to the externalization and internalization of sounds. *PLoS ONE*, **8**, e83068. 54, 181

Brimijoin, W. O., McShefferty, D., and Akeroyd, M. A. (2010). Auditory and visual orienting responses in listeners with and without hearing-impairment. *J. Acoust. Soc. Am.*, **127**, 3678–3688. 19

Brimijoin, W. O., McShefferty, D., and Akeroyd, M. A. (2012). Undirected head movements of listeners with asymmetrical hearing impairment during a speech-in-noise task. *Hear. Res.*, **283**, 162–168. 19, 180

Brindza, J., Szweda, J., Liao, Q., Jiang, Y., and Striegel, A. (2009). Wiilab: Bringing together the nintendo wiimote and matlab. In *Frontiers in Education Conference, 2009. FIE '09. 39th IEEE*. 259

Brungart, D. S. (1999). Auditory localization of nearby sources. iii. stimulus effects. *J. Acoust. Soc. Am.*, **106**, 3589–3602. 31

Brungart, D. S., Durlach, N. I., and Rabinowitz, W. M. (1999). Auditory localization of nearby sources. *J. Acoust. Soc. Am.*, **106**, 1956–1968. 31

Brungart, D. S. and Rabinowitz, W. M. (1999). Auditory localization of nearby sources. head-related transfer functions. *J. Acoust. Soc. Am.*, **106**, 1465–1479. 13, 31

Brungart, D. S. and Simpson, B. D. (2002). The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal. *J. Acoust. Soc. Am.*, **112**, 664–676. 105

Brutti, A., Omologo, M., and Svaizer, P. (2008). Comparison between different sound source localization techniques based on a real data collection. In *Hands-Free Speech Communication and Microphone Arrays (HSCMA)*. 83, 87

Buus, S. and Florentine, M. (2002). Growth of loudness in listeners with cochlear hearing losses: Recruitment reconsidered. *J. Assoc. Res. Otolaryngol.*, **3**, 120–139. 59

Byrne, D. and Noble, W. (1998). Optimizing sound localization with hearing aids. *Trends in Amplification*, **3**, 51–72. 19, 22

Byrne, D., Noble, W., and LePage, B. (1992). Effects of longterm bilateral and unilateral fitting of different hearing-aid types on the ability to locate sounds. *J. Am. Acad. Audiol.*, **3**, 369–382. 19, 20, 21, 22

Cabrera, D. and Gilfillan, D. (2002). Auditory distance perception of speech in the presence of noise. In *Proceedings of the 8th International Conference on Auditory Display*. 34

Carlile, S., Leong, P., and Hyams, S. (1997). The nature and distributions of errors in sound localization by human listeners. *Hear. Res.*, **114**, 179–196. 14

Catic, J., Santurette, S., Buchholz, J. M., Gran, F., and Dau, T. (2013). The effect of interaural-level-difference fluctuations on the externalization of sound. *J. Acoust. Soc. Am.*, **134**, 1232–1241. 5, 31, 48, 50, 144, 163, 164, 165, 166, 167, 176, 177, 243

Chatlani, N. (2011). *Advanced signal enhancement techniques with application to speech and hearing*. PhD thesis, Centre for excellence in Signal and Image Processing, Department of Electronic and Electrical Engineering, University of Strathclyde. 67, 80, 81, 214

Chatlani, N., Fischer, E., and Soraghan, J. J. (2010). Spatial noise reduction in binaural hearing aids. In *18th European Signal Processing Conference (EUSIPCO-2010)*. 80, 81, 214

Chung, K. (2004). Challenges and recent developments in hearing aids: Part i. speech understanding in noise, microphone technologies and noise reduction algorithms. *Trends in Amplification*, **8**, 83–124. 67, 77

Cochran, P., Throop, J., and Simpson, W. E. (1968). Estimation of distance of a source of sound. *Am. J. Psychol.*, **81**, 198–206. 26, 31

Coleman, P. D. (1968). Dual role of frequency spectrum in determination of audiitory distance. *J. Acoust. Soc. Am.*, **44**, 631–634. 30, 32

Cord, M. T., Surr, R. K., Walden, B. E., and Dyrlund, O. (2004). Relationship between laboratory measures of directional advantage and everyday success with directional microphone hearing aids. *J. Am. Acad. Audiol.*, **15**, 353–363. 76

Cord, M. T., Surr, R. K., Walden, B. E., and Olsen, L. (2002). Performance of directional microphone hearing aids in everyday life. *J. Am. Acad. Audiol.*, **13**, 295–307. 76

Day, R. H. (1968). Perceptual constancy of auditory direction with head rotation. *Nature*, **219**, 501–502. 52

de Boer, K. (1946). The formation of stereophonic images. *Philips Technical Review*, **8**, 51–56. 15

de Boer, K. (1947). A remarkable phenomenon with stereophonic sound reproduction. *Philips Technical Review*, **9**, 8–13. 15

de Boer, K. and Vermuelen, R. (1939). On improving defective hearing. *Philips Technical Review*, **4**, 316–319. 39

den Bogaert, T. V., Klasen, T. J., Moonen, M., Deun, L. V., and Wouters, J. (2006). Horizontal localization with bilateral hearing aids: Without is better than with. *J. Acoust. Soc. Am.*, **119**, 515–526. 22

Desloge, J. G. and Rabinowitz, W. M. (1997). Microphone-array hearing aids with binaural output - part i: Fixed-processing systems. *IEEE Transactions on Speech and Audio Processing*, **5**, 529–542. 79

DiBiase, J. H. (2000). *A High-Accuracy, Low-Latency Technique for Talker Localization in Reverberant Environments Using Microphone Arrays*. PhD thesis, Division of Engineering at Brown University. 83, 87

DiCarlo, L. M. and Brown, W. J. (1960). The effectiveness of binaural hearing for adults with hearing impairments. *J. Aud. Res.*, **1**, 35–76. 22

Dillon, H. (1996). Compression? yes, but for low or high frequencies, for low or high intensities, and with what response times? *Ear Hear.*, **17**, 287–307. 60

Dillon, H. (2001). *Hearing Aids*. New York: Thieme. 19, 21, 58, 61, 64, 66

Doclo, S., Moonen, M., and den Bogaert J. Wouters, T. V. (2009). Reduced-bandwidth and distributed mwf-based noise reduction algorithms for binaural hearing aids. In *IEEE Transactions on Audio, Speech, and Language Processing*, volume 17, pages 38–51. 81

Doerbecker, M. and Ernst, S. (1996). Combination of two-channel spectral subtraction and adaptive wiener post-filtering for noise reduction and dereverberation. In *Proceedings of the 8th European Signal Processing conference (EUSIPCO '96)*. 81

Domnitz, R. (1973). The interaural time jnd as a simultaneous function of interaural time and interaural amplitude. *J. Acoust. Soc. Am.*, **53**, 1549–1552. 12, 13, 166

Dreschler, W. A. (1992). Fitting multichannel-compression hearing aids. *Audiology*, **31**, 121–131. 61

