Development of the *AnimalSeek* Method to Evaluate the Localisation Ability of Children under Five

Damon Andrew McCartney BEng(Hons) MRes

Thesis submitted to The University of Nottingham

for the degree of Doctor of Philosophy

January 2013

Abstract

This thesis describes the development of a novel game-like method, the *AnimalSeek* method, which can be used, along with motion tracking technology, to measure the localisation ability of a child under five years of age. For the game-like task to be successful, a high number of responses (in particular correct head turn responses) was required. Previous studies, although not all looking at localisation ability, have used many different techniques to obtain the maximum number of responses from a child. The children were engaged inside a custom-built environment inside an anechoic chamber. Three large video screens onto which backgrounds and animated characters were projected and manipulated and used to engage the child in the game-like task. Behind the video screens were loudspeakers from which the auditory stimulus where presented. A correct response to the auditory stimulus i.e. a head, hand or eye movement towards the target speaker was rewarded with a animated character presented on the screens (incorrect responses were presented with a static character). The location of the reward in relation to the auditory stimulus was a point of interest and was investigated to see how it affected the number of responses. The method shows it was possible to engage the child with the visual environment and obtain responses, however, the results showed generally fewer head turn responses responses than expected, especially in the younger age groups. Motion tracking technology was used to measure the localisation ability of the children, as well as measuring the responses, the motion tracking data was used and programs developed which could automatically classify the responses the children made to the sounds. The thesis has shown that it is possible to devise a new method which can be used to engage the child in the task and extract and classify their responses to auditory stimuli in order to measure their localisation ability.

Acknowledgements

Contents

List of Abbreviations

6DOF	6 Degrees of Freedom
AC	Auditory Cortex
AEC	Anechoic Chamber
AM	Amplitude Modulation
ANN	Artificial Neural Network
Az	Azimuth
BA	Behavioural Audiometry
BM	Basilar Membrane
CN	Cochlear Nucleus
COR	Conditioned Reflex Audiometry
DCN	Dorsal Cochlear Nucleus
DTW	Dynamic Time Warping
DVD	Digital Versatile Disc

El	Elevation
FPS	Frames Per Second
GELP	God's Eye Location Pointing
HD	High Definition
IC	Inferior Colliculus
ILD	Interaural Level Difference
IP	Internet Protocol
ITD	Interaural Time Difference
JND	Just-notable difference
LAN	Local Area Network
LM	Levenberg Marquardt
LSO	Lateral Superior Olive
M	Mean
MAA	Minimal Audible Angle
MGB	Medial Geniculate Body
MSO	Medial Superior Olive
NLL	Nucleus of Lateral Lemniscus
OPP	Oberserver-based Psychoacoustic Procedure
OSC	Open Sound Control

RMS	Root-Mean Square
Ro	Rotation
SD	Standard Deviation
SOC	Superior Olivary Complex
SOFE	Simulated Open-Field Environment
SPL	Sound Pressure Level
SVM	Support Vector Machine
UDP	User Datagram Protocol
VE	Visual Environment
VRA	Visual Reinforcement Audiometry

List of Figures

List of Tables

Chapter 1

Introduction and Motivation

1.1 The Auditory System

The auditory system in humans is a complex structure consisting of many parts that are responsible not only for converting sound waves into brain waves but also the processing and interpretation of these sounds.

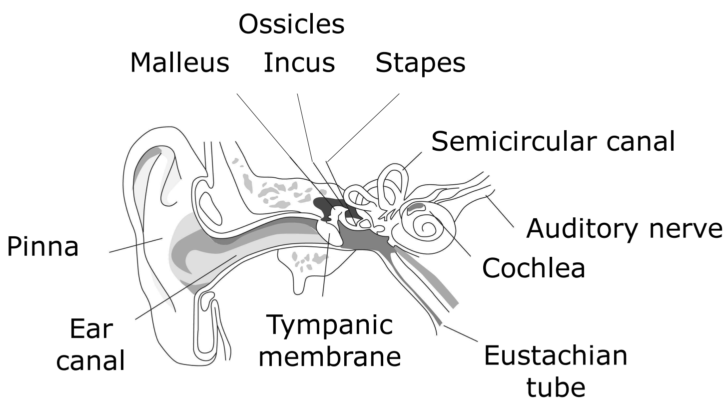


Figure 1.1: General structures of the outer and middle ear.

1.1.1 Peripheral Auditory System - Outer and Middle Ear

Sounds first arrive at the outer ear (pinna) where they are funneled into the ear canal. At the far end of the ear canal is the tympanic membrane (ear drum) that is connected to three small bones (ossicles: malleus, incus and staples) which are inside the middle ear. The lateral movement of the tympanic membrane in response to sound waves moves these three bones which perform an impedance transformation converting the mechanical movement of the tympanic membrane due to sound waves in the air (low impedance) into movement of the oval window. The movement of the oval window then causes pressure waves inside the liquid in the cochlea (high impedance) that cause the basilar membrane (BM), which lies insides the cochlea, to move. The movement of the BM excites hair cells, which in turn transmit neural signals to the brain stem via the auditory nerve.

1.1.2 Ascending Auditory Pathway

The auditory nerve passes signals along to the first nuclei, the cochlear nucleus complex. At this point the signal is split into a number of ascending pathways. The cochlear nucleus (CN) has connections to the superior olivary complex (SOC), which in turn projects up to the nuclei of lateral lemniscus (NLL) and the inferior colliculus (IC). IC then projects up to the final stage before auditory cortex (AC), the medial geniculate body (MGB). These nuclei are responsible for processing and encoding sound information.

1.2 Binaural Hearing

The description so far has been given with respect to listening to sound with one ear; humans (along with most other species) have two ears. The reasons we posses and use two ears are numerous. This section will discuss the binaural system and how it can help us in communication and navigation.

1.2.1 Binaural Cues

Interaural Cues

Figure 1.2 shows an overhead view of a listener and a speaker placed to their right. If this speaker speaks, we can see that sound will arrive at the listener's right ear before the listener's left. This is due to the propagation of sound only having a finite speed and results in a time difference between the two ears, this is referred to as the interaural time difference (ITD). The human auditory system is very sensitive and is capable of detecting an ITD as short as 10μs (700μs being the physical maximum i.e. the time taken for a sound to propagate through the air from the left to the right ear, typical human head radius of 8.75cm). The sound will also be smaller in amplitude (intensity) by the time it reaches the left ear, which is caused by the head attenuating the sound, causing an acoustic shadow at the left ear. This level difference is called the interaural level difference (ILD). These physical cues were first proposed at the beginning of the 20th century by Lord Rayleigh [1]. The use of these cues in sound localisation are called the Duplex Theory that states that ILDs are dominant at high frequencies since their wavelength is small in comparison to an adult human head and therefore is diffracted less by the head. At high frequencies the wavelength of the sound compared to the head is small, this results in large differences in amplitude between the two ears (as much as 35dB for a 10kHz tone [2]). At low frequencies (below around 1kHz) the wavelengths of sound are several times larger than the diameter of the human head, this makes ITDs more dominant due to the absence of substantial ILDs. The Duplex theory holds true for pure tones but not complex sounds. Complex or more real world stimuli have a combination of both high and low frequency energy. It has been shown that the perceived direction of a complex stimulus is determined by their low frequency ITD, when presented alone in silence [3]. When background noise is present, it was found that subjects always use the most optimal cue available to them, be that the ITD or ILD [4]. The superior olive is critical for sound localisation [5]. It is believed that the lateral superior olive complex (LSO, part of the SOC) is the area where the computation of high frequencies takes place, suggesting it is responsible for the processing of ILDs [6]. The medial superior olivary complex (MSO, part of the SOC) however, has a large proportion of cells which respond only to low frequencies, suggesting that it has a role in the processing of ITDs [7].

Spectral Cues

Besides ITDs and ILDs, spectral cues are also available to us. Spectral cues are caused by the filtering effects of the pinna (outer ear) as well as the listener's torso and shoulders. Spectral cues are useful as they allow the ambiguity of front/back confusion to be resolved as well as helping to locate sounds along the vertical plane (medial plane) [8]. Spectral cues are processed in the dorsal cochlear nucleus (DCN), that is part of the cochlear nucleus [9].

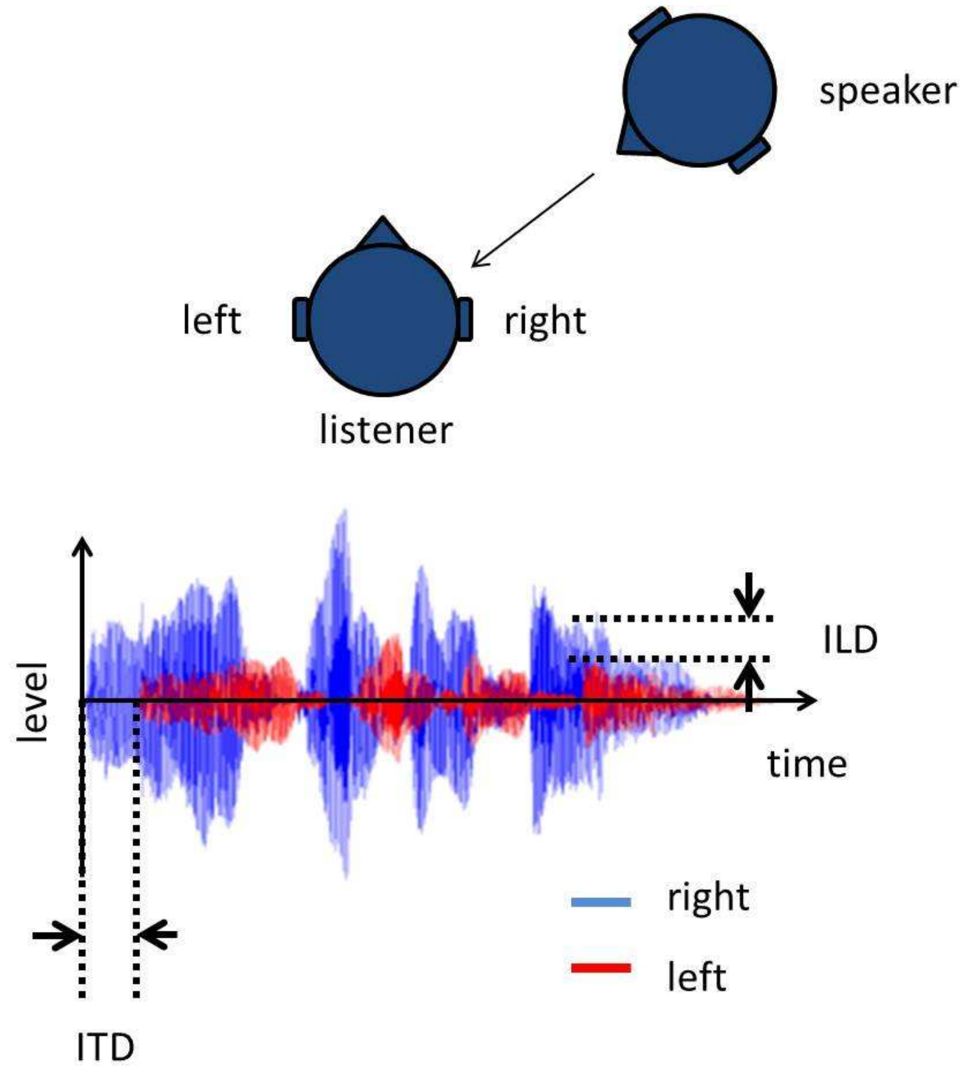


Figure 1.2: **If a speaker is to the right of the listener, the voice will reach the listener's right ear before the listener's left ear. This time delay is known as the interaural time difference (ITD). The head will also cause an acoustic 'shadow' on the left ear. This is caused by the head attenuating the speech. This drop in level between the two ears is called the interaural level difference (ILD). Combinations of these cues as well as spectral properties of an auditory stimulus enable us to localise sounds.**

1.3 Spatial Hearing In Normal Hearing Adults

1.3.1 Sensitivity of the Binaural System

Delivering stimuli over headphones allows the experimenter to vary the sound presented at each ear. The smallest changes (just-noticeable difference, JND) detected in listeners for ILDs is around 0.5-1dB [10]. ILD-JNDs are almost independent of frequency (0.2kHz to 10kHz) although there is a slight increase of the JND at 1kHz [11]. ITD-JNDs are frequency dependent, the JND of ITDs can be as low as 10μs for sinusoidal tones up to 1.5kHz. Above 1.5kHz, measurement of ITD-JNDs is impossible due to the auditory system not developing this ability at these frequencies [12]. ITD-JNDs slightly increase to around 15μs for non-sinusodial stimuli (noise band, tones and clicks) with frequencies between 0.5kHz and 1kHz [13]. Interestingly, although no ITD-JNDs can be measured for sinusoidal tones above 1.5kHz, subjects can detect ITDs for noise stimuli which contain energy above 1.5kHz. It is thought this ability is due to amplitude modulations (AM) in the envelopes of the stimuli [14, 15].

1.3.2 Horizontal Sound Localisation in Adults

Azimuthal Spatial Acuity in Adults

Spatial acuity can be measured by the minimal audible angle (MAA). This simple method involves the use of two spatially separated sound sources placed along the horizontal (azimuth) or vertical (elevation) plane. The subject is asked to discriminate between the two sources. The angle separating the two sources is then reduced until the subject can no longer hear directional changes between them. If the reference source lies ahead of the listener (0° azimuth), for frequencies below 1kHz the MAA is around 1° (at one meter) for sinusoid tones [16]. MAAs are largest at frequencies between 1.5kHz and 1.8kHz. As reported earlier, the Duplex theory states that we use ITDs at low frequencies and ILDs at high. However, ambiguity occurs in the middle, as was shown in headphones studies by Mills [17]. As the reference source moved from the midline, MAAs rise considerably and when the reference was placed at 60° or greater along on the azimuth, the MAAs become around seven times larger. When the frequency of the stimulus is greater than 1.8kHz, the MAAs are unmeasurable due to their wavelength [18]. Although MAAs are useful at demonstrating the spatially acuity of the auditory system, they are in fact not true measures of localisation ability, i.e. the exact pinpointing of the sound in space. The reason for this is that discrimination of spatial separated sounds can be achieved using a single acoustic parameter, and does not require the computation of an actual spatial image by the auditory system [19].

Absolute Sound Localisation in Adults

To fully evaluate localisation ability, methods are required in which the sound can only be truly located by using a combination of all the binaural cues available to the listener. The easiest and best way to do this is to play a sound in an environment with no acoustic reflections (free-field) and ask the listener, in some way, to indicate where they hear the sound. Methods of indication are numerous, this can be as simple as asking the subject to verbally report where the sound was [20, 21, 22]. To create truly anechoic conditions, Stevens et al. [20] placed their subjects on the roof of a building in order to eliminate the possibilities of reflections from nearby structures. A loudspeaker was placed approximately four meters away from the participant and moved in 15° increments along the azimuth. The participants were asked to report the location of the loudspeaker. The study found that participants would make front/back confusions on pure tone stimuli. The error was at its largest when the tone was around 3kHz. The root mean square (RMS) error for 'clicks' was reported as 8° and 5.6° for a 'hiss'. Verbally reporting the stimulus position in a localisation test has drawbacks. Firstly, both participant and experimenter errors can occur. The method is also slow, with some methods only collecting two to three locations per minute [22]. Similar problems exist with methods which ask the participants to draw the location of a stimulus on paper [23]. In recent times, technology has been utilized to try to overcome the problems of verbal reporting and other transcribing methods. A study by Gilkey et al. [24] used a system referred to as 'God's eye location pointing' (GELP). Subjects' were asked to place an electromagnetic sensor on a plastic sphere in reference to where they thought the location of the sound was. The subjects' heads were restrained and the average errors in localisation for broadband stimuli was found to be around 18.2°. Interestingly, when the same experiment was conducted but the subject was asked to report the location verbally, an error of 9° was found. These results showed a conflict with results found by Wightman and Kistler [22], who reported much larger errors (20°). It was hypothesised that acoustic reflections from the plastic sphere cause discrepancies in the findings. It was also reported by the authors, that the use of such a method was 'two to eight times faster' than either verbal reporting or head pointing [24]. The method also allows the user to point to any location in the room and it can be used to assess localisation ability across 360°. Djelani et al. [25] questioned methods such as GELP and in particular raised concerns regarding errors in judgments and the need for training on the task. Less extrinsic and more accurate ways of indication include using the participant's body such as their finger, head or eyes [26]. Issues with finger pointing methods ([27, 25]) are that methodological errors occur. Soechting et al. [28] found that pointing at targets causes errors due to disparities between visual system and the sensorimotor system. Accuracy was improved when a 'pointer' was used instead of the finger. A number of studies have used a laser pointer as a visual indicator [29, 30, 31]. Issues with pointing are that the disparities found by Soechting et al. [28] can still occur. Work by Seeber [30] overcame their issues by using a roller ball which in turn moved a visual pointer. This could evaluate localisation ability accurately and quickly (mean absolute errors (MAE) of 1.6°). Seeber's method removed the issues of using a laser by having the subject's pointing from an egocentric frame of reference, decoupling the motorical system.

Eye movements towards sounds have also been studied in humans [32, 33, 34]. The advantage of using eye pointing is that it can be done with no training and eye location can be measured to high accuracy (1/10th of a degree as reported by Populin et al. [35]). The downside of the procedure, however, is that eye tracking equipment is expensive and can require lengthy calibration. The main downside is that eye movement is restricted by how far a subject can look without moving their heads.

Head tracking

Although the propensity of head movements varies between adults, it has been shown that most adults will move their heads towards a sound source [36]. It is believed that the orientation is done to align the visual system to the sound source. The reasons for this are numerous, for example, it can alert the listener to potential dangers or enable orientation towards the face of a person whilst they are speaking for better understanding [37]. Using head movements to assess localisation ability is advantageous because head movements towards sounds are natural and therefore require no training. Also, in the modern day, motion tracking devices and software are readably available. A number of studies have used a variety of methods to track head motion for the purposes of auditory research [38], and to attempt to evaluate sound localisation abilities [19, 39, 40, 41]. An early study by Thurlow et al. [38] investigated the sorts of head movements made towards auditory stimuli. The study used a wooden frame with marks on it, which was placed on the participant's head. The experiments were recorded on 16-mm video film. Head movements were then measured after the experiment using a microfilm reader. Though these experiments did not specifically evaluate localisation ability, they recorded head movements in relation to sound for the first time. Using film cameras to record head movements makes measurement of the head movement and processing difficult.

One way around this is to use motion tracking technology from which head motion data is easily obtained. The first study to use motion tracking technology to evaluate absolute localisation ability was by Makous and Middlebrooks [19]. The study used electromagnetic markers placed on a participant's head which measured the rotation of the head in response to a presented sound. The method showed high localisation accuracy for brief, wideband sounds (RMS of 2°-3.5°) with an increase in the error as the stimuli moved away from midline and to the sides of the participants (maximum of 20°). This can be explained by the limited movement of the head at more peripheral angles as well as reduced localisation ability at the periphery. Other studies by Recanzone et al. and Carlile et al. [39, 40], have used motion tracking to evaluate localisation ability, and found similar findings to the work by Makous and Middlebrooks [19].

Using head pointing as a method of evaluating absolute localisation ability has a number of advantages. As previously discussed, head movements towards sounds occur naturally, and so no training is required to use them to evaluate localisation, unlike GELP or pointing tasks. Although eye pointing also shows natural responses to auditory stimuli, head pointing is an easier metric to use as it does not require complicated setups like eye tracking experiments. With todays technology, head tracking is easily implemented and provides an accurate evaluation of sound localisation ability. Head pointing is, however, not the fastest method of evaluating localisation ability. Each head turn takes time to complete, making it potentially slower than other methods of indication such as selecting from a display of possible locations. Measurements of head turn would also require some form of signal processing to extract salient information from the motion tracking data or a method of indication of the location being localised, i.e. a button press.

Interim Summary of Adult Localisation Ability

Adult humans are very sensitive to interaural cues and as a result are able to pinpoint sounds in space to a high degree of accuracy. Localisation ability for sinusoids varies as a function of frequency, whilst wideband, modulated stimuli provide us with the best cues with which to localise. The ways in which sound localisation is measured are numerous making comparisons of results difficult. Methods range from simple verbal reporting to the use of motion tracking technology to evaluate a participant's response to a hidden source. While methods report different magnitudes of errors in localisation ability, they all show a high level of accuracy when the sound is presented in front of the listener, and decreasing accuracy as the sound moves out to the periphery.

1.4 Behavioural Testing in Infants

Before findings can be discussed regarding localisation ability in children, the methods used to obtain localisation data must be outlined. The methods which are presented in this thesis are all forms of behavioural audiometry (BA). BA is an umbrella term under which a large number of procedures and techniques exist, all of which allow the assessment of a child's auditory system and perception. All BA techniques follow a simple paradigm, a stimulus is presented and a change in behaviour is looked for. The complexity of the change in behaviour varies and will be addressed along with each method. Another difference between procedures is the way in which responses are encouraged, or rewarded. Most methods use conditioning and reinforcement to encourage the child to respond to the stimuli. The ways in which this is achieved vary and will be addressed when each method is discussed. All BA techniques used two people who conduct the test, the first engages the child in the task and the second controls the experiment, i.e. present the sounds and rewards.

Early work in BA was not related to sound localisation, but was rather used to measure hearing thresholds in young children. Pure-tone audiometry used by audiologists today to measure hearing thresholds in adults involves the presentation of a tone and the listener indicating if they heard it. Dix and Hallpike [42] noted that these techniques were not applicable to young children. They comment that in pure tone audiometry, the relation between a tone and what it signifies is non-existent in young children. With these restrictions, Dix et al. [42] proposed a technique referred to as 'peep-show' audiometry. Like the popular child's game *peek − a − boo*, on which its name is based, the procedure involves an animated puppet which appears when the child responds correctly (pressing a button) to the presented sound. Children are first conditioned using both the visual and auditory stimuli and once conditioned, the auditory stimulus is presented alone and the visual reward presented if the child responds correctly to the presented sound. The method was reported by Dix et al. [42] to reliably obtain hearing thresholds children as young as three years. The downside to this method is the method of indicating the sound, i.e. a button press. This would not be possible with younger infants.

One of the most important factors in using BA is to obtain a large number of responses from the child before they get bored and stop responding. This is particularly important if the audiologist is trying to measure thresholds, as a certain number of responses are needed in order to determine a threshold. Suzuki et al. [43] presented an improved, but similar, method to peep-show audiometry called conditioned orientation reflex (COR) audiometry. Suzuki et al. as well as Statten et al. [44] noted that as for younger age children, the number of responses obtained using peep-show drops off drastically and that the response criteria (i.e. what type of response to the sound is marked as correct) due to the young child's age, needed to be made much simper. In the simplified set up the child is placed between two loudspeakers under which soft toys are placed. A sound is presented from one of the loudspeakers. The method relies on the child responding using their natural orientation reflex response and that such a response can be conditioned. By conditioning the child's natural reflex behavior to novel sounds it was hoped that more trials could be obtained from younger children. Using such a method, thresholds can be obtained in 85% of the children aged 1-3 years, [43]. The COR method was an improvement on the peep-show method as it conditioned natural behaviour towards novel stimuli, these are elicited by children who are too young to be able to press a button as required in the peep show method.

Issues with just conditioning natural reflex responses is that the children soon habituate. Liden and Kankkunen [45] saw this problem with the COR method and so devised a new method called visual reinforcement audiometry (VRA). VRA marks a trial correct if one of the following types of responses are seen to the stimulus: reflexive or investigatory behaviour, orientation or spontaneous responses. The set up used is similar to that used in COR methods with the child placed between two loud speakers and a stimulus then played from one of them. The child was presented with a reward when they responded with one of the responses discussed above. Liden and Kankkunen concluded that hearing thresholds were easier to obtain in young children using the VRA method. The most common form of BA used by most audiologists at a clinical level is VRA [46]. Today VRA is implemented using a variety of setups with modern day practices using video screens and DVDs instead of animated puppets [47]. One of the main issues with VRA is that it looks for a wide range of responses to the sounds. This makes classification difficult and as a result is dependent on the abilities of the audiologists at judging responses and is subject to experimenter bias. If the audiologist can not distinguish between a response and random or inattentive behaviour then false positive trials will occur (i.e. the audiologist scoring a correct response when it was not, and incorrect diagnostics on the child's hearing ability will be obtained). One way of reducing experimenter bias on the judgment of responses is to make the audiologist blind to the presentation of the sound. This can be implemented when the audiologist presents the sound via a button press, however, unbeknown to the audiologists, the button press will not always present a sound. If the audiologist goes to present the reward when no sound was presented, a false positive is scored and no reward is given. If too many false positives are scored the experiment can be stopped with no results for the experimental block. This form of VRA is called the observer-based psychoacoustics procedure (OPP) [48]. By tracking the false positive rates, the skill of the audiologist can be monitored and if needs be the testing session can be stopped, however, this still does not address or eliminate the issue of experimenter bias but simply present a method of overcoming it. The method also requires that the audiologist has a high degree of training so that responses can be observed correctly and their false positive rate not so high that a test can not actually be completed. This makes reducing the bias in BA techniques very difficult and their application on large scale studies difficult because highly trained experimenters are required.

1.5 Spatial hearing in normally-hearing children

Development of the Auditory System

The auditory system undergoes a large process of development and maturation during the first twelve years of life [49]. Early development during the embryonic stages shows the formation of the basic structures, i.e. inner ear, cortex and brainstem pathways. The development and full maturation (adult-like state) of the cochlea occur by the 14^{th} week. By the perinatal period, a peak rate of development is achieved and the brainstem becomes mature. It has been shown that auditory processing is taking place at this stage of development in studies such as those by Kisilevsky et al. [50]. This study found that a babies heart rate would change differently depending upon whether it was presented with a recording of its mother's or a stranger's voice. These findings suggest an ability to process and discriminate the difference between voices. By birth the auditory system is not fully mature but has been shown to process binaural cues [51]. The full maturation of the auditory system is a point of dispute, some studies believe that it becomes adult like by the ages of 6-7 years [52, 53], whilst others believe it continues to develop into the teenage years [54].

1.5.1 Sound Localisation in Children

Assessment of Spatial Acuity in Children

Measurements of MAAs work well with young infants. The method can be adapted to look for behavioural changes rather than orientation towards sound sources, something which requires development of muscles and motor skills. Morrongiello et al. [55] have measured MAAs in infants as young as 8 weeks (an error of 27°, measured as the angular difference between the actual location of the stimulus and where the child indicated). Using a classical MAA setup [16], the infant sat with their parents and sounds were presented from two loudspeakers. The stimuli used were wideband noise burst (500ms) and the experimenter was blind to when a trial began. The experimenter scored a trial as positive if a change in behaviour was observed. The work found a decrease in MAAs as the children got older, with 24 week old infants having MAAs of around 18°. Follow up work by Morrongiello [56] found that from six months to eighteen months, MAAs again reduce by a third (12° to 4°). Litovsky [57] found approximately the same MAAs in children aged 18 months (5.7°). Litovsky also shows that adult levels of MAA are reached at around five years of age [57]. Similar to the adult data, it has also been shown that MAA performance in children decreases as the reference is moved away from midline [58]. Differences in MAAs can be explained by the nuances of the procedures used in different studies, however, all the studies show that as the children get older, their MAAs decrease. The changes in MAAs show that the auditory system goes through a process of maturation as discussed in the previous section.

Absolute Localisation in Children

Measurements of MAA are not considered a true evaluation of localisation ability but similar methods can be used to evaluate absolute localisation ability in children. Moore et al. [59] used VRA, but instead of conditioning a wide range of responses, just conditioned head turns. Although the study was not looking at evaluating localisation ability but looking at reinforcement conditions, it showed that BA techniques could be used in a localisation paradigm. Furthermore, one study has looked into the role of localisation in VRA [60]. The study found that even when using VRA as a method of measuring threshold, the child is still localising a sound. If this localisation aspect is taken away from VRA, the number of responses reduces.

Early attempts to conduct absolute localisation tests with young children were undertaken by Clarkson et al. [51], who measured head movement towards the stimuli using surgical tape placed on top of a child's head as an indicator. Stimuli were wide-band pulse trains (created using a gated rattle sound). Infants were awake and alert, they were held by the experimenter between two loudspeakers which were placed at either side ($\pm 90°$) of them. Although only two speakers are used, the task is not a discrimination task but one in which measurements of the accuracy of the childs' head turn toward a sound sources.

Localisation was measured by measuring the rotation of the head to the sound source via the tape placed on the child's head. It was found that for sounds played at $\pm 90°$, the absolute average rotation along the azimuth was around 35-40°. During this procedure no rewards were used to encourage or condition responses. The method was criticized by Morrongiello et al. [61], who argued that errors of the strip placement would carry on over to the measurements and so devised an improved method, involving placement of strips on the infant's (one to two days old) face and head with the movement being videoed and scored later using a protractor on the screen. The method also used four speaker locations (36°, 54°, 72°, and 90°) and a 500ms noise burst was used. The study found that the average error was around 26°. Morrongiello and Roca [62] also used this method on slightly older children of between six to eighteen months. They found localisation RMS errors of 16° and 6° for children aged six and eighteen months respectively. Data for children aged two and three years is not present in the literature to the authors knowledge.

Attempts to evaluate children aged four and above was undertaken by Van Deun et al. [63]. Using an identification task, they found RMS errors of around 10° for children aged four years. This dropped to 4° (RMS) for the oldest children tested (six years). The four year olds showed significant differences from the other groups (five and six years old and adults). The other groups showed no significant differences from each other. These findings, as well as those found by another study, suggest that the binaural system is at adults levels by the age of five [57].

Interim Summary of Child Localisation

The auditory system goes through large changes during early infancy and the first few years of life. Spatial acuity in infants is shown to increase with age. Measurements of MAAs in 8-week old infants found they are around 27° [55], decrease as infants get older until they reached adults levels (1°) at 5 years of age [57]. This development is also reflected in the improvement of absolute localisation ability. Clarkson et al. [51] found localisation ability in new-born infants to be around 35-40°. Van Deun et al. [63] found RMS errors to be equal to adult levels (4°) when the children were aged 5 years.

The section has described how many of the approaches towards sound localisation measurements in younger children are, in the authors opinion, quite crude (protractor on a video screen), prone to errors and also slow. It is believed that by using more modern approaches to this problem, these issues can be overcome.

1.6 Research Objectives

The binaural system is an important part of our everyday lives, it allows us to understand speech in a noisy environment and alerts us to the location of potential dangers. Assessment of a subject's binaural system can be achieved by evaluating their localisation ability. A number of localisation studies have been carried out in adults, but few have been conducted with children under five because evaluation of a child's localisation ability is difficult. It has been discussed in this chapter that the binaural system is seen to go through a process of maturation during the first five years of life.

One of the first problems to overcome is, how do you know when a child has heard a sound? In adult localisation studies, this can be achieved, for example, with a button press or the subject verbally reporting. These approaches are not possible to deliver consistent results in young children. To overcome this problem, previous studies have employed BA techniques which look for changes in behaviour as an indication that the child has heard the sound. Such an approach can be performed with very young children, however, it requires at least two highly skilled experimenters to keep the child interested in the task, control the experiment and judge responses. If used in large scale studies this would be both labour intensive and time consuming.

Another issue with the current approaches discussed in this chapter is, how do you know where the child has turned to in order to evaluate their localisation ability? Previous methods discussed analysed the responses once the experiment was over, this again is both labour intensive and time consuming for use in large scale studies.

Development of a method which could simply and accurately evaluate the localisation ability in children under five would make further research into the development of the binaural system easier and allow for a large population of children to be tested. As well as normal hearing children, having a tool to evaluate the binaural system would allow for better evaluation of, and also the development of, bilateral cochlear implants in children. This might be a way to conduct larger studies with BiCi quicker and in a cost effective way. It could also be used to look at new coding schemes and ways to maximise the effectiveness of BiCi in children.

The overall aim of the project is to develop a new method which can be used to obtain localisation responses from young children and develop techniques which can analyse and score these responses. Such a technique would be valuable as it would allow research into the binaural system of young children. This would not only inform how the binaural system develops through childhood (especially in children aged between two and three, where no literature is presented) but also could be used with binaural cochlear implant children to look at ways of improving binaural hearing in cochlear implants. For such a method to be developed a number of objectives must be accomplished. The objectives are:

1. Develop a game-like method which is engaging to the child and from which a large number of response can be obtained,

2. Collect data on the localisation ability of normal hearing children ranging

from one to five years,

3. Develop and accurate and fast method of measuring children's localisation ability so that response can be processed on a trial by trial basis,
4. Develop a way of measuring a response to a sound, from a child, based on a set of criteria.

1.7 Thesis Layout

Chapter two will discuss the experimental and laboratory setup which will be used to develop and test the new method. **Chapter three** will present and discuss results of the localisation method in terms of how successful it is at keeping the child engaged in the task. **Chapter four** will look at how to measure localisation ability using the new method and the results obtained during testing. **Chapter five** discuss methods of classifying head turn responses automatically using the raw motion tracking data. **Chapter six** is a discussion and draws conclusions regarding the findings of the thesis.

Chapter 2

Experimental Setup

2.1 Introduction

Chapter one discussed the methods which have been used to evaluate localisation ability in young children and also proposed the development of a new method which would improve upon previous methods. Before the new method can be presented, a full overview of the laboratory in which the method is implemented must take place.

2.2 The Laboratory

2.2.1 The Anechoic Chamber Setup

All experimental procedures described within this thesis were carried out in an anechoic chamber (AEC) situated in the Psychology Department of the University of Nottingham. The AEC is a 5m x 2m room with triangular acoustic foam (0.8m long) placed on all walls, including the floor and ceiling. The AEC contains a custom-built setup (sound play back, visual environment etc) referred to as the 'Simulated Open-Field Environment' (SOFE). A full, in depth overview of the system can be found in Seeber et al.[64]. This chapter will discuss parts of AEC and SOFE developed or used specifically by this project. Figure 2.1 shows parts of the SOFE systems inside the AEC.

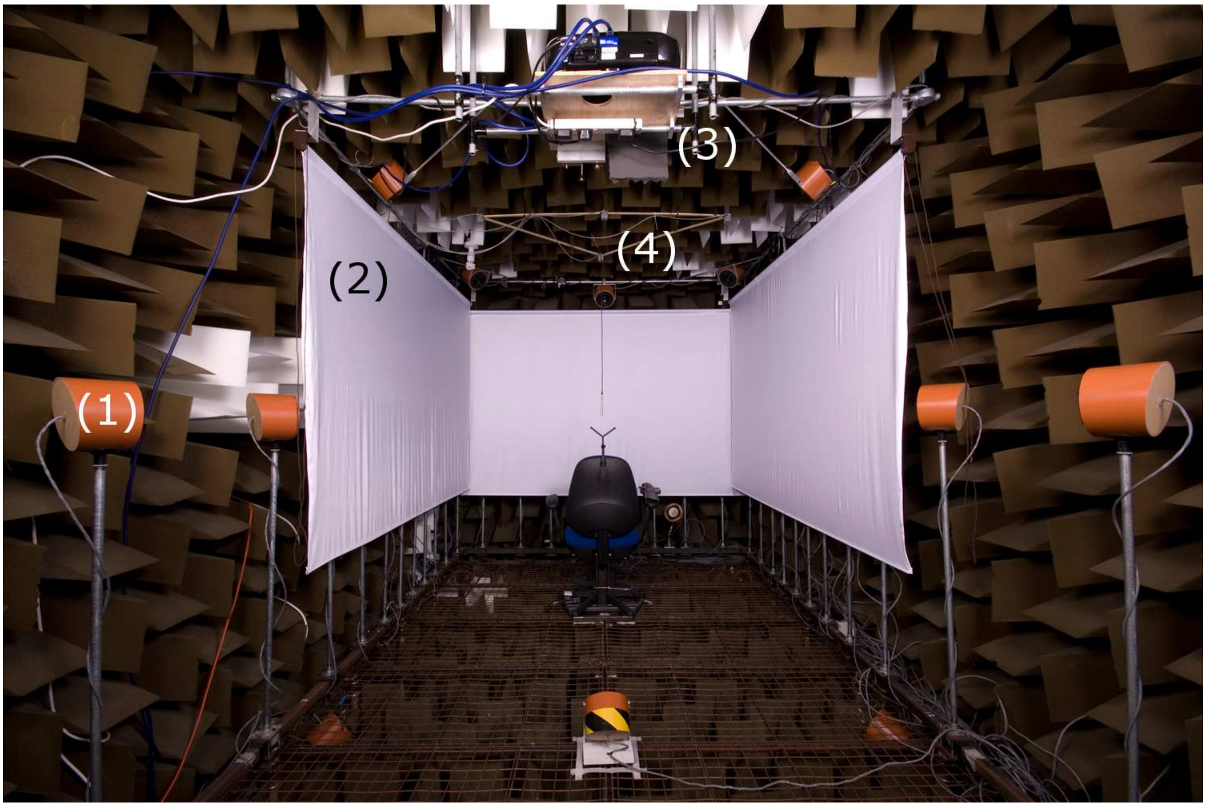


Figure 2.1: **Systems within the AEC. 1 - 36 speakers arrange in 10° intervals around the participant. 2 - Three video screens (2.1m wide) made of acoustic transparent material. 3 - Three full high definition projectors, two either side also placed slightly above side screens (not visible in picture). 4 - Four motion tracking receivers placed on the roof of the AEC, four more are also placed around the seat.**

2.2.2 The Control Room

The control room is situated adjacent to the AEC. The control room contains two independent computer systems, these are referred to as the audio PC and video PC. The primary function of the audio and video PC is to respectively control the audio playback and the visuals which appear on the projector screens respectively . The audio PC and video PC are connected to a high speed network inside the control room which allows for fast communication between the computers and devices. Also situated in the control room are the motion tracker unit and the amplifiers used for sound playback. Figure 2.2 shows the layout of the control room.

