A mechatronic approach to supernormal auditory localisation for telepresence

C.S. Harrison*, G.M. Mair

The Transparent Telepresence Research Group, Department of Design, Manufacture, and Engineering Management, James Weir Building, University of Strathclyde, 75 Montrose Street, Glasgow, G1 1XJ, Scotland, UK

Keywords: mechatronics, supernormal audition, telepresence, teleoperation, head-tracking.

Abstract

Remote audio perception is a fundamental requirement for telepresence and teleoperation in applications that range from work in hostile environments to security and entertainment. The following paper presents the use of a mechatronic system to test the efficacy of audio for telepresence. It describes work to determine whether the use of supernormal inter-aural distance is a valid means of approaching an enhanced method of hearing for telepresence. The particular audio variable investigated is the azimuth angle of error and the construction of a dedicated mechatronic test rig is reported and the results obtained. The paper concludes by observing that the combination of the mechatronic system and supernormal audition does enhance the ability to localise sound sources and that further work in this area is justified.

* Corresponding author

E-mail address: c.harrison@strath.ac.uk (C.S. Harrison)

Tel: +44 (0)141 548 2258

Fax:+44 (0)141 552 0775

1. Introduction

The sense of hearing is essential for the experience of telepresence and it is extremely useful for detecting warning signals or alerting the system user to an event or movement outside the visual field. Thus for remotely operated surveillance systems the ability of the operator to detect the source of a sound will rapidly indicate where to direct visual attention. “Ears tell eyes where to look” [1]. Our own auditory system can do this using our normally separated ears, however supernormal separation of remote “ears” may improve the localisation of a sound. The design of an experimental mechatronic test structure to investigate this and the test results obtained are described here.

1.1. Background

Many research workers have developed a number of anthropometric “binocular heads” in order to investigate *visual* aspects of telepresence, [2,3], and more recently as integrated elements of more complex humanoid teleoperators incorporating a “three dimensional” microphone system [4]. However a common feature of tracked binocular sensor platforms is that they rarely have focussed attention on the audio capability (though an attempt has been made to emulate localisation by barn owls [5]), and the literature suggests that there is a relative lack of research work in audio for telepresence as it relates to the *remote* perception of human spatial hearing [6]. Further to this, an extension can be made into testing for auditory localisation and to investigate how this could contribute to knowledge by enhancing the sense of presence by varying the interaural distance. Work has been done in this specific area and termed “supernormal audio” [7,8,9], but there is still a relative paucity compared with the array of systems developed for visual telepresence.

The research challenge facing telepresence is that the sound should be transduced in what is essentially a live recording and simultaneously reproduced using an instantaneous method that will generate the most realistic sensation, as if the remote surrogate was actually in the place of the users' own head. In the case of the synthesised studies there is often knowledge of source sound position and this is frequently linked to a head tracker and convolved with the relevant Head Related Transfer Function (HRTF) to generate a virtual sound based on the apparent location of a particular sound [10].

Results obtained from the addition of a microphone pair to a pre-existing binocular platform have been published [11], however there were a number of possible improvements to this work. The design of this earlier structure was intended for video rather than audio and that aspects of the motion control needed improvement particularly in the area of feedback and noise reduction. Thus a more correct approach would be to design a structure with the specific objective of having low audible noise and this paper describes such an attempt.

1.2. Spatial Hearing

As part of the design process for evaluating the audio sensation it was necessary to consider some of the mechanisms that humans use to hear spatially [12,13]. This information informed the mechatronic equipment and experiment design.

Human hearing is characterised by a series of mechanisms which affect our ability to localise sound, including *interaural differences* (in time and intensity), *screening effects* and *distortion effects* of the pinnae and head, and the ability to *move the head* to sample repeatedly from the acoustic field. The first of these is that of the *interaural* differences. In

duplex theory, a century ago, Lord Rayleigh [14] described the phenomenon of interaural differences between the two ears based on differences in both *intensity and time.* A sound wave to one side of the vertical median plane strikes the near ear first and travels around the head [15]. Since the speed of sound in air is a constant this gives temporal information for a waveform provided that the wavelength is sufficiently long that there is no phase ambiguity. Once this starts to happen at a frequency approximately equal to 1500 Hz a different mechanism is needed [16,17].

The limitation provided by using the *interaural phase* method for localisation is that as the frequency of the sound increases, the phase solution becomes ambiguous. This is because a time difference between cycles that produces a half cycle or more of phase lead is indistinguishable from an opposite time difference that produces a phase lag. At frequencies above approximately 1500 Hz the acoustical path exceeds a full wavelength and there are possible solutions on both sides of the head for pure tones having equal intensity.

The additional method of localisation at higher frequencies uses the differences in amplitude as a localisation method, otherwise known as *interaural level* differences. The difference in intensity between left and right has been measured for the human head [16]. There are other mechanisms acting, one of which is the existence of the ability to sense the sound by moving the head in order to generate multiple samples.

Moving ears or motional theories [12,13] describe one of the ways of resolving front-back confusions. A common result of localisation experiments where the head is held still and only interaural factors are present, is that the geometric solution could be directly in front, or its analogue directly behind the sensor platform, whether it be human or artificial [8,16].

However rotating the head about any axis generates a set of changed interaural differences which uniquely specify the location of the sound as the intersecting locus of two or more possible “cones of confusion” [18]. Studies have also been done with anthropomorphic manikins possessing high quality binaural equipment [19,20,21,22]. However there has been much less work on tying this to a sensor platform that is capable of motion in both azimuth and elevation and particularly with a view to investigating the sense of presence that results.

The outer ear, *the pinna*, has a subtle yet noticeable effect on our ability to localise and attend to sounds. The effect of two pinnae on a remote hearing system has been widely reported [12,17]. Whilst interaural differences are preserved by bare microphones, the standard result of fixed position experiments report front-back confusions, and up-down confusions. Note that front-to-back confusion is different from back-to-front confusion, the most difficult sound to convincingly reproduce is one where the listener perceives the sound to be in front of his (or the remote) head. Commonly, sounds which are actually in front are perceived to be in the rear, whilst those in the rear are (usually) correctly placed in the rear. Another mechanism is that provided by the pinna which modifies the spectral content of sounds, and attenuates those from the rear which aids localisation. It should be noted that any sound localisation system should reproduce faithfully the original high frequency components of the sound, as the pinnae are thought to react strongly only above 4000-10000 Hz.