Dreschler, W. A., Verschuure, H., Ludvigsen, C., and Westermann, S. (2001). Icra noises: artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment. international collegium for rehabilitative audiology. *Audiology*, **40**, 148–157. 164, 169, 202

Duda, R. O. and Martens, W. L. (1998). Range dependence of the response of a spherical head model. *J. Acoust. Soc. Am.*, **104**, 3048–3058. 10, 13

Durlach, N. I., Rigopulos, A., andW. S. Woods, X. D. P., Kulkarni, A., Colburn, H. S. C. H. S., and Wenzel, E. M. (1992). On the externalization of auditory images. *Presence*, **1**(2), 251–257. 35, 36

Durlach, N. I., Shinn-Cunningham, B. G., and Held, R. M. (1993). Supernormal auditory localization i. general background. *Presence*, **2**, 89–103. 23

Durlach, N. I., Thompson, C. L., and Colburn, H. S. (1981). Binaural interaction in impaired listeners: a review of past research. *Audiology*, **20**, 181–211. 19

Edwards, A. S. (1955). Accuracy of auditory depth perception. *J. Gen. Psychol.*, **52**, 327–329. 25

Edwards, B. (2007). The future of hearing aid technology. *Trends in Amplification*, **11**, 31–46. 2

Elliott, R. S. (1981). *Antenna Theory and Design*. Engelwood Cliffs, Prentice Hall. 65

Farina, A. (2000). Simultaneous measurement of impulse response and distortion with a swept-sine technique. In *Audio Engineering Society Convention 108*. 253

Fischer, H. G. and Freedman, S. J. (1968). The role of the pinna in auditory localization. *J. Aud. Res.*, **8**, 15–26. 16

Franssen, N. V. (1960). *Some considerations of the mechanism of directional hearing*. PhD thesis, Institute of Technology, Delft. 39

Frost, O. L. (1972). An algorithm for linearly constrained adaptive array processing. In *Proceedings of the IEEE*, volume 60, pages 926–935. 65

Füllgrabe, C., Baer, T., Stone, M. A., and Moore, B. C. J. (2010). Preliminary evaluation of a method for fitting hearing aids with extended bandwidth. *Int. J. Audiol.*, **49**, 741–753. 22

G. Kidd, J., Abrogast, T. L., Mason, C. R., and Gallun, F. J. (2005). The advantage of knowing where to listen. *J. Acoust. Soc. Am.*, **118**, 3804–3815. 21

Gamble, E. A. (1909). Intensity as a criterion in estimating the distance of sounds. *Psychol. Rev.*, **16**, 416–426. 25

Gardner, M. B. (1968). Proximity image effect in sound localization. *J. Acoust. Soc. Am.*, **43**, 163. 25

Gardner, M. B. (1969). Distance estimation of 0 degrees or apparent 0 degree-oriented speech signals in anechoic space. *J. Acoust. Soc. Am.*, **45**, 47–53. 27, 31

Gardner, W. and Martin, K. (2000). Hrtf measurements of a kemar dummy-head microphone. 259

Gatehouse, S. (1999). Glasgow hearing aid benefit profile: Derivation and validation of a client-centered outcome measure for hearing-aid services. *J. Am. Acad. Audiol.*, **10**, 80–103. 148

Gatehouse, S. and Noble, W. (2004). The speech, spatial and qualities of hearing scale (ssq). *Int. J. Audiol.*, **43**, 85–99. 34, 144, 146, 178

Ghazanfar, A. A., Neuhoff, J. G., and Logothetis, N. K. (2002). Auditory looming perception in rhesus monkeys. *Proc. Natl. Acad. Sci. U. S. A.*, **99**, 15755–15757. 33

Gilbert, G., Akeroyd, M. A., and Gatehouse, S. (2008). Discrimination of release time constants in hearing-aid compressors. *Int. J. Audiol.*, **47**, 189–198. 144, 163, 164

Goetze, S., Rohdenburg, T., Hohmann, V., Kollmeier, B., and Kammeyer, K.-D. (2007). Direction of arrival estimation based on the dual delay line approach for binaural hearing aid microphone arrays. In *Proceedings of the International Symposium on Intelligent Signal Processing and Communication Systems*. 83, 88

Good, M. D. and Gilkey, R. H. (1996). Sound localization in noise. the effect of signal-to-noise ratio. *J. Acoust. Soc. Am.*, **99**, 1108–1117. 20

Gordy, J. D., Bouchard, M., and Aboulnasr, T. (2008). Beamformer performance limits in monaural and binaural hearing aid applications. In *Canadioan Conference on Electrical and Computer Engineering (CCECE)*. 79

Goupell, M. J. and Hartmann, W. M. (2007). Interaural fluctuations and the detection of interaural incoherence. iii. narrowband experiments and binaural models. *J. Acoust. Soc. Am.*, **122**, 1029. 48

Grant, K. W. (2001). The eeffect f speechrspeech on masked detection thresholds for fifilter speech. *J. Acoust. Soc. Am.*, **109**, 2272–2275. 2

Grantham, D. W. (1984). Interaural instensity discrimination: Insensitivity at 1000 hz. *J. Acoust. Soc. Am.*, **75**, 1191–1194. 13

Gravel, J. S., Fausel, N., Liskow, C., and Chobot, J. (1999). Children's speech recognition in noise using omni-directional and dual-microphone technology. *Ear Hear*, **20**, 1–11. 75

Greenberg, J. M. and Zurek, P. M. (1992). Evaluation of an adaptive beamforming method for hearing aids. *J. Acoust. Soc. Am.*, **91**, 1662–1676. 79

Griffiths, L. J. and Jim, C. W. (1982). An alternative approach to linearly constrained adaptive beamforming. *IEEE transactions on Antennas and Propagation*, **30**, 27–34. 65, 66

Hafter, E. R., Dye, R. H., Nuetzel, J. M., and Aronow, H. (1977). Difference thresholds for interaural intensity. *J. Acoust. Soc. Am.*, **61**, 829–834. 14

Hanson, R. L. and Kock, W. E. (1957). Interesting effect produced by two loudspeakers under free space conditions. *J. Acoust. Soc. Am.*, **29**, 145. 40, 177

Hart, J., Onceanu, D., Sohn, C., Wightman, D., and Vertegaal, R. (2009). *INTERACT 2009, Part I*, chapter The attentive hearing aid: Eye selection of auditory sources for hearing impaired users, pages 19–35. Springer Berlin. 96

Hartmann, W. M. (1996). On the externalization of sound images. *J. Acoust. Soc. Am.*, **99**(6), 3678–3688. 37, 46, 47, 117

Hartmann, W. M., Dunai, L., and Qu, T. (2013). *Basic Aspects of Hearing: Physiology and Perception*, chapter Interaural Time Difference Thresholds as a Function of Frequency, pages 239–246. Springer New York. 12

Haustein, B. G. (1969). Hypothesen über die einohrige entfernungswahrnehmung des menschlichen gehörs (hypotheses about the perception of distance in human hearing with one ear). *Hochfrequenztech. Elektroakust*, **78**, 46–57. 26, 32