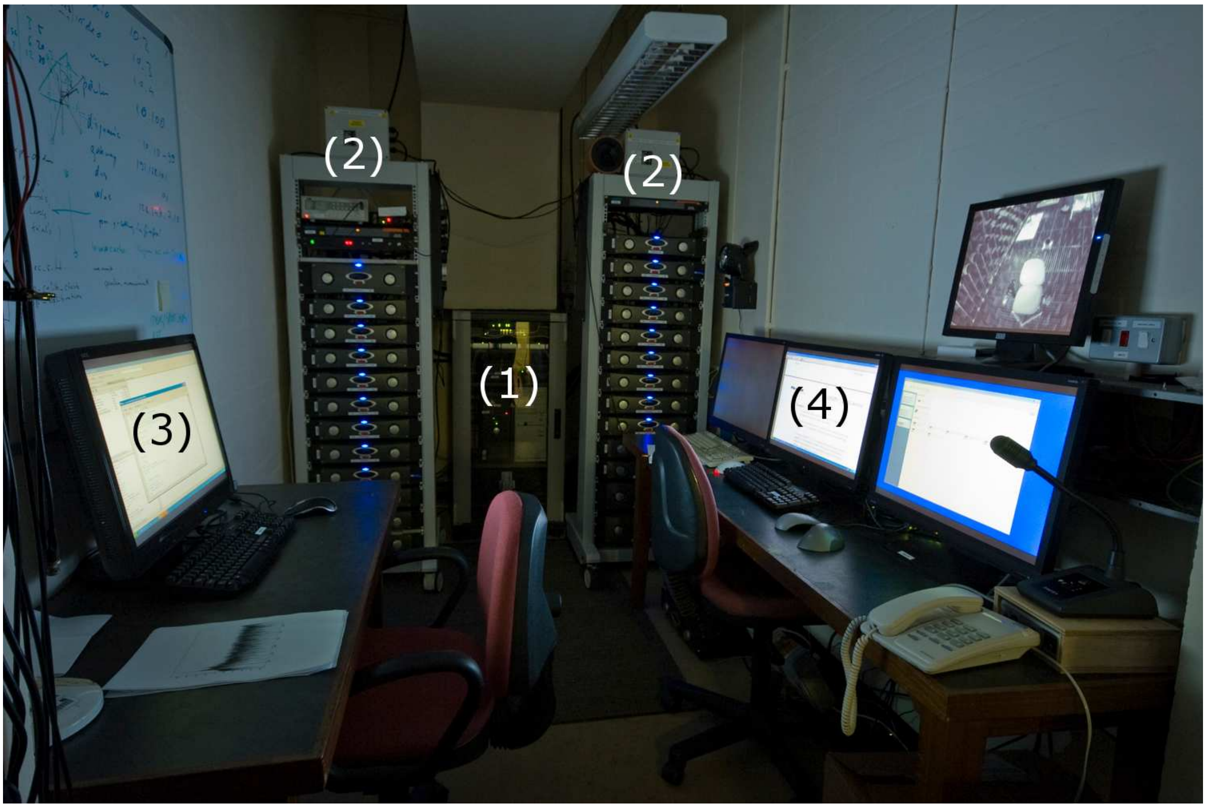


Figure 2.2: Systems within the control room of the AEC. 1 - Computer rack containing the video PC and audio PC. 2 - Amplifier racks containing 48 amplifiers (two channels per rack mount amplifier) which drive the speakers inside the AEC. 3 - Control point of the audio PC. 4 - Control point of the video PC.

2.2.3 Connections within and between the AEC and Control Room

Figure 2.3 shows the connections between the systems inside both the AEC and control room and the connections between the two rooms, the schematic also shows how the systems are linked and over what protocols. This diagram is used to show how the different systems communicate and interact when they are discussed in this thesis.

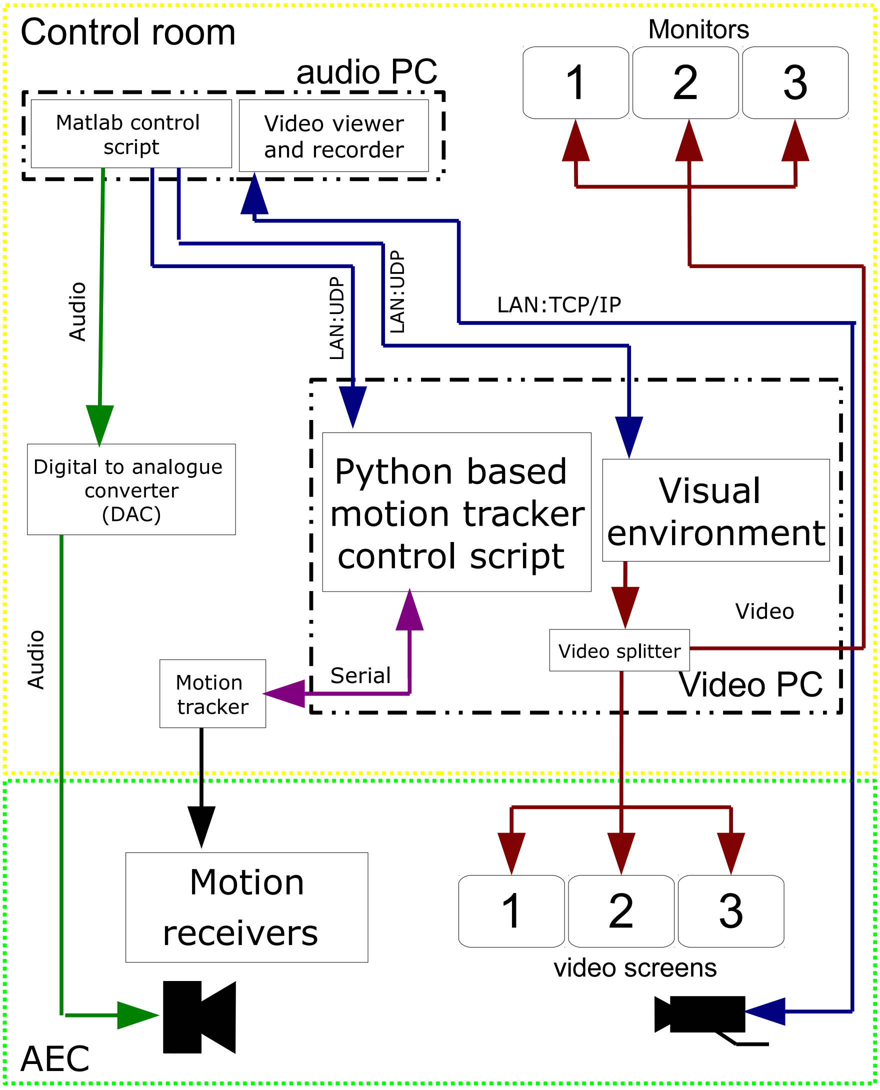


Figure 2.3: **Connections between the systems of the AEC and the adjacent control room. Also labeled are the types of connection between each system and in the case of devices connected to the local area network (LAN), the type of protocol used.**

2.3 Motion Tracking

The new localisation method requires the measurement of localisation responses to auditory stimuli. A method of doing this is to use motion tracking technology.The motion tracking system used is the Liberty Latus tracking system [65]. This system was used in the study because it was already part of the AEC and the SOFE system.

The system works using electromagnetic motion tracking markers which are placed on the object or person wishing to be tracked. The markers track in both polar (azimuth (Az), elevation (El) and roll (Ro)) and Cartesian (X,Y,Z) coordinates. This type of tracking is referred to as six degrees of freedom (6DOF). These degrees are shown in Figure 2.4 in reference to a head and torso. This thesis will mainly look at the rotation of the head along the azimuth to a sound source.

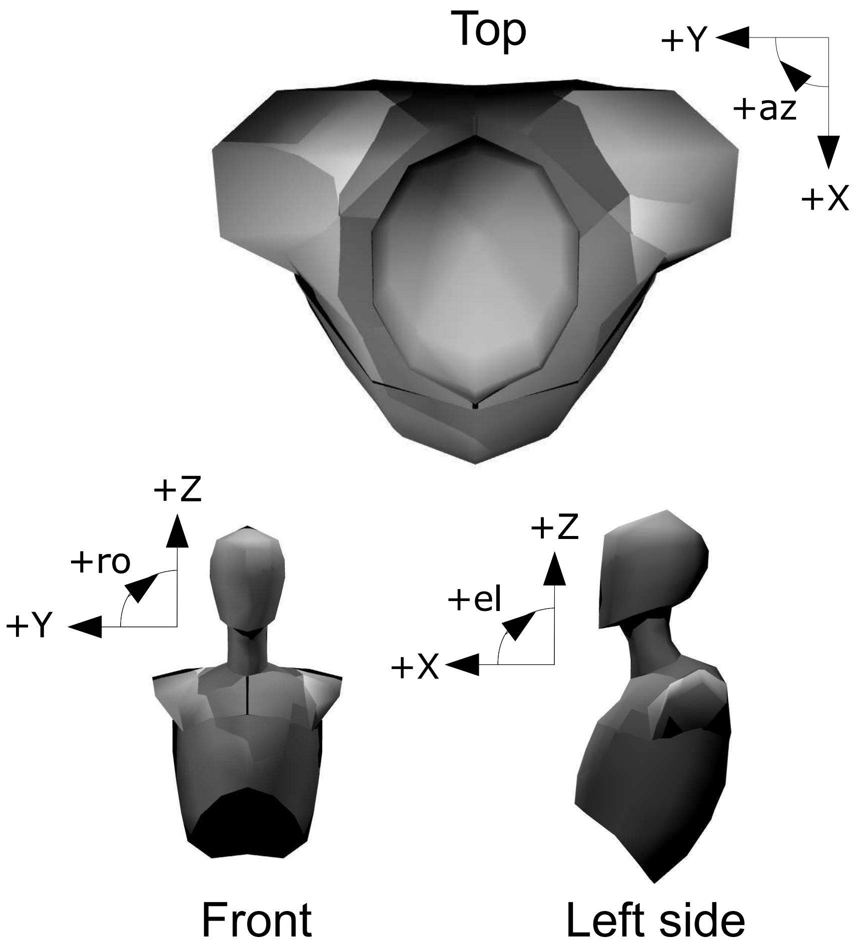


Figure 2.4: Co-ordinate system to describe responses to the auditory stimuli.

Motion tracker receivers

The motion tracking system consists of eight tracking receivers which are placed at defined locations within the AEC. These are in turn connected to the video PC via the motion tracking unit inside the control room. The receivers are placed on both the ceiling and floor of the AEC to allow for maximum tracking coverage around the participant. The layout of the receivers placed on the ceiling and floor are shown in Figure 2.5 and Figure 2.6 respectively.

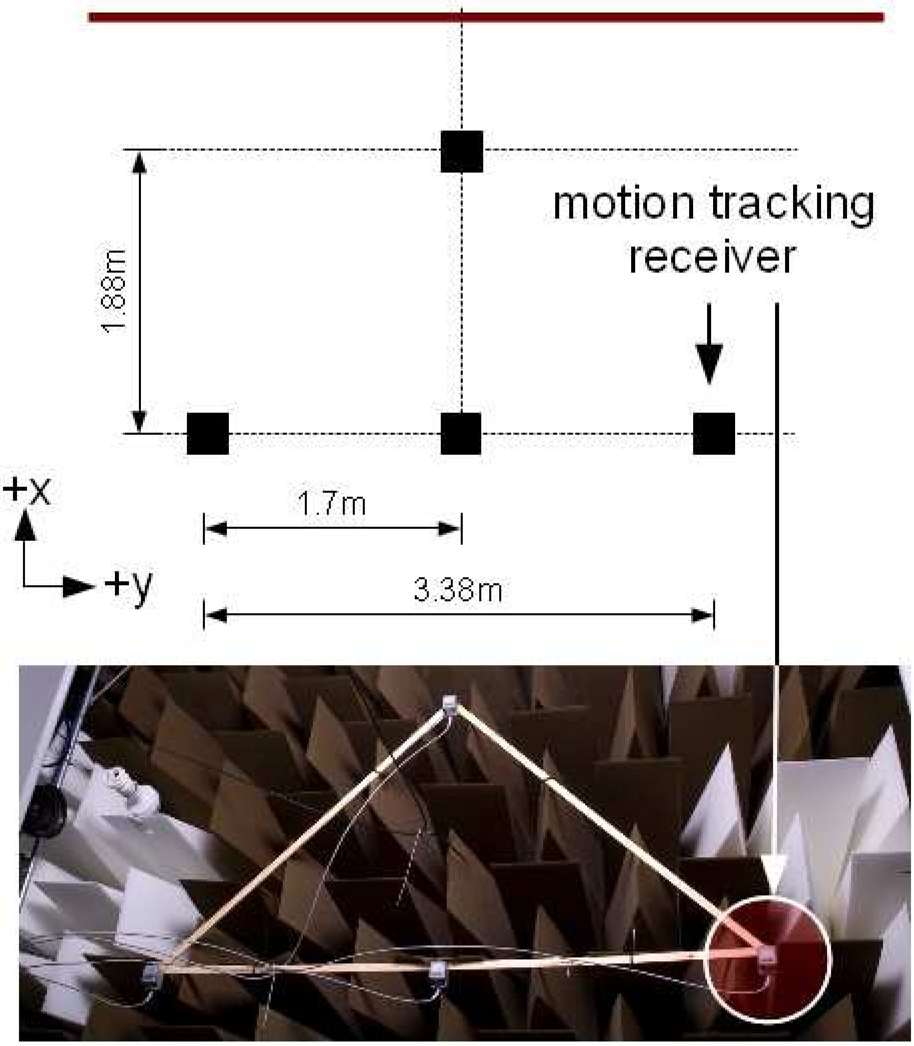


Figure 2.5: Schematic and picture of the four motion trackers placed on the ceiling of the AEC. Markers placed on floor and ceiling to increase coverage.

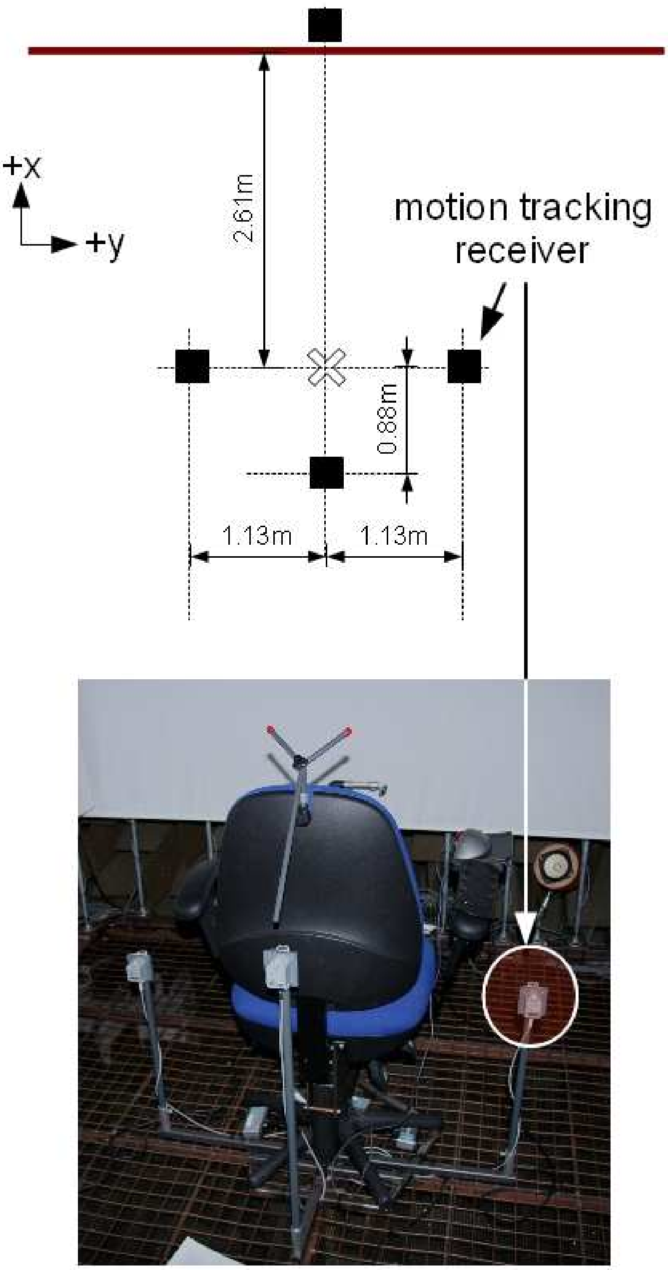


Figure 2.6: **Schematic and picture of the four motion trackers placed on the floor and around the seat of the AEC. The markers, along with those placed on the ceiling provide dense coverage around the participant.**

The motion tracker is specified to track with an accuracy better than 1° and with a quoted spatial resolution of under 1cm (manufactures specification ([65]) and also verified by a member of the MRC Institute of Hearing Research).

Motion Tracker Markers

Along with the receivers, the motion tracker has four motion tracking markers which are used to track the response of the participant to the auditory stimuli. Each marker weights approximately 80grams and has the dimensions, $7.4x4x2.2$cm (length, width and height).

During the development of the method it was unclear exactly what sort of responses would be made to the sounds presented so all four markers were used and placed on the participants head, body and arms. Placement of the markers on the participant is shown in Figure 2.7.

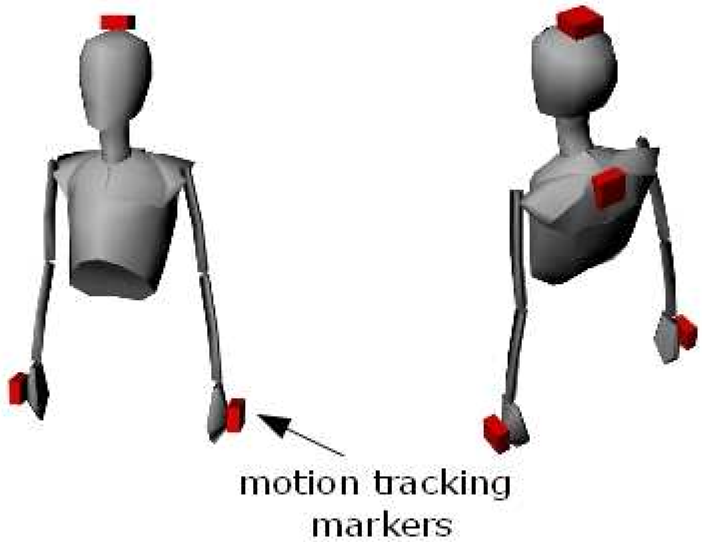


Figure 2.7: Red boxes show the locations of the motion tracking markers. The motion markers are held in place via the motion tracking gear worn by the child.

The markers were attached onto clothes (motion gear) which the child would wear over their own clothes. The 'motion gear' worn is shown in Figure 2.8.

Figure 2.8: **Picture showing the motion gear which the child wears over their own clothes. The motion trackers are attached to the motion tracking gear using industrial Velcro.**

The small size and weight of the makers made it possible to fit them easily to the participants (placed upon special garments) without obstruction to their movement. The motion tracking markers are held in place using Velcro attached to the top of the crown and the motion tracking marker respectively. This allows for the motion trackers to be removed for calibration and battery changing whilst at the same time providing a reliable method of attachment to the motion tracking gear. The head markers are placed forward facing onto straps on the crown. They are attached to the top of the jacket, between the subjects shoulder blades. The hand markers are attached by placing them into small pockets sewn onto the wrist bands. Issues did arise from certain children who did not enjoy wearing the garments to which the motion trackers were attached. This is discussed in depth later in the thesis.

Control of the Motion Tracking System

The tracking system is controlled via an independent program (developed by the author in Python [66]. The program was developed in Python as it was faster than MATLAB at reading the motion trackers serial port. The program is controlled via open sound control (OSC) messages. OSC is a protocol commonly used for the control of music instruments, however, the language allows for the control of almost any device over a network and was used because it was already used by the visual environment as developed by Seeber et al. [64]. OSC works by sending a key word along with a message. The keyword tells the receiving device what to expect and the message contains the data which is being passed. The OSC messages are sent from MATLAB (on the audio PC) to the motion tracking controller (video PC) over UDP to a specific IP and port. The Python script interprets the OSC message and in turn controls the part of the program called by the OSC message. For example, to collect head data, the OSC message $'/startcollecting'$ as well as an integer telling the Python program how long to collect motion tracking data for. The Python program connects to the motion tracker via a serial port. As well as controlling the motion tracker, the motion tracker controller also creates folders and files on the video PC to contain the motion tracking data on a session basis. This information is again sent via a OSC message from the MATLAB script. A full list of the OSC messages and functions they control is shown in Appendix C.

Calibration of the Motion Trackers

When the motion tracking system is first switched on, the motion tracking markers are not seen by the computer and they must first be initialised. Once initialised the motion trackers are calibrated. The launch and calibration of the markers is done via the motion tracking controller and controlled from the MATLAB experimental script. The calibration involves placing each marker in a set location (where the participants head is during the experiment) and then aligning them to the midline speaker until they read zero degrees. This is done because sometimes the marker will be 180° out. To overcome this, the marker was powered down and powered back up again. This procedure was introduced to reduce the issues regarding the motion tracking changing their polarities during a testing session or during a trial. Although correct launching and calibration of the markers reduced this happening, the issue still remained on some occasions and the data had to be post processed to eliminate these unwanted effects.

2.4 Visual Environment

The visual environment is the name given to the program which controls the three video screens inside the AEC. Three projectors (Epson EMP TW-2000 [67]) mounted 2.2m above the AEC floor project images onto three 2x2m acoustically-transparent screens. The images are also mirrored, using an image splitter, onto three screens inside the AEC so they can be observed by the experimenters in the control room. Figure 2.9 shows the three screens onto which characters and backgrounds are projected.

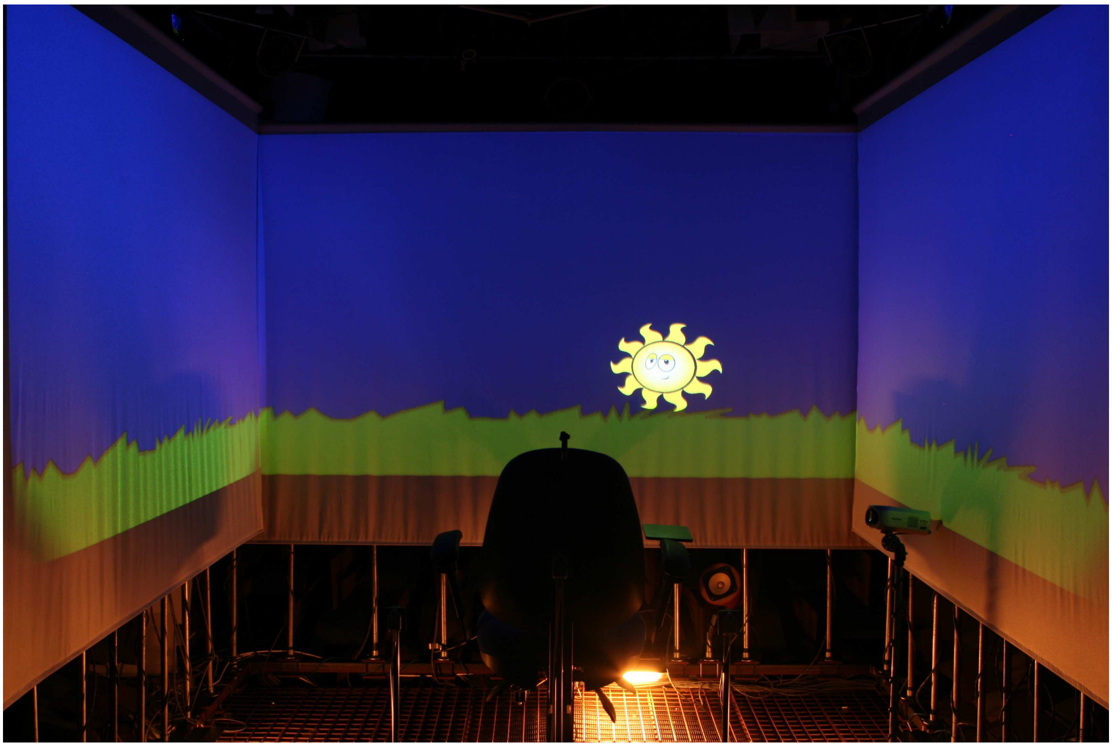


Figure 2.9: The three video screens inside the AEC with an example background and a character (sun).

The visual environment (VE) is a custom-built program written in the C programming language. The original VE was developed by Seeber [64] and later adapted to three screens for this thesis by the author, B.Seeber and V.Chilekwa. The program is built using the OpenGL and GLUT libraries [68] which allow for the creation of screens and images along with their manipulation (movement, rotation etc) in real time. Added to the visual environment code were a number of functions which allowed for the loading of bitmaps into the visual environment and functions which allow for the movement and manipulation of images. The visual environment code was already able to interpret OSC messages which could be sent from a number of devices or programs, e.g. MATLAB or a roller-ball mouse. The code was expanded by the author to include the interpretation of OSC messages which could be used to manipulate the images on screen.

The images are rendered on the screen using 2D 'sprites'. The 2D sprites are flat squares onto which the images are rendered. A function is used to upload all of the possible images available into a structure when the program is first run (a total of around 20 such images are currently used). A maximum of four sprites can be on the screen at any one time (including the background image), addition of more sprites is possible, however, the speed of the visual environment slows down noticeably. All of the properties regarding the image are stored in the code via structures, i.e. the images size, location, type of image being displayed, and if it is being displayed or not. Parts of the image structures are changed via OSC messages. For example, if the size of the image is wished to be changed, the UDP message '/object1size', followed by a scaling factor, is issued. A set of structures and OSC messages are also used to control the background images in a similar way. A full list of OSC messages and what they control can be found in Appendix B.

The characters used in the visual environment are shown in Figure 2.10. Characters were obtained from a number of sources, the dog and sun characters were obtained from the MRC Institute of Hearing Research's, the sheep was obtained from [69], the pig from [70] and finally the goat from [71]. These characters were not tested beforehand to check if children could identify the animals. They were also not checked for any bias which they might introduce, for example, the child being frightened of a particular animal presented.

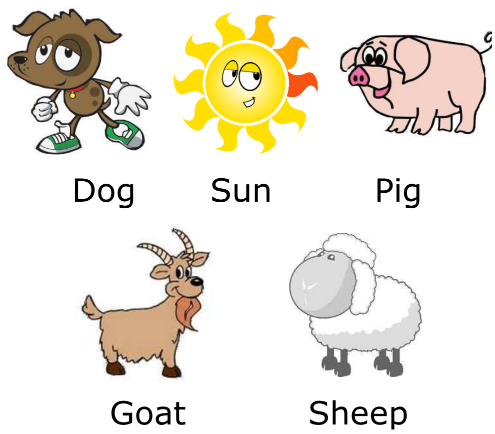


Figure 2.10: The characters which are used within the visual environment. The characters can be projected and moved on the screen independently. The visual environment allows up to around four such characters on the screen at any one time.

2.5 Video Cameras for Response Judgments

Chapter one discussed a range of different methods which are used, or have been used, to measure localisation ability in young children. All of these methods required at least two experimenters, one who controlled the experiment, i.e. presented the auditory stimulus and visual rewards, and a second experimenter who kept the child engaged. The method to be developed in this thesis will also use two experimenters, however, both will control the experiment whilst the visual environments and screens keep the child engaged. In order to monitor the child and their responses three video cameras are placed inside the AEC. The camera placement is shown in Figure 2.11

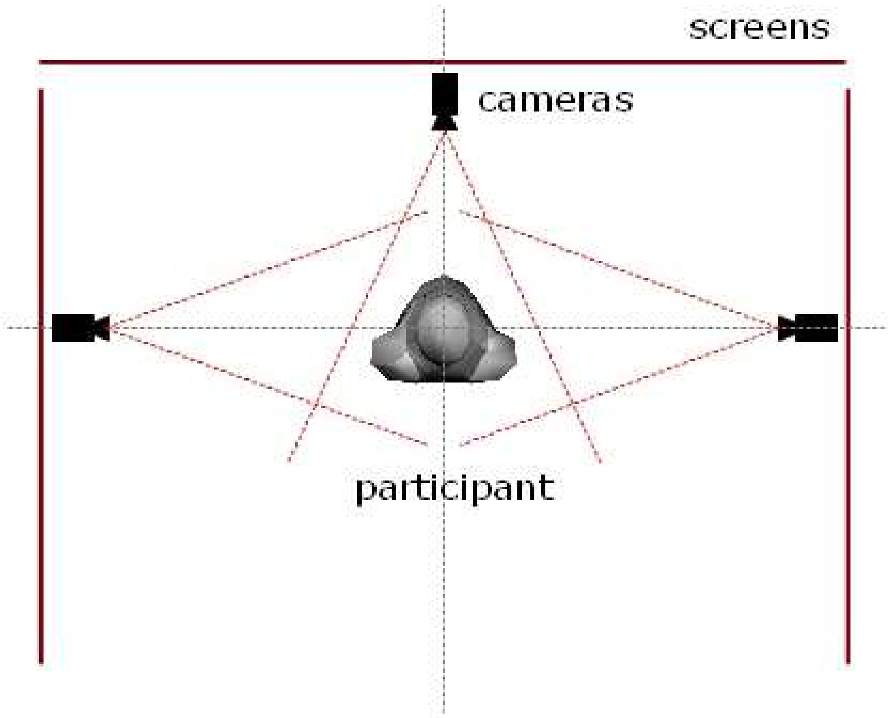


Figure 2.11: **Top view of the camera layout and coverage within the AEC. The cameras were placed so that full coverage of the participants was achieved. The cameras were placed on mounts which allowed full three dimensional adjustment allowing the cameras to be set up for maximum coverage for each participant.**

The AEC contains three HD video cameras, Sanyo VCC-HD4000 [72]. The cameras are mounted on custom-built mounts (constructed by the Institutes workshop) and are placed in front and to the sides of the participant (see Figure 2.11). The cameras are visible to the participants and are fixed about a meter away from them just below the projection screens. Placing the cameras in this location allows for the clear view on the video feed of the participants head, eyes and torso.

The cameras are controlled using Sanyo-VMS software [72]. This software controls the cameras over the network and records their outputs onto the audio PCs hard-drive. The software also allows viewing of all three cameras simultaneously on one screen at a rate of 25 frames per second.

2.6 Speakers and Sound Playback

The AEC contains 48 custom-built loudspeakers, 36 speakers are placed at ear-height (1.24m above the floor) along the azimuth plane with a separation of 10°. A further twelve speakers are placed in the floor and ceiling to allow for the simulation of rooms, this feature is not used in this thesis.

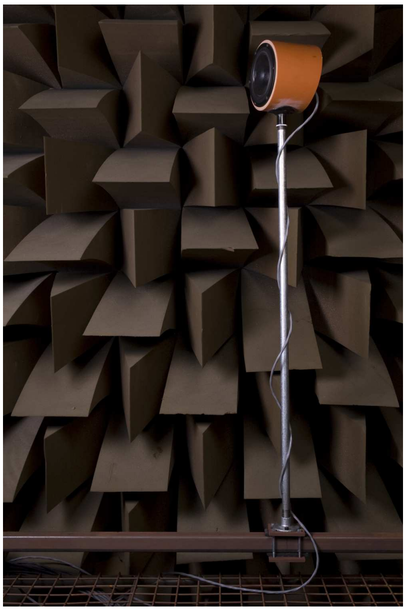


Figure 2.12: **One of 48 custom built speakers inside the AEC. The speakers are placed along the horizontal in 10° increments. Full description and schematic of the speakers is provided in [64]**

Each speaker is driven by a Alesis® RA500 amplifier [73] located in the control room. Sound playback is achieved via Matlab which is run from the audio PC. Such a system allows for the 48-channel system to be equalized in real time. A more in depth description of sound playback system can be found in [64]. Five speakers are used in this thesis, the auditory stimuli and it's parameters will be discussed where appropriate.

Chapter 3

Description of the *AnimalSeek* Method Used To Evaluate Localisation Ability in Children Under Five

3.1 Introduction

This chapter proposes a new method called *AnimalSeek* which can be used to evaluate localisation ability in young children aged between one and five years of age. For the method to be successful, it must be able to keep the child engaged in the task which is being used to evaluate their localisation ability.

The children's localisation ability was measured via motion tracking technology. Analysis of the motion tracking data is discussed in chapter four (see page 87).

This chapter will firstly discuss the procedural aspect of the BA techniques, their issues and how parameters of these have been changed in order to increase the number of responses obtained during testing session. Once key points from these studies are highlighted, the *AnimalSeek* method is discussed in depth in a step by step manner. The number of responses the *AnimalSeek* method obtains are then presented and discussed.

3.1.1 Methods for Increasing the Number of Responses in Behavioural Audiometry

BA has been used to evaluate both localisation ability and to obtain hearing thresholds with most BA techniques usually involving at least two experimenters. The first experimenter controls the experiment whilst the second sits in front of the child and engages them in the task. Being able to reduce the BA technique to one experimenter would make the method more efficient and therefore more beneficial in a clinical setting. However, reducing the method to one experimenter does increase the possibility of experimenter bias which affects the results. The method presented uses two experimenters to judge the responses the children makes the the auditory stimulus.

During a BA trial two key events take place; firstly the presentation of the auditory stimuli and secondly the presentation of a visual reward. Both of these events can effect the number of responses obtained from the child being tested.

Effects of Stimuli on Number of Responses

Early studies of infant audiometry by Hoversten et al. [74] found that auditory stimuli and their intensities had an effect on the number of head turn responses obtained from young infants. The study found that infants responded more consistently to human speech as compared to a 4kHz tone. Other studies such as Thompson et al. [75] and Weiss et al. [76] showed similar results to Hoversten et al. [74], with more responses obtained when a voice-like stimulus was used. This suggests that the best stimuli for eliciting responses are those which are either speech or which possess speech characteristics. Similar findings are seen in animal behavioural studies with species-specific sounds [77].

As well as its frequency content, the duration of a stimulus has an effect on the eliciting of head turns. Clarkson et al. [51] found that newborns orientate to wideband rattle sounds which are both short (one second) and long (twenty second). Stimulus which are shorter than one second have shown to lead to fewer responses [78].