It is therefore clear that the pinnae form an important design aspect for the remote system and that it would be highly desirable to somehow include pinna factors in the design, either by incorporating an HRTF approach [23,24], or by physical modelling. It would also be

useful to attempt to mimic the screening effect of the human head since this has shadowing/dispersion effects on the ears.

2. Specification of the Supernormal Audio System

2.1. Overview

A number of the key mechanisms for human hearing have been described and these are now used to formulate requirements for a telepresence system design specification. Generally the sound detected by the remote listener should be compact, correctly externalised and localised in space such that the listener can be convinced that their own location is remote and not local. The key practical considerations are that a binaural system should incorporate the following features:

Binaural Sound. Use two good quality microphones located at the analogue position of the human ears. This will provide the time and level differences required to help locate sounds correctly and to facilitate the human "Cocktail party effect" [25] (the ability to follow one conversation in a situation where multiple conversations are occurring).

Motional Theory. Include the ability to remotely incorporate the motion of the user's head in order to eliminate front-back confusions, and to aid in localisation [26]. In the case of a binaural approach this means head tracking.

Low Noise. Have the ability to transduce sound at the remote end without introducing mechanical motor noise.

Pinnae effects. The effect of the pinnae has been previously described though this has been rarely implemented in the development of telepresence heads/testgear. Pinnae filtering can give important localisation and externalisation cues.

Reconfigurablility. The ability to vary interaural distances and orientations easily by having some sort of modular or adjustable framework will be useful when it comes to exploring how humans actually localise. It should be possible to design a system that investigates supernormal auditory localisation by using such a structure [27].

Anthropometric features. These will be useful since spatial distances are critical to how humans locate in a binaural context, though monaural localisation is possible [28]. At the limit there is some evidence in the literature that a truly isomorphic system is the correct way to experiment for the ultimate audio presence system [19].

Vision Capability. Often the perception of sound is aided by good quality vision perception and to a certain extent audio work has often been the poor relation. Ideally some capability to include stereo vision should be included in order to enhance the overall telepresence experience. It has been suggested [29] that the sense of presence can be enhanced by up to 20% by combining auditory and visual senses. The stereoscopic vision capability should be able to extend all around the user and not just in the frontal region.

Externalisation A common feature of headphone localisation is that sounds tend to be perceived inside the head. Ideally the sound should be manipulated in a way that the user perceives it to be externalised, often through the use of pinnae and dummy heads. An alternative is to use B-Format with head tracking [30] and slaved to a large projection screen.

Bone Conduction There is a certain amount of sound that is carried via the bones of the human body to the ear transducers and this means that the sound perceived is different from that which we hear on replay [31].

Since the objective is a personal, single user, telepresence system the use of speaker systems that are located at a distance from the user is ruled out. For example home cinema systems are not practical since they require the user to be looking straight ahead. It was therefore decided to use a head mounted display and binaural sound. This binaural approach attempts to recreate the soundfield at the entrance to the ear canal when listening to a sound source [12,13] in as faithful a way as possible by using a dummy head which may be either synthesised or real. One of the key features of the binaural method is that the sound is intended to be replayed over headphones which means that externalisation is relatively less effective than with the ambisonic approach. However it is generally easier to control the sound presentation and practical considerations mean that isolating the user from the external world is more straightforward.

Essentially, the two microphones are placed as they would be on the human head and screened from each other either by the use of a spherical head or a polymer disc in order to mimic the shadowing effects [32]. The presentation of the sound is then over conventional good quality headphones which may be attached to a head mounted display or worn separately. An alternative analysis can be performed by using HRTFs [33] and provides one way of spatialising sound and relating it to head-tracked positions. Usually a large number of HRTFs are used to give good coverage (often the azimuth) using generic HRTFs to synthesise a widely applicable solution [34,35], also earlier work indicated that a mechatronic approach to telepresence may give good results, see Figure 1. The approach reported in this paper has been to develop an adjustable framework capable of reconfiguration to accommodate not just binaural sound with a tracked motion platform, but also to investigate supernormal auditory localisation with the addition of vision. One

important feature was that the gearmotors were selected to ensure very low mechanical noise characteristics as the noise from the steppers was reported by users of the earlier system.

Figure 1 Anthropometric head with aural capability [11]. *[Figure 1 near here]*

From the elements of the specification the following system comprised "remote site" equipment in the anechoic chamber, and "home site" VR headset sound equipment in adjacent room.

A short video clip is attached showing the system as it was used in practice for clarity.

The test system consisted of the following apparatus:

2.2. Equipment List

General Reality CyberEye VR headset having 22.5 degree horizontal and 16.5 degree vertical field of view (181,000pixels)

Intersense I-300 Inertia-Cube head tracker (RS232) having 3 Degrees of freedom (Yaw, pitch, roll), dynamic accuracy 0.02 degrees, resolution 3 degrees

Purpose written Visual Basic 6 Data collect/Motion Control/position sense software

Flexlink modular structure machine frame

JRKerr ServoMotion PIC CMC motion control cards (RS485 bus + RS232/485 interface)

Harmonic Drive RH14-2702-E050AL "Pan" motor + harmonic 110: 1 gearhead

Max continuous torque,	18Nm
Gearing	110:1
Backlash	<2arcmin
Encoder	500 line with quadrature encoding

2 opposed Maxon Motor 110212 with GP26B planetary gearhead motors for tilt axis

Max continuous torque	2Nm

Gearing	231:1
Backlash	<120 arcmin
Encoder	500 line with quadrature encoding

2 off Sennheiser MKE102 miniature microphones

2 off Sennheiser K6 Microphone pre-amplifier

Amplifier: Behringer UB1204 (to raise level to that suitable for VR headset)

Artificial pinnae moulded at University of Strathclyde (National Centre for Prosthetics and Orthotics)

Two Hitachi KP-D8 Cameras 410,000 pixel CCD, resolution 460H x 350V

Ono Sokki CF210 Field FFT analyser

Bruel & Kjaer 4231, source, Bruel & Kjaer 2669 microphone, Bruel & Kjaer 5935L pre-amplifier, for use with above FFT analyser

2.3. Overall Performance

The final assembly of the test rig is shown in the annotated photograph Figure 2. A brief description of the software required is given in the following sections as is an attempt to quantify the operating parameters particularly as they relate to the audio features.