Hawkins, D. B. and Yacullo, W. S. (1984). Signal-to-noise ratio advantage of binaural hearing aids and directional microphones under different levels of reverberation. *J. Speech Hear. Disord.*, **49**, 278–286. 74, 75, 78

Heyes, A. D. and Ferris, A. J. (1975). Auditory localization using hearing aids. *Br. J. Audiol.*, **9**, 102–106. 22

Hirsch, H. R. (1968). Perception of the range of a sound source of unknown strength. *J. Acoust. Soc. Am.*, **43**, 373–374. 30

Ho, W. K. D., Koo, W. H., and Tan, B. H. (2008). Hearing apparatus with automatic alignment of the directional microphone and corresponding method. 97

Hoffman, M., Trine, T., Buckley, K., and Tasell, D. V. (1994). Robust adaptive microphone array processing for hearing aids: realistic speech enhancement. *J. Acoust. Soc. Am.*, **96**, 759–770. 63

Hofman, P. M., Riswick, J. G. A. V., and Opstal, A. J. V. (1998). Relearning sound localization with new ears. *Nat. Neurosci.*, **1**, 417–421. 23

Hohmann, V. (2002). Frequency analysis and synthesis using a gammatone filterbank. *Acustica/Acta Acust.*, **88**, 433–442. 49

Holt, R. E. and Thurlow, W. R. (1969). Subject orientation and judgment of distance of a sound source. *J. Acoust. Soc. Am.*, **46**, 1584–1585. 31

Hunter, M. D., Griffiths, T. D., Farrow, T. F. D., Zheng, Y., Wilkinson, I. D., Hegde, N., Woods, W., Spence, S. A., and Woodruff, P. W. R. (2003). A neural basis for the perceptions of voices in external auditory space. *Brain*, **126**, 161–169. 55

Ingard, U. (1953). A review of the influence of meteorological conditions on sound propogation. *J. Acoust. Soc. Am.*, **25**, 405–411. 32

Iwaya, Y., Suzuki, Y., and Kimura, D. (2003). Effects of head movement on front-back error in sound localization. *Acoustical Science and Technology*, **24**, 322–324. 17

Jack, C. E. and Thurlow, W. R. (1973). Effects of degree of visual association and angle displacement on the "ventriloquism" effect. *Perceptual and motor skills*, **37**, 967–979. 25

Jastreboff, P. J. (1990). Phantom auditory perception (tinnitus). mechanisms of generation and perception. *Neurosci. Res.*, **8**, 221–254. 38

Jeffress, L. A. (1948). A place theory of sound localization. *J. Comp. Physiol. Psychol.*, **41**, 35–39. 12, 88

Jeffress, L. A. and Taylor, R. W. (1961). Lateralization versus localization. *J. Acoust. Soc. Am.*, **33**, 482–483. 37

Jesteadt, W., Wier, C. C., and Green, D. M. (1977). Intensity discrimination as a function of frequency and sensation level. *J. Acoust. Soc. Am.*, **61**, 169–177. 26

Junius, D., Riedel, H., and Kollmeier, B. (2007). The influence of externalization and spatial cues on the generation of auditory brainstem responses and middle latency responses. *Hear. Res.*, **225**, 91–104. 55

Kates, J. M. (1988). A computer simulation of hearing aid response and the effects of ear canal size. *J. Acoust. Soc. Am.*, **83**, 1952–1963. 58

Kates, J. M. (2008). *Digital Hearing Aids*, chapter Directional Microphones, pages 75–111. Plural Publishing. 57, 65, 126

Kates, J. M. (2010). Understanding compression. modeling the effects of dynamic-range compression in hearing aids. *Int. J. Audiol.*, **49**, 395–409. 59

Kayser, H., Ewert, S. D., Anemuller, J., Rohdenburg, T., Hohmann, V., and Kollmeier, B. (2009). Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses. In *"Digital Signal Processing for Hearing Instruments" EURASIP Journal on Advances in Signal Processing*. 63, 64, 67, 164, 191

Keidser, G., Rohrseitz, K., Dillon, H., Hamacher, V., Carter, L., Rass, U., and Convery, E. (2006). The effect of multi-channel wide dynamic range compression, noise reduction, and the directional microphone on horizontal localization performance in hearing aid wearers. *Int. J. Audiol.*, **45**, 563–579. 22, 62

Keyrouz, F., Maier, W., and Diepold, K. (2006). A novel humanoid binaural 3d sound localization and separation algorithm. In *IEEE-RAS International Conference on Humanoid Robots*. 89

Keyrouz, F. and Saleh, A. A. (2007). Intelligent sound source localization based on head-related transfer functions. In *IEEE International Conference on Intelligent Computer Communication and Processing*. 89

Kidd, G. and Favrot, S. (2013). Design and preliminary testing of a visually guided hearing aid. *J. Acoust. Soc. Am.*, **133**, EL202–EL207. 97

Kiesling, J., Brenner, B., Jespersen, C. T., Groth, J., and Jensen, O. D. (2005). Occlusion effects of earmolds with different venting systems. *J. Am. Acad. Audiol.*, **16**, 237–249. 59

Kietz, H. (1953). Das räumliche hören [spatial hearing]. *Acustica*, **3**, 73–86. 39

Killion, M. C., Wilbur, L. A., and Gudmundsen, G. I. (1988). Zwislocki was right... *Hearing Instruments*, **39**, 14–18. 59

Kim, C., Mason, R., and Brookes, T. (2013). Head movements made by listeners in experimental and real-life listening activities. *Journal of the Audio Engineering Society*, **61**, 425–438. 18

Kim, J., Choi, M., and Lee, S. (2012). In-situ sensor information filtering for pedestrian localization. In *9th International Conference on Ubiquitous Robots and Ambient Intelligence*. 95

Kim, S. and Choi, W. (2010). On the externalizatrion of virtual sound images in headphone reproduction: A wiener filter approach. *J. Acoust. Soc. Am.*, **117**(6), 3657–3665. 48, 114, 241

Kinsler, L. E., Frey, A. R., Coppens, A. B., and Sanders, J. V. (2000). *Architectural Acoustics*, chapter 12, page 336. Wiley. 28

Klasen, T. J., den Bogaert, T. V., and Moonen, M. (2007). Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues. *IEEE Transactions on Signal Processing*, **55**, 1579–1585. 80

Klatt, D. (1980). Software for a cascade/parallel formant synthesizer. *J. Acoust. Soc. Am.*, **67**(3), 971–995. 46

Klensch, H. (1948). Beitrag zur frage der lokalisation des schalles im raum (a contribution to the study of the localization of sound in space). *Pflügers Arch.*, **250**, 492–500. 52

Klumpp, R. G. and Eady, H. R. (1956). Some measurements of interaural time difference thresholds. *J. Acoust. Soc. Am.*, **28**, 859–860. 12

Knapp, C. H. and Carter, G. C. (1976). Generalized correlation method for estimation of time delay. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **24**, 320–327. 83, 84

Köbler, S. and Rosenhall, U. (2002). Horizontal localization and speech intelligibility with bilateral and unilateral hearing aid amplification. *Int. J. Audiol.*, **41**, 395–400. 20