Effects of Rewards on Number of Responses

Rewards are used to condition a child to respond to a sound. If a sound is presented and the child responds correctly, a visual reward is presented. Moore et al. [59] showed that the number of responses obtained from a child were influenced by the types of reward presented. The study used four types of reward;

no reward, social reward, simple visual reward (i.e. a blinking light) and a complex visual reward (i.e. an animated toy). The study found that a complex visual reward provided the most trials before the child stopped responding. More recently, several studies have investigated the effects of a complex reward i.e. using a video screen showing a children's DVD instead of a mechanical puppet [79, 80, 79, 81]. These studies found that the video screen obtained a higher number of responses compared to the traditional animated puppet.

Another aspect of the reward which can be altered is the duration of presentation. Culpepper and Thompson [82] found that with pre-term two year olds presenting the visual reward for four seconds produced fewer responses (before habituation) as compared to presenting the reward for half a second. The number of rewards also has an effect on the number of trials. Thompson et al. [83] found that if several different rewards were presented to the child during a testing session, the longer the child would take before responding. Moreover, the study found that introducing a ten minute break between testing sessions led to an increase of five trials per session.

Primus et al. [60] found that by presenting the auditory stimulus away from the visual reward affected the number of responses. The study used three locations: the reward location itself, above the child's head and directly opposite the reward location (i.e. if the stimuli was presented at +50°, then the reward would be presented at -50°). Presenting the auditory stimulus at the opposite side to the reward produced the fewest trials. This is because the two are spatially separated and the meaning of the reward in relation to the sound cannot be established by the child.

3.1.2 Motivation and Proposal of the *AnimalSeek* Method

This chapter will present the *AnimalSeek* method which can be used in the evaluation of the localisation ability of a child under five. The method is developed, based on the studies above, to obtain the highest number of responses before the child becomes disengaged with the task and stops responding. Unlike previous BA methods, no experimenter is required to be placed in front of the child - instead the child is engaged in the task via the three large video screens of the visual environment discussed in chapter two (see page 41).

Aspects of the method will be altered in order to see how the number of responses can be increased; in particular, where the reward is presented in relation to the auditory stimulus.

3.2 The *AnimalSeek* Method

The procedure is split into three main stages: the start of the trial and sound presentation, the presentation of a reward and the end of the trial. A trial shall be discussed in a step by step process.

Beginning of a Trial and Stimuli Presentation

Responses were judged based on the responses the children gave. By doing this, no button presses or other forms of indication were required and as a result the method can be used with a wide age range of children. One of the issues with developing a method which can be used with children of a wide age range is the types of instructions given to them. For this reason the procedure was required to produce head turn responses towards auditory stimulus without giving instructions to the participant. During pilot testing it was found that when testing normal hearing older children (three plus years) some instructions were required in order to make the children at ease when first visiting the AEC. As a result a simple instruction of *'find the animal'* was given before the children began their first testing session. For all children under three years of age, no instructions were given. This approach proved successful. The procedure also did not use any form of conditioning so that exposure to the task was limited and responses were as natural as possible.

When ready, the child put on the motion tracking equipment as described in chapter two. The child then sat on their parent's/guardian's lap unless old enough to sit on their own. If sat on their own, in order to reduce any anxiety the child might have had about the AEC, the parent was to take the child into the AEC and then stand at the back, out of sight of the child, and asked to remain silent throughout testing.

The method uses no experimenter inside the AEC. The child is engaged in the task using cartoon characters projected in front of and to the sides of the child. A trial begins with a character flashing twice (on and off) at midline for a duration of two seconds (on/off time of 500ms). After the second disappearance of the character, a short random delay (approximately 0-100ms) was added so that the child could not learn the exact time the stimulus were presented. After this delay, the sound was presented. The sounds were designed to be as easy to localise as possible. This required them to be above threshold, have a wide frequency range and be modulated in accordance with the literature. The stimulus was a child speaking animals name (dog, pig, sheep, goat and sun). The animal names related to the rewards which would be presented. The stimuli were around 1500ms long and presented at 60dB SPL with a random level-rove of $\pm$6dB SPL. The stimuli were presented randomly from one of five speaker locations (0°, $\pm$30° and $\pm$70°) along the azimuth, positive speaker locations were to the right hand side of the child (see Figure 3.1).

Response Judgment

After presentation of an auditory stimulus a response by the child to the stimulus was looked for. Responses were judged from a live video feed from inside

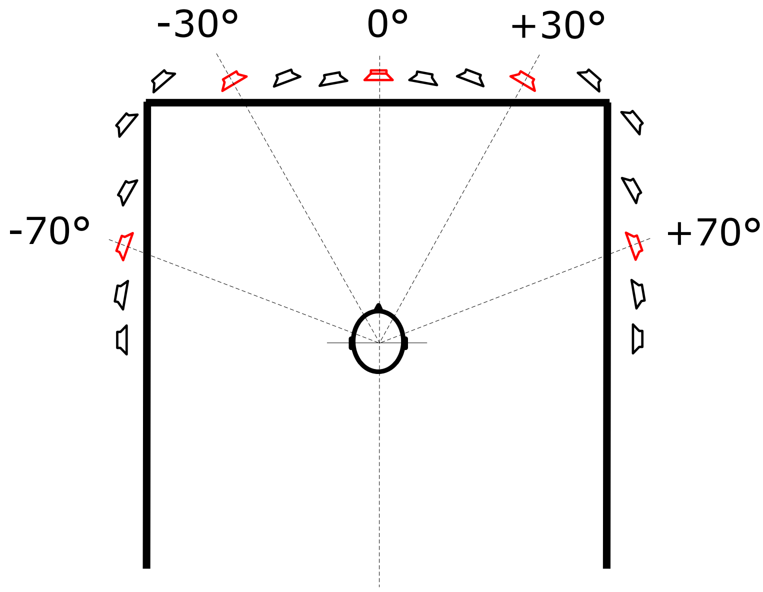


Figure 3.1: An overhead schematic of the AEC. The five red speakers correspond to the directions used in the experiment.

the AEC. The video feed allowed the judges to see the front and both sides of the children's head, eyes and torso. Sound presentation was initiated by the experimenters and the direction of the auditory stimuli were known to them. A response was marked as either correct or incorrect and then categorised to allow for the evaluation of the method. If a response was correct then it categorised into one of the following criteria:

- Head turn towards the sound source from which the stimulus was presented
- Eye movement towards the sound source from which the stimulus was presented
- Hand pointing towards the sound source from which the stimulus was

presented

If a response was scored as incorrect it was then categorised into one of the following:

- Head turn not towards the sound source from which the stimulus was presented
- Eye movement not towards the sound source from which the stimulus was presented
- Hand pointing not towards the sound source from which the stimulus was presented
- Null response.

To reduce errors in the response judgments, the number of response criteria was kept low so that correct and incorrect trials were easy to identify and categorise. The number of speaker locations were kept small for similar reasons. Using the frontal camera, it was possible to judge which side the child turned to and using the side cameras it was possible to see which speaker the child was turning to. Experimenter bias was reduced by keeping the types of responses classified low (head, hand and eye) and by using two experimenters.
Judgments were made shortly after sound presentation by the two experimenters independently and recorded on sheets containing the trial number and response options. If there was a conflict in judgment a 'null response' was scored, this was to maintain the flow of the testing and not hold up a trial by deciding what

response was observed. A trial was also scored a 'null response' when the child did not seem to be trying to localise the sound, for example, when a sound was presented and the child turned to their parent/guardian (if sat holding the child or sat at the back of the AEC), the child started crying, or tried to get off their parents/guardians lap.

Blocks were terminated early if five no trials were observed in a row, this was to reduce the stress on the child if they were not responding and also provided an indication of how effective the method was without trying for excessive amounts of time.

Reward Presentation

A correct response to the sound resulted in an animated (spinning or waving) visual reward being presented. If the child produced an incorrect responses, the visual reward was still presented but remained static, i.e. it was not animated. Where the visual reward was presented in the visual environment in reference to the auditory stimulus is a topic of interest and will be discussed in the following sections. The visual reward was presented (static or animated) for 2000ms. A short time was used so that exposure to the reward would not reduce the number of responses (as shown by Culpepper et al. [82]). Thompson et al. [83] showed that using more visual rewards obtained more responses before habituation. Therefore, a total of five different animated characters were used (see Figure 2.10). The number of characters which could be used was limited by the number of animals present in the speech corpus (Institute's in house corpus). The rewards presented are dependent on the auditory stimulus, i.e. the word dog, resulted in a picture of a dog being presented.

End of a Trial

At this point in the trial, it is assumed the child has turned to the sound presented off midline (zero degrees). With each trial starting at midline, it is important to get the child looking back there before the next trial can begin. To encourage the child to move back to midline the visual reward, after being presented, scrolls back across the screens to the midline (front).

During the testing of children aged between one to five years, it was found that all children tracked the character back to the midline and were ready for the next trial. It is also noted that even if the child responded incorrectly, the scrolling visual reward is still useful for drawing the child's attention back to the start of a trial. This allowed the experiment to flow trial by trial without the need for a second experimenter inside the AEC capturing the child's attention at the start of each trial.

3.3 Experimental Design: Effect of Reward Location on the Number of Responses

It is common for most BA techniques to present the visual reward in the same location as the auditory stimulus. Issues with this are that the child can learn

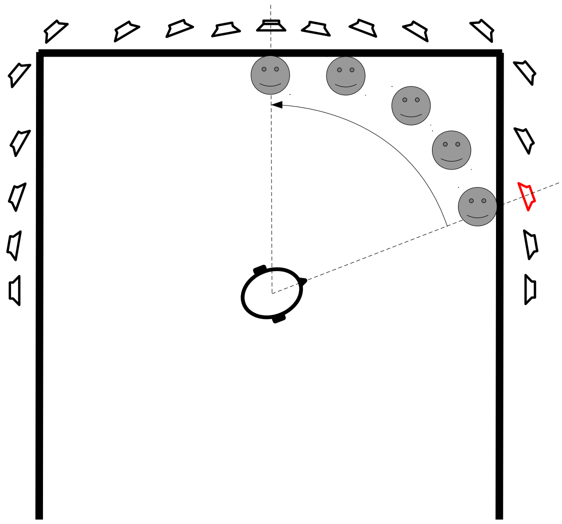


Figure 3.2: **Figure shows an over view of the AEC and the child in the center. In order to re capture the child's attention and start each trial from midline, the character presented as a reward is scrolled slowly back to midline (taking approximately five seconds). It was found that with all of the age groups this scrolling action was very good at drawing their attention back to midline without the use of an experimenter inside the AEC or instruction.**

where the auditory stimulus is being presented from by simply learning where he/she sees the visual reward. This is particularly true if only a few auditory locations are tested. One way to overcome this would be to spatially separate the two, however, as discussed, Primus [60] found that separating the reward and the auditory stimuli by too much results in fewer responses. To overcome this, the visual reward was presented close to but not at the stimulus location at a random location within 20° of the auditory stimulus. By doing this, the child still had to turn to the direction of the auditory stimulus to see the reward i.e. the two are related, but the visual reward did not show where the auditory stimulus was presented. It was hoped that by doing this, the task would still remain game-like, reduce any learning effects and not affect the number of responses obtained. To test that spatially separating the two does effect response performance, another type of reward location was tried where the reward was presented always at zero degrees. In accordance to the Primus [60] study, this was believed to be the least interesting for the children and that the number of responses would be less than both presenting the reward at, or close by, the stimulus location.

The three reward types are defined as follows:

- Reward type *zero*: Reward always presented at zero degrees (i.e. midline)
- Reward type *at location*: Reward presented at the stimulus location
- Reward type *jittered*: Reward presented at a random location $\pm 20^{\circ}$ about the stimuli location.

The study was broken up into three visits for each participant. Visits were spaced around one week apart (depending on time commitments of the parent/guardian and child). Each visit contained three blocks, with each block consisting of 30 trials. Each block would last around five to ten minutes depending on the compliance of the child.

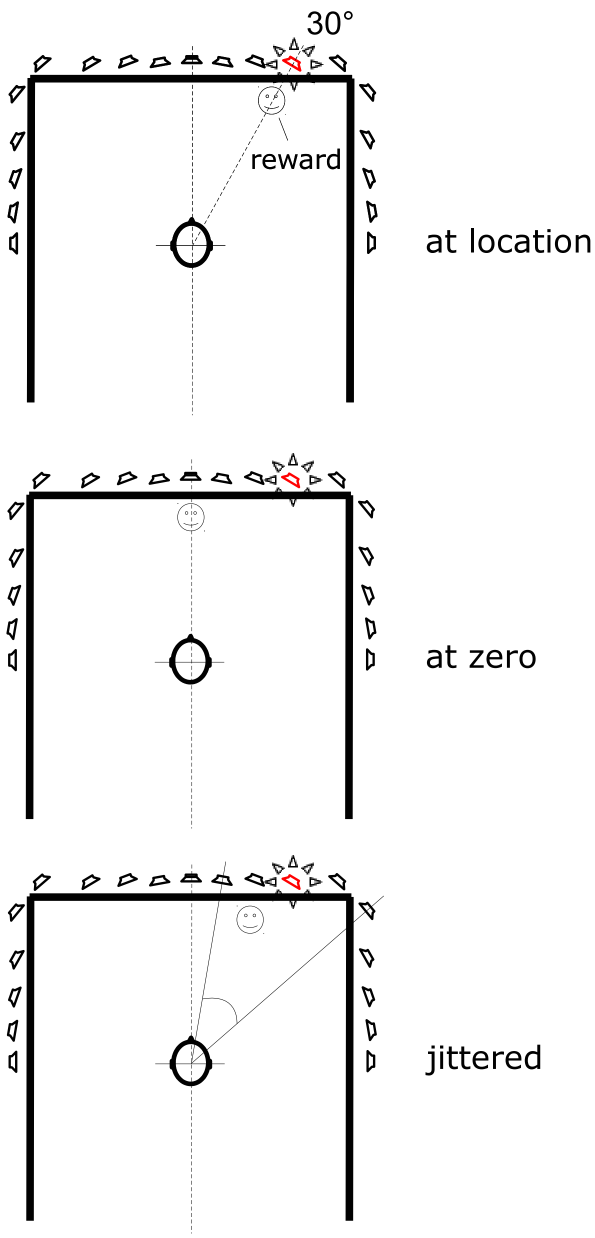


Figure 3.3: **Overview of the three reward types used. The red speaker corresponds to where the sound was presented from, the smiley face shows the location of the reward. The 'jittered' condition presents the reward randomly at a location equally 20° within where the auditory stimulus was presented.**

With each child visiting three times, there was a total of nine blocks (270 trials). During each visit, each reward type was tested. A Latin square design [84] was used to balance the rewards over the three visits. This resulted in each reward being the first, second or third block in the visits. Each child was assigned a random Latin square upon the first visit using a script developed in MATLAB. An example can be see below in Table 3.1.

	Block 1	Block 2	Block 3
Visit 1	*zero*	*at location*	*jittered*
Visit 2	*jittered*	*zero*	*at location*
Visit 3	*at location*	*jittered*	*zero*

Table 3.1: Example of how reward types are randomly assigned to each child using a Latin square design. Each square was generated for the child upon their first visit. The latin square was generated using functions contained within the LATSQ toolbox [85]

Breaks of approximately ten minutes were given between each block, following the work of Thompson et al. [86], who showed that ten minute breaks between blocks increased the number of responses in subsequent blocks compared to no breaks being given. During the breaks, the children were free to move around, play with the toys in the control room and to have refreshments.

3.4 Participants

Participants were recruited using several methods. These included the distribution of fliers and posters at local community and play centers, postings on websites (i.e. http://www.netmums.com) and via direct email contact. The information pack given to interested parents/guardians is reproduced in Appendix D.

In total around 35 normal-hearing children were recruited into the study. Of these seven children did not complete all visits; two due to ear related problems, and five due to other commitments. According to parental report, the children had normal hearing and had not recently suffered from a cold or other sickness. The children, also according to parental report, had no disabilities or learning difficulties. Older children (three years plus) were given hearing screens (20dB SPL) up to 8kHz on their first visit, using play audiometry. Approval for the study was obtained from the Research Ethics Committee of the Department of Psychology of the University of Nottingham. Parents gave written informed consent. The parents of child participants were given an inconvenience allowance to cover their travel and time costs.

Participants were split into three age groups, the age groups were decided based on the number of participants available and the age groups boundaries were created to maximize and balance the number of children in each age group. This was so that analysis and comparisons of the age groups were possible.

The three age categories are show in Table 3.2 along with the standard deviation and mean of their ages.

Age Group	N	Mean age in years (SD)
1.0-2.2 years	11	1.68 (0.26)
2.2-4.0 years	9	2.86 (0.35)
4.0-5.0 years	8	4.24 (0.36)

Table 3.2: Total number of participants, N, for each age group. The mean and standard deviation (SD) of each groups age is also shown.

3.5 Results

The success of the procedure was judged by looking at both the total number of trials (sum of both correct and incorrect trials) obtained by the method, that gives an indication of how long the child is willing to sit before becoming bored and stops responding, and also the number of correct head turn responses obtained, i.e. the number of head turns to the target stimulus which were used to evaluate localisation ability. Also discussed are the numbers of incorrect responses including incorrect head turns and also null responses. Null responses are a good indication of how effective the method is at keeping the child's attention and levels of boredom. The areas reported are the effects of age on the method, the effect of exposure to the task and finally the effect of the reward type on the number of responses obtained.

3.5.1 Effect of Age on the Number of Responses

Age (years)	Total Responses M(SD)	Total Correct M(SD)	Correct Heads M(SD)
1.0-2.2	20.4(5.5)	12.6(5.6)	9.2(5.2)
2.2-4.0	23.4(5.1)	14.9(5.5)	10.4(4.3)
4.0-5.0	26.9(4.1)	19.8(6.0)	12.5(3.8)

Table 3.3: Mean (M) and standard deviation (SD) for the total number of responses and the correct number of head turns for each age group per block. Data is averaged over the three rewards types and across the three visits.

Age (years)	Total Incorrect M(SD)	Incorrect Head M(SD)	Null M(SD)
1.0-2.2	7.9(2.0)	0.4(0.4)	7.2(1.7)
2.2-4.0	8.4(2.0)	0.8(1.0)	6.9(2.9)
4.0-5.0	7.1(2.9)	0.6(0.7)	5.9(2.4)

Table 3.4: Mean (M) and standard deviation (SD) for the total number of incorrect responses, the total number of incorrect head turns and the number of null responses for each age group per block. Data is averaged over the three reward types and across the three visits.

This section performs a number of statistical tests to compare the different age groups, reward types and task exposure. These statistical tests shall be discussed at the start of each section where applicable. The statistical tests used in this section are Kruskal Wallis tests. The test is used to see if the data being compared is from the same distribution i.e. are the age groups the same. If differences are seen between the groups, a Mann-Whitney U-test is performed in what is known as, post hoc analysis. The Mann-Whitney U-test is used to see if the two sets of data are the same e.g. differences between age group 1.0-2.2 and 2.2-4.0 years. A Bonferroni correction is used to remove the effect of the multiple comparisons of the groups. Table 3.3 shows the mean number of total responses, the mean number of total correct responses and the mean number of correct head turn responses averaged for all three age groups and averaged over all reward types and visits. As expected, the data shows an increase in all the responses as the children get older. For the total number of responses, this increase is significant between the age groups (Kruskal-Wallis test, $H(2) = 7.3, p < 0.05$). Post hoc tests show a statistically significant difference only between age groups 1.0-2.2 and 4.0-5.0 years (Mann-Whitney U-test with Bonferroni correction, $z = 0, p < 0.017$).

Table 3.4 shows the total number of incorrect responses, number of incorrect head turn responses and the number of null responses. The number of total incorrect responses is highest for the middle age group (2.2-4.0 years) with the least numbers of correct responses being the oldest age group, 4.0-5.0 years, statistically there is no difference, however, between the age groups.

It was expected that the youngest group would have the most incorrect responses out of all of the groups, however, this is not the case. When looking at the null responses however, i.e. ones which are due to inattentiveness of the child and cannot be categorised, the youngest age group (1.0-2.2 years) shows the most. Differences between the age groups are not significant for null responses.

When evaluating localisation ability, we require a high number of correct head turns, across all age groups we see an increase, however, statistically there is no difference. The biggest difference between the correct number of head turns and the total number of responses is for age group 4.0-5.0 years, with correct head turns making up 46.6% of the total number of trials. The smallest difference is for age group 2.2-4.0 years, with head turns making up 44.5% of the total trials. The lowest number of incorrect head turns are for age 1.0-2.2 years (45%), the most incorrect head turns are seen in the age group 2.2-4.0 years, difference of the number incorrect head turns between age groups are not significant.

The method shows that it is possible to obtain responses from children as young as one years old even without an experimenter keeping the child's attention. As predicted it can be seen that the number of responses is age dependent. In the youngest age group, the mean number of correct and total response was less as compared to the other two groups. The oldest group, 4.0-5.0 years, showed the most responses in terms of both correct and total response, they also showed the lowest number of null responses. This was expected, with the younger children not willing to sit for as long as the older children.

Other BA studies use innumerable different stimulus and testing conditions. Such studies have the number of trials ranging from approximately 10 to 30 trials for children aged between one and two years [87, 88, 89, 90, 91]. Studies using VRA to evaluate localisation ability do not comment on the number of trials obtained. If the total number of responses is taken as the measure, then the *AnimalSeek* method is comparable to other studies. This suggests the child is interested in the task and is willing to sit and engage in it, but the amount of correct head turns is only around 50% of the total number of trials. This could be due to the motion tracking gear the child wears causing them to be distracted or the fact the children were not conditioned to the auditory stimulus in an attempt to obtain natural head turn responses.

Total correct responses made up 61.6% of all of the responses made by age group 1.0-2.2 years, 63.8% were correct responses made by 2.2-4.0 years and 73.7% of the responses were correct for the age group 4.0-5.0 years.

The next section will look at the number of responses with respect to the different visits and also the blocks within those visits. Such analysis will show how well the *AnimalSeek* method is at engaging and motivating the children.

3.5.2 Effects of Exposure to the Procedure on the Number of Responses Obtained

A major issue when testing children using a behavioural task is that they become bored with the task and stop responding. For this study a large set of data was required so that localisation ability could be investigated. For this reasons it was understood that to increase the number of responses, the testing must be divided into a number of different visits (with time between each visit) and each visit divided into a number of different blocks (with breaks of at least ten minutes between each block).

Responses Over Three Visits

The statistical tests used in this section are the Friedman test and also the Wilcoxon signed-rank test. The Friedman test is used to look at the differences between the age groups when they have repeated measures i.e. multiple visits or blocks. The Wilcoxon signed-rank test is used as a post hoc test used to look for differences between individual measures e.g. the difference between visits one and two. Table 3.5 shows the total number of responses, total number of correct responses and the number of correct head turn responses averaged over all blocks and reward types. Table 3.6 shows the total number of incorrect trials, the number of incorrect head turn responses and the number of null responses averaged over all blocks and reward types.

Testing was broken down into three visits to limit the effect of the game be-

Visit	Total Number of Responses M(SD)			Total No. Correct Responses M(SD)			No. Correct Head Turns M(SD)		
	1	2	3	1	2	3	1	2	3
Age									
1.0-2.2	20.3 (5.6)	20.7 (8.3)	20.3 (5.3)	13.8 (5.5)	12.0 (7.6)	11.8 (6.1)	10.1 (5.1)	9.5 (7.0)	8.0 (5.6)
2.2-4.0	24.0 (9.3)	23.7 (5.3)	22.3 (5.6)	17.6 (9.0)	14.7 (5.9)	12.4 (5.1)	13.1 (7.1)	9.6 (5.2)	8.5 (4.8)
4.0-5.0	27.3 (4.1)	27.7 (4.6)	25.5 (5.2)	12.9 (6.2)	20.2 (6.9)	17.2 (6.3)	15.1 (4.5)	12.1 (5.2)	10.4 (3.8)

Table 3.5: Total and correct number of responses obtained during each visit for each age group. Averaged over blocks and reward types.

Visit Age	Total No. of incorrect responses M(SD)			No. incorrect head turns M(SD)			No. Null M(SD)		
	1	2	3	1	2	3	1	2	3
1.0-2.2	6.5(2.9)	8.7(2.6)	8.4(2.9)	0.2(0.4)	0.5(0.7)	0.6(0.5)	6.2(2.5)	7.8(4.0)	7.7(3.7)
2.2-4.0	6.5(4.3)	9.0(2.8)	9.9(2.3)	0.4(0.5)	0.7(0.9)	1.5(2.6)	5.8(4.3)	7.1(3.4)	7.8(3.5)
4.0-5.0	5.4(3.0)	7.5(4.4)	8.3(2.8)	0.5(0.9)	0.9(1.0)	0.3(0.4)	4.7(2.8)	5.5(3.6)	7.5(2.6)

Table 3.6: Total number of incorrect, incorrect head turns and null responses obtained during each visit for each age group. Averaged over blocks and reward types.

coming boring to the children and the number of responses dropping between visits, this was particularly important for the number of correct head turns. The oldest age group (4.0-5.0 years) shows a significant difference between the three visits (Friedman, $\chi^2(2) = 7.0, p < 0.05$). Post hoc tests were only significant for visits one and three (Wilcoxon signed-rank test with Bonferroni correction, $z = 0, p < 0.017$).

The other response types show no significances between visits for all age groups.

Responses within Each Visit

With the number of responses decreasing for subsequent visits, in order to investigate how the number of responses changed between blocks (i.e. does exposure to the task during a single visit effect the number of the responses) the data will need to be corrected to remove the effect of the visit. Before this can be investigated it was shown that the more visits the children attend we see an overall drop in the number of responses obtained. For analysis of the effect different blocks have within a visit, the effects of the visit must be corrected for. This is to correct for the drop in responses due to each visit and just see the drop in blocks within each visit. The relative number of responses within each block are corrected for by subtracting the mean across the three blocks per visit from the data within each block. The equation for this is shown in Equation 3.1;

$$BC_{ij} = B_{ij} - \bar{V}_j, \tag{3.1}$$

where BC is the corrected block, B is the block to be corrected, i is the block number and j is the visit number. $\bar{V}$ is the mean number of trials during visit j. Table 3.7 shows the total number of relative responses, the total number of relative correct responses and the number of relative correct head turns for each age group and block, averaged over each reward type and corrected for each visit. Table 3.8 shows the total number of relative incorrect responses, the relative number of incorrect head turns and relative null responses for each age group and block, averaged over each reward type and corrected for each visit. All three age groups show a drop in responses across the three blocks on each day's visit, with block three having the biggest drop in responses for all three age groups. When looking at the correct head responses, age group 1.0-2.2 years shows a significant drop between visits (Friedman, $\chi^2(2) = 14.6, p < 0.05$). Post hoc analysis shows significance between visits one and two (Wilcoxon signed-rank test with Bonferroni correction: $z = 0, p < 0.017$) and also visits one and three (Wilcoxon signed-rank test with Bonferroni correction: $z = 0, p < 0.017$). Age group 1.0-2.2 years also show a significant drop in the total number of correct responses (Friedman:$\chi^2(2) = 8.7, p < 0.05$). Post hoc analysis shows a difference between blocks one and two (Wilcoxon signed-rank test with Bonferroni correction: $z = 0, p < 0.017$) and also blocks one and three (Wilcoxon signed-rank test with Bonferroni correction: $z = 0, p < 0.017$). Age group 2.2-4.0 years also shows a significant drop in the total number of correct responses (Friedman:$\chi^2(2) = 12.7, p < 0.05$). Post hoc analysis shows a drop between visits one and two (Wilcoxon signed-rank test with Bonferroni correc-

	Relative No. Total			Relative No. Correct			Relative No. Correct head		
Blocks	1	2	3	1	2	3	1	2	3
Age Group									
1.0-2.2	4.0 (6.9)	-0.8 (8.7)	-3.0 (8.8)	6.5 (7.5)	-2.4 (7.2)	-4.3 (7.3)	3.2 (6.3)	-0.4 (6.0)	-2.9 (4.3)
2.2-4.0	3.3 (4.5)	0.5 (5.6)	-3.7 (6.9)	4.0 (5.2)	-0.1 (6.4)	-3.9 (7.0)	3.1 (4.3)	-0.4 (4.6)	-2.7 (6.1)
4.0-5.0	1.2 (2.4)	0.3 (6.8)	-1.4 (6.7)	1.9 (5.1)	-0.1 (7.0)	-1.7 (7.2)	1.4 (3.0)	0.3 (4.8)	-1.7 (4.3)

Table 3.7: Relative total number of responses per block (30 trials per block) and correct head turns per block corrected for visits using Equation 3.1. Signs are kept to show an increase (positive) or decrease (negative) in the number of responses per block.

	Relative No. Incorrect			Relative No. Incorrect Head			Relative No. Null		
Blocks	1	2	3	1	2	3	1	2	3
Age Group									
1.0-2.2	-1.2(2.6)	+0.5(2.4)	+0.7(3.4)	+0.1(0.5)	0.0(0.6)	-0.1(0.4)	-1.0(2.8)	+0.5(3.6)	+0.5(3.3)
2.2-4.0	-0.8(3.8)	+0.6(4.4)	+0.2(1.7)	0.0(1.1)	0.0(1.3)	0.0(1.0)	-1.1(2.8)	+0.5(3.6)	+0.5(3.3)
4.0-5.0	-0.7(3.6)	+0.4(4.5)	+0.3(3.7)	0.0(0.8)	0.0(0.8)	0.0(0.6)	-0.7(2.9)	+0.2(3.0)	+0.6(2.7)

Table 3.8: Relative number of total incorrect responses per block (30 trials per block), relative number of incorrect head turns per block and relative number of null responses, corrected for visits using Equation 3.1. Signs are kept to show an increase (positive) or decrease (negative) in the number of responses per block.

tion: $z = 0, p < 0.017$) and also visits one and three (Wilcoxon signed-rank test with Bonferroni correction: $z = 0, p < 0.017$).

Age group 4.0-5.0 years also shows a significant drop in correct head turns between visits (Friedman:$\chi^2(2) = 6.7, p < 0.05$). Post hoc analysis shows a difference between blocks one and three (Wilcoxon signed-rank test with Bonferroni correction: $z = 0, p < 0.017$).

The results suggest for all of the children, the total number of responses and the number of correct head turns is highest during block one and the number of incorrect responses lowest during block one. The drop in total and correct responses and increase in incorrect responses is seen to affect the younger age groups more (1.0-2.2 years and 2.2-4.0 years). Age group 4.0-5.0 years does show a change in the number of responses, however, this change is more gradual compared to the other two age groups.

Overall we see that increasing exposure to the task leads to a reduction in the number of responses and increase in the number of incorrect responses. This can be seen in terms of visits and also within blocks. It was hoped that by spreading testing over a number of visits and having breaks between blocks that this could be eliminated. However, the results that even with such breaks and visits, the children still become bored with the task.

3.5.3 Effect of Reward on Responses

The following section discusses the effect of reward type on the number of responses obtained by the *AnimalSeek* method. Full descriptions on the reward types can be found on page 58. Table 3.9 shows the total number of responses, the total number of correct responses and the number of correct head responses for each age group and reward type, averaged over all blocks and visits. It was hypothesised that reward type *zero* would produce the fewest responses because the visual reward and the auditory stimulus are spatially separated. This does not encourage head movements and it is believed that the children will soon become uninterested in the procedure. For all the data, across all of the age groups, we see only a significant for age group 1.0-2.2 years and the total number of correct responses (Friedman, $\chi^2(2) = 6.19, p = 0.05$). Post hoc analysis shows a difference between reward type *zero* and reward type *at location* (Wilcoxon signed-rank test with Bonferroni correction, $z = 0, p < 0.017$).

All of the other responses types for all of the age groups show no significant differences between reward types. Although statistically no differences are seen, due possibly to the low numbers of children testing, trends in the data are still observed.

If using the reward type to remove learning effects, the age of the child being tested would have to be taken into consideration. The older children seem to produces more correct head turn responses and less incorrect / null responses when presented with this reward type. The middle age groups however, (2.2-4.0 years) show a lower number of head turns and higher numbers of incorrect and null responses. For the youngest age group, performance of the children when the reward is jittered is similar to when simply presented at zero, in terms of numbers of correct head turns and total responses.

	Total Number of Responses M(SD)			Total No. Correct Responses M(SD)			No. Correct Head		
Reward Type	1	2	3	1	2	3	1	2	3
Age									
1.0-2.2	19.4 (5.9)	22.8 (7.0)	20.1 (6.0)	11.3 (5.9)	15.1 (6.9)	12.3 (5.4)	8.1 (5.4)	11.0 (6.1)	8.6 (5.1)
2.2-4.0	21.2 (6.8)	23.9 (5.7)	25.0 (4.5)	13.3 (5.7)	16.0 (6.3)	15.3 (5.2)	9.6 (4.4)	11.3 (5.2)	10.3 (3.9)
4.0-5.0	25.8 (5.2)	27.0 (4.0)	27.8 (3.5)	18.7 (6.8)	19.4 (6.7)	21.3 (5.3)	11.7 (4.6)	12.2 (4.3)	13.7 (4.0)

Table 3.9: Mean (M) and standard deviations (SD) for the total number of responses and number of correct head turns for each age group and reward type per block.

Reward Type	Total No. of incorrect responses M(SD)			No. incorrect head turns M(SD)			No. Null M(SD)		
Age	1	2	3	1	2	3	1	2	3
1.0-2.2	8.1(1.8)	7.8(2.2)	7.8(3.3)	0.4(0.4)	0.6(0.7)	0.3(0.5)	7.5(1.5)	6.9(1.7)	7.2(3.0)
2.2-4.0	7.9(3.0)	7.8(1.6)	9.6(2.6)	0.9(1.2)	1.0(1.6)	0.7(0.8)	6.5(3.4)	6.3(2.5)	7.9(3.8)
4.0-5.0	7.1(2.5)	7.6(4.0)	6.5(3.2)	0.5(0.6)	0.5(0.8)	0.6(0.8)	6.2(2.5)	6.3(3.3)	5.3(2.5)

Table 3.10: Mean (M) and standard deviations (SD) for the total number of incorrect responses, incorrect head turns and null responses for each age group and reward type per block.

3.6 Discussion

The chapter set out to present and evaluate a new method for evaluating the localisation ability of children under five. The method was required to be engaging to the child so that an adequate number of trials could be obtained so that their localisation ability could be evaluated. The new method showed that head turn responses to an auditory stimulus could be elicited in children as young as one years old.

Previous studies have found that the number of responses obtained from children aged two years range from 10 to 30 and 16 to 20 for children aged 1 years [91, 87, 88, 92, 89, 90, 93]. In comparison, the new method presented in this chapter showed the mean total number of correct head turns (i.e. the response used to evaluate localisation ability) for the age group 1.0-2.2 years was 9.2, for the age groups 2.2-4.0 and 4.0-5.0 was 10.4 and 12.5 respectively. The overall number of responses was 20.4 for the age group 1.0-2.2, and 23.4 and 26.9 for the age groups 2.2-4.0 and 4.0-5.0 respectively. Direct comparisons with other studies is difficult because most of the studies use BA to obtain hearing thresholds and look for a large number of different responses to the auditory stimulus (i.e. head, eye movement and other gestures), and not just head turns. Also the conditions and experimental parameters are numerous, with a wide range of stimulus types (wide-band and pure tone) as well as the use of video screens and mechanical puppets.

The total number of responses (i.e. total number of responses correct and incorrect responses) are comparable to previous studies. This suggests the method is engaging for the child and that they are willing to sit, however, the number of correct head turn responses which are used to evaluate the child's localisation ability is generally lower than expected per block. The number of incorrect head turns per block are quite low with the results showing each of the age groups has less than one per block. The number of null responses is quite high for each age group.

The results also showed that exposure to the task, i.e. the more blocks undertaken, there is a decrease in the number of correct head turns for all of the age groups. If the main aim is to use this new method as a tool to evaluate localisation ability in children with bilateral cochlear implants, then testing will take place in a clinical environment. The time allowed for testing in a clinic is generally shorter than that for a research study. On the basis of the present study, however, if three or more blocks could be obtained per visit and around 30 correct head turn responses obtained (i.e. within an hour visit) this would be adequate to begin to evaluate the childs' localisation ability. With more than one visit more likely in a research study, the method could be used to gather data regarding a children's binaural development and also research into the improvement of bilateral cochlear implants.