Figure 2 Supernormal Audio Test Rig for Telepresence *[Figure 2 near here]*

The overall performance of this system is as shown below:

	Range	**Speed**	**Acceleration**
Pan	+/- 170 degrees	44 degrees/sec	550 degrees/sec^2
Tilt	+/- 45 degrees	23 degrees/sec	290 degrees/sec^2

Note that there are suggestions [36] that “roll” is less important in telepresence platforms than pan or tilt, although fewer studies have specifically investigated this.

2.4. Experimental Procedure

A 15 second white noise sound burst [37] was presented to the subject from three different positions, two real from matching speakers at 1.5m distance and +36 and –144 degrees from the centre of the structure, and one synthesised. Each remote head width corresponded to 10 individual data points and a randomised order was used for each 10 point data collection. At the end of each data collection run the subjects were asked to remove the HMD to prevent them hearing what was happening within the anechoic chamber. The interaural distance was then varied according to the following 4 patterns:

Interaural Distance Variation

1st Run	170mm	1 x head width
2nd Run	340mm	2 x head width
3rd Run	510mm	3 x head width
4th Run	680mm	4 x head width
5th Run	170mm	1 x head width + with binocular stereoscopic vision

The subjects were asked to locate the sound in space by moving their head to where they perceived the sound to be coming from and then pressing a button. The time at which the sound was started and the instantaneous position of the remote head were recorded automatically by the data collect procedure. Limit switches on the remote head were used to zero centre the position of the remote apparatus each time in order to start from the same physical remote head location of the remote system at each test. Thus the user moved through a relative angle but there was no requirement for the user to start from an identical position each time, though in practice they started from a similar position. Each experimental run at a given head width consisted of two target values +36 degrees and

-144 degrees. The data collection started with a conventional interaural separation distance and gradually increased to 2 x head width, 3 x head width and 4 x head width, before finally reverting to conventional 1x head width with the binocular vision added. *This final check was so that it could be demonstrated that the subject was capable of using the equipment.*

2.5. Noise testing

As part of the assessment procedure the level of noise inherent in the system was measured. The method used was to attach the Bruel & Kjaer calibrated source to the ear canal position on the telepresence head (at the position of the human analogue) and feed this output to the Field FFT Analyser.

The results of the testing showed that background noise in the chamber was 25dB. This is therefore a very low noise environment, somewhat lower than a whisper (30dB), and much lower than would be encountered in a busy office environment which is typically 50dB (A weighted). The noise that does exist penetrates to some extent round the door, due to the cable routing, and through the walls of the chamber itself, whose physical location is adjacent to an ante-room bordering a busy street, although a sound proofed room would be preferable [38,39]. When the system was powered and the motors driven the overall level of noise was 44.7dB.

Also the noise was measured to see that the attenuating polymer disc representing the human head was acting in a way to be expected. This was measured by placing the calibrated Bruel & Kjaer test microphone on each side of the disc at the position of the human ears and measuring the sound detected using the Ono Sokko measuring gear. The

response indicated that there was an approximately 10dB overall reduction in the signal between each side based on the use of the polymer disc. See figures 3,4. This is particularly noticeable at the higher frequencies, which emulates the way that the head works as essentially a low pass filter.

Figure 3 Left "ear" microphone unscreened by polymer disc.

[Figure 3 near here]

Figure 4 Left "ear" microphone screened by polymer disc illustrating attenuation.

[Figure 4 near here]

3. Results

3.1. Overview

In general, listeners perceived the sound as being *external* to the head (another requirement of telepresence) rather than in the head which is the usual result of a conventional pair mixed stereo approach [40]. Thus, simply by the addition of artificial pinnae two of the design goals of a telepresence audio system could be met. The conversion of internal sounds reproduced within the head to *externalised* sounds and the *localisation* of the sound are greatly improved.

The head track, and final position data recorded via the users' button presses, represents a series of paths, one of which is shown in Figure 5. Note that the chart represents a measured "head waggle" based on the instantaneous time and measure both of the rotary encoder position and the head tracker attached to the user's head. The target value in this case is the broken line positioned at $+36^{o}$ on the chart. The master track is the value obtained from the head tracker and the slave track is the response of the slave system as taken from the encoder. Here it should be noted that there is a variable tracking delay as

indicated and that the system exhibits a delayed behaviour but catches up when the user stops moving. A typical value of this delay is approximately 0.5 seconds at the indicated point. This indicates the lag in the system based on local operation.

Figure 5 Head track for localising a sound at $+36^0$. *[Figure 5 near here]*

Also from the chart it will be seen that there is a clear, two cycle repeat in the "head waggle" process. During the first cycle (0-6.0s) there is a hunting and overshoot procedure and it appears that the subject is about to localise to 15 degrees. Then the user contra-rotates to –28 degrees, there is then a rapid reorientation back to an overshoot and finally the value of 16 degrees is chosen. Note that the azimuth error here is approximately 20 degrees since the target value is 36 degrees. This is within expected limits for a system [41, 12] including head movement which can be 20 degrees and sometimes more. The process by which the final value is obtained consists of hunt, overshoot, correction, dwell and decision. Note here that the complete cycle time for this process was 12.5 seconds which compares with the duration of the source sound of approximately 15 seconds. In all cases there was sufficient time during which to localise the sound. Also note that the "overshoot" is quite large compared to the final value that was selected by the user. These characteristic overshoot and other localisation features have not been widely emphasised in the literature and may reflect the order. Also from Figure 6 it will be seen that there is a characteristic "dwell" cycle after the sound commences while the user of the system attempts to consciously localise before moving, and that this is approximately 1 second.