Kock, W. E. (1950). Binaural localization and masking. *J. Acoust. Soc. Am.*, **22**, 61–62. 39

Koehnke, J., Bessing, J., and Breitenbach, J. (2000). The effect of distance on localization of complex stimuli. In *Proceedings of the Midwinter Meeting of the Association for Research in Otolaryngology*. 31

Koenig, W. (1950). Subjective effects in binaural hearing. *J. Acoust. Soc. Am.*, **22**, 61–62. 39

König, G. and Sussman, W. (1955). Zum richtungshören in der median-sagittal-ebene (on directional hearing in the medial-sagittal planes). *Arch Ohren-Nasen-Kehlkopfheilk*, **167**, 303–307. 18, 214

Kopčo, N. (2010). Speech localization in a multitalker mixture. *J. Acoust. Soc. Am.*, **127**, 1450–1457. 21

Kopčo, N., Schoolmaster, M., and Shinn-Cunningham, B. (2004). Learning to judge distance of nearby sounds in reverberant and anechoic environments. In *Proceedings of the Joint Conference of CFA/DAGA*. 30

Krumbacher, G. (1969). Leistungsfähigheit der kopfbezuglichen stereophonie. *Acustica*, **21**, 288–293. 40, 44, 177

Kuhn, G. F. (1977). Model for the interaural time differences in the azimuthal plane. *J. Acoust. Soc. Am.*, **82**, 157–167. 12

Kuhn, G. F. (1987). *Physical acoustics and measurements pertaining to directional hearing*, pages 3–25. New York: Springer-Verlag. 12, 13

Kuk, F., Baekgaard, L., and Ludvigsen, C. (2000). Design considerations in directional microphones. *Hearing Review*, **7**, 58–63. 77

Kuk, F. K. (2000). *Audiology Treatment*, chapter Recent approaches to fitting nonlinear hearing aids, pages 261–289. New York: Thieme Medical Publishers. 60

Kuk, F. K., Kollofski, C., Brown, S., Melum, A., and Rosenthal, A. (1999). Use of a digital hearing aid with directional microphones in school-aged children. *J. Am. Acad. Audiol.*, **10**, 535–548. 75

Kulkarni, A. and Colburn, H. S. (1998). Role of spectral detail in sound-source localization. *Nature*, **396**, 747–749. 37, 47

Kulkarni, A. and Colburn, H. S. (2000). Variability in the characterization of the headphone transfer-function. *J. Acoust. Soc. Am.*, **107**, 1071–1074. 46, 111

Kwon, B., Park, Y., and Park, Y. (2008). Sound source localizationfor robot auditory system using summed gcc method. In *International Conference on Control, Automation and Systems*. 89

Kwon, B., Park, Y., and Park, Y. (2010). Analysis of the gcc-phat technique for multiple sources. In *International Conference on Control, Automation and Systems*. 84

Lamarche, L., Giguere, C., Gueaieb, W., and amd H. Othman, T. A. (2010). Adaptive environment classification system for hearing aids. *J. Acoust. Soc. Am.*, **127**, 3124–3135. 95

Laugesen, S., Jensen, N. S., Maas, P., and Nielsen, C. (2011). Own voice qualities (ovq) in hearing-aid users. there is more than just occlusion. *Int. J. Audiol.*, **50**, 226–236. 59

Lavandier, M. and Culling, J. F. (2010). Prediction of binaural speech intelligibility against noise in rooms. *J. Acoust. Soc. Am.*, **127**, 387–399. 2

Laws, P. (1973). Entfernungshoren und das problem der im-kopf-lokalisiertheit von hörereignissen [auditory distance perception and the problem of "in-head localization" of sound images]. *Acustica*, **29**, 243–259. 36, 45, 47, 137, 242

Lee, L. and Geddes, E. R. (1998). Perception of microphone noise in hearing instruments. *J. Acoust. Soc. Am.*, **104**, 3364–3368. 78

Leeuw, A. R. and Dreschler, W. A. (1991). Advantages of directional hearing aid microphones related to room acoustics. *Audiology*, **30**, 330–344. 78

Lehman, E. A. and Johansson, A. M. (2010). Diffuse reverberation model for efficient image-source simulation of room impulse responses. *IEEE Transactions on Acoustics, Speech, and Language Processing*, **18**, 1429–1439. 251

Levy, E. T. and Butler, R. A. (1978). Stimulus factors which influence the perceived externalization of sound presented through heaphones. *Journal of Auditory Research*, **18**, 41–50. 44

Li, X., Liu, H., and Yang, X. (2011). Sound source localization for mobile robot based on time difference feature and space grid matching. In *International Conference on Intelligent Robots and Systems*. 89

Little, A. D., Mershon, D. H., and Cox, P. H. (1992). Spectral content as a cue to perceived auditory distance. *Perception*, **21**, 405–416. 32

Liu, C., Wheeler, B., O'Brian, D., Bilger, R., C.R.Lansing, and Feng, A. (2000). Localization of multiple sound sources with two microphones. *J. Acoust. Soc. Am.*, **108**, 1888–1905. 83, 87

Lloyd, L. S. (1970). *Music and Sound*. Ayer Publishing. 27

Löllmann, H. W. and Vary, P. (2012). Beamformer for driving binaural speech enhancement. In *International Workshop on Acoustic SIgnal Enhancement*. 80, 84, 86

Looijestijn, J., Diederen, K. M. J., Goetkoop, R., Sommer, I. E. C., Daalman, K., Kahn, R. S., Hoek, H. W., and Blom, J. D. (2013). The auditory dorsal stream plays a crucial role in projecting hallucinated voices into external space. *Schizophr. Res.*, **146**, 314–319. 55

Loomis, J. M. (1990). Active localization of virtual sounds. *J. Acoust. Soc. Am.*, **88**, 1757–1764. 36

Loomis, J. M. (1999). *Mixed Reality: Merging Real and Virtual Worlds*, chapter Auditory Distance Perception in Real, Virtual, and Mixed Environments, pages 201–214. Tokyo: Ohmisha. 53

Loomis, J. M., Klatzky, R. L., Philbeck, J. W., and Golledge, R. G. (1998). Assessing auditory distance perception using perceptually directed action. *Percept. Psychophys.*, **60**, 966–980. 32

Lorenzi, C., Gatehouse, S., and Lever, C. (1999a). Sound localization in noise in hearing-impaired listeners. *J. Acoust. Soc. Am.*, **105**, 3454–3463. 20

Lorenzi, C., Gatehouse, S., and Lever, C. (1999b). Sound localization in noise in normal-hearing listeners. *J. Acoust. Soc. Am.*, **105**, 1810–1820. 14

Lotter, T. and Vary, P. (2006). Dual-channel speech enhancement by superdirective beamforming. *EURASIP Journal on Applied Signal Processing*, **2006**, 1–14. 80

Lurquin, P. and Rafhay, S. (1996). Intelligibility in noise using multimicrophone hearing aids. *Acta Otorhinolaryngol Belg*, **50**, 103–109. 75