The Effect of Visual Reward Location

The method presented a new way of presenting the visual rewards to the children. Usually BA techniques present the reward at the same location as the auditory stimulus. It was believed that by doing this with a small number of speaker locations then the child could learn the auditory stimulus location by remembering the location of the visual reward. Taking such an approach might not be ideal for the younger age groups (1.0-2.2 years and 2.2-4.0 years) because the *jittered* location produces few correct head turn responses as compared to the *at location* reward type. Caution would also have to be taken because for middle age group the *jittered* reward type produces a higher number of incorrect head turn and total responses. The findings follow those of Primus [60] in that the reward type which was most spatially separated from the stimulus (presented at zero degrees) produced the fewest results.

When looking at the correct number of head turns and also the total correct number of responses, the data show reward type *at location* as having the highest means for age groups 1.0-2.2 years and 2.2-4.0 years. For age group 1.0-2.2 years, the difference between *at location* and *zero* is significant. For the younger two age groups (1.0-2.2 years and 2.2-4.0 years), the number of incorrect head responses is highest for the *at location* reward type, this might be the children being engaged by this reward type more but it not eliciting more correct head turn responses.

The most correct head turn and also the total correct number of responses for the age group 4.0-5.0 years are seen by reward type *jittered*. This suggests that presenting the reward away from the auditory stimulus location makes the task more interesting for the child resulting in more head movement and turns during trials. Null responses are also lowest for this reward type for this age group. The differences between the reward types for this age group (4.0-5.0 years) however, are not significant.

It was hypothesised that presenting the reward away from the speaker location (i.e. *jittered* reward type) could reduce the effect of the child learning the speaker location, as in the *at location* reward type and provide a better method of evaluating localisation ability. The results suggest that for age group 1.0-2.2 the *jittered* location is not ideal because it produces few correct head turn responses as compared to the *at location* reward type. For age groups 2.2-4.0 years and 4.0-5.0 years, the *jittered* reward type produces a higher number of total responses, statistically, however, there is no difference between the reward types. For the oldest age group tests (4.0-5.0 years) the most correct head turns are produced by the *jittered* reward type, however, this reward type also produces the most incorrect head turn responses.

With all of the results discussed, although trends are evident in the data, due to the small sizes of the groups statistical significance is not shown in the data apart from the total number of correct responses for age group 1.0-2.2 years.

If using the reward type to remove learning effects, the age of the child being tested would have to be taken into consideration. The older children seem to produces more correct head turn responses and less incorrect / null responses when presented with this reward type. The middle age groups however, (2.2-4.0 years) show a lower number of head turns and higher numbers of incorrect and null responses. For the youngest age group, performance of the children when the reward is jittered is similar to when simply presented at zero, in terms of numbers of correct head turns and total responses.

If the method was to be used in a clinical or research setting, jittering the reward would be advantageous because it would reduce the learning effects of presenting the reward at the speaker location. This has been shown to suggest that this could effect the number of responses, however, the differences for all of the response types (correct, incorrect etc) were shown to be only a few trials difference for all of the age groups.

3.7 Conclusion

The new method showed that it was possible to replace the experimenter inside the AEC with three large video screens. Even though the experimenter was not inside the AEC, the overall number of responses was comparable to studies which did use an experimenter, suggesting using the visual environment does engage the child well and can keep the child's attention in the task. The downside of the method was that from the collected dataset the number of correct head responses was lower than expected. This could be partially due to the motion tracking gear worn by the child with some children refusing to wear the gear or consistently tried to remove it during testing. The children were also not conditioned to the auditory stimulus and as a result could have been confused by the nature of the task. The method did find that the number of correct head turns were lower than expected with only around ten obtained per block, however, with most children taking part in three blocks per visit this is potentially enough response to begin to evaluate a child's localisation ability. This approach yields caution, however, due to the drop in the number of responses observed over several blocks and visits.

The next chapter will discuss how the method was used to obtain localisation results from the children.

Chapter 4

Development of Head Motion Analysis Methods and Their Use To Evaluate Localisation Ability in Children Under Five.

4.1 Introduction

Propensity of head movements can vary between adults, but it is known that adults will automatically move their heads towards a novel sound [36]. Such head movements have also been found in infants who are just a few days old [94]. Discussed in chapter one (see page 8) were a number of studies and methods which have been used to evaluate absolute sound localisation ability in adults. One of the large variations between these studies is how participants indicate where a sound source is. Methods of indication include, but are not exclusive to, verbal reports of the target location [20, 21, 22], abstract pointing methods, i.e. indication of the target on a globe [24], pointing using a laser pointer [29, 30, 31], finger pointing [27, 25] and head pointing [19, 39, 40, 41]. All these methods have advantages and disadvantages which are discussed in chapter one.

The main downside of these methods of indication is that they are impossible to use with young infants. Using head pointing as a method of evaluating localisation ability has a number of advantages. Firstly the spatial coordinate system of the head is the same as that of the stimulus. Also, we turn our head to novel sounds in our everyday lives, this means no special training is required before conducting localisation experiments. Head pointing also has appeal since it can be performed by very young children because it does not need instructions as it is a natural response [94].

This section will look at a number of studies which had used head pointing as a method of evaluating absolute sound localisation ability in both children and adults and the difficulties and issues these approaches have to overcome. The chapter will then present an analysis technique which can be used in order to avoid some of the issues these studies had.

4.1.1 Measuring Head Turns in Adults and Children

Measurement of Head Turns in Adults

Measurements of head motion in adult localisation experiments have become ever more popular over the past two decades due to the increased availability of motion tracking technology. The first study, to the authors knowledge, to use electromagnetic motion tracking technologies to evaluate absolute sound localisation was Makous et al. [19]. The study used a Isotrak Polhemus electromagnetic tracking system [65]. The setup consisted of a motion tracker and a receiver capable of sampling the head movement at a rate of up to 20Hz. The tracker was placed onto the participant's head using a specially designed hat. Indication that the target had been localised was done by the participant via a button press. The study looked at localising sound sources along both the horizontal (azimuth) and vertical (elevation). The study did not separately analyse the horizontal and vertical axis, with all locations along the azimuth being elevated with a minimum of plus or minus five degrees. For the frontal target speaker locations a high localisation accuracy was found with an error ranging from 1.5°-2.2° RMS, for stimuli presented along the horizontal and vertical at (0°,±5°).

The study showed that measuring the head was possible using motion tracking and that accurate data could be collected. Analysis of the data from this study was easily done due to the nature with which the participant indicated they had localised the sound i.e. a button press. A number of other studies also used this approach [40, 39]. Issues with this method are that it cannot be easily used with children, especially young infants (less than one year), who cannot be instructed to press a button when they have turned to a sound. Other studies have used different techniques to analyse the data. Brimijoin et al. [41] post-processed the results by marking the point in each head trace where the target was localised using a special software GUI (head location taken as $\pm 3^{\circ}$ of the final head resting position). This approach could be used with children, however, the task is very labour intensive and would not be practical in a clinical setting. Recanzone et al. [39] used a time window which found the point of localisation as 1950-2000ms after the onset of the sound. Again, although computationally simple and faster than the human analysis demonstrated by Brimijoin [41], the method relies on each subject responding consistently each time over many trials, something which is not possible when working with children.

Goldring [95] devised a method in which signal processing techniques were applied to the collected data so that the start and end of the head turns as well as the point at which the sound was localised could be extracted. A method for extracting the start and end of the head turn based on the velocity of the head turn towards the target sound source was developed. This study did not use motion tracking technology but a more rudimentary method involving a hockey helmet attached to a potentiometer via a universal joint to measure the rotation of the head. The voltages from the potentiometer were collected and stored during each trial and processed after the experiment to define the start of the head turn and the fixation point to the target sound source. Velocities exceeding 15°/s were classified as the start of a head turn. The author reports *'head and gaze movements were determined, based on position and velocity criteria'*, however, these criteria are not explained. The study was not investigating localisation ability but instead the eye-head gaze shifts towards visual and auditory targets. Populin [35] used a similar approach to Goldring [95], measuring head rotation using small metal loops and electromagnetic receivers (identical to those used in eye tracking experiments) to measure the rotation of the head to a target sound source. The start of the head movement was defined by first measuring the 'baseline' velocity, which is defined as the mean velocity of the head 100ms before and 10ms after stimulus presentation. The point at which the sound was localised (also marked as the end of the head turn) is defined as the point where the head velocity returned to within two standard deviations of the baseline velocity. Using this method, the end of head turn could be extracted quickly and easily from a large data set. None of the methods discussed errors which might occur from measuring localisation ability this way.

Measurement of Head Turns in Children

Adult studies have shown that it is possible to motion track head responses to target sound sources and use signal processing techniques to evaluate localisation ability from this data. One of the advantages of using head motion as a method of pointing to a target sound source is that it requires no training as it is a natural orientation response and has been shown to be present in children who are just a few hours old [94]. For this reason head pointing is an ideal method of measuring absolute sound localisation ability in younger children. Clarkson et al. [51] devised a method which involved placing surgical tape on the newborns forehead. Printed on to the surgical tape were vertical dashed lines spaced 0.5 inches apart. During the procedure the childs' face were recorded with a video camera, which faced the children head on. Responses were measured by looking at how many of the dashed lines were obscured (out of the cameras view) as a result of the infants movement. Using this method the study found that for sounds played at $\pm 90°$, the average error was $33.6°$. This method was criticized by Morrongiello et al. [96] who saw the measurement technique as inaccurate. The argument for this was that the placement of the strip in the center of the head could be off center, any amount of off centering of the head would also translate into measurement errors. To overcome this, the same technique was used but instead of multiple marks being placed on the strip, just one was used which marked the middle of the child's head. The same procedure and measurement technique as Clarkson et al. [51] was then undertaken. This technique was used and data collected from both newborns [61] and children aged six to eighteen months [96]. The method shows an innovative and non-invasive method of measuring absolute localisation ability in young children. Nevertheless, the method of measurement is not ideal because it is very time consuming. The way of measuring the response is slow as each trial is measured by hand using a protractor and a TV screen playing the trails back. This would allow factors such as the distance of the child from the screen to influence the accuracy of the measurements. However, accuracy is reported by Morrongiello et al. as being around 2°.

4.1.2 Development of Techniques to Evaluate the Localisation Ability of Young Children

Chapter three discussed the development of a behavioral test (*AnimalSeek* method) that could be used to evaluate localisation ability in young children. This chapter will discuss the development of signal processing techniques which shall be used to extract salient information from the raw motion tracking data. Discussed in this section were methods, which have been used to attempt to evaluate localisation ability. Although many of the adult studies have methods which can collect and evaluate localisation ability, none of them are practical with young children. Furthermore, the methods used with children require analysis on a trial by trial basis using crude and time intensive measurements. This chapter will discuss the development of techniques which can evaluate localisation ability in a fast and more accurate way. It is hoped to develop methods which can ultimately be used in real time, providing the localisation error on a trial by trial basis.

4.2 Experimental Data for Head Turn Analysis

4.2.1 Child Participants

The participants discussed in this section were the same as those used in chapter three. Whilst the method was being assesed, localisation data was being collected. During testing, assessment of the method as well as the collection of motion tracking data was made. The following section will discuss the use of the motion tracking data.

The motion tracking gear consisted of a hat, a jacket and two wrist bands. Placed onto these garments were motion tracking markers which provided six degrees of tracking (see Figure 2.4). The garments did cause issue with some of the children, especially those in the age group 1.0-2.2 years. Some of the children would refuse to wear the hat or insist on taking it off after only a few trials. As a result limited motion tracking data is available for those age groups. Collection of the motion tracking data and the development of the *AnimalSeek* method (Chapter three) took place at the same time. If the tracking gear worn by the children caused them distress and made them not able to perform the *AnimalSeek* game, the garments containing the motion trackers were removed and the child continued with the *AnimalSeek* method without the collection of the motion tracking data. Table 4.1 shows the number of children in each age group from which usable data was obtained.

Age (years)	N(total)	Mean age in years (SD of ages)
1.0 to 2.2	4 (11)	1.65 (0.36)
2.0 to 4.0	8 (9)	2.90 (0.33)
4.0 to 5.0	8 (8)	4.22 (0.38)

Table 4.1: Number of participants from which adequate tracking data was obtained. It can be seen that compliance issues were with the youngest age group, 1.0-2.2 years.

Response screening

Initial analysis was on responses marked as a correct head turn towards the target stimulus. This analysis showed large errors and variance across all age ranges, something that was not expected after the testing sessions. For this reason the data were screened so that erroneous responses which were marked as correct responses were rejected. During the piloting stage of the experiment, data were collected for just over two seconds after the onset of the auditory stimulus. This approach was taken as the childs' natural response to the sound was wanted and be within the first few seconds after sound presentation. For some of the children this two second window was too short and as a result they would start their head turn either after the two seconds or near the end of the two seconds. These responses were still scored as correct by the human observer but the data does not contain adequate information (due to the tracking time) because the head turn is not fully present within the time window. This makes analysis with the developed algorithms impossible. The criteria for a response is that both the start and end of the head turn is clearly visible within the two second time window. By removing responses which did not meet this criterium and also trials where data from the tracker was simply noise, the results showed lower variance and more accurate responses as expected. Unfortunately, this did reduce the overall amount of responses available for analysis, especially when looking at individual children.

Number of responses

As a result of the children not wearing the hat and also after the post processing of the erroneous responses, equal numbers of responses were not gathered from each age group, direction and child. Table 4.2 shows the total number of collected head turns and the number of usable responses after screening them by hand. For some of the children the number of responses per sound source was very low. Some, after screening, showed zero responses to particular sound sources. For this reason analysis was performed on pooled data across all participants of the particular age range.

Direction (○)	Age (years)			
	1.0-2.2	2.2-4.0	4.0-5.0	Adult
-70	22	38	45	24
-30	10	14	13	24
+30	17	22	19	24
+70	12	33	46	24
Total Responses	61	107	123	96

Table 4.2: Number of correct responses for each direction and age group.

4.2.2 Adult Participants

Data was collected from four adult participants (three males, one female, mean age (SD): 26(3.3) years). The adults were tested so that they could be used as a comparison to the child data. Subjects' hearing were screened using conventional audiometry for hearing thresholds within 20dB SPL of audiometric zero at frequencies between 0.5kHz and 8kHz. Adult participants were naive to the study and took part in the same task as the children. A simple instruction of 'turn your head to where you hear the sound' was given. The adults were told to follow the animated character back to midline after which the next trial started. The reward type used was 'at location'. The same speaker locations were used as in the children's experiment ($0°$,$\pm30°$,$\pm70°$) stimuli were presented from each location six times during a single block. The adults did only one block, this resulted in 24 response for each direction.

4.3 Analyzing and Extracting Head Turn Information

Discussed in the chapter's introduction (see page 87) were methods used to evaluate absolute localisation ability in both adults and children. Although most of these methods used motion tracking technology to evaluate responses, their are three main methods used to extract the point at which the sound was localised. These methods are, the participant reporting the location via a button press, manual observation (someone hand judging responses) and via signal processing techniques based on criteria of the motion. This section will discuss the development of an extraction algorithm which is designed to take in the raw head turn data from the motion trackers and extract the point at which the participants localised the sound. The extraction algorithm MkII developed will also be used in the automatic classification of responses, discussed in chapter five (see page 138).

Ideally, the extraction algorithm would operate in real time and allow the analysis of responses on a trial by trial basis. In order to develop the method, all results presented in this section were all analysed offline, i.e. the trials were not processed whilst the children were making responses but instead recorded and then processed at a later date. Data was collected and processed offline so that an understanding of what the responses data would look like could be established and the method of extracting them developed fully.

4.3.1 Description of a Head Turn

An example of a 'correct' head turn is seen in Figure 4.1. The parameters which are to be extracted by the algorithm are also shown in Figure 4.1.

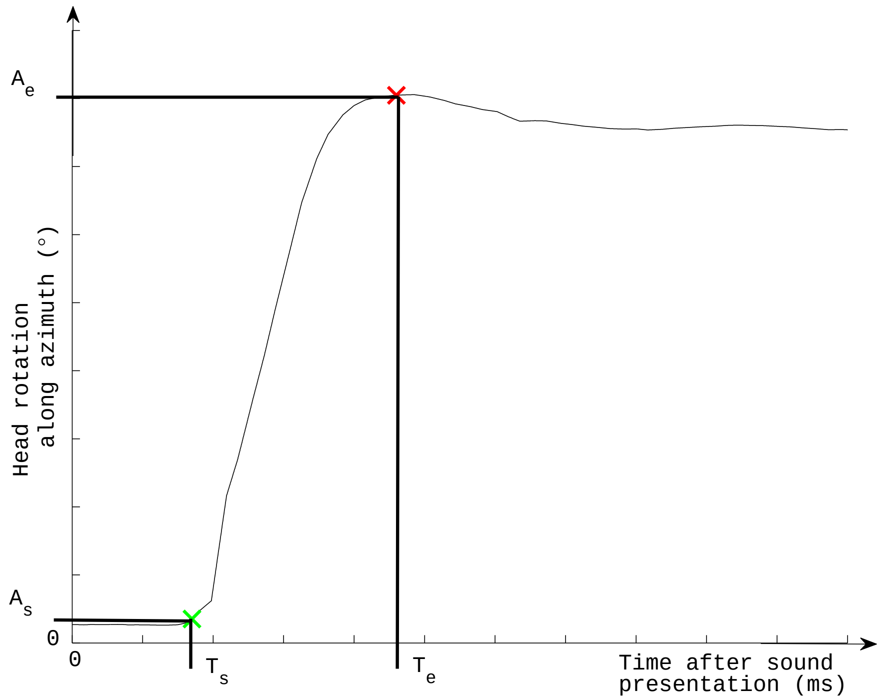


Figure 4.1: A typical head response as seen in adult participants. The x-axis shows the time, in milliseconds, after the presentation of the sound. The y-axis shows the rotation of the head along the azimuth from midline. The start and end of the head turn (green and red crosses respectively) can be visibly seen on the responses. The points which are to be found are the times of the start and end of head turn, T_s and T_e respectively, and also the direction the head is pointing at the start and end of the head turn, A_s and A_e respectively. Using these parameters, the localisation ability of the child can be measured.

T_s and T_e show the start and end of head turn respectively. A_s and A_e define the value of the head rotation along the horizontal (azimuth) at times T_s and T_e respectively. Extraction of these points would allow for the evaluation of the start and end of head turn, these can be in turn used to evaluate where the sound was localised.

4.3.2 Proposed Extraction Algorithm

The motion data was read from the motion trackers at a rate of 100Hz. Due to the nature of the serial port however, data could not be streamed at this speed to a text file. Instead the data was read for approximately two seconds, stored into a buffer, and then placed into a text filed at the end of the trial. Each sample was placed onto a new line of the text file. For analysis, the data was transformed using MATLABs interpolate function (interp1.m) to transform the data from the samples in the text file into milliseconds so it could be used for analysis. Each head turn was tracked for exactly 2262 milliseconds. All the data points are made absolute (sign removed) before they are processed, to prevent any issues which resulted from the instabilities of the motion tracking gear the polarities they reported.

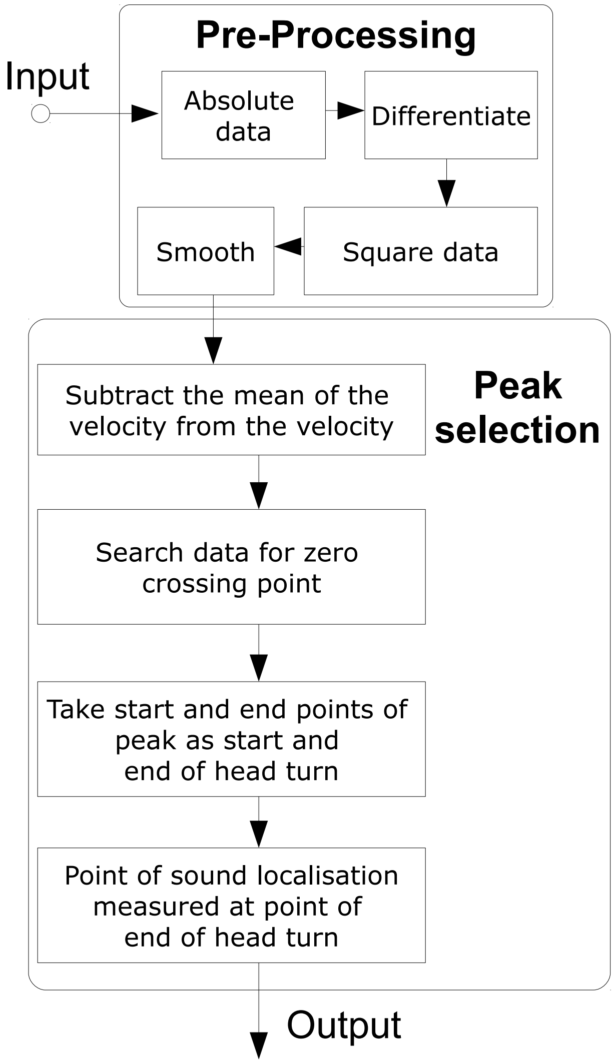


Figure 4.2: Flow diagram on the method of extraction used. The data is first 'pre-processed' in which the raw head turn data is differentiated to obtain the velocity of the head turn, it is then squared so that the head turn peaks in the velocity vector are amplified and the smaller, random head movements are suppressed. After this the data is filtered using a moving average filter to suppress any further noise. The peaks in the data are then used as the determinants of the start and end of the head turn. A zero crossing algorithm is used to find the start and the end of the head turn by determining when the head velocity goes over and then falls back down over zero, these points are taken as the start and end of the head turn respectively. The angle of the start and end of head, A_s and A_e respectively, is taken as the head location along the azimuth (°) at T_s and T_e respectively.

To extract the start and end of the head turn, the velocity of the head turn is required. To do this, the first derivative (approximated by the difference between adjacent samples divided by the sampling interval) of the head position over time is found. The peak of the velocity vector represents the middle of a head turn, i.e. the fastest point of the movement. The T_s and T_e of the head turn can be defined as the point at which a change in velocity goes above (start of head turn) or below (end of head turn) zero, these points can be found by using zero crossing techniques.

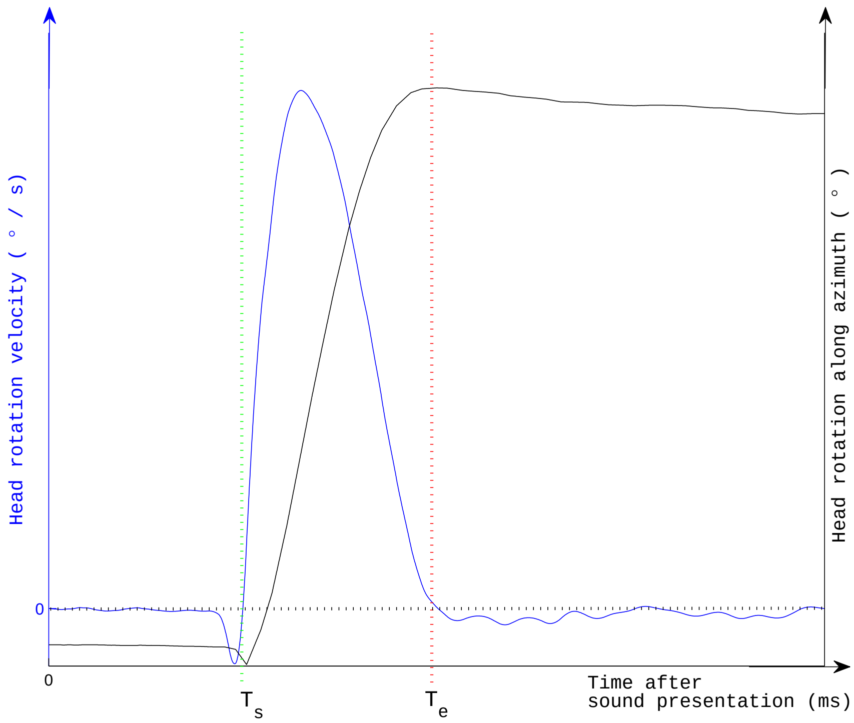


Figure 4.3: Typical head turn showing both the rotational movement (black) and also the velocity of the movement (blue). The start and end of head turn (T_s and T_e respectively) are seen at points when the velocity vector crosses zero. These points are used as the start and end of the head turn. The data shown has been through the pre-processing stage of the algorithm i.e. it has been squared and also smoothed.

Before zero crossing is preformed on the data, it is first squared and smoothed. Ideally, a correct responses should exhibit one peak in the velocity vector corresponding to a turn to the sound source. More peaks in the velocity vector are due to random head movements, the response to the sound source is usually the the largest peak in the velocity vector. The data is squared to enhance the biggest peak corresponding to the head turn and suppress any of the smaller peaks present in the data. The data is smoothed using a moving average filter (-3dB at 5Hz, 'filtfilt.m' function used in MATLAB providing a zero phase delay). The data was smoothed because, although the data is already low pass filtered by the head tracker using an inbuilt hardware filter, it was found that noise was still present in the raw traces. Once squared and smoothed, the mean velocity of the data is taken away from the velocity vector, this removed any offset and is required for the method to work successfully. Zero crossing is implemented by looking for the point where the velocity changes magnitude, i.e. crossing from a negative to a positive velocity, or positive to negative, representing the start and end of head turn respectively. If no peaks could be found in the data, T_s and T_e were set at one millisecond and 2262 milliseconds respectively. This stopped the analysis program from crashing when analysing large data sets and did not get rid of any more trials because numbers were already low. The number of trials scored this way was 0.8% (two trials) for the correct head turn data.

4.3.3 Evaluation of the Extraction Algorithm

Accuracy at finding the start and end times of the head turn

To evaluate the effectiveness of the extraction algorithm, all raw data from all age groups (see Table 4.2 for numbers) were judged by hand by the author to find the start and end of a head turn. This was done by visual inspection of all the motion tracking data trial by trial and marking T_s and T_e, this took several seconds per trial. A_s and A_e were obtained by finding the corresponding head angle at T_s and T_e respectively. This data was then used as a comparison to the extraction algorithm. The difference between the start ($T_{s,ha}$) and the end ($T_{e,ha}$) were calculated as shown in Equation 4.1 and Equation 4.2 respectively. It is important to get the extraction algorithm as close to the human judged data as possible so that accurate measurements of the head angle can be made.

$$T_{e,ha} = T_{e,h} - T_{e,a} \tag{4.1}$$

$$T_{s,ha} = T_{s,h} - T_{s,a} \tag{4.2}$$

Where $T_{e,h}$,$T_{s,h}$ are the end and start of the head turns as picked by a human observer and $T_{e,a}$,$T_{s,a}$ are the end and start of the head turn as picked by the algorithm. It is noted that the data is not made absolute as the sign of $T_{e,ha}$ and $T_{s,ha}$ indicate if the algorithm is either under- or overshooting the hand picked location.

Age Group	Start(ms) M(SD)	End(ms) M(SD)
1.0 to 2.20 yrs	-647.2 (+644.4)	-131.2 (+540.9)
2.20 to 4.0 yrs	-703.0 (+684.9)	-189.5 (+508.3)
4.0 to 5.0 yrs	-659.1 (+632.9)	-116.0 (+484.8)
Adult	-140.6 (+173.5)	+225.4 (+135.2)

Table 4.3: The time differences between the hand picked start and end of head turns and the ones extracted using the algorithm. The results show large mean and standard deviations for the start of head turn extraction. This error is large across all of the children but lower in the adult data. The end of head turn can be seen to be relatively small of the adult and children, however, the standard deviation of the data is large for the children and lower for the adults.

Table 4.3 shows the mean and standard deviation of the time difference between the hand judged responses and the extraction algorithm. The algorithm is required to be as close to the hand judged responses as possible. The data shows that the algorithm can find the end of the head turn for the children within 200ms, however, the consistency of the responses, i.e. the standard deviations, are large at around 500ms. This suggests that the extraction algorithm is not consistent in finding the end of the head turn. For the start of the head turn the accuracy is very poor with the average error for the children age groups being greater than -600ms.

For the adult data, the detection of the end of the head turn accuracy is poor (+225.4ms), the adult data however, does show a lower standard deviation, showing that the judgment of the extraction algorithm is more consistent. $T_{s,a}$ for the adults is low (-140.6ms) suggesting it is accurate at finding the start of the head movement in adults. The reason for the slightly better performance in the adult data is due to the consistency of the head turn response data, i.e. less noise, unlike the child data.

By looking at the magnitudes of $T_{s,ha}$ and $T_{e,ha}$, the algorithm can be seen to be either under- or overestimating the point where the human observer judged the response. For all the child age groups, the algorithm is overestimating the end of head turn. However, for the adults the algorithm is underestimating. For the start of a head turn, all groups tested show an overestimation. The adult data shows that the algorithm is consistent in misclassification (high errors and low standard deviations) suggesting it is not good at finding the start and end of the head turn.

Accuracy at finding the start and end angles of the head turn

To find the angular difference between the human judged and algorithm judged responses, Equation 4.3 and Equation 4.4 are used.

$$A_{e,ha} = A_{e,h} - A_{e,a} \tag{4.3}$$

$$A_{s,ha} = A_{s,h} - A_{s,a} \tag{4.4}$$

The main reason the extraction algorithm was developed was so that it could automatically evaluate the localisation ability of the child. The point at which the sound is localised, A_e, is taken as the head's location at time T_e. Using Equation 4.3, the difference between the extracted end of head turn and the hand picked end of head turn can be computed. The differences in the data is plotted on a histogram showing $A_{e,ha}$ against the number of occurrences.

Figure 4.4 shows a histogram showing the accuracy between the two methods and if the algorithm is under- or overshooting the hand picked locations.

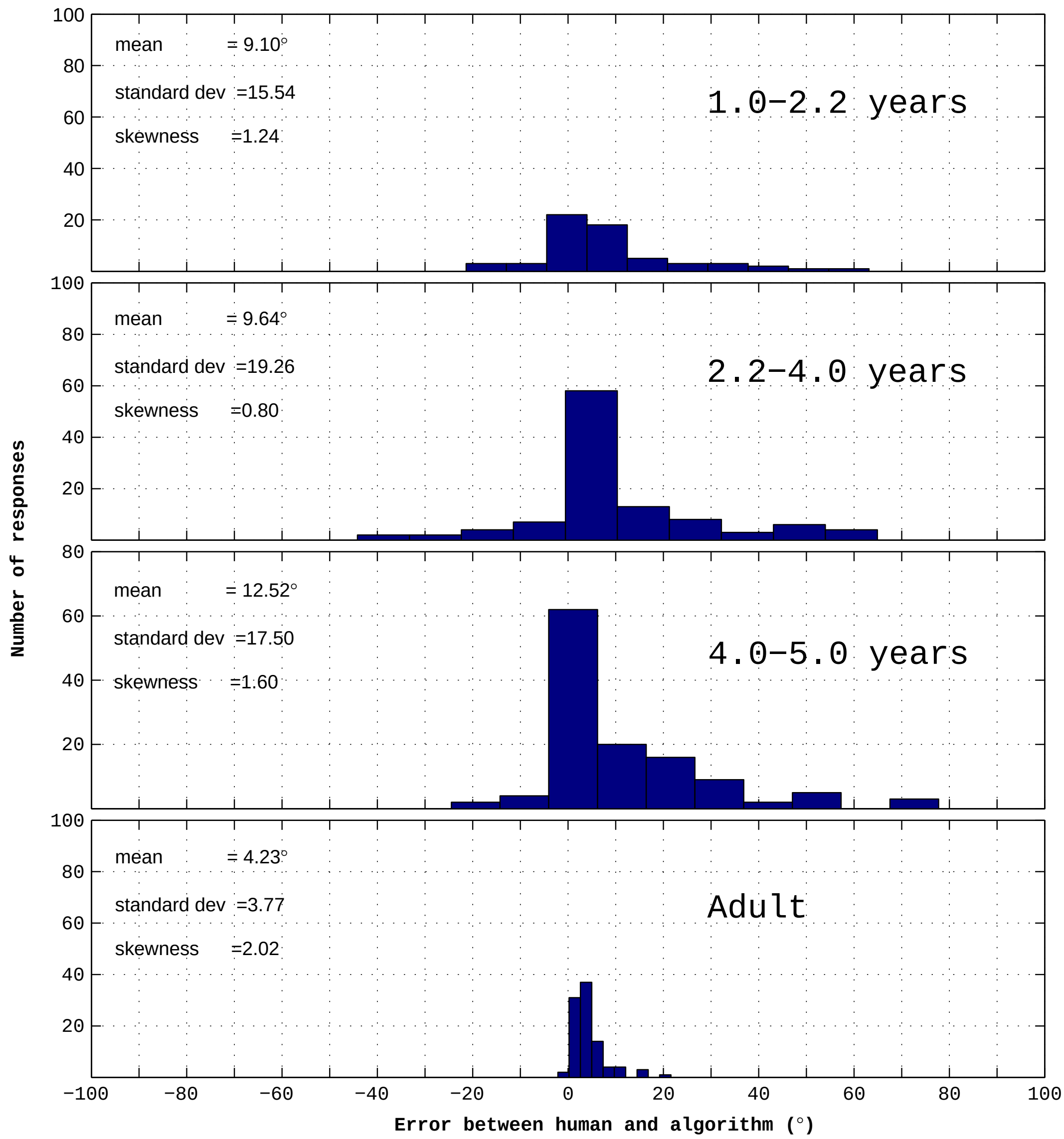


Figure 4.4: Distribution of the difference between the manual and algorithm estimation. Results show large differences (mean and standard deviation) across all of the age groups. The positive skewness of the histograms about zero show that the algorithm is underestimating the point at which the child localizes the sound.

For the child data we see the algorithm gives large errors in finding the end of the head turn. The largest difference is seen in the the oldest age group of 4.0-5.0

years (mean - 12.5°). The other age groups show an error of just under 10° but with large standard deviations (19.3° and 15.5° for the 2.2-4.0 years and 1.0-2.2 years respectively). The adults show the lowest mean error and also the lowest standard deviation, 4.3° and 3.8° respectively. One way of understanding why the algorithm is going wrong is to look at the skewness of the data, i.e. is the algorithm consistently over- or undershooting the hand picked values (skewness calculated using the skewness.m function in MATLAB). All the data shown in Figure 4.4 had positive skew, i.e. the algorithm is consistently undershooting the hand picked location.