Figure 6 Head track for localising a sound at -144^0. *[Figure 6 near here]*

Another sample track for the alternate location at -144^0 is shown in Figure 6. There are several notable features from this sample track, which is again an early test run for an

interaural distance of 170mm. The most immediate feature is that there is very little overshoot in this particular case -approximately 7^0. This pattern is repeated throughout the data collected and from observation most users tended to slowly rotate their trunk and head to locate the rear position, whereas the front location was much more rapid. Therefore the designer of a telepresence system to reflect human performance capabilities must have a physical platform that is capable of very high speeds in the front orientated "shoulder to shoulder" head rotation region.

However the tracking delay in the system is still evident and is illustrated by the difference between the master side attached to the user and the slave side measured from the position of the encoder. The tracking delay is a constant feature of every recorded localisation track that was generated by the system. Note that there is a similar dwell cycle at the start of the process before motion begins, typically 1 second, before localising to the final value, which in this case is -151 degrees. In this particular case the final azimuth tracking error is 7^0 since the actual location was -144^0.

Figure 7 (Tracking error for panning axis) *[Figure 7 near here]*

The tracking error for a typical location cycle (the same track as illustrated in Figure 6) which illustrates the value of the difference between the instantaneous position in time between the master track and the slave track. This demonstrates that there is indeed a latency in the system which can be clearly seen by the instantaneous following error of the remote platform which in this instance ranges up to about 12.5 degrees. This tracking error merely illustrates that there was indeed a portion of time during which the system caught up with the movement of the master side. Additionally there is a small offset steady state

error in the position of the tracking system which can be seen in the 0-1 second portion of the chart. A typical value of this error is 0.5 degrees or less, once a steady state position has been reached. The idea behind using the button to indicate position was that a steady state position should be reached and the above sample tracks show that this was the case. Worst case latency at the maximum error point was approximately 500ms, with a typical value being 150ms or less.

3.2. Supernormal auditory localisation and head width

The variations in head width that were performed and the ability to localise are presented for each of the 8 subjects below. The results for each track (angle of error) were calculated and then averaged for each run at the various settings of head diameter. Thus each experimental run at a given head width consists of two target values +36 degrees and -144 degrees. Therefore a cone of confusion error would be said to exist at 180-36 degrees (namely +144 degrees) and in the case of –144 degrees at –36 degrees. Note the critical importance of the sign conventions here.

Actual location	+36 degrees	-144 degrees
Cone of confusion location	+144 degrees	-36 degrees

Here the convention in the literature is to fold cone of confusion errors into the correct location and report on the number of reversals separately. In this case the results are easy to present since there were no cone of confusion errors detected based on the +36 degree value. Here the authors take a cone of confusion error to be based on the mirror image

value but bounded by the average error value of the overall experimental run. So, for example, since the average error is 11 degrees unless a confusion error is within 11 degrees of +144 it is not a confusion error; it is taken to be simply a wrong value. In practice no occurrences of these errors in the data were found, this was due to the freedom of the subjects to move their heads, null out ITDs (interaural time differences), and hence eliminate front/back confusions in the movement regime.

There are a number of factors that mitigate against the occurrence of cone of confusion errors, namely the very long broadband signal, the ability to move the head and the possession of pinnae in order to interact strongly with the broadband signal and the high frequency content [42]. Broadband sound can occur in a real telepresence condition, such as a hissing pipe occurring outside the field of view, in a nuclear application for example.

In summary, excluding the anomalous results from test subject SFC, front to back confusions were completely eliminated.

Also there is a general pattern of the front right position (+36 degrees) being located considerably more accurately in terms of angle of error than the position at the rear (-144 degrees). This reflects test results in the literature which give greater values of angle of error in the rear hemifield than the front [43], particularly for a broadband long duration source and where head movement is allowed.

The ability to use increased head width to localise is one that has been reported as improving the resolution but distorting the mean response 44,45].

3.3. Comments on Results for Telepresence Subjects

Some of the users were chosen from the telepresence research group but were not told the precise nature of the tests, and in particular none were told the location of the sound sources. In particular SJEC and SGM had considerable experience of using VR headsets and trackers, being members of the telepresence research group. Their individual results illustrate angles of error of about half of the average or in the case of SGM up to 20% of the average error. This suggests that as users become more familiar with the VR systems then the angle of error can be reduced quite markedly. In the case of subject SFC it was noticeable that despite being instructed in detail on how to localise the sounds this subject simply rotated to (approximately) the 90^0 position and tilted her head. This behaviour was so odd that it was recorded, however the data was excluded from the calculations of average angle of error - this subject was also reluctant to have her eyes covered by the VR headset.

It is also clear from the data that all the subjects had more difficulty locating the sounds placed in the rear quarter, that is at -144^0 from straight ahead. It can be seen from table I that 5 of the 6 averages, including the vision test (apart from the test at 340mm), reflect a slightly greater angle of error for the -144^0 position than the $+36^0$ position. In order to achieve this location each subject tended to firstly rotate to approximately the 90^0 position and then the remaining movement was via the trunk. Since in effect the users were absolutely free to rotate to any position and in any combination of head/ shoulder/ chair movements this suggests that it is the range and pattern of movement itself that determines how easily the sound can be located. In effect each subject was free to move their head to a new effective “head centre” position and then locate the sound. The movement traces for

the rear sounds (see Figure 6) show much less rapid movement to the rear location and result in a less precise result on average. The implication is that localizing in the “rear” for telepresence would appear to be less efficient.

If the results of the supernormality are considered, then for the forward facing source at +36 degrees there is an improvement in the average angle of error from 13.2 degrees down to 7.2 degrees or 46% reduction in the angle of error (Table I). The trend of the reduction also appears to be quite clear and the check of using vision in assistance gives some idea of the minimum possible result that could be expected at about 2.6 degrees. The average reduction is about 2 degrees per head diameter (170mm) for the particular front location. The 80mm width of the loudspeaker at the 1.5m distance from the head is also approximately 3.0 degrees, so even traversing from one side to the other of the seen object is approximately the angle of error.