Lybarger, S. (1980). *Acoustic Factors Affecting Hearing Performance*, chapter Earmold venting as an acoustic control factor, pages 197–217. University Park Press. 57

Macrae, J. H. and Dillion, H. (1996). An equivalent input noise level criterion for hearing aids. *J. Rehabil. Res. Dev.*, **33**, 355–362. 78

Mahony, R., Hamel, T., and Pflimlin, J.-M. (2008). Nonlinear complementary filters on the special orthogonal group. *IEEE Transactions on Automatic Control*, **53**, 1203–1218. 93

Makous, J. M. and Middlebrooks, J. C. (1990). Two-dimensional sound localization by human listeners. *J. Acoust. Soc. Am.*, **87**, 2188–2200. 14

Markides, A. (1977). *Binaural Hearing*. London: Academic Press. 21

Mershon, D. H., Ballenger, W. L., Little, A. D., McMurtry, P. L., and Buchanan, J. L. (1989). Effects of room reflectance and background noise on perceived auditory distance. *Perception*, **18**, 403–416. 29, 34

Mershon, D. H. and Bowers, J. N. (1979). Absolute and relative cues for the auditory perception of egocentric distance. *Perception*, **8**, 311–322. 29

Mershon, D. H., Desaulniers, D. H., Amerson, T. L. J., and Kiefer, S. A. (1980). Visual capture in auditory distance perception: Proximity image effect reconsidered. *J. Aud. Res.*, **20**, 129–136. 25

Mershon, D. H. and King, E. (1975). Intensity and reverberation as factors in the auditory perception of egocentric distance. *Percept. Psychophys.*, **18**, 409–415. 26, 29, 33

Middlebrooks, J. C. (1999). Individual differences in the external-ear transfer functions reduced by scaling in frequency. *J. Acoust. Soc. Am.*, **106**, 1480–1492. 11

Middlebrooks, J. C. and Green, D. M. (1991). Sound localization by human listeners. *Annu. Rev. Psychol.*, **42**, 135–159. 15, 16

Miller, G. A. (1947). Sensitivity to changes in the intensity of white noise and its relation to masking and loudness. *J. Acoust. Soc. Am.*, **19**, 609–619. 26

Mills, A. W. (1960). Lateralization of high-frequency tones. *J. Acoust. Soc. Am.*, **32**, 132–134. 13, 15, 37, 166

Mills, A. W. (1972). *Foundations of Modern Auditory Theory*, chapter Auditory localization, pages 303–348. New York: Academic Press. 10, 14

Mohrmann, K. (1939). Lautheitskonstanz im entfernungswechsel (constancy of loudness with changes in distance). *Zeitschrift fur Psychologie*, **145**, 146–199. 26

Molino, J. (1973). Perceiving the range of a sound source when the direction is known. *J. Acoust. Soc. Am.*, **53**, 1301–1304. 30

Moore, B. C. J. (1985). Frequency selectivity and temporal resolution in normal and hearing-impaired listeners. *Br. J. Audiol.*, **19**, 189–201. 21

Moore, B. C. J. (2000). Use of a loudness model for hearing aid fitting. iv. fitting hearing aids with multi-channel compression so as to restore “normal” loudness for speech at different levels. *Br. J. Audiol.*, **34**, 165–177. 164

Moore, B. C. J. (2003). *An Introduction to the Psychology of Hearing*. Emerald Group Publishing Ltd. 17

Moore, B. C. J., Stainsby, T. H., Alcántara, J. I., and Kühnel, V. (2004). The effect on speech intelligibility of varying compression constants in a digital hearing aid. *Int. J. Audiol.*, **43**, 399–409. 61

Morency, L.-P., Sidner, C., Lee, C., and Darrell, T. (2005). Contextual recognition of head gestures. In *Proceedings of the 7th International Conference on Multimodal Interfaces*. 16

Morikawa, D., Toyoda, Y., and Hirahara, T. (2013). Head movement during horizontal and median sound localization experiments in which head-rotation is allowed. In *Proceedings of Meetings on Acoustics*, volume 19. 18

Musa-Shufani, S., Walger, M., von Wedel, H., and Meister, H. (2006). Influence of dynamic compression on directional hearing in the horizontal plane. *Ear Hear.*, **27**, 279–285. 62

Nielsen, S. H. (1993). Auditory distance perception in different rooms. *Journal of the Audio Engineering Society*, **41**, 755–770. 29

Noble, W. and Byrne, D. (1991). Auditory localization under conditions of unilateral iftting in different hearing aid systems. *Br. J. Audiol.*, **25**, 237–250. 22

Noble, W., Byrne, D., and Lepage, B. (1994). Effects on sound localization of configuration and type of hearing impairment. *J. Acoust. Soc. Am.*, **95**, 992–1005. 19

Noble, W. and Gatehouse, S. (2006). Effects of bilateral versus unilateral hearing aid fitting on abilities measured by the speech, spatial, and qualities of hearing scale (ssq). *International Journal of Audiology*, **45**, 172–181. 1, 2, 37, 51, 143, 161, 162

Noble, W., Sinclair, S., and Byrne, D. (1998). Improvement in aided sound localization with open earmolds: observations in people with high frequency hearing loss. *J. Am. Acad. Audiol.*, **9**, 25–34. 22

Noble, W., Ter-Horst, K., and Byrne, D. (1995). Disabilities and handicaps associated with impaired auditory localization. *J. Am. Acad. Audiol.*, **6**, 129–140. 21

Ohl, B., Laugesen, S., Buchholz, J., and Dau, T. (2009). Externalization versus internalization of sound in normal-hearing and hearing-impaired listeners. Master's thesis, Department of Electrical Engineering, Technical University of Denmark. 2, 29, 37, 51, 54, 55, 110

Orton, J. F. and Preves, D. A. (1979). Localization ability as a function of hearing aid microphone placement. *Hearing Instruments*, **30**, 18–21. 21, 22

Palmer, C., Bentler, R., and Mueller, R. G. (2006). Evaluation of a second-order directional microphone hearing aid:ii. self report outcomes. *J. Am. Acad. Audiol.*, **17**, 190–201. 76

Pavlidi, D., Puigt, M., Griffin, A., and Mouchtaris, A. (2012). Real-time multiple sound source localization using a circular microphone array based on single-source confidence measures. In *Proceedings of the International Conference on Acoustcis, Speech, and Signal Processing*. 89

Perrett, S. and Noble, W. (1997). The contribution of head motion cues to localization of low-pass noise. *Perception & Psychophysics*, **59**(5), 1018–1026. 17

Perrott, D. R. and Saberi, K. (1990). Minimum audible angle thresholds for sources varying in both elevation and azimuth. *J. Acoust. Soc. Am.*, **87**, 1728–1731. 14

Petersen, J. (1990). Estimation of loudness and apparent distance of pure tones in a free-field. *Acustica*, **70**, 61–65. 26

Petersen, P. M., Durlach, N. I., Rabinowitz, W. M., and Zurek, P. M. (1987). Multimicrophone adaptive beamforming for interference reduction in hearing aids. *J. Rehabil. Res. Dev.*, **24**, 103–110. 65