4.3.4 Issues with the Proposed Extraction Algorithm

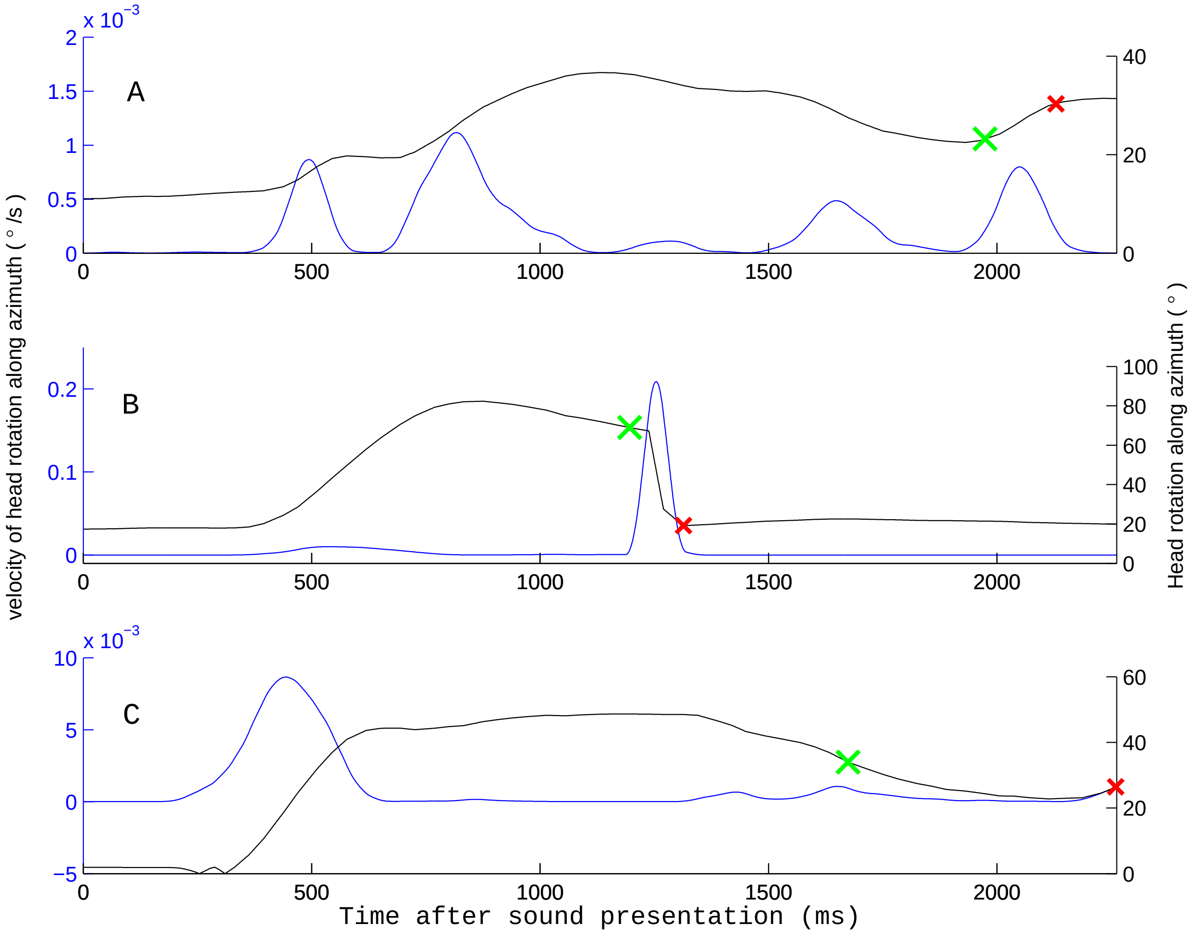


Figure 4.5: Three examples of the extraction algorithm failing to classify the correct head turn. Shown on the figure is the head position over time (black) and also the head velocity (blue). A: shows multiple head movements to and away from the target stimulus. B: shows a fast response to mid-line and C: shows a slow drift. Because the method is only assuming one head turn, all the data is analysed and the last peak chosen as the head turn. This is obviously not correct if multiple head turns are in the data.

Figure 4.5 shows three correct head turn responses, each from a different age group (**A** = 1.0 - 2.2 years, **B** = 2.2 - 4.0 years and **C** = 4.0 - 5.0 years). The figure shows both the head displacement along the azimuth (black) and the head velocity (blue). By looking at the head rotation along the azimuth (black) it can be seen that the responses have a number of movements most of which do not correspond to the localisation of the sound. Marked in green and red crosses are the calculated start and end of head turn respectively. Response **A** shows several different head movements which makes choosing the correct head turn to the sound source difficult, the current method simply looks for the largest peak, **B** shows a fast return to midline after sound localisation, whilst **C** shows a slow return to midline after sound localisation. The algorithm is failing because it scans all the data and picks the start and end of head turn based on the zero crossing. The algorithm is storing the last peak it sees as the head turn towards the target. When differentiated, a head turn response consists of several small head movements which do not correspond to the child turning to the sound source (as shown in Figure 4.5). When processing a trial it is necessary to choose which of these head turn are correct, something the current extraction algorithm (MkI) fails to do. These errors in processing is also why Figure 4.4 shows the algorithm overestimating the hand picked location.

4.3.5 Improving the Extraction Algorithm - Extraction Algorithm MkII

To overcome the problem of the algorithm selecting the wrong peak, the peaks are selected based on a criterium which corresponds to movement towards the stimulus. Figure 4.5 showed three typical head turns seen in the child data and depicts how the algorithm failed to classify these because the correct peak was not chosen. To choose the correct peak a set of criteria must be used. Panel **B** in Figure 4.5 shows a large velocity peak as the head returns back to midline. It can be seen in panel **B** that such a head movement can cause misclassification of a head turn. To suppress such peaks, criteria based on their width can be used. The data presented in this thesis used a cut-off of 100ms. It was found that if the cut-off was made too small, the extraction algorithm would continue to misclassify such head movement peaks, however, if made too big, all head movements would be suppressed. To eliminate the issues seen in panels **A** and **C** the data must be processed further based on the difference in angle between the start and end of head turn. For panels **A** and **C** it can be seen that the largest peak corresponds to the head turn. Based on this, the peak that is chosen is the largest peak. It can be seen this works well for panel **C**. However, for data in panel **A** the start of the head turn is still not 100% accurate due to the child's head rotation stopping and then turning again. It is notable that the end of head turn is picked out correctly, which is important when trying to evaluate localisation ability.

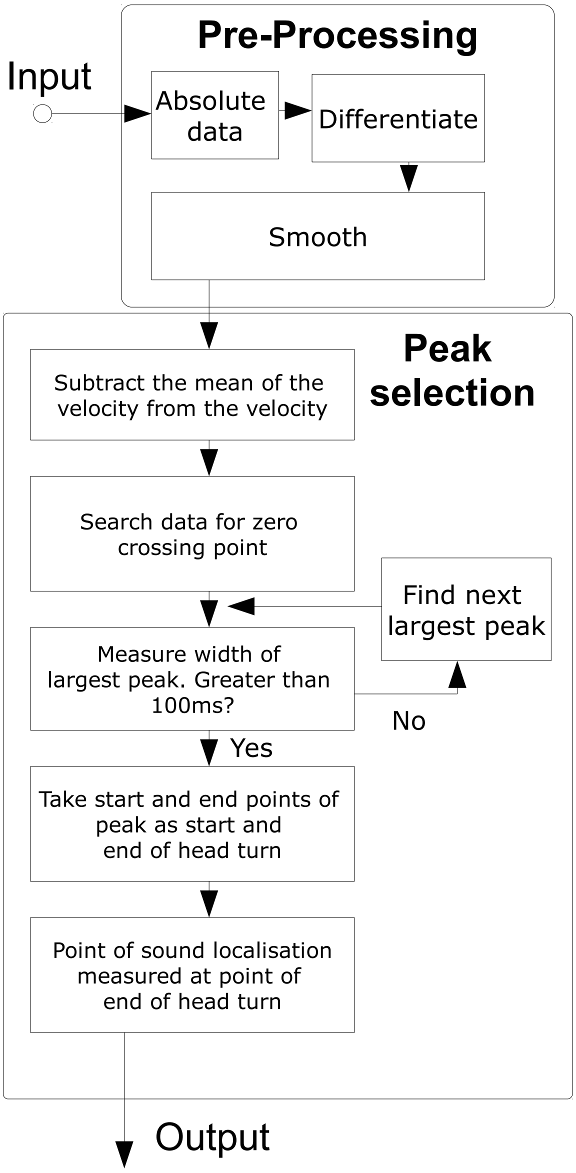


Figure 4.6: Flow diagram on the refined extraction method. The algorithm uses a different pre-processing to version one of the algorithm, instead of squaring the data to suppress the small head movements and amplify the big ones, all peaks are kept as they may correspond to a head turn towards the sound source. One the data is pre-processed and the zero crossing points found, the width of each peak is measured, using zero crossing, and the largest peak is measured to see if it is wide enough. If it is not, then the next biggest peak is found and so on until the largest peak with a width of greater than 100ms is found. This peak is then analysed for the start and end of it, which corresponds to the start and end of head turn respectively.

Figure 4.6 shows a flow diagram of the modified extraction algorithm which searches for the correct peak based on their widths and magnitudes. The data is pre-processed the same way as the previous algorithm. However, instead of squaring the data, it is simply made absolute, differentiated and then smoothed. The data is no longer squared as all the peaks in the data want to be preserved for analysis. Zero crossing is used to find the start and end of each peak. Once all the peaks in the data are found, as well as their corresponding start and ends, the largest peak is found and the difference in time between the start and end measured. If this is greater than 100 ms then it is taken as the start and end of the head turn towards the sound source. If it is not then it is rejected and the next largest peak used, the width of this is checked and so on until the largest peak with a width of more than 100ms is found. Figure 4.7 shows the results obtained using extraction algorithm MkII.

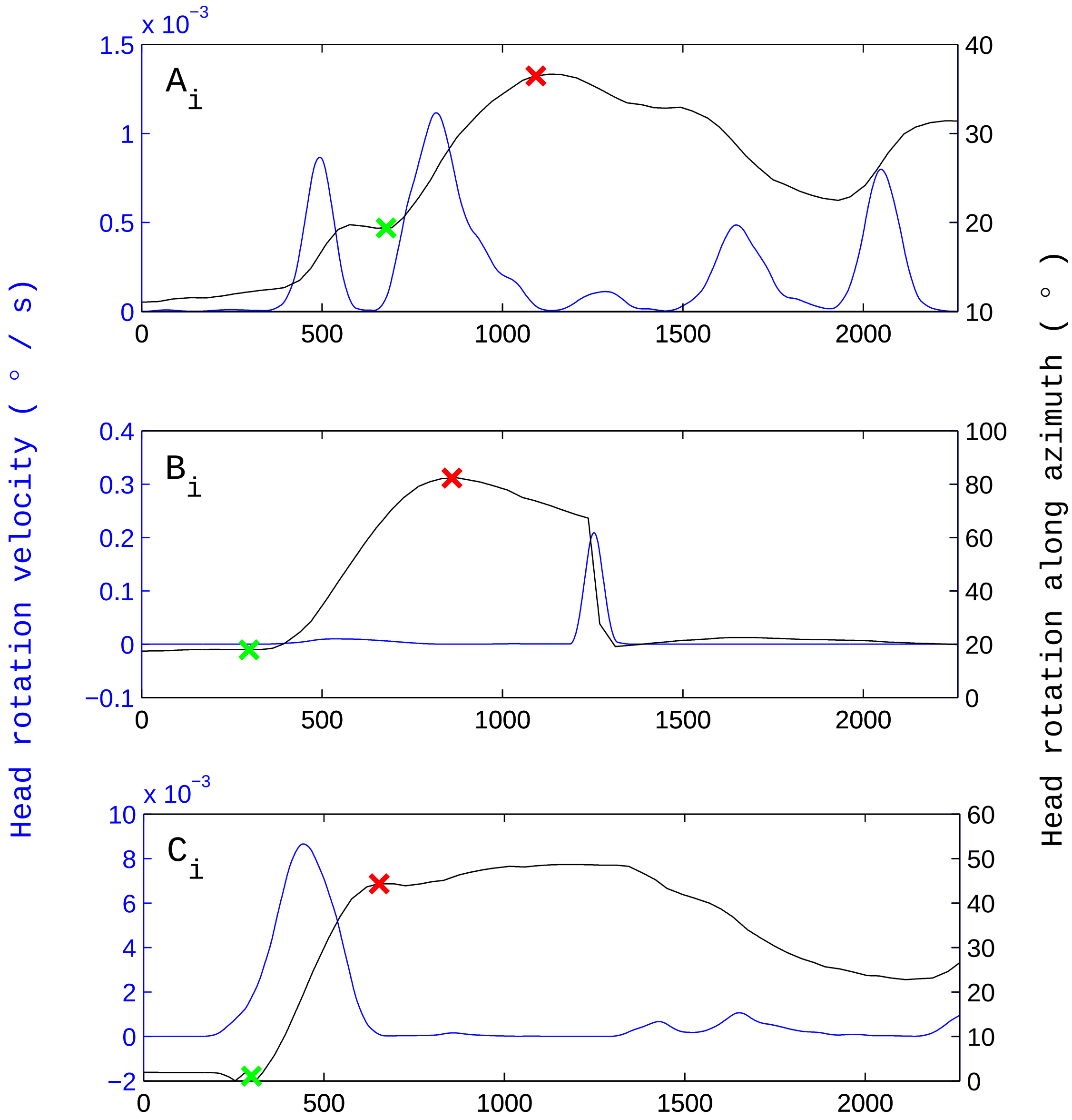


Figure 4.7: The improved algorithm showing the start (green) and end (red) points (×) of the head turn.

4.3.6 Evaluation of Extraction Algorithm MkII

Age Group	Start(ms) M(SD)	End(ms) M(SD)
1.0 to 2.2 yrs	-87.9 (+344.9)	+119.9 (+330.2)
2.2 to 4.0 yrs	-40.2 (+398.7)	+24.9 (+317.5)
4.0 to 5.0 yrs	-103.8 (+289.0)	+69.2 (+304.2)
Adult	-77.4 (+60.7)	+126.0 (+94.4)

Table 4.4: The table shows the differences between the human judged and the improved extraction algorithm. The results show that both the start and end of head turn can be found with reasonable accuracy with reduced means and also reduced standard deviations. The adult data shows very good agreement with the start of head turn mean and standard deviation being less than 100ms. The end of head turn both are within 150ms. The child data show low means with acceptable levels of standard deviation.

Table 4.4 shows the improvement in the results at finding the start and end of the head turn. The table shows that there is decrease in difference in time between the human observer and the refined extraction method (<150ms across all age groups) and also a reduction in the size of the standard deviation (<350ms across all age groups) for both the start and end of head turn. The adult data show good accuracy for the start and end of head turn with all means and standard deviations showing a large improvement over the first algorithm, however, it can be seen that the end of head time is larger than the child data. The standard deviations, however, are the lowest of all the age groups. The increase in accuracy of finding the end of head turn can also be seen in an increase in accuracy at finding the point at which the sound is localised.

Figure 4.8 shows a histogram for each age group calculated using Equation 4.3 and Equation 4.4. Compared to Figure 4.4 the results show a reduction in error between the human observer and the extraction algorithm, the adults show a mean error of 1.° and the worst performance (age group 1.0-2.2 years) having a mean error of 3.4°. The standard deviations are also smaller across all of the age groups, <16° for the children and only 1.5° for the adult data. From measuring the skewness of the data we see that the data for the children is overestimating (negative skewness) the responses angle. For the adult data, the algorithm is underestimating the angle (negative skewness). The level of skew is reduced when using extraction algorithm MkII.

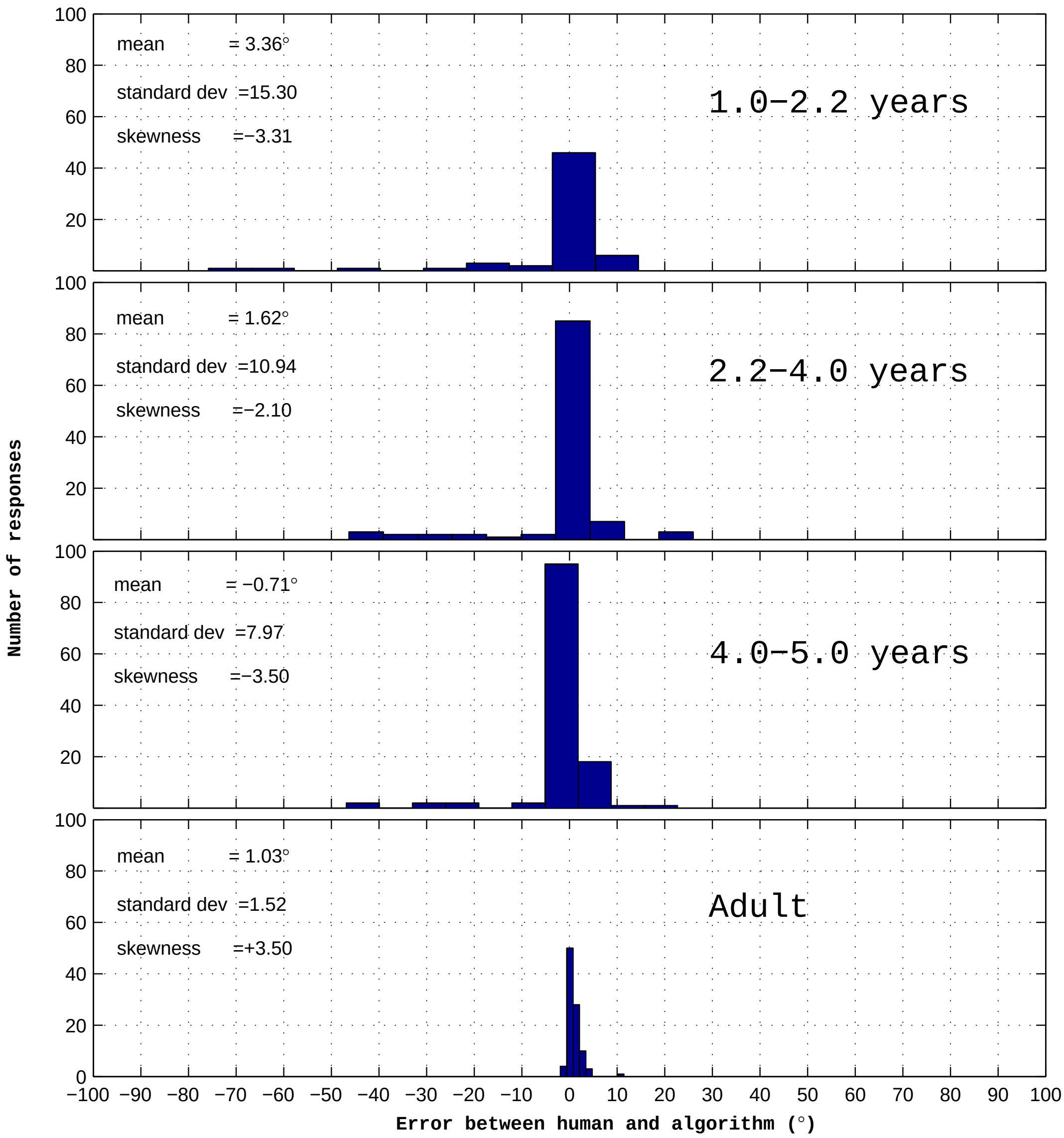


Figure 4.8: **Figure shows the difference between the human judged responses and the new, improved extraction algorithm for all four age groups.All results show high accuracy with means less than 6° across all age groups and considerably smaller standard deviations than the first extraction algorithm.**

4.3.7 Discussion of Extraction Algorithms

This section sets out to discuss the automatic extraction of parameters from the raw head tracking data. The extraction algorithm is capable of extracting the start and end of head turn as well as the point at which the sound is localised. Due to the motion tracking data containing small head movements not necessarily to the target sound, this was not a trivial task. With the accuracy of the algorithm showed a significant improvement once the peak selection was incorporated. The accuracy across all age ranges improved. The algorithm still shows some errors and in some cases complete misjudgment, this can be seen when the error between the two is large i.e. greater than 30°. The errors are quite low as compared to the time saved by the method, the method takes around twenty milliseconds per trial where as the hand picked method takes around four seconds per trial.

The next section will apply extraction algorithm MkII to the localisation data and use the results to assess the localisation ability of the four age groups.

4.4 Evaluation of Localisation Abilities Using Extraction Algorithm MkII

Error Measures

Results in this section are discussed in terms of medians, interquartile ranges (lower quartile - 25%, upper quartile - 75%) and mean absolute error (MAE, Equation 4.5). MAEs are used for the child data instead of root-mean-square errors (RMS, see Equation 4.6) due to the nature of the child data. The RMS error calculation squares the data before it is averaged, this gives large weighting to one-off, large errors [63]. Such errors exist often in the child data due to moments of inattentiveness and as a result give large errors in the measurments. For this reason, MAEs are used in all the calculations presented in this results section.

$$MAE(^\circ) = \frac{1}{n}\sum_{i=1}^{n} |target_i - response_i| \tag{4.5}$$

$$RMS(^\circ) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(target_i - response_i)^2} \tag{4.6}$$

4.4.1 Child Localisation Ability

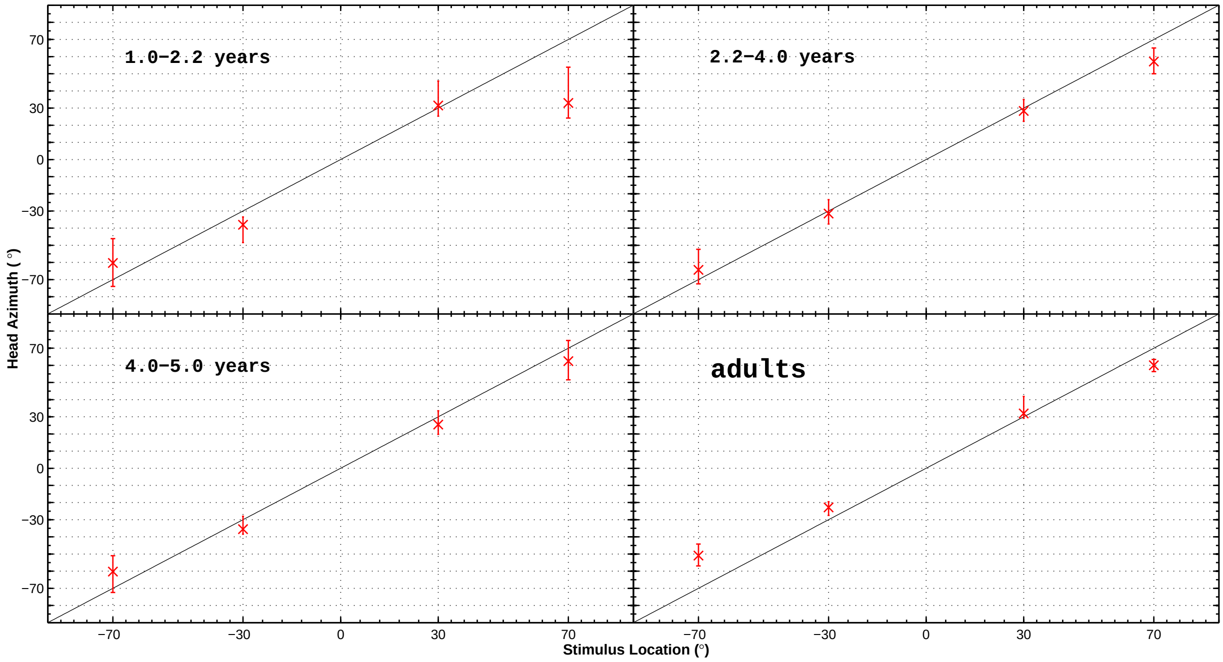


Figure 4.9: Localisation ability for each age group, 1.0-2.2 years, 2.2-4.0 years, 4.0-5.0 years and adults. Localisation ability expressed in terms of medians (×) and interquartile ranges (lower quartile - 25%,upper quartile - 75%) for each speaker location.

Age	Speaker location			
	-70°	-30°	+30°	+70°
	MED (IQR)	MED (IQR)	MED (IQR)	MED (IQR)
1.0-2.2 years	-60.3 (27.8)	-38.0 (14.9)	+31.5 (20.4)	+33.0 (29.7)
2.2-4.0 years	-64.3 (20.2)	-31.5 (14.1)	+28.3 (12.7)	+57.2 (15.0)
4.0-5.0 years	-60.2 (21.5)	-35.5 (10.5)	+25.5 (13.8)	+62.4 (22.9)
Adults	-50.9 (12.8)	-22.8 (8.0)	+32.0 (12.7)	+60.0 (7.1)

Table 4.5: Medians (MED) and interquartile ranges (IQR) obtained by the extraction algorithm for each age group and speaker location.

Age	Speaker location			
	-70°	-30°	+30°	+70°
	MAE(°)	MAE(°)	MAE(°)	MAE(°)
1.0-2.2 years	15.6	15.3	12.1	32.0
2.2-4.0 years	13.1	6.6	7.0	13.3
4.0-5.0 years	12.9	7.5	8.1	14.2
Adults	19.4	8.3	7.6	10.1

Table 4.6: Mean absolute errors (MAE) of the data obtained by the extraction algorithm. The MAEs are shown for each age group and speaker direction.

Age group, 1.0-2.2 years

The top left panel of Figure 4.10 shows the localisation responses of the age group 1.0-2.2 years. The responses to the inner angles target locations, -30° and +30°, have response medians of -38.0° and +31.5° respectively. Both these medians show an overshoot of the sound source, especially responses to the negative speaker locations. The quartile ranges (+30°=14.9°,-30°=20.4°) and also the MAEs (+30°=12.1°, -30°=15.3°) for these two locations are large, indicating uncertainty in the children's responses.

Responses to the outer speaker locations, -70° and +70°, show medians of -60.3° and +33.0° respectively. The results for the outer angles indicate the expected undershoot. Response to the inner angles show that the localisation accuracy is poor, this is shown in terms of both the median of the responses and also the large interquartile ranges. However, this could also be due to the low numbers of trials collected not measuring localisation ability effectivly.

The MAEs for the outer angles (-70°=+15.6° and +70°=+32.0°) show that for the +70° sound source the MAE is the largest of all the age groups, the MAE for the -70° location, however, is lower than that found in the adult data. A reason for this could be due to the child turning their whole body to the sound source, rather than just the head. If the children were rotating their whole bodies, a smaller undershoot to the target and smaller MAEs would be seen.

The MAEs and quartile ranges are particularly large for the positive sound locations (+30° and +70°), the difference between the medians on the outer and inner angles on the positive side is around 1°. A line of best fit can be drawn between the medians of the outer angle and the inner angle on both sides of responses to gauge the level of undershoot. If responses were 'ideal' to the positive angle locations, the line would have a gradient of one. The line of best fit shows that for the positive angles the gradient is very shallow, +0.17. The large errors, and also the shallow gradient, suggest that the children of this age group may not be localising but simply discriminating sound sources on their right-hand (positive speaker locations) side.

This can also be seen by performing statistical tests (Mann-Whitney U-test - non-parametric independent two sample test) because of the pooled data on the positive and negative speaker pairs to see if they are significantly different, i.e. if the distribution of response to +30° difference to the distribution of responses to +70°. The tests confirm that the response to the negative speaker locations (-30°,-70°) are significantly different ($U = 91, z = 3.0, p < 0.05$). The positive angles (+30°,+70°) do not differ significantly ($U = 190, z = 0.4, p > 0.05$). This suggests that for the sources on the right, the children were simply discriminating and not localising.

A number of studies have looked at MAEs in infants who are just a few days old. Clarkson et al. [51] found MAEs ranging between 35-40° for targets presented at ±90°. As expected, these errors are larger than those of the current study. In a study with slightly older children (0.5 years) (Morringello et al. [62]), MAEs of 16.9° and 16.5° were measured for responses to sounds presented at 36° and 72° respectively. Errors were averaged over the positive and negative stimulus locations. The results for the target locations in this thesis were 14.8° for the inner angles and 22.1° for the outer angles. As the children got older, the MAEs dropped, for the oldest age group, 1.5 years, the MAEs were 5.9° and 6.6° for sounds presented at 36° and 72°, respectively.

The results show that the MAEs found in this study are comparable to those found for the 0.5 years age groups. The reason for this could be due to the accuracy of the data collected due to the low number of participants used in the 1.0-2.2 years age group and as a result the total number of trials collected being low. However, interestingly the results shown by Morriengello et al. [62] do not show an undershoot for the outer speaker locations. The MAEs for sounds presented at 18° and 90° are the same. This suggest the children turned their heads fully round, even to the outer speaker locations. For the results collected and presented in this study we did see the undershoot expected from adult studies even with such large errors present in the data.

Age group, 2.2-4.0 years

The top right panel of Figure 4.10 shows the localisation responses of the age group 2.2-4.0 years. The results for the inner speaker locations (±30°) show appropriate localisation accuracy with medians of +28.3° and -31.5° for sound sources at +30° and -30° respectively. This age group shows a smaller quartile range (-30°=14.1° and +30°= 12.7°) for the inner angles as compared to the 1.0-2.2 years age group, but it is larger than in the adult data. The MAEs for the inner angles (-30°=6.6° and +30°=7.0°) are considerably lower than those found in the 1.0-2.2 years age group and are within a few degrees of the adult data for the corresponding angles.

The outer speaker locations show medians of -64.3° and +57.2° for sound sources at respectively +70° and -70° respectively. The quartile range for the outer speaker locations (-70°=20.2° and +70°=15.0°) are smaller than the 1.0-2.2 years age group but larger than those found in the adult data. This suggests that this age group's localisation ability is less accurate for the outer target sound sources. The MAEs for the outer angles (-70°=13.1° and +70°=13.3°) are some what higher than the 1.0-2.2 years age group, but lower than the adult data. For the +70° sound source however, the MAE is lower than the 1.0-2.2 years age group but higher than the adult data.

The gradient of the least square-linear line-fit through the medians of the -30° and -70° sound sources is high, -0.84. One reason for this could be due to children of this age turning their whole bodies and not just their heads. This could explain why the MAE is lower (1°-2°) than the adult data. For the positive angles the gradient is lower, 0.56. A Mann-Whitney U-test on the responses for each location shows a statistical difference between the responses to the speaker locations on each side, positive ($U = 269, z = 6.0, p < 0.05$) and negative ($U = 128, z = 5.0, p < 0.05$).

The gradient on the negative side shows a consistent offset from the 'ideal' line. This suggests that the children were consistently undershooting the negative locations. As discussed in the adult data, this is not due to the extraction algorithm as it is also seen in the raw data. Issues may lie in the calibration of the motion tracking system, although during each session the head tracker was fully calibrated. A child moving the hat or adjusting it could introduce errors in the measurements. However, the children were monitored and if the hat was removed, the experiment was halted and the hat and marker placed back on correctly.

Comparisons with other studies for children aged 2.2-4.0 years is difficult due to the lack of literature on the this age range, this study has been the first to attempt such measurements. Overall the results show lower MAEs and quartile ranges than the 1.0-2.2 years age group. Comparison with the adult data show MAEs for the inner and outer angles the responses are comparable and in some cases lower. The quartile ranges compared to the 1.0-2.2 years age group are smaller, an absolute averaged difference across all directions of $8.7°$. Compared to the adult data, the quartile ranges are higher, an absolute averaged difference of $5.4°$. This suggests that although the children are, on average, as accurate as the the adults, the consistency of their responses is not. This could be due to the children having much shorter attention spans and as a result responding less consistently when being tested.

Age group, 4.0-5.0 years

The bottom left panel of Figure 4.10 shows the localisation responses of the age group 4.0-5.0 years. The medians of the responses for the inner angles show good agreement with the sound sources, with median responses of -35.5° and +25.5° for sound sources at -30° and +30° respectively. The quartile ranges (-30°=10.5° and +30°=13.8°) are only slightly larger than those found in the adult data. The +30° quartile range is slightly larger than the 2.2-4.0 years but comparable to the adult data.

Responses to the -30° sound source are comparable to the results found for the 2.2-4.0 years age group, however, the responses to this angle are generally larger than the adult data. The MAEs (-30°=7.5° and +30°=8.1°) for the inner sound sources are comparable to adult data and also the 2.2-4.0 years age group, with both MAEs for both sound sources being only a few degrees larger.

The outer angles show median responses of -60.2° and +62.4° for sound sources at +70° and -70° respectively. The quartile ranges (-70°=21.5° and +70°=22.9°) are higher than those found in the adult data and the +70° being higher than the same angle in the 2.2-4.0 years age group. The MAEs for the outer angles are -70°=12.9° and +70°=14.2°.

The MAEs for the 4.0-5.0 years age groups are the same as those found in the 2.2-4.0 years age group and the adult data, with the difference between the averaged MAEs across all directions being less than 1°. The mean quartile ranges are higher than the 2.2-4.0 years (2.3°) and higher than the adult quartile range, with a difference of 9.5°. The gradients for both sides show good agreement with the sound sources (positive gradients = 0.88, negative gradients = -0.75). The positive angle shows a consistent undershoot for +30° and +70°, this can be seen in the offset of the linear-line fit between the two sound sources. Comparisons with the 1.0-2.2 years age group show differences of around 8° for the MAEs and around 7° difference for the quartile ranges. Overall, the results show that the age groups 2.2-4.0 and 4.0-5.0 years have similar responses both in terms of MAE and quartile ranges. The age group 4.0-5.0 years, although showing similar MAEs to the adult data, has larger quartile ranges suggesting inaccuracies in their responses.

A number of studies have looked at sound localisation ability in children of this age range. However, many were not what has been defined in this thesis as 'absolute' sound localisation. Van Deun et al. [63] performed an identification task which consisted of nine visible loud speakers from which the child was to choose where a sound came from. The study found RMS errors of around 10° for children aged four years. The results presented above show slightly higher values than this. Litovsky et al. [97] conducted a similar identification task to Van Deun et al. [63] using speech stimuli. The results from seven five year olds found a mean RMS across all the subjects of 18.3°. This results is higher than those found in the 4.0-5.0 and 2.2-4.0 years age group. The results are more comparable to the 1.0-2.2 years age group, however, results are presented as

RMS and not MAEs so therefore not directly comparable. Another study by Litovsky [57], found that children aged five had performance levels equal to those of adults in a minimal audible angle (MAA) task. The results above suggest that the MAEs are equal and sometimes lower than the the adult data for the 2.2-4.0 and 4.0-5.0 years age groups. However, the quartile ranges are usually higher for the children, suggesting a lower consistency between responses, this could be due to the lower number of trials collected.

4.4.2 Adult Localisation Ability

The bottom right panel of Figure 4.10 shows the localisation responses of the adult participants. The results for the inner angle speaker locations, sound source at -30° and +30°, show medians of -22.8° and +32.0° respectively. The medians show that response to -30° show a relatively large undershoot. The MAEs for the inner speaker locations (-30°=+8.3° and +30°=+7.6°) are also large suggesting errors in the adults localisation ability. Even with such errors, the interquartile ranges for both inner sound source locations (-30°=+8.0° and +30°=+12.7°) are low suggesting the adults are making consistent responses to the target sound sources.

The outer speaker locations, sound sources at -70° and +70°, show median responses of -50.9° and +60.0° respectively. The interquartile ranges for the outer angles again show consistency in the responses (-70°=+12.8° and +70°= +7.1°). The same is true for the interquartile ranges of the inner angle sound sources. The MAEs for the outer angles are larger than those found for the inner angles

(-70°=+19.4° and +70°=+10.1°). The medians for the outer angles show a consistent undershoot to the target, which is seen in the larger MAEs values. The results for the outer speaker locations show an undershoot which is expected and seen in all localisation data in which head turns are measured [19, 39, 40, 41]. The undershoot is due to the head only being capable of rotating so far due to physiology, with the eyes making up the azimuth angular difference between the maximum head rotation and the sound source.

The results show a consistent offset between the positive and negative speaker locations, i.e. difference between ±30° is around nine degrees as is the difference between ±70°. Looking at the raw data and the processing method it seems that this is part of the results and not a systematic error in the processing of the results.

The results always undershoot the negative target locations. The adults have more undershoot to the negative angles; from looking at the raw data this is not due to the extraction algorithm but present in the raw data. One reason for this could be due to the instructions given to the participants. To try and obtain natural responses the adults were simply instructed to 'turn to face the sound'. Some of the adults may have not turned fully and used their eyes instead in some of the trials resulting in large undershoots. Most of the adults in previous studies also had training, in the task, this was avoided so that natural responses were obtained from the participants and was an important part of the *AnimalSeek* method. Also, the number of participants used in this study was lower than those used in previous investigations.

The overall errors are quite large compared to other sound localisation studies. Makous et al. [19] showed small errors for brief, wide band noise stimulus, with RMS errors of around 0.6° for stimulus presented from midline and RMS errors of 7.5° for stimulus presented between 60° and 80° (70° was not tested). The average MAE found using the new localisation test method for the outer speaker locations was 14.7°, twice that of the RMS error found in Makous et al. study [19]. Interestingly, the results from the Makous et al. [19] show an increase in errors to 9.7° and 9.4° for sound sources at (+60°,+5°) and (-60°,-5°). This then drops to 5.2° and 5.5° for sound sources at (+80°,+5°) and (-80°,-5°), i.e. localisation accuracy gets worse as the participants localise sounds away from midline but then performance gets better. The size of the error then increases again up to a maximum at sounds sources placed at +160°. Other studies found similar results to this, i.e. Recanzone et al. [39]. For noise stimuli, MAEs of 2°-4° were found for speaker locations ranging from 0° to 48°. No undershoot was seen even at 48°, however, the standard deviation increased in the responses.