The advantage of using supernormality for the rear location is less definite since the average angle of error reduces from 14.0 degrees and then flattens out at approximately 11 degrees despite the head diameter increasing from 340 to 680mm. So for the sound localising in the rear there appears to be less benefit in having an increased interaural separation. However in all the tests the range of individual differences in the ability to localise with increasing separation was stark. Those familiar with the headset from other tests (SJEC, SGM) showed improvement in the ability to localise at greater interaural separation distances, particularly SGM. The other subjects had almost no prior experience with using VR headsets and one in particular (SFC) found the equipment difficult to use.

4. Conclusions

This paper has described the testing of a mechatronic research apparatus incorporating supernormal interaural distance for telepresence in order to test if there is an increased capability to localise sound based on increased head diameter. The inclusion of both data from the head tracking master side and the inherent position control of the motor slave structure has enabled detailed data on the localisation process to be gathered from this apparatus. Other studies into telepresence have concentrated on the visual aspects of telepresence and relatively little work has been performed on the audio side, or even the adoption of a multimodal element, as has been done here.

The aim has been to quantify the results of audio for telepresence using a supernormal approach to remote audio, via angle of error, and to do that based on experimental data. The results of this analysis have been presented in summary form in figure 8 and numerically in table I. These illustrate that for this source and where head movement (pan and tilt) is allowed, subjects are able to localise broadband sources in the forward direction with an accuracy that increases almost linearly with head diameter up to 680mm, and at an average rate of about 2 degrees per head diameter. The evidence for ability to localise the rear sound is less obvious and the average angle of error remains constant at approximately 11 degrees for 2 head diameters and more, however there are large individual differences and it may be that with concentrated training the subjects would be able to localise up to the standard of SGM, an experienced telepresence user. However all subjects were able to localise the sounds to the correct position within 11 degrees and front to back confusions were eliminated, and even relatively inexperienced users were able to track the sound to within about 14 degrees. For the developer of a telepresence system therefore, the audio

location sense is essential and can help to focus the attention on “rear based” sounds as well as accurately pinpoint noises where vision is obscured.

One clear cut outcome from the data is that once the location of the actual sound source is tied to the vision sense then the angle of error reduces for all the users. Additionally for the complete set of experiments cone of confusion errors were eliminated. The importance of attempting to screen the remote “ears” with an attenuator such as a polymer disc was also demonstrated, since this mimics the human head action of a low pass filter. Finally the addition of binaural hearing gives a clear ability to localise a sound at 14 degrees or better for all users and can be reduced to as low as 7 degrees on average, or with experienced listeners to less than 2 degrees in one case. There is a demonstrated reduction in the average angle of error with increased interaural distance of approximately 2 degrees per head diameter up to 680mm. Designers and users of future telepresence systems can therefore conclude that the inclusion of supernormality, can give a measurable improvement in localisation ability, particularly where vision is obscured.

References

[1] Wenzel, E.M. (**1992**). “Localisation in virtual acoustic displays”. Presence, **1**, No. 1. 80-107.

[2] Mair, G.M., Fryer R., Heng J. (**1994**). “The design of mechatronic anthropometric sensor platforms”. British International Mechatronics Conference, 291-296.

[3] Sharkey P.M., Murray D.W., McLaughlan P.F., Brooker J.P., (**1998**). “Hardware development of the Yorick series of active vision systems”. Microprocessors and Microsystems J. **21**, 363-375.

[4] Tachi, S., Komoriya, K., Sawada, K., Nishiyama, T., Itoku, T., Kobayashi, M., Inoue, K. (**2003**) “Telexistence cockpit for humanoid robot control”. Advanced robotics, **17**, 199-217.

[5] Natale, L., Metta, G., Sandini, G. (**2001**) "Visuo-acoustic cues integration in an artificial developing agent, workshop in developmental cognition". DECO, 1-6.

[6] Begault, D.R. (**1999**) "Auditory and non-auditory factors that potentially influence virtual acoustic imagery". AES 16th Int. Conf. on Spatial Sound Reproduction.

[7] Durlach, N.I., Pang X.D. (**1986**) "Interaural Magnification". J. Acoust. Soc. Am. **80**, 1849-1850.

[8] Durlach, N. (**1991**) "Auditory localisation in teleoperator and virtual environment systems: ideas, issues, and problems". Perception, **20**, 543-554.

[9] Durlach, N.I., Shinn-Cunningham B.G., Held, R.M. (**1993**) "Supernormal auditory localisation". Presence, **2**, No. 2, 89-103.

[10] Minaar, P, Olesen, SK, Christensen F, Moller H.(**2001**) "The importance of head movements for binaural room synthesis". Proc. 2001 International Conference on Auditory Display, 21-25.

[11] Harrison, C.S., Mair, G.M. (**1999**). "Mechatronics applied to auditory localisation for telepresence". Mechatronics Journal, **9**, 803 – 816.

[12] Blauert, J. (**1996**) "Spatial hearing: The psychophysics of human sound localisation", Cambridge MA:MIT press.

[13] Begault, D.R. (**1994**) "3D-sound for virtual reality and multimedia". AP Professional.

[14] Rayleigh L., (**1907**) "On our perception of sound direction". Philosophical magazine, **13**, 214-232

[15] Shaw, E.A.G. (**1974**). "The external ear". In Handbook of sensory physiology, Keidel WD, Neff WD.Springer, Berlin, 474.

[16] Mills, A.W. (**1972**) "Auditory localisation, foundations of modern auditory theory". Ed:Tobias J.V., Academic Press.

[17] Batteau, D.W. (**1962**) "Localisation of sound, Characteristics of human localisation of sound". US Naval Ordinance Test Station Report, TP 3109, part 1.

[18] Wenzel, E.M., Arruda, M., Kistler, D.J., Wightman, F.L. (**1993**). "Localisation using non-individualised head related transfer functions". J. Acoust. Soc. Am. **94**, 111-123.