Pickles, J. O. (1988). *An Introduction to the Physiology of Hearing*. London: Academic Press. 12

Plenge, G. (1974a). On the differences between localization and lateralization. *J. Acoust. Soc. Am.*, **56**(3), 944–951. 12, 36, 40, 41

Plenge, G. (1974b). Über das problem der im-kopf lokalisation (on the problem of in-the-head localization). *Acustica*, **26**(5), 241–252. 40

Pörschmann, C. (2007). 3-d audio in mobile communication devices: Methods for modbile head-tracking. *Journal of Virtual Reality and Broadcasting*, **4**, N/A. 95

Premerlani, W. and Bizard, P. (2009). Direction cosine matrix imu theory. White paper. 93

Preves, D. A., Sammeth, C. A., and Wynne, M. K. (1999). Field trial evaluations of a switched directional/omnidirectional in-the-ear hearing instrument. *J. Am. Acad. Audiol.*, **10**, 273–284. 77

Purkyne (1859). Cited by Eiselt in the Vierteljahresschrift für praktische Heilkunde Med. Fak., Prag 17 (1860). 38

Rakerd, B. and Hartmann, W. M. (1986). Localization sound in rooms, iii: onset and duration effects. *J. Acoust. Soc. Am.*, **80**, 1695–1706. 43

Rakerd, B. and Hartmann, W. M. (2010). Localization of sound in rooms v. binaural coherence and human sensitivity to interaural time differences in noise. *J. Acoust. Soc. Am.*, **128**(5), 3052–3063. 14

Raspaud, M., Viste, H., and Evangelista, G. (2010). Binaural source localization by joint estimation of ild and itd. *IEEE transactions on Acoustics, Speech and Language Processing*, **18**, 68–77. 88

Rayleigh, L. (1894). *The theory of sound*. Macmillan and Co. 13

Rayleigh, L. J. W. S. (1907). On our perception of sound direction. *Philos. Mag.*, **13**, 214–232. 14

Recommendation, I. (2003). Recommendation itu-r bs.1534-1 method for subjective assessment of intermediate quality level of coding systems. 116

Reed, C. M., Braida, L. D., and Zurek, P. M. (2009). Review of literature on temporal resolution in listeners with cochlear hearing impairment: a critical assessment of the role of suprathreshold deficits. *Trends in Amplification*, **13**, 4–43. 178

Regulations, B. (2003). Building bulletin 93. 114

Reichardt, W. and Haustein, B. G. (1968). Zur ursache des effekts der 'im-kopf-localisation'. *Hochfrequenztech. Elektroakust*, **77**, 183–189. 39, 41

Richards, D. G. and Wiley, R. H. (1980). Reverberations and amplitude fluctuations in the propogations of sound in a forest. implications for animal communication. *American Naturalist*, **115**, 381–399. 28

Ricketts, T. and Dahr, S. (1999). Aided benefit across directional and omni-directional hearing aid microphones for behind-the-ear hearing aids. *J. Am. Acad. Audiol.*, **10**, 180–189. 74

Ricketts, T., Henry, P., and Gnewikow, D. (2003). Full time directional versus user selectable microphone modes in hearing aids. *Ear Hear.*, **24**, 424–439. 75

Ricketts, T. and Mueller, H. G. (1999). Making sense of directional microphone hearing aids. *American Journal of Audiology*, **8**, 117–127. 63

Ricketts, T. A. and Hornsby, B. W. Y. (2003). Distance and reverberation effects on directional benefit. *Ear Hear*, **24**, 472–483. 75

Riesz, R. R. (1933). The relationship between loudness and the minimum perceptible increment of intensity. *J. Acoust. Soc. Am.*, **4**, 211–216. 26

Robinson, K. and Gatehouse, S. (1995). Changes in intensity discrimination following monaural long-term use of a hearing aid. *J. Acoust. Soc. Am.*, **97**, 1183–90. 20

Rohdenburg, T., Goetze, S., Hohmann, V., Kammeyer, K.-D., and Kollmeier, B. (2008a). Combined source tracking and noise reduction for application in hearing aids. In *ITG-Fachtagung Sprachkommunikation*. 87

Rohdenburg, T., Goetze, S., Hohmann, V., Kammeyer, K.-D., and Kollmeier, B. (2008b). Objective perceptual quality assessment for self-steering binaural hearing aid microphone arrays. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*. 83, 88, 90

Rothbucher, M., Durkovic, M., Habigt, T., Shen, H., and Diepold, K. (2012). Hrtf-based localization and separation of multiple sound sources. In *IEEE International Symposium on Robot and Human Interactive Communication*. 89

Roy, O. and Vetterli, M. (2006). Rate-constrained beamforming for collaborating hearing aids. In *IEEE International Symposium on Information Theory*. 81

Sakamoto, N. (1976). On 'out-of-head localization' in headphone listening. *Journal of the Audio Engineering Society*, **24**, 710–716. 42

Samuels, H. R. (2011). Hearing instrument controller. accel one. 97

Schaefer, K. L. (1890). Zur interauralen lokalisation diotischer wahrnehmungen [on the interaural localization of diotic perceptions. *Z. Psychol. u. Physiol. Sinnesorg*, **1**, 300–309. 38

Schirmer, W. (1966). Zur deutung der übertragungsfehler bei kopfbezuglicher stereophonie. *Acustica*, **17**, 228=233. 39, 40, 177

Schwartz, A. H. and Shinn-Cunningham, B. G. (2013). Effects of dynamic range compression on spatial selective auditory attention in normal-hearing listeners. *J. Acoust. Soc. Am.*, **133**, 2329–2339. 62

Shaw, E. A. G. (1974). Transformation of sound pressure level from the free field to the eardrum in the horizontal plane. *J. Acoust. Soc. Am.*, **56**, 1848–1861. 13

Shaw, E. A. G. and Vaillancourt, M. M. (1985). Transformation of sound-pressure level from the free field to the eardrum presented in numerical form. *J. Acoust. Soc. Am.*, **78**, 1120–1123. 13

Sheeline, C. W. (1983). *An investiation of the effects of direct and reverberant signal interaction on auditory distance perception*. PhD thesis, Stanford University. 30

Shinn-Cunningham, B. G. (2000). Distance cues for virtual auditory space. In *Proceedings of the First IEEE Pacific-Rim Conference on Multimedia*. 29

Shinn-Cunningham, B. G., Durlach, N. I., and Held, R. M. (1998a). Adapting to supernormal auditory localization cues i. bias and resolution. *J. Acoust. Soc. Am.*, **103**, 3656–3666. 23

Shinn-Cunningham, B. G., Durlach, N. I., and Held, R. M. (1998b). Adapting to supernormal auditory localization cues ii. constrainsts on adaptation of mean response. *J. Acoust. Soc. Am.*, **103**, 3667–3676. 23

Shinn-Cunningham, B. G., Santarelli, S., and Kopčo, N. (1999). Tori of confusion. binaural localization cues for sources within reach of the listener. *J. Acoust. Soc. Am.*, **107**, 1627–1636. 16