4.4.3 Comparison of Localisation Ability between Hand Judged and Extraction Algorithm MkII

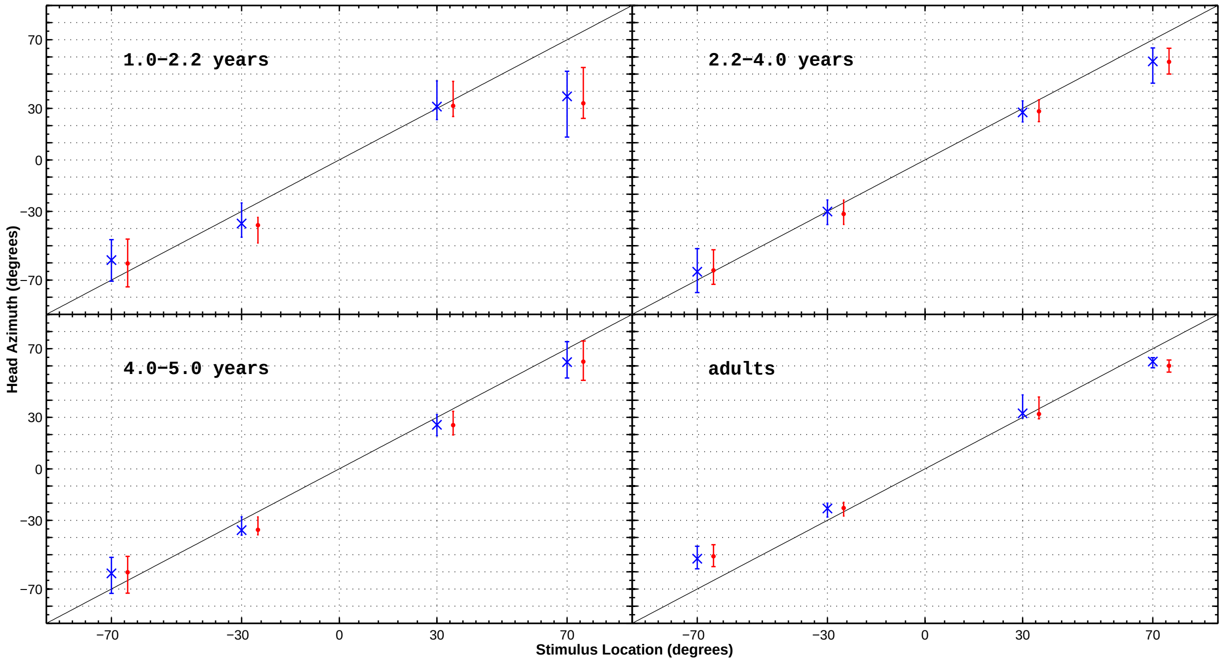


Figure 4.10: Localisation results for each age group. Localisation ability expressed in terms of medians (×) and interquartile ranges (*whiskers*). The two sets of responses for each age group show those obtained by hand picking (blue) and the extraction algorithm (red). The two are offset by 5° along the x-axis to show the two on one figure.

Age	Speaker location (°)			
	-70°	-30°	+30°	+70°
	MED (IQR)	MED (IQR)	MED (IQR)	MED (IQR)
1.0-2.2 years	1.9 (3.6)	1.0 (5.1)	0.5 (2.1)	4.1 (8.7)
2.2-4.0 years	0.9 (5.4)	1.5 (0.1)	0.7 (0.4)	0.3 (5.5)
4.0-5.0 years	0.7 (0.5)	0.2 (0.2)	0.2 (1.5)	0.2 (1.8)
Adults	1.4 (0.4)	0.3 (0.1)	1.3 (0.8)	2.5 (1.2)

Table 4.7: Difference between the human picked and the extraction algorithm of the medians and interquartile ranges as obtained by Equation 5.1 and Equation 5.1 respectively.

$$\Delta m = m_h - m_a \tag{4.7}$$

$$\Delta iqr = iqr_h - iqr_a \tag{4.8}$$

The adults show errors of under 2.5°. Results for the age groups, 2.2-4.0 years and 4.0-5.0 years show excellent agreement both in terms of their medians and also interquartile ranges. For the age group 2.2-4.0 years we see slightly higher interquartile ranges for the lateral locations for the age group 4.0-5.0 years, although they are within 5°.

The youngest age group, 1.0-2.2 years, shows the largest differences between both approaches in terms of medians and interquartile ranges. This is especially true for the outer target locations. One of the reasons for this could be due to the types of responses and quality of responses seen for this age group making accurate extraction difficult. Also, the number of participants and data points in this age group are low.

4.5 Discussion

The results show that the methods of extraction can be used to evaluate localisation ability. The results from the adult participants show low quartile ranges and MAEs. The MAE for the -70° sound source shows a large error. However, looking at the raw motion data we see this is present in all of the participants. The quartile ranges for the adults show consistent responses with the mean of the quartiles being 10.2° across all directions. The average MAE across all directions is 11.4°. Comparisons with other studies show that the errors in the responses are slightly higher than expected. Makous et al. [19] found average MAEs of 6.6° for targets average across 20°-40° and 60°-80°. The difference in the results could be due to the accuracy of the methods. Makous et al. instructed the participants to press a button when the sound was localised, something with more direct instruction and a defined head turn point.

Looking at the child data, the difference between 2.2-4.0 years and the 4.0-5.0 years age group MAEs is only 0.8°. The largest MAEs are seen in the 1.0-2.2 years age group. The smallest quartile ranges are also found in the 2.2-4.0 years age group, 2.3° lower than the 4.0-5.0 years age groups average quartile range. The largest average quartile range is the 1.0-2.2 years age group with a mean quartile range of 26.1° across all directions. It could also be due to the lack of data available due to there being only a small number of children who were willing to wear a hat.

The point at which the child localised the sound was taken as the point in which the head stopped moving (T_e). This approach was taken as it was easily identifiable and as expected in an ideal responses, i.e. the response flattens out when the sound is localised. This measured point might not have been where every child localised the sound and they may have moved their heads after. However, finding a point where the child had localised the sound is difficult due to the amount of trials available from the head tracker. Overall however, this method seems to get reasonable measures of localisation ability in children.

Comparison between the algorithm and the hand picked data shows good agreement across all age ranges with the biggest difference being 4.1° and an average difference of 1.1°. This shows that the method could be used as a fast alternative to the hand picking method and would allow the rapid evaluation of localisation ability in even the youngest of children.

4.6 Conclusion

The chapter set out to develop a method which could rapidly and accurately evaluate localisation ability in young children from raw motion tracking data. The chapter has shown, for this first time in a child localisation study, that such a method can be developed and that by using signal processing techniques it is possible to evaluate localisation ability to within 1° of that measured by hand but around 100 times faster (approximately four seconds by hand, around twenty milliseconds using the algorithm). At this processing speed it is possible to use the algorithm in real time to asses localisation ability on a trial by trial basis.

Improvement of the method and future proposals

The results show responses can be automatically extracted from the raw motion tracking data using the technique proposed. The method showed that localisation ability could be measured with similar accuracy to hand judging the data but faster. The main issue with the results is the amount of data available. This was due to the motion tracking gear and how it was mounted on the child. The hat in particular, used to measure head rotation, caused distress to the child who refused to either wear the hat or would only keep it on for a few trials, this resulted in a poor yield of trails. One way around this would be through the use of more 'remote' tracking. Although not as transparent as electromagnetic tracking, reflective tape could be placed onto the child's head / forehead, similar to the studies by [51, 96, 61], and the motion tracked using infrared cameras. This would allow for signal processing techniques to be applied to the data and the localisation ability of the child assessed in real time. This would require camera tracking software however. This poses a lot of issues and problems which would need to be overcame such as the development and implementation of software which could extract and analyse the movement from camera data. An even more unobtrusive method would be the use of facial recognition software. A number of studies [98, 99, 100] have used remote tracking via a video camera to determine the types of head movement and also the magnitude of the head movement in free space. Such methods use a number of different approaches. Zivkovic et al. [99] uses a polygon model based paradigm onto which the face is superimposed. From this the rotation of the head can be extracted from the video in real time. This sort of method is common and is easily implemented if using the correct equipment. Applying such a method to this study would allow for less distractions to the child during testing, i.e. no need to wear a hat or tape on the head.

Another aspect of the study which would be interesting would be the tracking of the subjects eyes. Conventional head tracking equipment tends to be large and bulky and apparatus available to us during this study was not applicable to the children. Being able to study the eye movement would allow experimenters to collect a lot more information on the types of strategies used when localising the sounds. Analysis of the data would be similar to those used in the head movement as data is similar. Work by Frankchak et al. [101] has shown tracking of the infant's eyes is possible. Another approach would be to track the eye remotely using cameras, something which has been shown to be possible in adults [102].

The current method processed the results after they had been collected. For the method to be used in a clinical setting real time analysis of the head turns would be required. This would require recoding of the experiment to pass the head tracking data to the analysis software (MATLAB), although such an implementation is not complicated due to time constraints it was not implemented.

Chapter 5

Development of an Artificial Observer to Automatically Classify Head Turn Responses in the *AnimalSeek* method

5.1 Introduction

So far a method and a set of analysis techniques have been discussed and presented which can be used to evaluate the localisation ability of children under five. This chapter will look at ways in which the judgment of the children's responses can be made automatic.

In chapter three it was shown that by using the new localisation method, a relatively high number of responses could be collected from children without using a second experimenter inside the AEC. Instead, three large video screens displaying scenes and characters were used to engage the child and bring their attention back to midline. However, two experimenters were still used during the testing of the new method to reduce experimenter bias in the judgment of responses. Due to the use of humans in the procedure, and even with methods such as OPP, in the authors opinion, response judgment is subjective, due to the usual wide variety of responses which are used to classify the responses as opposed to random or inattentive behaviour and also the high level of training required by the experimenter to notice these differences. The *AnimalSeek* method attempted to obtain consistent results by classifying a small set of responses to a small number of speaker locations. The responses, which are based on a set of criteria, can then be used to train a classifier (based on artificial intelligence) to judge each response.

With the *AnimalSeek* method we wish to classify correct head turn responses which were clearly to an auditory stimulus and reject responses which are not, i.e. head movements not towards the target stimuli or other forms of responses. The other type of response witnessed during the testing of the method was what was judged as a null response. This usually involved the child failing to respond within five seconds or turning to their parents/guardians, crying or attempting to get down from their parents/guardian lap or the chair. These types of responses were commonplace, especially with the younger age group, so it is important that they can be classified and rejected correctly. If the trial is a correct head turn, it is important that the classifier can classify the direction of the head turn i.e. is the response to +30° or +70°. The classifiers must also be able to do this correctly and to the highest rate of classification possible. The more accurate the method is, the more reliable a tool it is to be used at judging responses.

This chapter will discuss two classifiers and compare their results. The comparisons will be made upon their rate of classification and also their false positive rates. False positives are trials which are scored as correct head turns when in fact they were incorrect. Such classification is unwanted as it could lead to confusion in the child by presenting them with a reward when they did not respond correctly and subsequent wrong analysis of the trial. The methods will be trained and implemented offline, ideally they would be implemented in real time on a trial by trial basis. The end of the chapter will discuss such implementation.

5.2 Template Matching

Gesture recognition is the identification of a predefined gesture (e.g. a person waving) from data which may contain a mixture of other types of gestures, arbitrary responses or noise. One way of identifying gestures in a set of data is through the use of a technique called template matching. Template matching uses a predefined template of the gesture which is to be recognised. This template is then compared to the data set for occurrences of that template, as measured by the error between the template and the data. The templates used in template matching can take many forms; these range from simple representations of the data such as a gesture in its raw motion form [103] or templates which are based on statistical representations of the gesture, such as values describing the mean or standard deviation of the data [103, 104]. This section will look at using a template matching method to see if response to an auditory stimulus can be identified. The rotation of the head towards a auditory stimulus (the gesture wishing to be classified) ideally has a sigmoid-like shape (see Figure 4.1). This means that a head turn can be categorised using a simple pre-defined template to see if the response is correct or not. The simplicity of the method and also the fact that only one type of response is to be classified makes template matching an attractive method to use for classification of head turn responses.

Dynamic Time Warping

When using template matching on raw motion data a problem can occur due to the temporal nature of movement; the time it takes to make a head turn towards a target stimulus will differ between subjects and also between trials. Dynamic time warping (DTW) is a form of template matching which attempts to overcome the issues of temporal mismatch between the template and the data being classified. First developed by Sakoe et al. [105] for use in speech recognition tasks, the method has been used and applied to a number of classification problems, including human gesture recognition [106, 107] and speech recognition tasks [108].

To illustrate the DTW algorithm, Figure 5.1 shows a template of the word, 'head' - and a response which is a stretched form of the word 'head', $HEEEADD$. We can represent the template and responses as shown in Equation 5.1 and Equation 5.2 respectively:

$$Xt = xt_1, xt_2......xt_i \tag{5.1}$$

$$Xr = xr_1, xr_2......xr_j \tag{5.2}$$

where Xt and Xr represent vectors containing the template and response data respectively. i and j are the lengths of the template and response data.

Figure 5.1 shows a choice of movement at the second E in the response $HEEEADD$. From the cell marked $(i-1, j-1)$, the algorithm has a choice of movement to one of three adjacent cells: the choice of movement are $(i-1, j)$,$(i, j-1)$ and

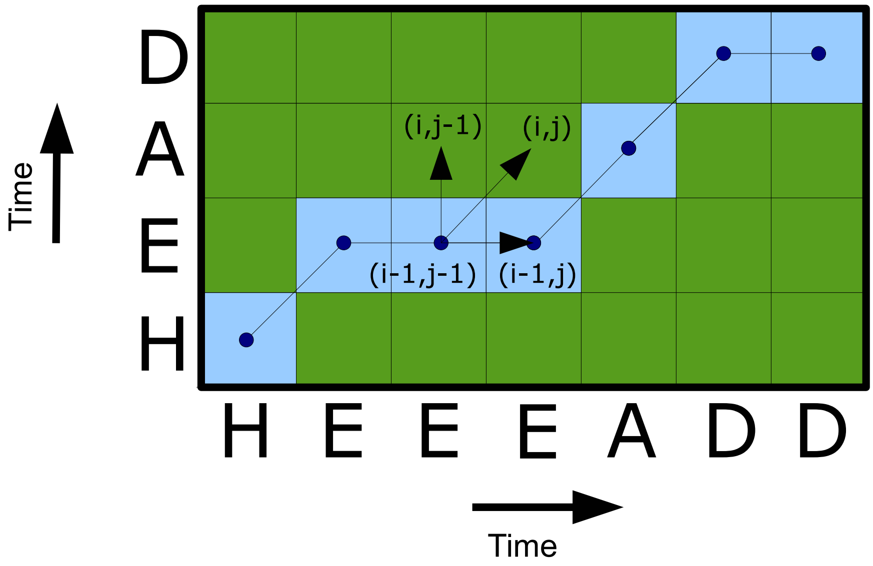


Figure 5.1: Example of the DTW algorithm illustrated by matching a word. The template, HEAD, is matched against the response HEEEADD, a longer (temporally stretched) version of the template. The blue squares show the path of the DTW algorithm. The second E in HEAD of the response shows a series of arrows corresponding to where the DTW algorithm can go for its next step. The next step is the step with the lowest distance. Once the full sequence is completed, the response is warped to the new sequence, the template is then compared against the response and an error measure obtained.

(i, j).

Movement between cells is constrained by several rules; firstly, the boundary conditions within which the movement can take place, defined as being between (xt_1,xr_1) and (xt_n,xr_m). Secondly, the algorithm can only move forward and to adjacent cells, defined as the monotonicity condition:

$$i_1 <= i_2 <= i_L \tag{5.3}$$

$$j_1 <= j_2 <= j_L \tag{5.4}$$

The final restriction is that the path can only move one cell at a time, this is defined as the continuity condition.

The DTW algorithm moves through the matrix shown in Figure 5.1 from left to right, starting at the origin. The size of each movement or the distance, is referred to as the local cost measure. As the algorithm moves, the local costs of each step can be added together to give the total cost of the algorithm's path $cost_p(Xt, Xr)$, as shown in Equation 5.5:

$$cost_p(Xt, Xr) = \sum_{l=1}^{L} cost(xt_{m_l}, xr_{n_l}) \tag{5.5}$$

for a warping path, p, of length L, with respect to the local cost measure, $cost(xt_{m_l}, xr_{n_l})$. The local cost measure for each point, l, on the warping path is the Manhattan distance (the absolute difference between the two points) between the xt value of point l (xt_{m_l}) and the xr value of point l (xr_{n_l}).

In the example given in Figure 5.1, the movement with the lowest cost would be to cell $(i-1, j)$, since the xt and xr values of that cell match (the letter E), giving a local cost measure of 0. Movement to other cells would result in a larger cost function, since the letters would not match. The DTW algorithm computes all paths through the matrix according to the rules discussed above. The path with the lowest overall cost across the whole response is then used as the warping path. Once complete the time axis of the data gets 'warped' onto the new time axis of the path with the lowest cost function.

The DTW algorithm was implemented in MATLAB using the DTW toolbox developed by Micó [109]. The use of DTW algorithm is ideal for the application required in this thesis as it allows the classification of a predefined shape which is subject to temporal differences, i.e. a head turn from a child.

Support Vector Machines

The output of the DTW algorithm is an error measure. This error measure is used to classify the data into either a correct or an incorrect responses to the auditory stimulus. One way of doing this is to draw a decision boundary between the values which represent correct and incorrect responses. The boundaries can be formed in a number of ways. One way would be to form a simple cut off by visual inspection, i.e. values greater than X are correct, less than X, incorrect. This method is quite a crude approach and would result in large errors of classification. The method used in this thesis are support vector machines (SVM). Figure 5.2 shows a graphical representation of a two-dimensional SVM. SVMs work by constructing a plane (refered to as a hyperplane in higher dimensions) which separates the values of two or more classes, respectively. Data can be described as a set of vectors, x_i, with each vector in x_i belonging to a class. Figure 5.2 shows the two classes represented by the symbols $*$ and $+$. The plane, p , separating the classes is described in Equation 5.6:

$$h(x) = x.w + w_0 \tag{5.6}$$

where $x.w$ describes the dot product of the data, x, and w, the vector normal to the separating plane. w_0 describes the distance between the origin (0,0) and the

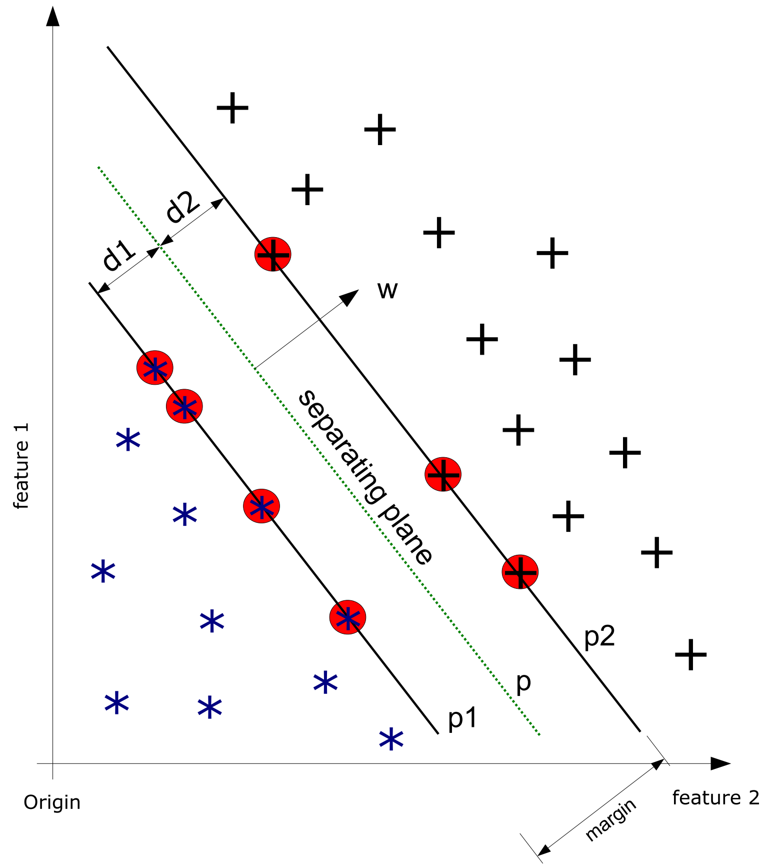


Figure 5.2: Application of a SVM to separate two classes of data, $*$ and $+$. The plane p is used to separate the two classes of data. This is achieved by using support vectors, highlighted by red circles. Support vectors are chosen so that the separation between the two classes, called the margin, is at its maximum.

vector w.

SVMs choose data points in each class, referred to as support vectors. The support vectors are used to form boundaries to separate classes. The two boundaries constructed in this example are defined as, $p1 : x.w + w_0 = +1$ and $p2 : x.w + w_0 = -1$ where $d1$ and $d2$ are the distances between the separating plane, p, and $p1$ and $p2$, respectively. These boundaries separate class $*$ and class $+$, respectively.

The distance between $p1$ and $p2$ is defined as the margin. Because the magnitudes of $d1$ and $d2$ are equal, we will define this distance simply as d. The distance is shown in Equation 5.7

$$d = \frac{|w.x + w_0|}{|w|} \tag{5.7}$$

The width of the margin can be written as shown in Equation 5.8

$$margin = \frac{2}{\|w\|} \tag{5.8}$$

SVM work by finding support vectors which maximize the margin separating the classes. The approach discussed above is the original 'linear maximised margin' definition of SVMs as defined by [110]. In current implementations the margin is maximised according to rules relating to 'soft margins' [111]. A soft margin allows classification of the classes even if some of the data belonging to class $*$ lies in the region of class $+$, and vice versa. SVMs can also be implemented to produce non-linear separating planes. By implementing SVMs, the most efficient boundary can be created allowing for accurate classification of the DTW error measures. The discussion of soft margins is out of scope of this thesis. Linear SVMs were implemented using the open source MATLAB toolbox 'OSU-SVM' [112] to separate the DTW results.

A different SVM was trained for each set of data wishing to be classified (correct/incorrect, correct/null and direction). The SVMs were trained on randomly divided data with 70% of the data and then tested on the remaining 30% to verify classification.

5.2.1 Implementation of a Template Matching Classifier

The data used for classification were from all age groups except 1.0-2.2 years, because the results from chapter four showed that this age group were not localising but discriminating the speaker locations. This makes fitting a template difficult when categorizing directions.

A correct head turn comprises of a sigmoidal shape with its maximum in the direction of the sound source. Although the child data was not always a perfect sigmoid like those seen in the adult data, the head turns are usually sigmoidal enough that a template can be reasonably fitted and an error measure obtained. The template was derived by averaging the head turns responses from the adult data. The template is shown in Figure 5.3.

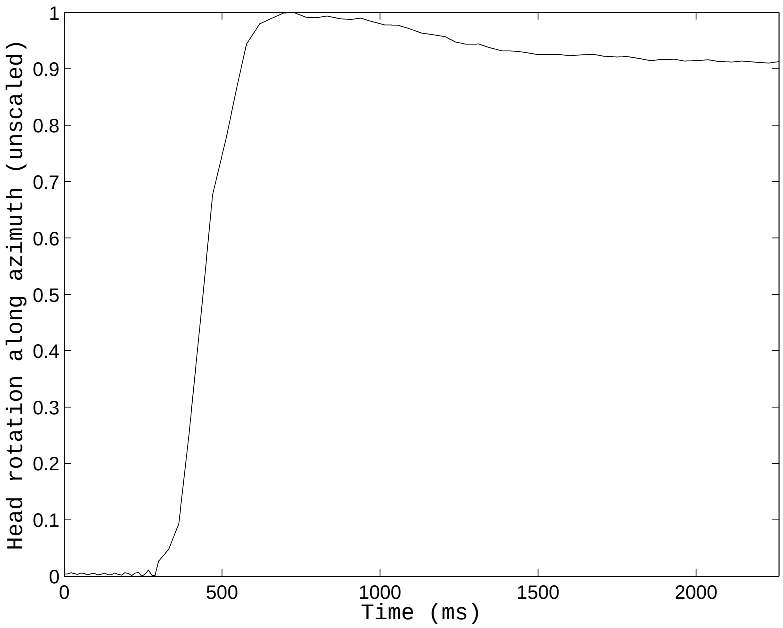


Figure 5.3: Template of head turn used in DTW algorithm.

Templates for the classification of correct / incorrect head turns

A template is required to decide if the head turn response is correct or incorrect. One issue with using template matching techniques is that they do not scale the template in terms of amplitude and the responses we wish to classify have two possible amplitudes (30° and 70°). One approach to overcome this would be to scale the template and the response data to the same values e.g. between zero and one. However, scaling responses would mean that small movements, e.g. in a null response trial, would be amplified. When the template is fitted to the data, the amplification of a non-directed head movement could produce a similar error measure to that of a directed head turn, and thus be incorrectly identified as a correct response. Another approach to this problem is to scale the template so that its maximum amplitude is the mean of the median response to the four stimulus directions. This gives template with a maximum amplitude of +42.5°.

In addition to the magnitude of the template, the polarity of the template must also be taken into consideration. The responses to the negative angles will give an inverted sigmoid and result in large errors if compared to a template with a positive magnitude. One approach to overcome this would be to use two templates which would also be useful for checking the direction of the responses. However, this approach was not possible, since the data had to be made absolute before they were passed through the DTW algorithm. This is due to errors made in the data collection stage and the motion trackers often switching polarity (see page 40 for details). Instead, by pooling all the data into just two directions, +30° and +70°, only one template is needed.

Templates for the classification of the target direction of correct head turns

For the classification of directions, one template was used which was scaled to two different magnitudes. The first is scaled, starting at 0°, to the mean of the median responses to +70° (56.0°) and the second to the mean of the median responses to +30° (29.4°). This approach is used as it is anticipated that error measures for the responses to the auditory stimuli placed at 30° and 70° will be less than the errors to incorrect responses. Correct responses to the angles should have smaller errors than null or incorrect head movements, which will contain either under- or overshoots to the auditory targets. SVMs are not used to classify these responses but instead simple comparisons to which template (30° or 70°) has the lowest error.

5.2.2 Classification Rates using Template Matching Classifier

The classification was achieved by randomly dividing the data up into a ratio of 70%/30%, which corresponded to the training and test data respectively. All test and training data also have a corresponding response type, i.e. correct or incorrect. Once divided, an error measure was obtained for each trial using the DTW algorithm and then classified using the SVM. The SVM was then applied to the test data on a trial by trial basis. The response judgement from the SVM, correct or incorrect and direction, was then compared to the outcome from test data response type. The same template was used to classify the data, apart from where scaled.

Classification of correct / incorrect head turns

A total of 341 trials were used, 230 correct head turn responses and 111 incorrect head turn responses. The incorrect trials were those marked as incorrect head, hand or eye responses to the auditory stimulus. All of the trials were randomly divided up into training and test data (70% and 30% respectively), along with their corresponding response type (1 = correct, 0 = incorrect). The results show a classification rate of 63.3%, with a false positive rate of 21.2%. The results show that the classification rate were quite low and the false positive rate high. The reason for the high false positive rate is due to the incorrect responses containing head turns which are being fit to the template. As discussed previously in Chapter four, this results in similar error measures as the correct head turns and makes them hard to classify in the SVM.

Classification of correct head turns and null responses

A total of 766 trials were analysed, 230 correct head turns and 536 null responses. Null responses were the responses which resulted from the child becoming inattentive, turning to their parent/guardian or not responding within five seconds. The data was randomly divided up into training and test data, along with their corresponding reward types (1 = correct, 0 = null). The classification rate was 76.0%, with a false positive rate of 6.5%. Once again, the results show a relativley low classification rate and but relativley good false positive rates.

Classification of the direction of correct head turns

The correct trials were used to classify the direction of the responses. A total of 230 trials were used, 68 response to 30° and 162 to 70° (1 = 30°, 0 = 70°). The results show a classification rate of 90.9%. False positives rates are 7.4%.

5.2.3 Discussion of the Template Matching Classifier

The DTW method is simple and easy to apply to the data. It is also fast, taking approximately 10ms to classify each trial (i.e. the time taken to fit the template and compare to the SVM). The classification rates of responses as correct, incorrect or null are low i.e. less than 80%, but the classification of the direction of correct responses which showed a high rate of classification. The main issue with the template method is that it fits a template to all of the recorded head movement and classifies it as a head turn; this is an issue when trying to classify data on just a single feature, i.e. the rotation of the head.
The method showed high rates of classification for the direction data, i.e. if the responses was to +30° or +70°. The method showed lower levels of performance however for the correct/incorrect and correct/null responses. Chapter four showed that statistically 30° and 70° were different and therefore discrimination between the two is possible. Issues arise with the other responses due to

the noise present in the data. The downside of using a template method on the raw data is that incorrect/null responses contain head turn components which the template will fit too. An approach to overcome this is to take the raw data and only use the salient information about the head turn to classify it. The next section will look at such an approach.

5.3 Artificial Neural Networks

The previous section showed that it is possible to classify head turns using a pre-defined template of a head turn. Noise and head turn components in the incorrect/null data made template matching produce low classification rates and high false positives. Another approach to classification is to break the data down into a set of features and then use these features for classification. By doing this, the issues of the template matching method should be reduced because the whole head trace is not presented to the classifier, only the salient information corresponding to a head turn.

This section will look at the use of features vectors and their classification using Artificial Neural Networks (ANN). It is hoped that using such an approach will increase the classification rates to around 90% with false positives below 5%.

Feature Vectors

Feature analysis involves breaking down raw gesture data into a set of features which can be used to describe a gesture. The features are placed into a feature vector which can then used to classify responses using one of many types of classifier. Before the feature vector can be classified, the features must be extracted from the data. Feature extraction methods can be done several ways - one approach is to describe the data using simple statistics, which can include values such as means, medians or standard deviations. More complex statistical representations of the data can also be used; examples from recent literature include principal components [113, 114], discrete cosine transforms [115], or discrete wavelet transforms [116].

Classification Techniques

Once feature vectors of the data are obtained, they are classified using any number of classification techniques. Such types of classification can include non-probabilistic classifiers such as support vector machines (SVM), Bayesian based probability classifiers such as Hidden Markov Models and statistical model classifiers such as Artificial Neural Networks (ANN).

The type of classifier used is dependent on the type of data which is to be classified. Hidden Markov Models are commonly used on classification and speech problems in which continuous, movement data is collected. The use of Bayesian probabilities allows the constant stream of data to be classified in real time using the rules of Bayesian probability. However, the classification problem addressed in this section is not continuous with data being collected on a trial by trial basis and processed as such. Non-probabilistic classifiers such as SVMs or nearest neighbour methods can show good performance, however, the type of classifier which was used during this thesis are ANNs since they provide a robust way of solving classification problems. They are also well supported in MATLAB, with a MATHWORKS toolbox available [117].

Artificial Neural Network Architectures

ANNs are made up of artificial neurons which are mathematical models of biological neurons. The first artificial neurons and artificial neural networks were suggested by McCulloch and Pitts [118], who showed that it was possible to compute functions using a suitably constructed network of neurons. The first implementation and usable ANNs were developed much later by Minksy et al. [119]; these first ANN were called perceptrons.

Today, ANNs are used in a large number of real world classification problems, including financial prediction [120], medical diagnosis [121] and in engineering problems [122]. Figure 5.4 shows the architecture of a simple, single-layer

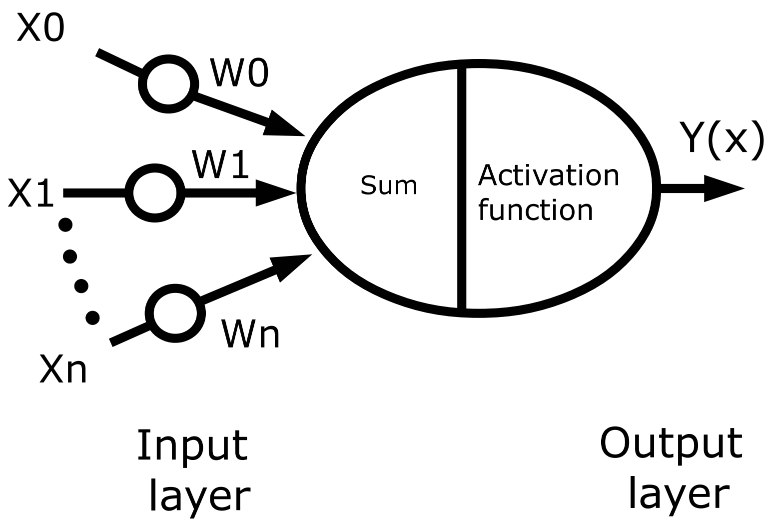


Figure 5.4: **A schematic of a perceptron. The perceptron consists of two layers, the input layer and the output layer. The inputs x_i are fed through their corresponding weights,w_i, the weights are adjusted during the training phase of the neural network so that the presented inputs give a corresponding output. The output of each weight is summed together and fed into the activation function. The output then changes depending on a threshold, θ. When the activation function level is reached, the output changes. The output of these perceptron is limited to a boolean expression (1 or 0, true or false)**

ANN (the perceptron). The ANN consists of two layers, the input layer and the output layer. Each i inputs, x_i, also has i corresponding weights, w_i. The inputs pass through their corresponding weight and are summed together to produce an output, $H(x)$, this is shown in Equation 5.9:

$$H(x) = \sum_{i=0}^{n} x_i w_i, \tag{5.9}$$

the summed values, $H(x)$, are then passed to an activation function, G. The activation function for a perceptron is a step function described as follows:

$$G(H) = \begin{cases} 1, H \geq \theta \\ 0, H < \theta \end{cases} \tag{5.10}$$

from the activation function we can see that the output of an ANN can only be true or false. The input also contains a node referred to as the 'bias' of the network. The bias node has a constant output of one and an associated weight, w_b. The bias allows the position of the decision boundary (θ) of the activation function to be moved i.e. the point which makes the output either correct or incorrect. The output of the ANN can thus be defined as shown in Equation 5.11:

$$y(x) = G(\sum_{i=0}^{n} x_i w_i), \tag{5.11}$$

where $y(x)$ is the output when supplied with the inputs, x_i, w_i the corresponding weights and G, is the activation function,

ANNs are trained using a supervised learning technique, i.e. a set of inputs and their corresponding outputs are provided to the learning network. The weights of the ANN are adjusted in an iterative process until the lowest error between the outputs of the training ANN and the expected outputs is obtained. Learning is generalised as backpropagation. When a simple form of backpropagation being the delta rule which is the method used in perceptron learning. The delta rule is defined as shown in Equation 5.12:

$$\Delta w_{ij} = \alpha(yt_j - ya_j)x_i, \tag{5.12}$$

where Δw_{ij} is the change of the i^{th} weights of the j^{th} neuron, yt is the target output, ya the actual output at each iteration, x_i the input and α the learning rate of the network. The delta rule is used to find a set of weights which give the smallest error between the actual output, ya, and the target output, yt, of the network.

The type of ANN discussed above is a very simple, linear ANN. As the area of ANN was researched more and computing power increased, several others architecture types emerged, including feed-forward networks, recurrent networks and probabilistic networks [123]. A simple, linear ANN like the perceptron is too simplistic for solving the head turn data feature vector, so instead this section will discuss the use of a feed-forward neural network.