[19] Noro, K., Kawai, T., Takao, H. (**1996**) "The development of a dummy head for 3-D audiovisual recording for transmitting telepresence". Ergonomics, **39**, No. 11, 1381-1389.

[20] Burkhard, M.D. and Sachs, R.M. (**1975**) "Anthropometric manikin for acoustic research". J. Acoust. Soc. Am. **58,** 214-222.

[21] Han, H.L. (**1994**). "Measuring a dummy head in search of pinna cues". J. of the Audio Engineering Society, **42**, 15-37.

[22] Aoki S., Cohen M., Koizumi N. (**1994**). "Design and control of shared conferencing environments for audio telecommunication using individually measured HRTFs.", Presence: .**3**, No.1, 60-69.

[23] Begault D.R. (**1992**) "Binaural auralization and perceptual veridicality". 93rd Audio Engineering Soc. Convention, 1-14.

[24] Begault, DR and Wenzel, E.M. (**1993**) "Headphone localisation of speech". Human Factors **35**(2), 361-376.

[25] Jeffress, L.A. **(1972)**. "Binaural signal detection: Vector theory", in Foundations of modern auditory theory. Ed. Tobias, J.V. Academic Press.

[26] Perret, S., Noble, W. (**1997**). "The effect of head rotations on vertical sound plane localisation". J. Acoust. Soc. Am. **102**, 2325-2332.

[27] Rabinowitz, W.M., Maxwell J., Shao Y., Wei M. (**1993**). "Sound localisation cues for a magnified head: Implications from sound diffraction about a rigid sphere". Presence, .**2**, No.2, 125-129.

[28] Wightman, FL, Kistler, D., (**1997**). "Monaural sound localisation revisited". J. Acoust. Soc. Am. **102**, 1050-1063.

[29] Barfield W., Hendrix C., Bjorneseth O., Kaczmarek, K.A., Lotens W. (**1996**). "Comparison of human sensory capabilities with technical specifications of virtual environment equipment". Presence, **4**, No.4, 329-356.

[30] Hollier, M.P., Rimell, A.N., Burraston, D. (**1997**). "Spatial audio for telepresence". BT Technology Journal, **15**, No.4, 33-41.

[31] Tonndorf., J, (**1972**) ed Tobias JV., "Bone conduction" in Foundations of modern auditory theory, **2**, Academic Press, 195-237.

[32] Alkin G., (**1991**). "Sound recording and reproduction". 2nd Edition, Focal Press.

[33] Abouchacra K.S., Breitenbach J. (**2001**). "Binaural helmet: Improving speech recognition in noise with spatialised sound", Human Factors, **43**, No.4, 584-594

[34] Langedijk, E.H.A., Bronkhorst, A.W. (**2000**). "Fidelity of three-dimensional sound reproduction using a virtual auditory display", J. Acoust. Soc. Am. **107,** 528-537.

[35] Bronkhorst, A.W. (**1995**) "Localisation of real and virtual sound sources", J. Acoust. Soc. Am. **98,** 2542-2553.

[36] Adelstein B.D., Ellis S.R. (**2000**). "Rotation and direction judgement from visual images head slaved in two and three degrees of freedom", IEEE Trans. on Systems Man and Cybernetics –Part A:Systems and Humans, **30**, No.2, 165-173.

[37] Brungart, DS, Durlach NI, Rabinowitz, (**1999**) "Auditory localisation of nearby sources. II localisation of a broadband source". J. Acoust. Soc. Am. **106,** 1956-1968.

[38] Jouppi, N.P. (**2002**). "First steps towards mutually immersive telepresence". CSCW 02, Nov 16-20, 354-363.

[39] Jouppi, N.P., Pan, M.J. (**2002**). "Mutually-immersive telepresence". 113th Audio Engineering Soc. Convention, Los Angeles, California, 1-6.

[40] Burns, R. (**1999**) "Blumlein and the birth of stereo". IEE review, 269-273.

[41] Wenzel, E.M., Arruda, M., Kistler, D.J., Wightman, F.L. (**1993**). "Localisation using non-individualised head related transfer functions". J. Acoust. Soc. Am. **94**, 111-123.

[42] Wightman, F,L., Kistler, D. (**1999**). "Resolution of front-back ambiguity in spatial hearing by listener and source movement". J. Acoust. Soc. Am. **105**, 2841-2853.

[43] Middlebrooks, J.C. (**1992**). "Narrow band sound localisation related to external ear acoustics". J. Acoust. Soc. of Am., **92**, 2607-2624.

[44] Shinn-Cunningham, B.G., Durlach, N.I., Held, RM, (**1998a**). "Adapting to supernormal auditory localisation cues, I Bias and Resolution". J. Acoust. Soc. Am. **103**, 3656-3666.

[45] Shinn-Cunningham, B.G., Durlach, N.I., Held, RM, (**1998b**) "Adapting to supernormal auditory localisation cues. II Constraints on adaptation of mean response". J. Acoust. Soc. Am. **103**, 3667-3676.

List of Figures

Table I

This shows the Target (actual) angle and the angle of error for each of the test subjects at various interaural widths. The last column shows the error when audition was supplemented with vision.

Width(mm)	**170**		**340**		**510**		**680**		**170 + vision**	
Target	**+36**	**-144**	**+36**	**-144**	**+36**	**-144**	**+36**	**-144**	**+36**	**-144**
SJOC	12.3	9.1	6.9	14.6	3.7	5.4	5.9	12.5	2.6	4.3
SXTY	8.6	12.5	5.9	14.1	3.1	9.2	1.5	10.5	1.1	2.9
SJEC	8.7	6.6	7.5	4.0	7.2	3.2	6.2	6.8	1.9	1.4
SAM	17.6	17.2	32.3	6.1	22.7	34.3	16.1	13.9	8.7	3.3
SUL	29.5	34.8	20.6	10.9	15.5	5.0	16.3	6.1	2.0	6.1
SAL	9.4	7.8	3.9	16.0	11.6	15.6	2.0	23.6	0.8	3.4
SGM	6.2	10.1	5.3	7.4	2.7	3.5	1.7	1.9	1.3	2.0
SFC	*9.1*	*63.2*	*8.2*	*62.9*	*7.4*	*77.3*	*4.6*	*59.5*	*3.9*	*63.3*
Av(ex SFC)	13.2	14.0	11.8	10.4	9.5	10.9	7.1	10.8	2.6	3.3