Silverman, H. F. and Kirtman, S. E. (1992). A two-stage algorithm for determining talker location from linear microphone array data. *Computer Speech and Language*, **6**, 129–152. 86

Simpson, W. E. and Stanton, L. D. (1973). Head movement does not facilitiate perception of the distance of a source of sound. *Am. J. Psychol.*, **86**, 151–159. 25, 26

Siqueira, M. B. and Alwan, A. (2000). Steady-state analysis of continuous adaptation in acoustic feedback reduction systems for hearing-aids. *IEEE Transactions on Speech and Audio Processing*, **8**, 443–453. 58

Soede, W., Berkhout, A. J., and Bilsen, F. A. (1993a). Development of a directional hearing instrument based on array technology. *J. Acoust. Soc. Am.*, **94**, 785–798. 80

Soede, W., Bilsen, F. A., and Berkhout, A. J. (1993b). Assessment of a directional microphone array for hearing-impaired listeners. *J. Acoust. Soc. Am.*, **94**, 799–808. 80

Sone, T., Ebata, M., and Nimura, T. (1968). On the difference between localization and lateralization. In *Proceedings of the 6th International Congress on Acoustics*. 40

Srinivasan, S. (2008). Low-bandwidth binaural beamforming. *Electronics Letters*, **44**, 1292–1293. 81

Srinivasan, S. and den Brinker, A. C. (2009). Analyzing rate-constrained beamforming schemes in wireless binaural hearing aids. In *17th European signal Processing Conference (EUSIPCO)*. 81

Stacey, P. C. and Summerfield, A. Q. (2007). Effectiveness of computer-based auditory training in improving the perception of noise-vocoded speech. *J. Acoust. Soc. Am.*, **121**(5), 2923–2935. 103, 115, 130, 164, 169, 190, 228, 231, 259

Stephenne, A. and Champagne, B. (1995). Cepstral prefiltering for time delay estimation in reverberant environments. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. 86

Stevens, S. S. and Guirao, M. (1962). Loudness, reciptocality and partition scales. *J. Acoust. Soc. Am.*, **34**, 1466–1471. 26

Strobel, N. and Rabenstein, R. (1999). Classification of time delay estimates for robust speaker localization. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. 86

Strybel, T. Z. and Perrott, D. R. (1984). Discrimination of relative distance in auditory modality: The success and failure of the loudness discrimination hypothesis. *J. Acoust. Soc. Am.*, **76**, 318–320. 26

Surr, R. K., Walden, B., Cord, M. T., and Olsen, L. (2002). Influence of environmental factors on hearing aid microphone preference. *J. Am. Acad. Audiol.*, **13**, 208–322. 76

Tessendorf, B., Bulling, A., Roggen, D., Stiefmeier, T., Feilner, M., Derleth, P., and Tröster, G. (2011a). *Pervasive Computing*, chapter Recognition of Hearing Needs from Body and Eye Movements to Improve Hearing Instruments, pages 314–331. Springer Berlin Heidelberg. 96

Tessendorf, B., Bulling, A., Roggen, D., Stiefmeier, T., Tröster, G., Feilner, M., and Derleth, P. (2010). Towards multi-modal context recognition for hearing instruments. In *International Symposium on Wearable Computers*. 95

Tessendorf, B., Debevc, M., Derleth, P., Feilner, M., Gravenhorst, F., Roggen, D., Stiefmeier, T., and Tröster, G. (2013). Design of a multimodal hearing system. *Computer Science and Information Systems*, **10**, 483–501. 2, 3, 96, 213

Tessendorf, B., Derleth, P., amd F. Gravenhorst, M. F., Kettner, A., Roggen, D., Stiefmeier, T., and Tröster, G. (2012). Ear-worn reference data collection and annotation for multimodal context-aware hearing instruments. In *IEEE International Conference on Engineering in Medicine and Biology*. 96

Tessendorf, B., Kettner, A., Roggen, D., Stiefmeier, T., Tröster, G., Derleth, P., and Feilner, M. (2011b). Indentification of relevant multimodal cues to enhance context-aware hearing instruments. In *Proceedings of the 6th International Conference on Body Area Networks*. 96

Teutsch, H. and Elko, G. W. (2001). First-and second-order adaptive differential microphone arrays. In *Proceedings of the 7th International Workshop on Acoustic Echo and Noise Control (IWAENC)*. 64, 66, 67, 82

Thompson, S. C. (1999). Dual microphones or directional-plus-omni: Which is best? *Hearing Review*, **6**, 31–35. 77

Thompson, S. P. (1877). On binaural audition, part i. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, **4**, 274–276. 38

Thompson, S. P. (1878). On binaural audition, part ii. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, **6**, 383–391. 38

Thompson, S. P. (1892). On the function of the two ears in the perception of space. *Philosphical Magazine*, **13**, 320–334. 25

Thurlow, W. R., Mangels, J. W., and Runge, P. S. (1967). Head movements during sound localization. *J. Acoust. Soc. Am.*, **42**, 483–493. 17

Thurlow, W. R. and Runge, P. S. (1967). Effect of induced head movements on localization of direction of sounds. *J. Acoust. Soc. Am.*, **42**, 480–488. 17

Toole, F. E. (1969). In-head localization of acoustic images. *J. Acoust. Soc. Am.*, **48**(4), 943–949. 41, 42, 44, 53

Turk, R. (1986). A clinical comparison between behind-the-ear and in-the-ear hearing aids. *Audiology Acoustics*, **25**, 76–86. 22

Urbanschitsch, V. (1889). Zur lehre von den schallempfindungen [on the science of sound sensations]. *Pfl ́ugers Archiv.*, **24**, 574–595. 38

Usagawa, T., Saho, A., Imamura, K., and Chisaki, Y. (2011). A solution of front-back confusion within binaural processing by an estimation method of sound source direction on sagittal coordinate. In *IEEE Region 10 Conference*. 83, 89

Usman, M., Keyrouz, F., and Diepold, K. (2008). Real time humanoid sound source localization and tracking in a highly reverberant environment. In *Proceedings of the IEEE International Conference on Speech Processing*. 89

Valente, M., Fabry, D. A., and Potts, L. G. (1995). Recognition of speech in noise with hearing aids using dual microphones. *J. Am. Acad. Audiol.*, **6**, 440–449. 74, 75

Veen, B. D. V. and Buckley, K. M. (1988). Beamforming: A versatile approach to spatial filtering. IEEE ASSP magazine. 63

von Békésy, G. (1938). Über die entstehung der entfernungssempfindung beim hören (on the origin of the sensation of distance in hearing). *Akustische Zeitschrift*, **3**, 21–31. 29

von Békésy, G. (1949). The moon illusion and similar auditory phenomena. *Am. J. Psychol.*, **62**, 540–552. 26, 33

Walden, B. E., Surr, R. K., and Cord, M. T. (2003). Real-world performance of directional microphone hearing aids. *Hearing Journal*, **56**, 40–46. 75

Walden, B. E., Surr, R. K., Cord, M. T., and Dyrlund, O. (2004). Predicting hearing aid microphone preference in everyday listening. *J. Am. Acad. Audiol.*, **15**, 365–394. 75, 76