The Feedforward Network

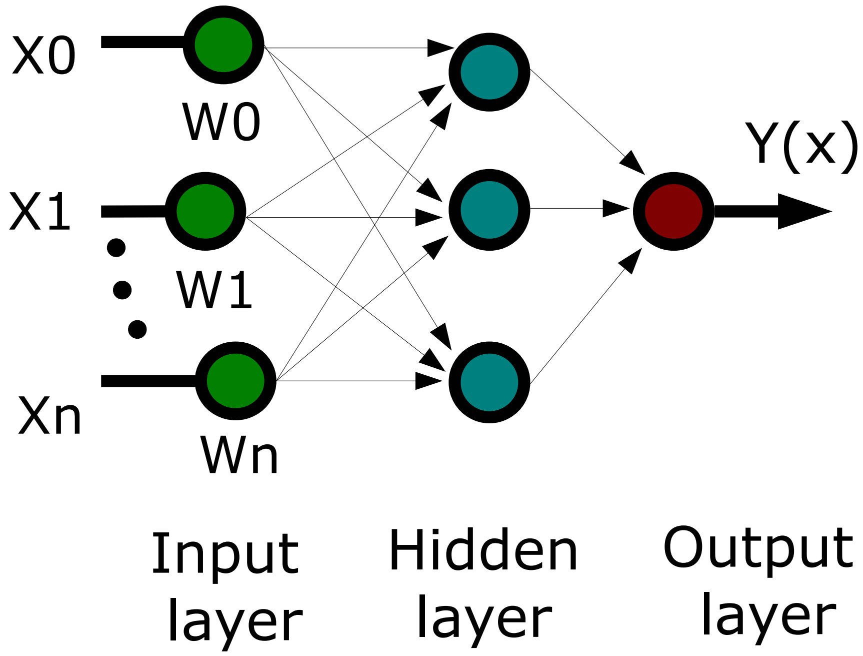


Figure 5.5: Schematic of a feedforward network. This feedfoward network contains three layers. This is one more than the perceptron and includes what is referred to as the hidden layer. The inputs are connected to the hidden layer via their corresponding weights. The hidden layer allows the ANN to perform more complex processing. The hidden layer is connected to the output layer. The hidden layer and the output layer have their own sets of activation functions. Learning in the feed-forward network is achieved via a backpropagation algorithm.

Figure 5.5 shows the architecture of a feed-forward network. The feed-forward network differs to the ANN discussed previously in that it contains an additional layer called the hidden layer. The hidden layer takes its name from the fact that its state cannot be observed from the input or the output layers. The hidden layer allows the network to perform more complex, nonlinear calculations compared to the linear ANN. The activation function (G) of the feed-forward network can be any of a varied set of mathematical functions (i.e. linear, sigmoid or hyperbolic tangent sigmoid). The hidden layer and the output layer each have their own activation functions which can be the same or different to one another. The more complex activation functions allows the output of the feed-forward network to be continuous, i.e. not restricted to a boolean expression like the one used in the perceptron. The type of activation functions used in this thesis are discussed when the network is designed in the following section.

The feed-forward network uses a backprorogation algorithm to learn. Unlike the perceptron, finding the minimum error, E, of the feed-forward network is slightly more complex due to the hidden layer and a non-linear learning rule must be applied. The learning rule used here is the Levenberg-Marquardt (LM) algorithm [124, 125]. The LM algorithm solves the sum of squares of the non-linear function relating the weights of the network, w, and the error observed at the output of the network E. The minimum of the function is found by using the LM algorithm to solve Equation 5.13:

$$(J^T J + \lambda I)w = J^T E, \tag{5.13}$$

where J is the Jacobian of the function being solved, λ the Levenbergs damping factor, w the weight update vector and E the error vector. During each iteration of the training procedure the error gradient, $J^T E$, is computed, and a new set of weights is found by solving Equation 5.13 for w. Using the new weights, the output error ($yt - ya$) of the network is found. The LM algorithm is used as it is the fastest (yet most memory intensive) method of solving the sum of squares of a non-linear function. The algorithm works by updating the weights by different amounts depending on the state of the error gradient. If the gradient is a constant slope, large changes are made so that the algorithm converges quickly. If the gradient is changing rapidly, the weights are changed by a smaller amount so that the minimum of the function is not missed and as a result, never converges [126, 127].

5.3.1 Implementation of a Classifier Using ANN

Chapter four and the previous section regarding template matching showed that the head turns are made up of a sigmoid shape. From these raw head turns a number of simple values can be extracted to form a feature vector which describes a head turn response. A correct head turn response will consist of a start and an end. The start and end of the head turn will have a value in time (milliseconds) and also a magnitude (°). The absolute difference between these points are used in the feature vector and are described as Δt and Δa. From these parameters, the average gradient of the head turn can be derived, this feature is referred to as $\Delta a/\Delta t$. A final feature which will be used in the feature vectors is the number of peaks seen in the velocity vector of the head turn. It can be seen in the examples (see Figure 4.7, page 110) that correct head turns towards sound sources are made up of one or two rotational movements of the head. Incorrect head turn responses usually contain a larger number of head movements. Therefore, the incorrect head turns will show more peaks in the velocity vector. The four features are therefore defined as:

1. Δt, the time difference between start and end of the head turn.

2. Δa, angular difference between start and end of the head turn.

3. Δa/Δt, gradient of response.

4. P, the number of peaks in a response.

Using these four simple features it will be possible to discriminate between a correct head turn responses and incorrect/null responses and also to identify the direction of a correct head turn. This section will discuss the design of a simple feed-forward network for the classification problem discussed in this chapter. The neural network must take in the feature vector presented in the previous section, process it and then decide on whether the response was firstly a correct head turn or not and secondly to which direction.

The network consists of three layers, firstly an input layer which is made up of four nodes, corresponding to one per feature. Connected to the input layer is the hidden layer which contains twenty nodes (large number chosen to allow for effective classification). There are no rules which determine how many hidden layers should be used, or how many nodes should be in a hidden layer. Since it is advised by Russell et al. [128] that a single layer is capable of solving all continuous classification problems, one layer was used. Finally, connected to the other side of the hidden layer is the output layer. The output layer contains two nodes. For the first neural network, these outputs will represent if the response was correct (1) or incorrect (0). For the second ANN, which will be trained to discriminate direction, one will represent a response to 30° and one to 70° (output of the ANN will classify a 1 as 30° and 0 and 70°. The activation functions used for the hidden and output layer are hyperbolic tangent sigmoid transfer functions.

The use of this activation function allows for a continuous output. The output is then processed further by simple rounding it to the nearest whole integer. A number of methods were used to make decision boundaries on correct/incorrect classification, however, this proved to get the highest rates of classification. Before the data were analysed they were separated randomly into a ratio of 70%/30% (training data/test data, as discussed in MATLAB Neural Network Toolbox [117]). The ANNs were trained three times, each time with the data being randomised and then presented to the ANNs. The results from these three runs were then averaged to obtained a classification and false positive rate.

The network was trained using the training data set and then simulated on the test data. The test data results were then compared with those outputted from the ANN. Classification rate was the percentage correct obtained by the ANN. The false positive rate, that is, the number of responses scored as correct when in fact it was not was also calculated, to make sure over fitting has not taken place.

5.3.2 Classification Rates Using ANN Classifier

Classification of correct / incorrect head turns using ANN

A total of 341 trials were used, 230 correct head turn responses and 111 incorrect head turn responses, these were randomly divided up into training and test data (70% and 30% respectivley). The results show a classification rate of 79.1% with a false positive rate of 11.5%. Although the classification rate is higher than the template matching method as is the false positive rate, however, this rate is still quite high. The reason for the false positive rate, like with the template matching methods, is that some of the incorrect head turns are being classified as correct and as such lead to misclassification. It was hoped that by classifying using feature vectors this could be overcome, but the rates are still high, however, the performance has increased as compared to the template matching method.

Classification of correct head turn and null data using ANN

The data contained a total of 766 trials, 230 correct head turns and 536 null responses, these were randomly divided up into training and test data (70% and 30% respectivley). The results show a classification rate of 84.6% with a false positive rate of 8.4%. The ANN shows fairly high classification rates, however, the false positive rate is again quite high. The use of feature vectors still show a resonably high rate of misclassification, suggesting that the null data contains many compoents present in a correct head turn.

Direction of head response

The correct trials were used to classify the direction of the responses. A total of 230 trials were used, 68 response to 30° and 162 to 70°. It was found that using all four feature vectors, a classification rate of 87.0% and a false positive rate of 13% could be achieved. These rates could be improved, however, with the removal of features, Δt and P. The new vector gave classification rates of 90.8% and a false positive of 7.3%. It seems the reason for this is that the inclusion of the peak and time data adds variance to the feature vectors which causes error in the classification rate. Although many of the correct responses have one peak, some also have two or three peaks; this results in errors when trying to find a boundary to classify. The same is true for the inclusion of the time vector.

5.3.3 Discussion of the ANN Classifier

The results showed that classification is possible with relatively simple networks and feature vectors which are easily derived from the response data. The time taken to process each trial is around 15ms. The results show a fairly high rate of classification (all greater than 80%), however, the false positive rate is also quite high because a high proportion of null responses and incorrect head turns contained head turn components and were classified as correct. This was hoped to be overcome by using the feature vector method as compared to the template matching method. The next section will discuss and compare the two methods.

5.4 Discussion and Conclusion

5.4.1 Comparison of the Two Classifier Methods

Method	Correct/Incorrect	Correct/Null	Direction
DTW	75.2 (17.1)	76.0 (6.5)	90.9 (7.4)
NN	79.1 (11.5)	84.6 (8.4)	90.8 (7.3)

Table 5.1: Comparison of the two classification methods for the child data. Data shown as percent correct with the percentage of trials which were false positive in brackets.

Discussed in this chapter were two different approaches for the classification of the head turn responses. The first was the template matching technique which used a predefined template of a correct head turn and fitted it to each trial in order to obtain an error measure which was then classified using a SVM. The second was an ANN, which classified the data using four feature vectors based on a number of parameters of each head turn.

The ANN classifier showed higher classification rates and lower false positive rates for both the classification of correct / incorrect head turns and null responses. When inspecting the raw misclassified data it can be seen that the template method will classify all types of head turn as correct, as only one feature is given to the classifier, the error measure. The ANN approach gives improved results because it provides more information to the classifier in the form of the four features in the feature vector, as predicted. The rate of false positives is comparable for the correct / null classification, with the template matching method producing better rates.

The ANN and the template method provide the same rate of classification of directions. The ANN provides slightly lower false positives but this difference is small (<0.1%). The reason for the lower than expected results when using the ANN is that the feature vectors are describing a head turn but such head turn components are present in the incorrect/null data. This is due to the quality of the raw data from which the classification is based. To increase the classification rates more information would be needed. This could take the form of other factors of the head turn i.e. the roll or elevation. However, when looking into this, such components were also noisy and even more varied between, and within, participants. Errors in the data scoring would also carry over into the classification results, i.e. the experimenters judgments. Although precautions and care were taken in the response judgment, this problem was unavoidable. The data from the motion trackers was also not ideal, with some of the null and incorrect traces containing noise or polarity inversions. Although attempts were made to overcome and screen out these responses, some errors still remained in the data which can effect the classification rates.

Looking at the two methods it can be seen that the ANN is advantageous because it provides higher classification results, apart from when classifying directions. If further development of the method was undertaken and the method used in real time, implementation of the ANN classifier would be more difficult than the template matching method. Development of the ANN classifier and feature extraction would ideally be done outside of MATLAB using a faster and more independent programming language (e.g. C++). This is so that the processing of the trials would be as fast as possible and the system would not be reliant on the MATLAB environment.

Implementation of the template matching classifier would be simpler, the raw motion tracking data could be fed into the classifier and an error measure obtained which could then be classified by the SVM. This could also be developed in a programming language such as C++. Although the template matching method would be easier to implement and is faster at processing the trials, the differences in percent correct and also the time taken to process between the two methods, show that the implementation of the ANN classifier would be advantageous.

5.4.2 Conclusion

The chapter has discussed two methods which could be used to classify responses children make to sounds during a new localisation test method. The issue with the localisation method and previous methods were that they required at least two experimenters to control the experiment and engage the child, and to remove experimenter bias in any measurements made. The chapter has shown that using a classifier approach, adequate rates of classification can be obtained with the child data. Such approach has not been attempted before in both an adult or a child localisation study. For the method to be used in a clinical setting, the false positive rates would have to be reduced. The high rates of classification are thought not to be due to the method but the quality of the raw data used. Even with a method (ANN) which didn't use the raw data but instead the salient information, noise in the data caused misclassification. To improve this, more data would need to be collected so that the amount of training data available to the classifiers was increased and boundaries easier to determine.

For the method to be effective in a clinical setting it would need to be implemented in real time so that responses could be assessed on a trial by trial basis. For the method to be implemented in real time, the way the system would need slight recoding so that data could be passed from the motion tracker and directly to MATLAB for analysis.

Chapter 6

Discussion and Conclusions

This chapter provides an overview of the findings presented in this thesis and also discusses the aims and possibilities for future work.

Development the *AnimalSeek* Method

The main aim of the thesis was to develop a new behavioural test method which could be used to evaluate the localisation ability of children under five years of age. The method was required to be intuitive so that even young children (one year old) could perform it without instruction. The children were engaged using the three large video screens, these were also used to return the child's attention back to midline at the end of each trial. By using the video screens in this way, it was also possible to reduce the need for an experimenter to be inside the AEC engaging the child. Such an approach has not been undertaken before in a child localisation study.

The method tested 28 participants ranging from one to five years of age. The method was evaluated by looking at how many responses (correct and incorrect) could be obtained from the child before they stopped responding. Responses were marked as correct or incorrect and the method evaluated on the total number of responses (correct and incorrect trials) and the number of correct responses which were the trials consisting of a head turn towards the target stimulus location. The number of responses were seen to be a function of age, i.e. the younger children produced fewer responses than the older children. This was to be expected because the attention span of the younger children is shorter. The total number of correct head turn responses for all of the age groups was generally lower than the total number of responses obtained (approximately 50%). Comparisons with other methods are difficult because many BA studies do not score just head turn responses, with other types of responses and gestures scored as correct.

The total number of responses (sum of correct and incorrect trials) are comparable to previous studies. However, the number of correct head turns (those used to evaluate localisation ability) were lower than expected with only around ten obtained per block. It was found that most children would sit for at least three blocks per visit, this can elicit enough trials so that localisation ability could be obtained. One of the main reasons for the low number of responses per block was the the motion tracking gear worn by the child, and especially the youngest age group (1.0-2.2 years), which caused them distress and as a result caused many blocks to be terminated early.

As well as looking at the methods performance, the effect of the reward location on the number of responses was also investigated. The visual environment allowed the experimenter to manipulate where the reward was displayed with respect to the auditory stimulus. Three reward locations were tested:

1. The reward was presented at the same location as the auditory stimuli.
2. The reward was presented at midline (zero degrees), no matter where the auditory stimuli was presented from.
3. The reward was presented at a random location $\pm 20^{\circ}$ about the target stimuli.

It was hypothesised that presenting the reward at midline (zero degrees) would make the game-like task uninteresting for the children and would produce the fewest number of responses. In regard to the other two reward types it was hypothesised that these rewards would produce the same number of responses, however, the *jittered* condition would be the preferred due to it taking away learning effects regarding where the sound was presented from. It was hypothesised that by using the *jittered* reward, the location of the auditory stimulus could not be learnt by the child by simply remembering where the visual reward was presented, this would be particularly problematic if only a few auditory stimulus locations were used. When the reward was presented at zero, as expected, fewer responses were seen. For age group 1.0-2.2 years reward type *at location* produced the most responses, however, statistically, the *jittered* location was no different. However, a statistical difference was observed between *at location* and *zero* for this age group. For the older age group, 4.0-5.0 years, the *jittered* condition was preferred, however, statistically there was no difference. Presenting the reward at *zero* produced the fewest responses across all the age groups but only showed a statistical significance for the youngest age group when looking at total correct responses. The amount of data available for comparisons was lower than expected so more testing would be required to confirm these differences.

Overall a good number of responses were obtained using the *AnimalSeek* method, however, the proportion of correct head turns, which is used for evaluating localisation ability, was found to be lower than that obtained in previous studies..

Measurement of Localisation Ability

As well as evaluating the method, during the testing of the children in chapter three, motion tracking data was also being collected. Chapter four investigated ways of evaluating and measuring the head motion of the children to an auditory stimulus and using the motion data to evaluate the child's localisation ability. Previous methods of evaluating a child's localisation ability have been, in the author's opinion, crude and both time and labour intensive. A method was required that would evaluate the children's responses both accurately and quickly. By having such a method, the analysis could be performed on a trial by trial basis.

The extraction algorithm (see page 110) worked by looking at the velocity of the head turn and taking the point of localisation as the point where the head stopped moving. To evaluate the method the start and end of the head turn were found manually and the times and magnitudes of the head turn at these points recorded, the same process was then performed using the extraction algorithm and the two compared. The extraction method showed good agreement with the hand picked data. Compared to hand picking the data this method was a lot faster and could analyse the head movements and extract the salient information in milliseconds rather than seconds. Using this processing technique allows for the fast evaluation of head turn responses, something which has not been seen in child localisation literature before.

The extraction algorithm was then used to evaluate the localisation ability of the children. The youngest age group of children (1.0-2.2 years) showed generally poor localisation results, this was seen by large undershoots to the target stimulus. The older children (age groups 2.2-4.0 years and 4.0-5.0 years) showed better performance. This data, in comparisons with other studies, showed large errors in the children's localisation ability. The data was also evaluated by hand and showed similar levels of performance suggesting that the raw data was the reasons for the errors. The reason for the poor performance could be due to the low number of children tested, the low number of trials and also issues with windowing of the data during collection. The point at which the sound was localised (end of the biggest head turn) could also have caused errors in the localisation performance, such a point is difficult to evaluate in young children however because there is not method of indication and so must be extracted from the raw head data.

Making the Evaluation of Responses Objective

Chapter five presented a method of automatically classifying the responses made by the children to the presented sounds. It is hoped that using such a method would allow for the consistent judgments of responses based on a set of criteria. Data in this section was done offline, i.e. the data was first collected and then analysis took place. The automatic classification of response will ideally take place in real time on a trial by trial basis.

Responses made by the children to the auditory stimulus were classified as either correct or incorrect, if correct they were also classified to a direction (30° or 70°). The chapter presented two methods of classification, template matching and artificial neural networks (ANN). When using the template matching approach, the raw head turn data was fitted to a template of a head turn and an error measure produced. Based on the error measure, the data was classified. This method produced a low classification rate due to the amount of variance present in the raw data. To try and overcome this, a feature vector and a ANN approach was used. This approach split the raw data into four features which described the head turn data. The ANN was then used to classify the feature vector. The ANN method was able to classify response as correct or incorrect/null at rates of around 85%. Directions were classified at a rate of around 90%. Issues with the classifiers were the high rates of false positive classification - around 9% for all response types. False positives are not desired as they could lead to confusion for the child who is responding to the sounds. One reason for the high false positive rates is the quality of the data being presented to the classifiers. Even when using a feature vector approach, head turn components and features are still present in the incorrect/null data and are classified as correct.

Although the classifier was not implemented in real time, it showed that it is possible to classify the head turn responses accurately even though the raw head traces were highly variable. Previous techniques needed a number of experiments, who were required to be highly trained, to judge response. This suffered from subjectively issues resulting from experimenter bias's in the judgments.

6.1 Future Research

6.1.1 Application of Results

This thesis looked at a new method which could evaluate the localisation ability of children under five. Although the *AnimalSeek* method is an adaptation of a BA technique, the setting in which the method took place (SOFE) was novel. The *AnimalSeek* method has shown that it is possible to engage a child in a BA task using a visual environment and without the need for an experimenter in front of the child. Although it was shown that the method did not obtain a high level of head turns as required for a localisation task, it did show that response could be obtained. This could mean that the method could be adapted and used for other BA tasks.

The visual environment allowed for the presentation of rewards anywhere on the visual environment in front, and to the side of child. Unlike previous BA techniques which regularly use a static reward, this allowed for research into how the location of the visual reward affected the motivation of the child. Due to the low numbers of children involved, the results showed no differences between the reward types, trends in the data, however, suggest that there could be an effect of reward location. Having a tool such as the visual environment, which can be altered easily with a few lines of code, would allow further research to look into a number of different factors regarding the visual rewards. With the simple recoding of the experiment a large number of different aspects of the visual reward could be altered easily (e.g. duration, type etc) to look into the effect of the visual rewards.

The thesis presented methods to evaluate the localisation ability of children and developed a method which automatically evaluates the responses children make to the presented sounds. These methods could be used in future research which attempt to evaluate localisation responses in children. Such methods could use more advanced methods of motion tracking (e.g. a camera based system) to overcome the issues discussed in this thesis but the algorithms would remain the same.

The methods, as well as the classification techniques could also be used in adult studies where the responses are more ideal and in which the processing techniques may work to a higher level of accuracy. This would save the lengthy evaluation techniques used in the evaluation of adult localisation ability discussed in the previous chapters.

6.1.2 Application of the ANN Classifier in Real Time

Presented in this thesis was a behavioural test method which could be used, along with motion tracking technology, to analyse and classify responses the children made to the auditory stimuli. For the method to be usable in a real world setting, i.e. in a audiological clinic, the head turn responses the children make to the auditory stimulus must be analysed in real time and feedback given no more than one to two seconds after their response. In chapters four and five, extraction algorithms were discussed which could analyse the data around 15ms, however, due to time constraints of the project, the extraction algorithms were never tested in real time.

For the extraction algorithms to run in real time, a number of additions would be needed to be added to the control script and motion tracker controller so that data could be sent to the control script for analysis, analysed, and then images presented on the video screen. To evaluate the method it would need to be also scored by a human observer, the accuracy of the automatic method could then be evaluated.

6.1.3 Using the Method to Evaluate Localisation Ability in Bilateral Cochlear Implant Patients

The method set out to be a test which could be used to understand the development of the binaural system in young children and to be used as a tool for investigating aspects of bilateral cochlear implants. The method was shown to be able of measuring the head turn responses. However, for the method to be fully verified it would need to be tested on children with bilateral cochlear implants. This could be achieved using the current method discussed in this thesis or if the method was developed to analyse response in real time, it could be performed as part of a larger study.

Adaptation of the Method for a Clinical Setting

The method presented in this thesis took place in a custom built laboratory. Ideally the method would be used by anyone wishing to investigate bilateral cochlear implants in children, be that in a research laboratory or in an audiological clinic. The method would need to be performed in a sound booth. The first issue with the setup is the size of the video screens. Research into how the size of the screens affects the number of responses would be required. If such a method could be performed with smaller screens or even TV screens, it would make the method a far more viable test method.

The use of motion trackers are becoming more common. There are many such systems on the market and even open source alternatives which could be used to provide accurate results of head motion. Ideally, a video based approach would be used to monitor the children's responses. This would also reduce the issues regarding children wearing the motion tracking gear. This is discussed in more depth in the next section.

6.1.4 Remote Methods of Tracking

One of the main issues with the method was the use of motion tracking gear which distracted the child and caused distress. This resulted in fewer trials being obtained from the children. One way to overcome this would be by using a remote tracking technique. This would be a form of camera setup which would record and analyse the movements of the child in real time. Application of such methods have been done, however, the development of such a method would be far from trivial [129, 130, 131, 100].

Remote tracking would reduce the stress on the child as they would simply take part in the game without the need of first placing on the motion tracking gear;

this might produce more responses from the child. This sort of setup would require a camera setup along with the processing software.

6.1.5 Introducing Eye Tracking

The thesis discussed a method which tracked and evaluated just the head motion to the auditory stimulus, however, the eye movement is also an important factor in the localisation of sounds. The undershoots of the head to the target stimulus shown in the results is due to the head not fully turning to the target location (due to physiology) with the eyes making up the angular difference between the end of the head rotation and the auditory stimulus. To investigate the eye movement, eye tracking would be required. Several studies have looked into tracking the eyes of infants [132, 133, 134].

Research into eye movement with respect to the localisation of an auditory stimulus, to the authors knowledge, has not been conducted and would add to the literature on child localisation ability. Tracking both the head and eyes could help improve the classification rates of the classifiers.

6.2 Overall Conclusion on the *AnimalSeek* Method

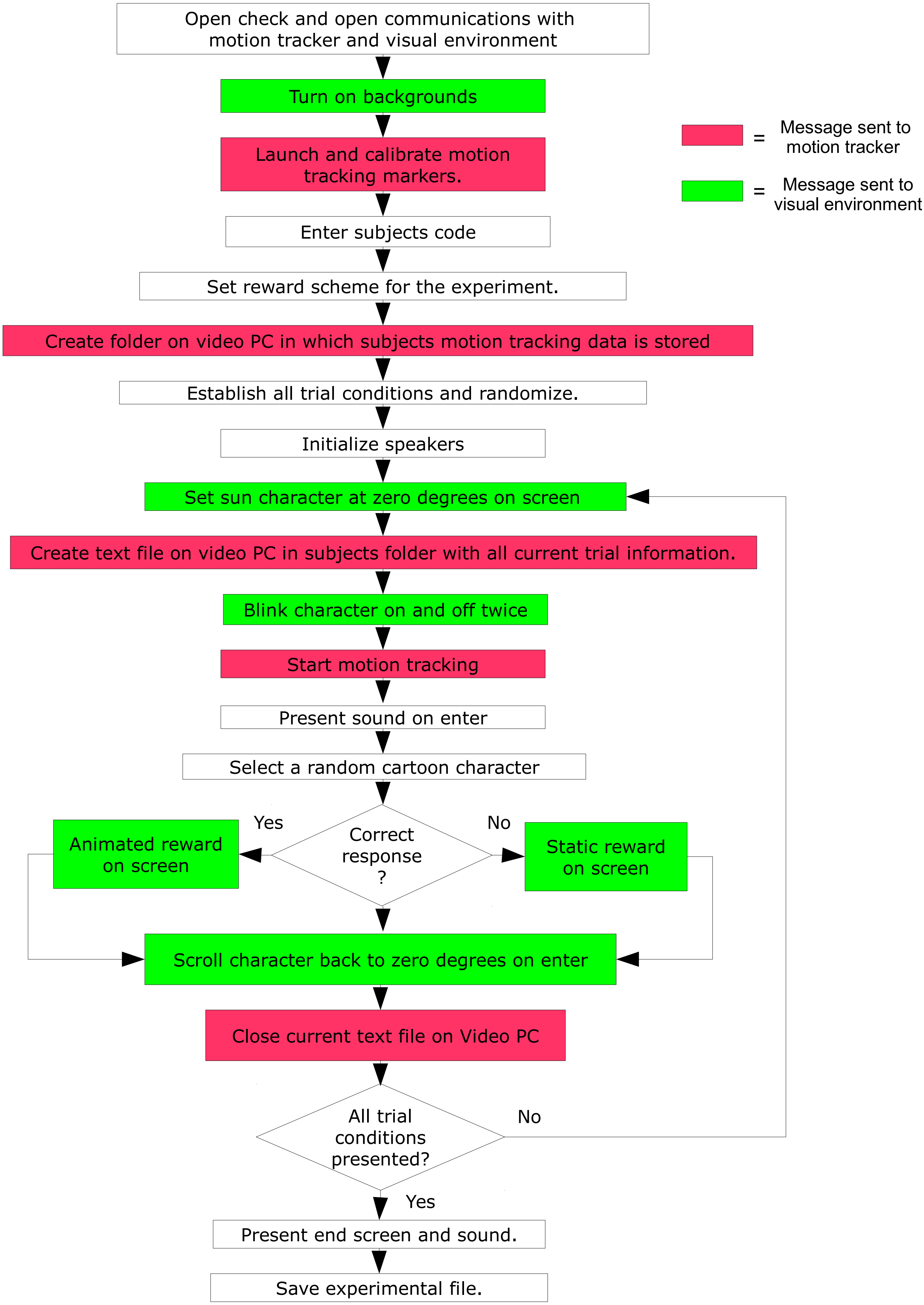
The thesis has proposed a new behavioural method which can be used to evaluate localisation ability. The method engages the child in a game-like task whilst evaluating their head turn response using motion tracking technology. It has been shown that the number of experimenters required to undertake the task theoretically can be reduced to one. This was achieved by using a combination of a visual environment, which engages the child and replaces the experimenter inside the AEC, and signal processing techniques which can judge responses made by the children, score them appropriately. This takes away any experimenter bias which might be caused by a human observer scoring the responses. Application of the method has not been undertaken in real time, however, the results suggest that the method could be used as discussed.

Appendices

Appendix A

Experimental script overview

Flow diagram of the experimental script developed in MATLAB. Boxes highlighted in pink represent that they send messages (over OSC) to the motion tracker. Boxes highlighted in green represent that they send messages to the visual environment. A full list of OSC messages to control the visual environment and the motion tracker can be found in Appendix B and Appendix C, respectively.


Open check and open communications with motion tracker and visual environment
Turn on backgrounds
Launch and calibrate motion tracking markers.
Enter subjects code
Set reward scheme for the experiment.
Create folder on video PC in which subjects motion tracking data is stored
Establish all trial conditions and randomize.
Initialize speakers
Set sun character at zero degrees on screen
Create text file on video PC in subjects folder with all current trial information.
Blink character on and off twice
Start motion tracking
Present sound on enter
Select a random cartoon character
Correct response ?
Yes
Animated reward on screen
No
Static reward on screen
Scroll character back to zero degrees on enter
Close current text file on Video PC
All trial conditions presented?
No
Yes
Present end screen and sound.
Save experimental file.
= Message sent to motion tracker
= Message sent to visual environment

Appendix B

Visual Environment Messages

A list of all of the OSC messages added to the visual environment as part of this thesis. The messages control parameters inside a series of structures relating to the objects and backgrounds. A maximum of four objects can be controlled on the screen at one time. Three backgrounds are used (one relating to each screen). All commands are sent over the OSC protocol and include a message containing datum. A list of the commands and their corresponding messages (including variable type) are shown on the next page.

/objectX	Select which object (X) wishing to be manipulated. Commands are sent as strings to turn the object 'on' and 'off''. Maximum of four objects on the screen at one time.
/objectX_texture_idx	Changes the character image of the objectX to 'idx'. Idx is an interger.
/objectX_pos	Change the position of ObjectX. Position sent as a 3x1 float array representing Az, El and Ro.
/objectX_size	Change the size of the object. Scaling factor sent as an integer

OSC messages to control the background and characters within the visual environment.

Appendix C

Motion tracker control messages

A list of all of the functions present inside the motion tracker controller along with the corresponding OSC commands which can be used to control them. Some of the OSC commands also can be used to pass data. The messages (*msg) are explained along with their variable type.

The motion tracker script must be activated on the V-PC before OSC messages can be sent to it.

OSC message	Function description
/close_serial	Closes the serial connection to the motion tracking unit
/create_file	Creates a file with the file name put as *msg. *msg is sent as a string.
/close_file	Close the file with the name specified in *msg. *msg is sent as a string.
/htstart	Start reading from the head tracker and writing to the last opened file. Read for the number of seconds specified in *msg (interger).
/calibration	Reads the motion tracker for one second. *msg contains the subjects information, sent as a string.
/set_path	Sets the home directory of the files to the directory specified in *msg. *msg sent as a string.
/create_folder	Creates a folder with the name *msg, *msg sent as a string.
/marker_launch	Launch the markers. Checking that the receivers identify all four motion tracking markers and make sure they are calibrated.
/marker_check	Check the markers. Making sure they all are calibrated correctly.

OSC messages to control the motion tracker.

Appendix D

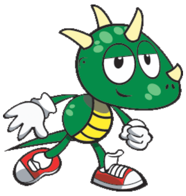
Participant Information documents

Six documents were sent out to interested parents / guardians. All documents had ethics approval from the University of Nottingham Psychology Department. The documents are as follows: 1) A contact letter explaining the study 2) Participant information form containing all the information which is needed by the parent regarding why the research was being done and why we would like them to take part. 3) General questionnaire regarding the child / children to be tested. 4) Consent form for the study. 5) Consent form for the video which are collected during the judgment of responses 6) Further contact forms for the study.

Before testing was undertaken, an overview of the project and why it was being undertaken was given along with what was to be expected during the testing sessions. The parent/guardian of the child was asked if they fully understood the study and had the oppitunity to ask questions.

As well as approaching the parents/guardians directly, posters were also put up in local community centers, libraries and post offices.

ADDRESS
ADDRESS
ADDRESS



Dear Parent,

Our ability to find where sounds are coming from, when we cannot see what is making the sound, is very important to our everyday lives. Just imagine walking in traffic if you couldn’t tell where the cars were. Babies are not able to do this at birth; this skill develops over the first few years of life. We currently have very little knowledge about when or how it develops in early childhood. One reason for this is that there has been no easy method to test for this skill with young children.

We at the Institute of Hearing Research are developing a new child-friendly method for testing how children find sounds and would like you and your child to help us with it.

We have enclosed an information sheet which provides more details about the study and what we would need you and your child to do if you are able to help with this work.

For this early part of the study, we need to work with children who have no known problems with their development. If this describes your child, we would like to invite you and your child to help us with our research. **If you think you are able to help us with the study, we need you to do 4 things:**

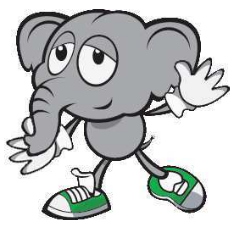
1) **Read the information sheet provided in this pack**
2) **Fill in the Consent Form**
3) **Fill in the Contact Form**
4) **Return both forms in the envelope provided (no stamp required)**

If your child has a hearing, eye sight, learning, physical or medical issues (even if it is only slight), we will not be able to work with them at this point. However, we may well need your help at a later stage when our method is better developed. If you and your child would like to help us with future studies, please complete the future studies form and return it to us in the envelope provided (no stamp required) so we can contact you in the future.

We have enclosed full information about the study. If you would like more details before making up your mind, please feel free to contact us via phone or email.

Thank you for reading this letter. We hope to meet you and your child in the near future.

Yours sincerely,



Dr Bernhard Seeber
Programme Leader Auditory Research
Tel.: 0115 951 8508 ext. 205
E-mail: seeber@ihr.mrc.ac.uk

INFORMATION ABOUT THE "FINDING SOUNDS"-STUDY

STUDY TITLE

How children find sounds

RESEARCHERS

Dr Bernhard Seeber	Programme Leader Auditory Research
Mr Damon McCartney	PhD Student

AIMS OF THE STUDY

- We need to develop a new method for testing how children find (locate) sounds that come from different directions. The test will take place in a room, which has sound absorbing material on the walls, ceiling and floor. The room also contains an array of loudspeakers and three large video screens. This will allow us to play sounds from all directions while presenting fun pictures or cartoons on the videos screens.
  Due to the "FINDING SOUNDS"-study being in its developmental stage we are currently looking for children who are free from hearing, vision, behavioural, physical and medical issues, even if only slight.
- To measure how your child locates sounds we have designed special items of clothing for your child to wear which have sensors built in. These are:

  1) wrist-bands
  2) a lightweight jacket
  3) a bicycle helmet.

  The sensors report where the hands are relative to the body and how far your child turns his or her head to the sound. We hope most children will be happy to wear the special clothing as this will help us to be sure our measures are accurate. However, if your child feels uncomfortable wearing the helmet he/she does not have to.

- With your permission, we will also video-tape your child when doing the study. This will give us the chance to check your child reactions to the sounds and see if they agree with the measures taken. If your child is not able to be video-taped they are still able to take part.
- From the measured positions of hands, head and eyes and the video recordings we will learn how your child responds to the sounds we play and how easy he or she is able to find from where the sound came. To make the task more enjoyable for your child we will present visual 'rewards' on the screens such as interesting pictures and cartoons. We believe this will make the study fun and game-like for your child.
- We are asking several questions in relation to finding sounds such as:
    1) Do the children like to see a picture where they think the sound is – a visual reward
    2) When is the best time to give a visual reward
    3) Where is the best place to show a visual reward
    4) Which reward to choose
    5) If there is any difference in the child's ability to find different types of sound (animal sounds/quiet noise)
    6) Do younger children find it more difficult to find sounds than older children?

WHY WE HAVE CONTACTED YOU

We have approached you because you have a child who is in the right age to help us with this study.

HOW YOU AND YOUR CHILD CAN HELP?

If you and your child are able to help in our study, we will need:

- You to sign the enclosed consent forms and return it to us in the envelope provided – No stamp required.
- You and your child to come to the Institute. We will make arrangement over the phone and make it a date and time convenient for you.
- To work/play with your child for between 1-3 hour. We will try to complete the work in one day with older children. It is unlikely young children will be able to complete the work in one visit. If necessary we might need you and your child to come for up to three 1 hour sessions which may need to be split over 3 days. Each test session will be broken up into small blocks with lots of rests between.