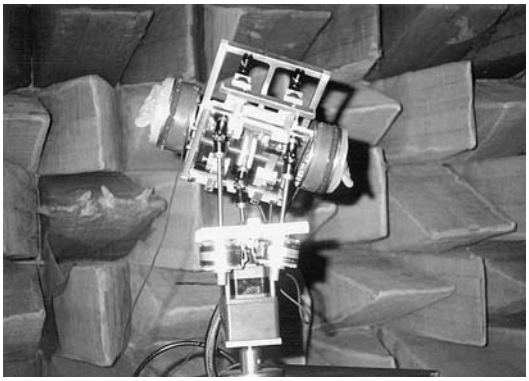

Figure 1 Anthropometric Head with Aural Capability [11]

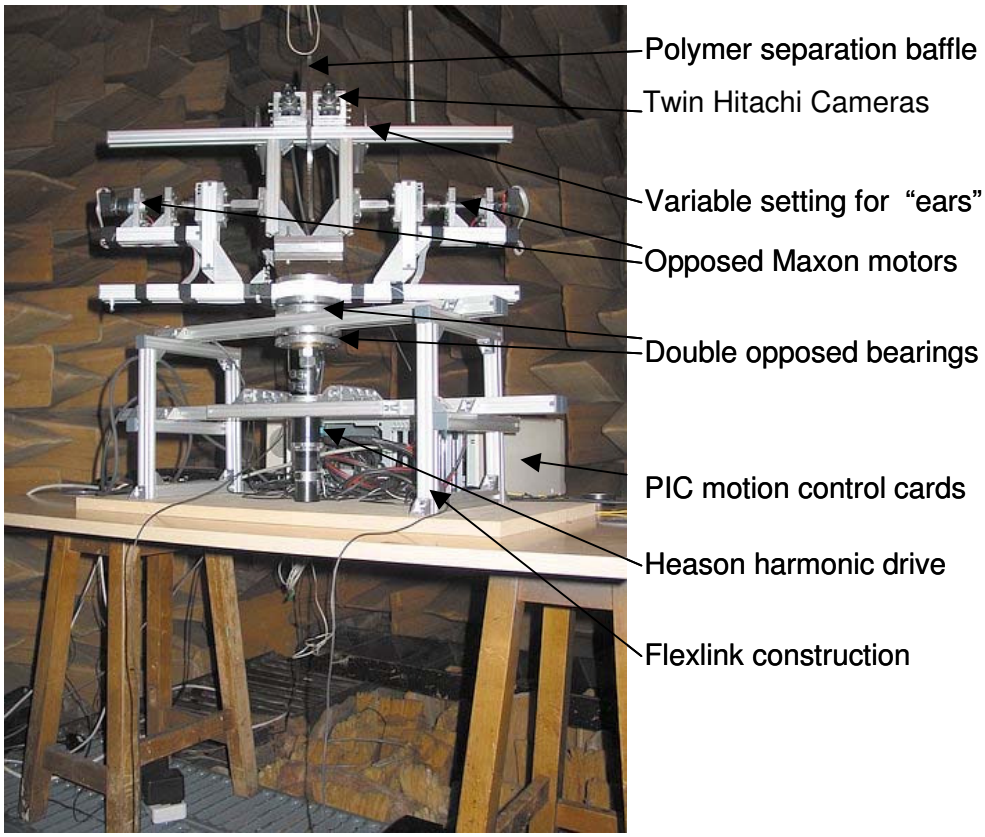

Figure 2 Supernormal Audio Test Rig for Telepresence

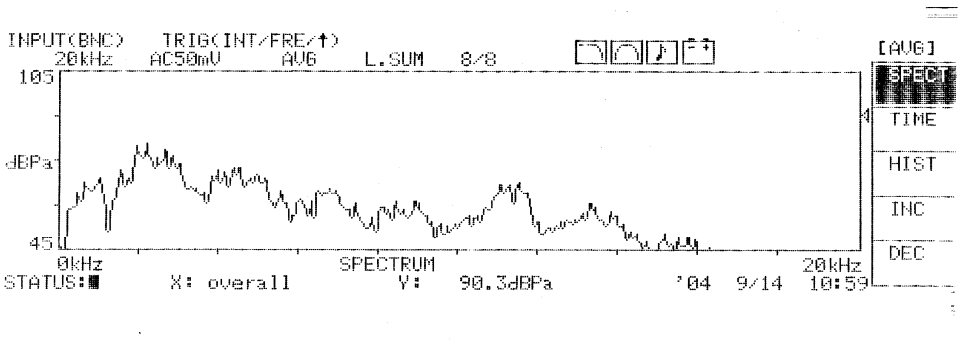

Figure 3 Left “ear” microphone unscreened by polymer disc