Walden, B. E., Surr, R. K., Cord, M. T., Edward, B., and Olsen, L. (2000). Comparison of benefits provided by different hearing aid technology. *J. Am. Acad. Audiol.*, **11**, 540–560. 76

Walden, R. H. (1999). Analog-to-digital converter survey and analysis. *IEEE Journal on Selected Areas in Communications*, **17**, 539–550. 76, 81

Walker, G. and Dillon, H. (1982). Compression in hearing aids: An analysis, a review and some recommendations. Technical report, National Acoustic Laboritories Report No. 90. 60

Wallace, J. M. and Hobbs, P. V. (2006). *Atmospheric Science: An introductory survey*. Academic Press. 92

Wallach, H. (1938). Über die wahrnehmung der schallrichtung [on the perception of the direction of sound]. *Psychol. Forsch.*, **22**, 238–266. 16

Wallach, H. (1940). The role of head movements and vestibular and visual cues in sound localization. *J. Exp. Psychol.*, **27**, 229–368. 15, 16

Wanrooij, M. M. V. and Opstal, A. J. V. (2005). Relearning sound localization with a new ear. *The Journal of Neuroscience*, **25**, 5413–5424. 137

Warren, R. M. (1958). A basis for judgments of sensory intensity. *Am. J. Psychol.*, **71**, 675–687. 26

Warren, R. M. (1973). Anomalous loudness function for speech. *J. Acoust. Soc. Am.*, **54**, 390–396. 33

Warren, R. M. (1981). Measurement of sensory intensity. *Behavioural and Brain Sciences*, **4**, 175–223. 27

Weber, E. H. (1848). On the circumstances under which one is led to refer sensations to external objects. In *Proceedings of the Royal Saxon Society for Science in Leipzig*, volume 2, pages 226–237. 38

Welker, D. P., Greenberg, J. E., Desloge, J. G., and Zurek, P. M. (1997). Microphone-array hearing aids with binaural output - part iii: A two-microphone adaptive system. *IEEE Transactions on Speech and Audio Processing*, **5**, 543–551. 79

Wersényi, G. (2009). Effect of emulated head-tracking for reducing localization errors in virtual audio simulation. *IEEE Transactions on Audio, Speech and Language Processing*, **17**, 247–252. 53

Westerman, S. and Toepholm, J. (1985). Comparing btes and ites for localizing speech. *Hearing Instruments*, **36**, 20–24. 22

Wettschureck, R., Plenge, G., and Lehringer, F. (1973). Entfernungswahrnehmung beim natürlichen hören sowie bei kopfbezogener stereophonie. *Acustica*, **29**, 260–272. 29

Wiggins, I. M. and Seeber, B. U. (2011). Dynamic-range compression affects the lateral position of sounds. *J. Acoust. Soc. Am.*, **130**, 3939–3953. 62, 126, 144, 165, 176, 177

Wiggins, I. M. and Seeber, B. U. (2012). Effects of dynamic-range compression on the spatial attributes of sounds in normal-hearing listeners. *Ear Hear.*, **33**, 399–410. 51, 62, 144

Wightman, F. L. and Kistler, D. J. (1989a). Headphone simulation of free-field listening. i: Stimulus synthesis. *J. Acoust. Soc. Am.*, **85**(2), 858–867. 36, 45, 48, 110, 125

Wightman, F. L. and Kistler, D. J. (1989b). Headphone simulation of free-field listening. ii: Psychophysical validation. *J. Acoust. Soc. Am.*, **85**(2), 868–878. 36, 46, 110

Wightman, F. L. and Kistler, D. J. (1992). The dominant role of low-frequency interaural time differences in sound localization. *J. Acoust. Soc. Am.*, **91**, 1648–1661. 14

Wightman, F. L. and Kistler, D. J. (1997). Monaural sound localization revisitied. *J. Acoust. Soc. Am.*, **101**, 1050–1063. 33

Wongwirat, O. and Chaiyarat, C. (2010). A position tracking experiment of mobile robot with inertial measurement unit (imu). In *International Conference on Control, AUtomation and Systmes*. 94

Woods, W. S. and Trine, T. D. (2004). Limitations of theoretical benefit form an adaptive directional system in reverberant environments. *Acoustics Research Letters Online*, **5**, 153–157. 75

Woodworth, R. S. (1938). *Experimental psychology*. New York: Holt. 10

Wouters, J., Litiere, L., and van Wieringen, A. (1999). Speech intelligibility in noisy environments with one- and two-microphone hearing aids. *Audiology*, **38**, 91–98. 75

Xiang, J., McKinney, M. F., Fitz, K., and Zhang, T. (2010). Evaluation of sound classification algorithms for hearing aid applications. In *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*. 189

Young, P. T. (1931). The role of head movements in auditory localization. *J. Exp. Psychol.*, **14**, 95–124. 52

Zahorik, P. (1997). Scaling perceived distance of virtual sound sources. *J. Acoust. Soc. Am.*, **101**, 3105–3106. 32

Zahorik, P. (2001). Estimating sound source distance with and without vision. *Optometry and Vision Science*, **78**, 270–275. 25

Zahorik, P. (2002a). Assessing auditory distance perception using virtual acoustics. *J. Acoust. Soc. Am.*, **111**, 1832–1846. 30, 33, 34, 102, 108

Zahorik, P. (2002b). Direct-to-reverberant energy ration sensitivity. *J. Acoust. Soc. Am.*, **112**, 2110–2117. 33

Zahorik, P., Brungart, D. S., and Bronkhorst, A. W. (2005). Auditory distance perception in humans: A summary of past and present research. *Acta Acustica United with Acustica*, **91**, 409–420. 28, 30, 33

Zahorik, P., Kistler, D. J., and Wightman, F. L. (1994). Sound localization in varying virtual acoustic environments. In *Proceedings of the Second International Conference on Auditory Display*. 33

Zahorik, P. and Wightmann, F. L. (2001). Loudness constancy with varying sound source distance. *Nat. Neurosci.*, **4**, 78–83. 26, 27

Zhang, C., Florencio, D., and Zhang, Z. (2008). Why does phat work so well in low noise, reverberative environments? In *IEEE International Conference on Acoustics, Speech and Signal Processing*. 84

Zhang, P. X. and Hartmann, W. M. (2010). On the ability of human listeners to distinguish between front and back. *Hearing Research*, **260**, 30–46. 16

Zurek, P. M. (1993). *Acoustical Factors Affecting Hearing Aid Performance*, chapter Binaural advantages and directional effects in speech intelligibility, pages 255–276. Boston, MA: Allyn and Bacon. 19, 78

Zwiers, M. P., Opstal, A. J. V., and Paige, G. D. (2003). Plasticity in human sound localization induced by compressed spatial vision. *Nat. Neurosci.*, **6**, 175–181. 23

Zwislocki, J. and Feldman, R. S. (1956). Just noticeable differences in dichotic phase. *J. Acoust. Soc. Am.*, **28**, 860–864. 12