These breaks are important and we will have toys on hand for your child to play with during these breaks and we will provide drinks and biscuits for you and your child. We also understand if your child decides he/she does not want to co-operate on the day. We will stop and arrange a date for another visit.

Your child will be invited to play a short 'game' which is in fact a test described below.

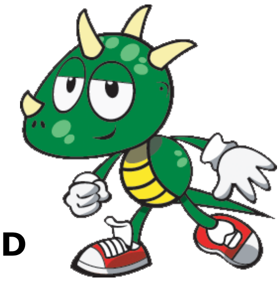
- We will ask you to sit with your child inside the test room. If your child feels confident to sit in the room alone that would be fine. People in the test room are in constant contact with the researchers who are in the next room. This is via a two-way intercom system and CCTV.
- Once ready we will play a sound, e.g. a voice speaking a name, at a safe and comfortable level. We are interested in measuring the movement of your child towards the sound. If your child locates the sound a short visual reward may be presented; this reward will either be a picture such as those shown in this letter or a short animation.
- By rewarding your child we hope to keep your child interested in the task for a longer period of time, making testing easier and more fun for your child.

If you and your child agree to participate, your child will receive a hearing test. If his/her thresholds do not fall within normal ranges, we will not be able to use your child in this research, but we will provide a treatment letter for your child's family doctor. We will reimburse your travel expenses. It is most helpful if you can let us have your bus/train/parking tickets and or taxi receipts as we need these for accounting. You will receive an inconvenience allowance to cover the time spent in the testing session. This is set by the MRC Institute of Hearing Research and is currently £5 per hour.

THE INFOMATION WE WILL COLLECT

We will make a record of the following:

- The date of the test
- Your child's date of birth and gender
- Any known previous hearing difficulties or ear infections your child may have had
- Which hand your child used to cut and draw
- Results of each test session
- **IF** consent given, video recordings of each experiment



WHAT HAPPENS IF YOU SAY 'YES', BUT CHANGE YOUR MIND

If you change your mind, at whatever stage of the study, and decide you no longer want to take part you are free to withdraw. Similarly, if your child decides he/she no longer wants to take part, then he/she is also free to withdraw.

You are free to withdraw at any point, and do not need to give a reason for doing so. We will not pressure you or your child to participate.

CONFIDENTIALITY

All the information we collect will be made anonymous and treated confidentially. It will not be linked to you or your family in any report or publication. None of the information we collect will be added to your child's medical notes or discussed with anyone outside the stated research teams. The Medical Research Council is covered by the terms of the Data Protection Act.

If consent is given for the experiment to be recorded on video then these recordings will be treated confidential to the degree you agreed them to be presented to third parties. In no case you or your child's name will be disclosed. Please see the video consent form for more information.

Thank you for reading this information sheet.

We hope you and your child will consider
taking part in our study!

Participant code:

Taking part in the study
HOW CHILDREN FIND SOUNDS

This questionnaire is to be completed by the parent/guardian of the child taking part in the 'how children find sounds' study.

1. Today's date: ……………./……………./…………….

2. Child's Date of Birth? (dd/mm/yyyy) ……………./……………./…………….

3. Child's Gender? **Male / Female**
   *(Please **circle** appropriate option)*

4. Which hand does your child predominantly use to e.g. write, draw, cut?
   *(Please **circle** appropriate option)*

   Left / Right

5. Has your child suffered from an ear infection within the past three months or had glue ear?
   *(Please **circle** appropriate option)*

   Yes / No

CONSENT FORM for the study **"How Children Find Sounds"**

All information provided will be treated confidentially. It will be kept on a computerised database. None of the information will be disclosed to anyone outside the research team. The Medical Research Council is covered by the terms of the Data Protection Act.

I .... (Parent's name in block capitals)

and my child

.... (Child's name in block capitals)

are able/are not able* to help with this research study. (*Please delete as appropriate)

.... (Parent's signature)

............/............/............ (Date - DD/MM/YYYY)

If you are not able to help, this is all the information we require. Please return the form in the envelope provided. We will try to ensure you are not contacted again over this study.

If you are able to help, please read the following information and continue to complete the form:

- I understand that I can change my mind at any point in the study and withdraw my child from the study without prejudice to any service I or my child may need now or in the future.
- I have read the information provided, and (where appropriate) have had my queries answered satisfactorily.

Your relationship to the child	....
Child's Date of Birth (dd/mm/yyyy)	............/............/............
Home Address	....
	....
Post code	....
Telephone number/s	....
Email	.... (if available)

THANK YOU FOR COMPLETING THIS FORM

Please return it as soon as possible in the enclosed envelope (no stamp required)

Video Consent Form

As part of the study videos recordings are to be made of your child which will include you if you are holding your child. Cameras will be placed discreetly inside the specially designed room. These recordings will be used by us for later analysis of your child's responses. The recordings will not be given to anyone outside the research team.

If you decide you do not want the experiment to be recorded please tell the experimenters. The refusal to be recorded does not affect your child's ability to participate in the study and a reason does not have to be given, however if you do not mind the experiment being recorded we would be grateful of the video for our research because it would help us check your child's responses to the sounds more accurately. Thank you for your help.

TO BE COMPLETED BY THE CHILD'S PARENT/GUARDIAN

I have read and understood the above information and give / do not give my permission for me and my child to be recorded during the experiment.

Signature of child's parent or guardian.

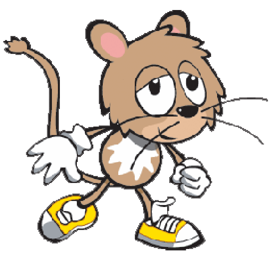
……………………………………………………………………………………

Printed Name

……………………………………………………………………………………

Date

……………………………



MRC | Institute of Hearing Research

Video Consent Form – Permission to present recordings at scientific meetings

Thank you for agreeing to video recordings being taken of your child and of you (if you are holding your child) during the experiment. We believe that it will be important to demonstrate the procedure of the experiment and responses of children to sounds to a scientific audience. We hereby ask you if the recordings of you and your child taken during the experiments may be presented to a scientific audience at scientific meetings. Please be assured that we will not disclose your or your child's name.

If you decide you do not want recordings to be presented at scientific meetings please tell the experimenters. The refusal to do so does not affect your child's ability to participate in the study; however, if you do not mind the recordings being presented, it may help other researchers learn better from our research results and experience. Thank you for your help.

TO BE COMPLETED BY THE CHILD'S PARENT/GUARDIAN

I have read and understood the above information and give / do not give my permission for video recordings of me and my child to be presented at scientific meetings.

Signature of child's parent or guardian.

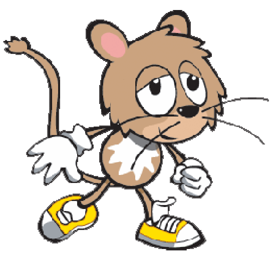
...

Printed Name

...

Date

....



MRC Institute of Hearing Research

Video Consent Form – Permission to present recordings to lay audience and on the internet

You have agreed to video recordings being taken of your child and of you (if you are holding your child) during the experiment. Our research is funded by the taxpayer and we frequently present our research results to the public, i.e. in schools, in interviews and on our web-pages. We anticipate that it will be important to demonstrate the procedure of the experiment and responses of children to sounds to the public to help explain the research and its results. We hereby ask you if the video-recordings taken during the experiments may be presented to the public in the ways described which include scientific areas on the internet. Please be assured that we will not disclose your or your child's name.

If you decide you do not want recordings to be presented to the public please tell the experimenters. The refusal to do so does not affect your child's ability to participate in the study; however, if you do not mind recordings to be presented it may help explain our research to the public. Thank you for your help.

TO BE COMPLETED BY THE CHILD'S PARENT/GUARDIAN

I have read and understood the above information and give / do not give my permission for video recordings of me and my child to be presented to the public, e.g. in interviews, presentations or on the internet.

Signature of child's parent or guardian.

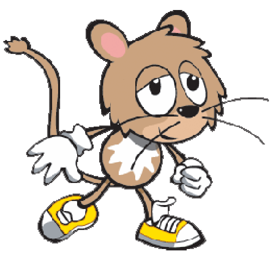
……………………………………………………………………………………

Printed Name

……………………………………………………………………………………

Date

……………………………



“**How children find sounds**”

Contact Form

I am interested in taking part with my child in this study at the MRC Institute of Hearing Research.

Please contact me with more information about the study.

Please fill out in **BLOCK CAPITALS.**

My first name: --

My Family Name: --

My Child’s First Name: --

My Child’s Family Name: --

Home Address: --

Town: --

Postcode: --

Phone (daytime): --

Phone (evening): --

Mobile: --

Email: --

Signature: --

Date (dd/mm/yyyy): ---------/---------/---------

“How children find sounds”

Future Studies Form

I am interested in taking part with my child in future studies when the methods have been developed further.

Please contact me with more information about future studies.

Please fill out in **BLOCK CAPITALS**

My first name: ---

My Family Name: ---

My Child’s First Name: ---

My Child’s Family Name: ---

Home Address: ---

Town: ---

Postcode: ---

Phone (daytime): ---

Phone (evening): ---

Mobile: ---

Email: ---

Signature: ---

Date (dd/mm/yyyy): ---------/---------/---------

Please continue to complete the next set of questions. These are designed to help us understand better what will be needed to ensure we make the test/ games fully appropriate for your child.

My child has some difficulty with the following:

Hearing? **YES / NO**
*(Please **circle** as appropriate)*

If YES, please give brief details

Eye sight? **YES / NO**
*(Please **circle** as appropriate)*

If YES, please give brief details

Learning? **YES / NO**
*(Please **circle** as appropriate)*

If YES, please give brief details

Medical? **YES / NO**
*(Please **circle** as appropriate)*

If YES, please give brief details

MRC Institute of Hearing Research

PARENTS!

PLAY A GAME WITH YOUR CHILD AND HELP RESEARCH!

Where?

MRC Institute of Hearing Research, University Park, Nottingham, NG7 2RD

Who?

Children aged 1-5 with no ear or hearing problems

Interested?
Please Contact:

Miss Carla Church
0115 951 8508 ext. 202
carla@ihr.mrc.ac.uk

Inconvenience allowance and travel costs paid

Carla Church
0115 951 8508 ext. 202
carla@ihr.mrc.ac.uk

Carla Church
0115 951 8508 ext. 202
carla@ihr.mrc.ac.uk

Carla Church
0115 951 8508 ext. 202
carla@ihr.mrc.ac.uk

Carla Church
0115 951 8508 ext. 202
carla@ihr.mrc.ac.uk

Carla Church
0115 951 8508 ext. 202
carla@ihr.mrc.ac.uk

Carla Church
0115 951 8508 ext. 202
carla@ihr.mrc.ac.uk

Carla Church
0115 951 8508 ext. 202
carla@ihr.mrc.ac.uk

Carla Church
0115 951 8508 ext. 202
carla@ihr.mrc.ac.uk

Carla Church
0115 951 8508 ext. 202
carla@ihr.mrc.ac.uk

Carla Church
0115 951 8508 ext. 202
carla@ihr.mrc.ac.uk

Carla Church
0115 951 8508 ext. 202
carla@ihr.mrc.ac.uk

Carla Church
0115 951 8508 ext. 202
carla@ihr.mrc.ac.uk

HOW CHILDREN FIND SOUNDS

MRC Institute of Hearing Research

MRC Institute of Hearing Research
How Children Find Sounds

MRC Institute of Hearing Research
How Children Find Sounds
MRC Institute of Hearing Research
How Children Find Sounds
MRC Institute of Hearing Research
How Children Find Sounds
MRC Institute of Hearing Research
How Children Find Sounds
MRC Institute of Hearing Research
How Children Find Sounds
MRC Institute of Hearing Research
How Children Find Sounds
MRC Institute of Hearing Research
How Children Find Sounds
MRC Institute of Hearing Research
How Children Find Sounds
MRC Institute of Hearing Research
How Children Find Sounds
MRC Institute of Hearing Research
How Children Find Sounds
MRC Institute of Hearing Research
How Children Find Sounds

Bibliography

[1] L. Rayleigh, "On our perception of sound direction," *Philos Mag*, vol. 13, pp. 214–232, 1907.

[2] W. Fedderson, T. Sandel, D. Teas, and L. Jeffress, "Localization of high-frequency tones," *J Acoust Soc Am*, vol. 15, pp. 149–161, 1957.

[3] F. Wightman and D. Kistler, "The dominant role of low-frequency interaural time differences in sound localization." *J Acoust Soc Am*, vol. 91, no. 3, pp. 1648–1661, 1992.

[4] C. Lorenzi, S. Gatehouse, and C. Lever, "Sound localization in noise in normal-hearing listeners." *J Acoust Soc Am*, vol. 105, pp. 1810–1820, 1999.

[5] G. Kavanagh and J. Kelly, "Midline and lateral field sound localization in the ferret (mustela putorius): contribution of the superior olivary complex." *J Neurophysiol*, vol. 67, no. 6, pp. 1643–1658, 1992.

[6] C. Tsuchitani and J. Boudreau, "Stimulus level of dichotically presented tones and cat superior olive s-segment cell discharge." *J Acoust Soc Am*, vol. 46, no. 4B, pp. 979–988, 1969.

[7] P. B. JM. Goldberg, "Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization." *J N Physiol*, vol. 32, no. 4, pp. 613–636, 1969.

[8] S. Carlile and D. Pralong, "Auditory spectral cues for the resolution of front back confusion in humans." *J Acoust Soc Am*, vol. 92, pp. 2297–2297, 1992.

[9] K. Davis, R. Ramachandran, and B. May, "Auditory processing of spectral cues for sound localization in the inferior colliculus." *J Assoc Res Otolaryngol*, vol. 4, pp. 148–163, 2003.

[10] R. Domnitz and H. Colburn, "Lateral position and interaural discrimination," *J Acoust Soc Am*, vol. 61, pp. 1586–1598, 1977.

[11] D. Moore and C. Plack, *The Oxford Handbook of Auditory Science: Hearing*. Oxford, 2010.

[12] R. Klumpp and H. Eady, "Some measurements of interaural time difference thresholds," *J Acoust Soc Am*, vol. 28, pp. 859–860, 1956.

[13] R. Domnitz and H. Colburn, "Lateral position and interaural discrimination," *J Acoust Soc Am*, vol. 61, pp. 1586–1598, 1977.

[14] G. Henning, "Detectability of interaural delay in high-frequency complex waveforms," *J Acoust Soc Am*, vol. 55, no. 1, pp. 84–90, 1974.

[15] S. Colburn, B. Shinn-Cunningham, J. Kidd, and N. Durlach, "The perceptual consequences of binaural hearing," *Int J Audiol*, vol. 45, pp. 424–430, 2006.

[16] W. A. Mills, "On the minimum audible angle," *J Acoust Soc Am*, vol. 6, no. 4, pp. 237–246, 1957.

[17] A. Mills, "Lateralization of high-frequency tones," *J Acoust Soc Am*, vol. 32, no. 1, pp. 132–134, 1960.

[18] B. Moore, *An Introduction to the Psychology of Hearing, Fifth Edition*. Academic Press, April 2003.

[19] J. Makous and J. Middlebrooks, "Two-dimensional sound localization by human listeners," *J Acoust Soc Am*, vol. 87, pp. 2188–2199, 1989.

[20] S. S. Stevens and E. B. Newman, "The localization of actual sources of sound," *Am J Psychol*, vol. 48, no. 2, pp. 297–306, 1936.

[21] S. Getzmann, "The effect of eye position and background noise on vertical sound localization," *Hear Res*, vol. 169, pp. 130–139, 2002.

[22] F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening *ii*: Psychophysical validation." *J Acoust Soc Am*, vol. 85, pp. 868–878, 1989.

[23] R. Butler, E. Levy, and W. Neff, "Apparent distance of sounds recorded in echoic and anechoic chambers," *J Exp Psychol Hum Percept Perform*, vol. 6, pp. 745–750, 1980.

[24] R. Gilkey, M. Good, M. Ericson, J. Brinkman, and J. Steward, "A pointing technique for rapidly collecting responses in auditory research," *Behav. Res. Meth. Instr. and Comp*, vol. 27, pp. 1–11, 1995.

[25] T. Djelani, C. Porschmann, J. Sahrhage, and J. Blauert, "An interactive virtual-environment generator for psychoacoustic research *ii*: Collection of head-related impulse responses and evaluation of auditory localization," *Acustica*, vol. 86, pp. 1046–1053, 2000.

[26] L. Haber, R. Haber, S. Penningroth, K. Novak, and H. Radgowski, "Comparison of nine methods of indicating the direction to objects: data from blind adults." *Perception*, vol. 22, no. 1, pp. 35–47, 1993.

[27] W. R. Thurlow and P. S. Runge, "Effect of induced head movements on localization of direction of sounds," *J Acoust Soc Am*, vol. 42, pp. 480–488, 1967.

[28] J. Soechting and M. Flanders, "Sensorimotor representations for pointing to targets in three-dimensional space," *J. Neurophysiol*, vol. 63, pp. 868–878, 1989.

[29] J. Lewald and W. Ehrenstein, "Auditory-visual spatial integration: A new psychphysical approach using laser pointing to acoustic targets," *J Acoust Soc Am*, vol. 104, no. 3, pp. 1586–1597, 1998.

[30] B. Seeber, "A new method for localization studies," *Acustica*, vol. 6, no. 88, pp. 446–450, 2002.

[31] B. Razavi, W. O'Neill, and G. Paige, "Auditory spatial perception dynamically realigns with changing eye position." *J Neurosci*, vol. 27, no. 38, pp. 10 249–10 258, 2007.

[32] H. Heuermann and H. Colonius, *Localization experiments with saccadic responses in virtual auditory environments. Psychophysics, Physiology and Models of Hearing*. World Scientific, 1999.

[33] L. Yao and C. K. Peck, "Saccadic eye movements to visual and auditory targets," *Exp Brain Res.*, vol. 115, pp. 25–34, 1997.

[34] P. M. Hofman and A. Opstal, "Spectro-temporal factors in two-dimensional human sound localization." *J Acoust Soc Am*, vol. 103, pp. 2634–2648, 1998.

[35] L. C. Populin, "Human sound localization: measurements in untrained, headunrestrained subjects using gaze as a pointer," *Exp Brain Res*, vol. 190, no. 1, pp. 11–30, 2008.

[36] J. Fuller, "Head movement propensity," *Exp Brain. Res.*, vol. 92, pp. 152–164, 1992.

[37] M. Middelweerd and R. Plomp, "The effect of speech-reading and the speech-perception threshhold of sentences in noise," *J Acoust Soc Am*, vol. 82, no. 6, pp. 2145–2147, 1987.

[38] W. Thurlow, J. Mangels, and P. Runge, "Head movements during sound localization," *J Acoust Soc Am*, vol. 42, pp. 489–493, 1967.

[39] G. H. Recanzone, S. D. Makhamra, and D. C. Guard, “Comparison of relative and absolute sound localization ability in humans,” *J Acoust Soc Am*, vol. 103, pp. 1085–1097, 1998.

[40] S. Carlile, P. Leong, and S. Hyams, “The nature and distribution of errors in sound localization by human listeners,” *Hear Res*, vol. 114, pp. 179–196, 1997.

[41] W. Brimijoin, D. McShefferty, and M. Akeroyd, “Auditory and visual orienting responses in listeners with and without hearing-impairment,” *J Acoust Soc Am*, vol. 127, no. 6, pp. 3678–3688, 2010.

[42] M. Dix. and C. Hallpike, “The peep show: New technique for pure-tone audiometry,” *Brit Med J*, vol. 2, p. 719, 1947.

[43] T. Suzuki and Y. Ogiba, “Conditioned orientation reflex audiometry,” *Arch Otolaryngol*, vol. 74, pp. 84–90, 1961.

[44] P. Statten and D. Wishart, “Pure tone audiometry in young children: Psychogalvanic-skin-resistance and peep-show,” *Ann Otol*, vol. 65, pp. 511–534, 1956.

[45] A. Kankkunen and G. Lidén, “Visual reinforcement audiometry,” *Acta Oto-laryngologica*, vol. 67, pp. 281–292, 1969.

[46] N. C. Group, “Visual reinforcement audiometry testing of infants: A recommended test protocol,” - hearing.screening.nhs.uk/getdata.php?id=10763, 2008, last viewed: 07/02/2011.

[47] M. Schmida, H. P. HJ, and A. Tharpe, "Visual reinforcement audiometry using digital video disc and conventional reinforcers." *Am J Audiol*, vol. 12, pp. 35–40, 2003.

[48] L. Olsho, E. Koch, C. Halpin, and E. Carter, "An observer-based psychoacoustics procedure for use with young infants," *J Acoust Soc Am*, vol. 23, pp. 627–640, 1987.

[49] S. E. Threhub and B. Schneider, *Auditory Development in Infancy*. Plenum Press, 1985.

[50] B. Kisilevsky, S. Hains, K. Lee, X. Xie, H. Huang, H. Ye, K. Zhang, and Z. Wang, "Effects of experience on fetal voice recognition." *Psychol Sci*, vol. 14, pp. 220–224, 2003.

[51] M. Clarkson, R. Clifton, and B. Morrongiello, "The effects of sound duration on newborns′ head orientation," *J Exp Child Psychol*, vol. 39, pp. 20–36, 1985.

[52] D. Moore, M. Ferguson, A. Edmondson-Jones, S. Ratib, and A. Riley, "Nature of auditory processing disorder in children." *Pediatrics*, vol. 26, no. 2, pp. 382–90, 2010.

[53] A. Holmes, A. Niskar, S. Kieszak, C. Rubin, and D. B. DJ., "Mean and median hearing thresholds among children 6 to 19 years of age: the third national health and nutrition examination survey, 1988 to 1994, united states." *Ear Hear*, vol. 25, no. 4, pp. 397–402, 2004.

[54] M. Devous, D. Altuna, N. Furl, W. Cooper, G. Gabbert, W. Ngai, S. Chiu, J. Scott, T. H. TS, J. Payne, and E. Tobey, "Maturation of speech and language functional neuroanatomy in pediatric normal controls." *J Speech Lang Hear Res.*, vol. 49, no. 4, pp. 856–66, 2006.

[55] B. Morrongiello, K. Kenwick, and G. Chance, "Sound localisation acuity in very young infants: An observer-based testing procedure," *Dev Psychol*, vol. 26, pp. 75–84, 1990.

[56] A. B. Morrongiello, "Infants localization of sounds along the horizontal axis: estimates the minimum audible angle," *Dev Psychol*, vol. 6, no. 24, pp. 8–13, 1988.

[57] R. Litovsky, "Developmental changes in the precedence effect: Estimates of minimum audible angle," *J Acoust Soc Am*, vol. 102, pp. 1739–1745, 1997.

[58] B. Morrongiello and P. Rocca, "Infants localization of sounds within hemifields : Estimates of minimum audible angle," *Child Dev*, vol. 61, pp. 1258–1270, 1990.

[59] J. Moore, G. Thompson, and M. Thompson, "Auditory localization of infants as a function of reinforcement conditions," *J Speech Hear Disord*, vol. 40, pp. 29–34, 1974.

[60] M. Primus, "The role of localization in visual reinforcement audiometry," *J Speech Hear Res*, vol. 35, no. 5, pp. 1137–1141, 1992.

[61] B. Morrongiello, K. D. Fenwick, L. Hiller, and G. Chance, "Sound localization in newborn human infants," *Dev Psychobiol*, vol. 27, no. 8, pp. 519–538, 1994.

[62] B. Morrongiello and P. Rocca, "Infants' localization of sounds in the horizontal plane: Effects of auditory and visual cues," *Child Dev*, vol. 58, pp. 918–927, 1987.

[63] L. V. Deun, A. van Wieringen, T. V. den Bogaert, F. Scherf, F. Offeciers, P. V. de Heyning, C. Desloovere, I. J. Dhooge, N. Deggouj, L. D. Raeve, and J. Wouters, "Sound localization, sound lateralization, and binaural masking level differences in young children with normal hearing," *Ear Hear*, vol. 30, pp. 178–190, 2009.

[64] B. Seeber, S. Kerber, and E. Hafter, "A system to simulate and reproduce audio-visual environments for spatial hearing research," *Hear Res*, vol. 260, no. 1-2, pp. 1–10, 2010.

[65] Polhemus, "Liberty latus," http://www.polhemus.com/, 2010, last viewed 07/02/2012.

[66] Python, "Python programming language," http://www.python.org/, last viewed 20/11/2012.

[67] Epson, "Epson emp-tw2000," http://www.epson.com.au/products/projector/emptw2000.asp, last viewed 07/11/2012.

[68] OpenGL, "Opengl," http://www.opengl.org, last viewed 07/11/2012.

[69] M. Pecyna, "Sheep cartoon," http://upload.wikimedia.org/wikipedia/commons/d/d3/Sheep_in_gray.svg, last viewed 10/11/2012.

[70] B. Windsor, "Pig cartoon," http://www.brentwindsor.co.uk, last viewed 10/11/2012.

[71] unknown, "Goat cartoon," http://www.picgifs.com/, last viewed 10/11/2012.

[72] Sanyo, "Sanyo," http://uk.sanyo.com/home/, last viewed 07/11/2012.

[73] Alesis, "Ra500," http://www.alesis.com/ra500, last viewed 07/11/2012.

[74] G. Hoversten and J. Moncur, "Stimuli and intensity factors in testing infants," *Hear Res*, vol. 12, pp. 687–702, 1969.

[75] M. Thompson and G. Thompson, "Response of infants and young children as a function of auditory stimuli and test methods," *J Speech Hear Res*, vol. 15, pp. 699–707, 1972.

[76] M. Weiss, P. Zerlazo, and I. Swain, "Newborn response to auditory stimulus discrepancy," *Child Dev*, vol. 59, pp. 1530–1541, 1988.

[77] C. Olmstead and J. Villablanca, "Development of behavioral audition in the kitten," *Physiol Behav*, vol. 24, pp. 705–712, 1980.

[78] M. Clarkson, R. Clifton, and B. Morrongiello, "The effects of sound duration on newborns' head orientation," *J Exp Child Psychol*, vol. 39, pp. 20–36, 1985.

[79] M. Schmida, H. Peterson, and A. Tharpe, "Visual reinforcement audiometry using digital video disc and conventional reinforcers," *Am J Audiol*, vol. 12, pp. 34–40, 2003.

[80] C. Hicks, A. Tharpe, and D. Ashmead, "Behavioral auditory assessment of young infants: Methodological limitations or natural lack of auditory responsiveness," *Am J Audiol*, vol. 12, pp. 35–40, 2003.

[81] L. Werner and B. Kopyar, "A procedure for developing effective visual reinforcers for infant psychoacoustics," www.aro.org/archives/1994/286.html, 1994, last viewed 07/02/2012.

[82] B. Culpepper and G. Thompson, "Effects of reinforcer duration on the response behavior of preterm 2-year-olds in visual reinforcement audiometry," *Ear Hear*, vol. 15, no. 2, pp. 161–167, 1994.

[83] G. Thompson, M. Thompson, and A. McCall, "Strategies for increasing response behavior of 1 and 2 year old children during visual reinforcement audiometry," *Ear Hear*, vol. 13, no. 4, pp. 236–240, 1992.

[84] D.H.Mcburney and T.L.White, *Research Methods*. Wadsworth/Thomson Learning, 2004.

[85] J. van der Geest, "Latin square toolbox," http://www.mathworks.com/matlabcentral/fileexchange/12315-latsq, 2009, last viewed 05/05/2012.

[86] G.Thompson, M.Thompson, and A. McCall, "Strategies for increasing response behavior of 1 and 2 year old children during visual reinforcement audiometry," *Ear Hear*, vol. 13, pp. 236–240, 1992.

[87] C. B and G. Thompson, "Effects of reinforcer duration on the response behavior of preterm 2-year-olds in visual reinforcement audiometry," *Ear Hear*, vol. 15, no. 2, pp. 161–167, 1994.

[88] M. Thompson, G. Thompson, and S. Vethivelu, "A comparison of audiometric test methods for 2-year-old children," *J Speech Hear Disord*, vol. 54, pp. 174–179, 1989.

[89] M. Primus and G. Thompson, "Response strength of young children in operant audiometry," *J Speech Hear Res*, vol. 28, pp. 539–547, 1985.

[90] G. Thompson, M. Thompson, and A. McCall, "Strategies for increasing response behavior of 1 and 2 year old children during visual reinforcement audiometry," *Ear Hear*, vol. 13, p. 1992, 236-240.

[91] G. Thompson and R.C.Folsom, "Reinforced and nonreinforced head-turn responses of infants as a function of stimulus bandwidth," *Ear Hear*, vol. 6, no. 3.

[92] M. Primus, "Response and reinforcement in operant audiometry," *J Speech Hear Disord*, vol. 52, pp. 288–294, 1987.

[93] M. Schmida, H. Peterson, and A. Tharpe, "Visual reinforcement audiometry using digital video disc and conventional reinforcers," *AJA*, vol. 12, pp. 35–40, 2003.

[94] D.Muir and J.Field, "Newborn-Infants Orient to Sounds," *Child Dev*, vol. 50, no. 2, pp. 431–436, 1979.

[95] J. Goldring, M. C. M. C. Dorris, B. Corneil, P. Ballantyne, and D. Munoz, "Combined eye-head gaze shifts to visual and auditory targets in humans," *Exp Brain Res*, vol. 111, pp. 68–78, 1996.

[96] B. Morrongiello and R. Clifton, "Effects of sound frequency on behavioral and cardiac orienting in newborn and five-month-old infants," *J Exp Child Psychology*, vol. 38, pp. 429–446, 1984.

[97] T. GriecoCalub and R. Litovsky, "Sound localization skills in children who use bilateral cochlear implants and in children with normal acoustic hearing." *Ear Hear*, vol. 31, no. 5, pp. 645–656, 2010.

[98] R. Allison and M. E. B. Cheung, "Combined head and eye tracking system for dynamic testing of the vestibular system," in *IEEE Trans Biomedical Engineering*, 2002.

[99] Z. Zivkovic and F. van der Heijden, "A stabilized adaptive appearance changes model for 3d head tracking," in *Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, 2001. Proceedings. IEEE ICCV Workshop on*, 2004.

[100] C. Wen-Chung and C. Chih-Wei, "Active head tracking using integrated contour and template matching in indoor cluttered environment," in *IEEE International Conference on Systems, Man and Cybernetics, 2006. SMC '06.*, 2006.

[101] J. Franchak, K. Kretch, K. Soska, J. Babcock, and K. Adolph, "Head-mounted eye-tracking of infants' natural interactions: a new method," in *Proceedings of the 2010 Symposium on Eye-Tracking Research 38 Applications*, 2010.

[102] F. Coutinho and C. Morimoto, "Head-mounted eye-tracking of infants' natural interactions: a new method," in *Computer Graphics and Image Processing, 2006. SIBGRAPI '06. 19th Brazilian Symposium on*, 2006.

[103] M. H. Ko, G. West, S. Venkatesh, and M. Kumar, "Using dynamic time warping for online temporal fusion in multisensor systems," *Inf Fusion*, vol. 9, pp. 370–388, July 2008.

[104] A. Bobick and J. Davis, "Real-time recognition of activity using temporal templates," in *Applications of Computer Vision, 1996. WACV '96., Proceedings 3rd IEEE Workshop on*, 1996, pp. 39–42.

[105] H. Sakoe, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 26, pp. 43–49, 1978.

[106] S. Sempena, N. Maulidevi, and P. Aryan, “Human action recognition using dynamic time warping,” in *International Conference on Electrical Engineering and Informatics (ICEEI)*, 2011.

[107] T. Vajda, “Action recognition based on fast dynamic-time warping method,” in *IEEE 5th International Conference on Intelligent Computer Communication and Processing, 2009. ICCP 2009.*, 2009, pp. 127–131.

[108] V. Vuckovic, “Dynamic time-warping method for isolated speech sequence recognition,” in *5th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Service. TELSIKS 2001*, 2001.

[109] P. P. Micó, “Dtw toolbox for matlab,” http://www.mathworks.com/matlabcentral/fileexchange/16350, last viewed 07/02/2012.

[110] V. Vapnik, *Estimation of Dependences Based on Empirical Data*. New York: Springer-Verlag, 1982.

[111] C. Corinna and V. Vladimir, “Support-vector networks,” *Mach Learn*, vol. 20, pp. 273–297, 1995.

[112] D. Eads, “Osu-svm toolbox for matlab,” http://sourceforge.net/projects/svm/, last viewed 07/02/2012.

[113] S. Cooray and N. O’Connor, “Facial feature extraction and principal component analysis for face detection in color images,” in *International Conference on Image Analysis and Recognition*, 2004, p. 1.

[114] Y. Yuan and K. Barner, "Hybrid feature selection for gesture recognition using support vector machines," in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008.*, 2008, pp. 1941–1944.

[115] V. Kohir and U. Desai, "Face recognition using a dct-hmm approach," in *Fourth IEEE Workshop on Applications of Computer Vision, 1998. WACV '98.*, 1998, pp. 226–231.

[116] A. Sarkaleh, F. Poorahangaryan, B. Zanj, and A. Karami, "A neural network based system for persian sign language recognition," in *IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, 2009, pp. 145–149.

[117] MATLAB, *Neural Network Toolbox: version 6.0.2 (R2009a)*. Natick, Massachusetts: The MathWorks Inc., 2009.

[118] W. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *B Math Biol*, vol. 5, pp. 115–133, 1943.

[119] M. Minksy and S. Papert, *Perceptrons*. MIT Press, 1969.

[120] H. Marzi, M. Turnbull, and E. Marzi, "Use of neural networks in forecasting financial market," in *IEEE Conference on Soft Computing in Industrial Applications, SMCia '08.*, May 2009, pp. 240–245.

[121] H. Kordylewski, D. Graupe, and L. Kai, "Ieee transactions on information technology in biomedicine," in *A novel large-memory neural network as an aid in medical diagnosis applications*, 2002, pp. 202–209.

[122] T. Yung, C. Chih-Ho, F. Hui-Yi, and C. Shin-Liang, “Apply computer vision and neural network to glue dispenser route inspection,” in *International Conference on Mechatronics and Automation, 2007. ICMA 2007.*, September 2007, pp. 3882–3887.

[123] P. Picton, *Neural Networks: Second Edition*. Palgrave, 2000.

[124] K. Levenberg, “A method for the solution of certain non-linear problems in least squares,” *Quarterly of Applied Mathematics*, vol. 2(2), pp. 164–168, 1944.

[125] D. Marquardt, “An algorithm for the least-squares estimation of nonlinear parameters,” *SIAM J. Applied Mathematics*, vol. 11, pp. 431–441, 1963.

[126] Matlab, “Levenberg-marquardt backpropagation,” http://www.mathworks.co.uk/help/toolbox/nnet/ref/trainlm.html, 2012, last viewed 12/05/2012.

[127] A. Suratgar, M. Tavakoli, and A. Hoseinabadi, “Modified levenberg-marquardt method for neural networks training,” *World Academy of Science, Engineering and Technology*, vol. 6, pp. 46–48, 2005.

[128] S. Russell and P. Norvig, *Artifical Intelligence: A Modern Approach*. Pearsons Education, 2001.

[129] S. Ohayon and E. Rivlin, “Robust 3d head tracking using camera pose estimation,” 2006, pp. 1063 – 1066.

[130] R. Wooju and K. Daijin, "Real-time 3d head tracking and head gesture recognition," in *Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on*, 2007.

[131] J. Lee, "Head tracking for desktop vr displays using the wii remote," http://johnnylee.net/projects/wii/, 14/10/2011, last viewed 07/02/2012.

[132] J. Franchak, K. Kretch, K. Soska, J. Babcock, and K. Adolph, "Head-mounted eye-tracking in infants' natural interactions: A new method." in *Proceedings of the 2010 Symposium on Eye Tracking Research and Applications.*, 2010, pp. 1–7.

[133] R. Aslin and B. McMurray, "Automated corneal-reflection eye tracking in infancy: Methodological developments and applications to cognition." *Infancy*, vol. 6, pp. 155–163, 2004.

[134] H. Yoshida and L. Smith, "What's in view for toddlers? using a head camera to study visual experience." *Infancy*, vol. 13, pp. 229–248, 2008.