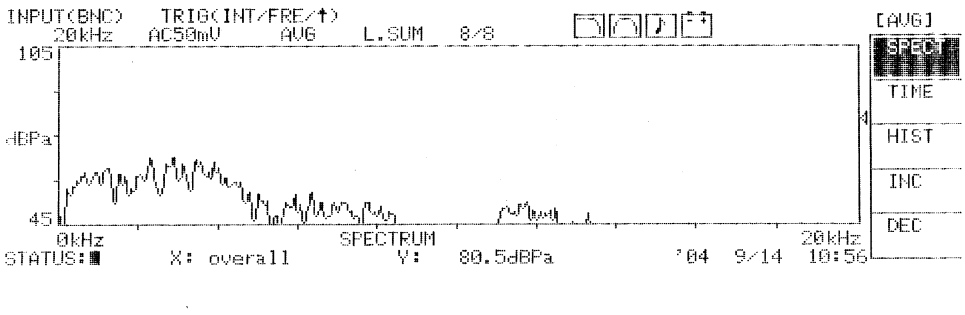

Figure 4 Left “ear” microphone screened by polymer disc illustrating attenuation

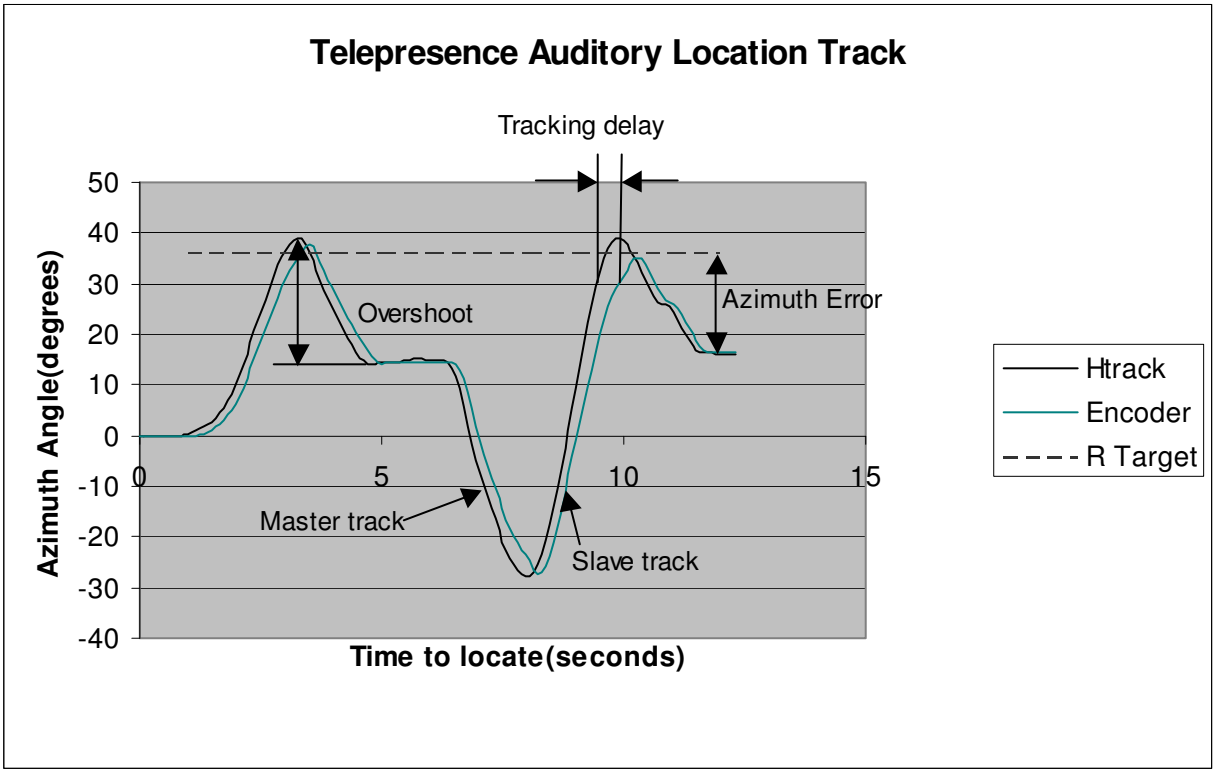

Figure 5 Head track for localising a sound at $+36^0$

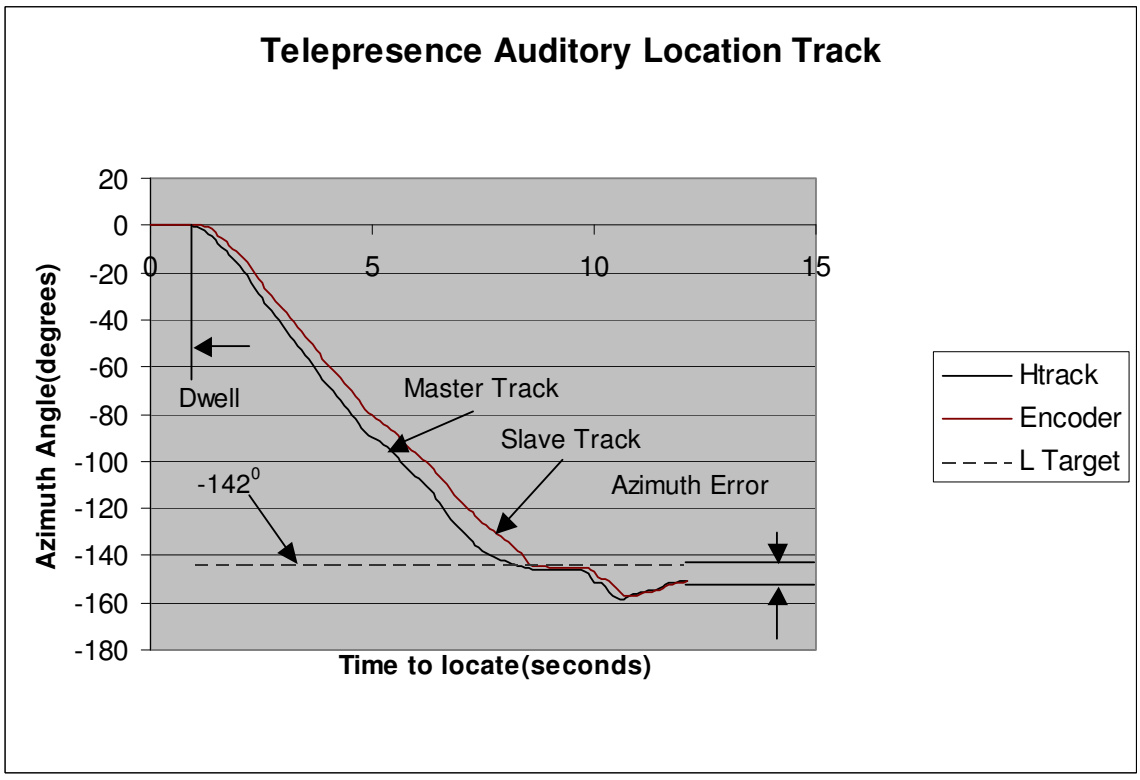

Figure 6 Head track for localising a sound at -144^0

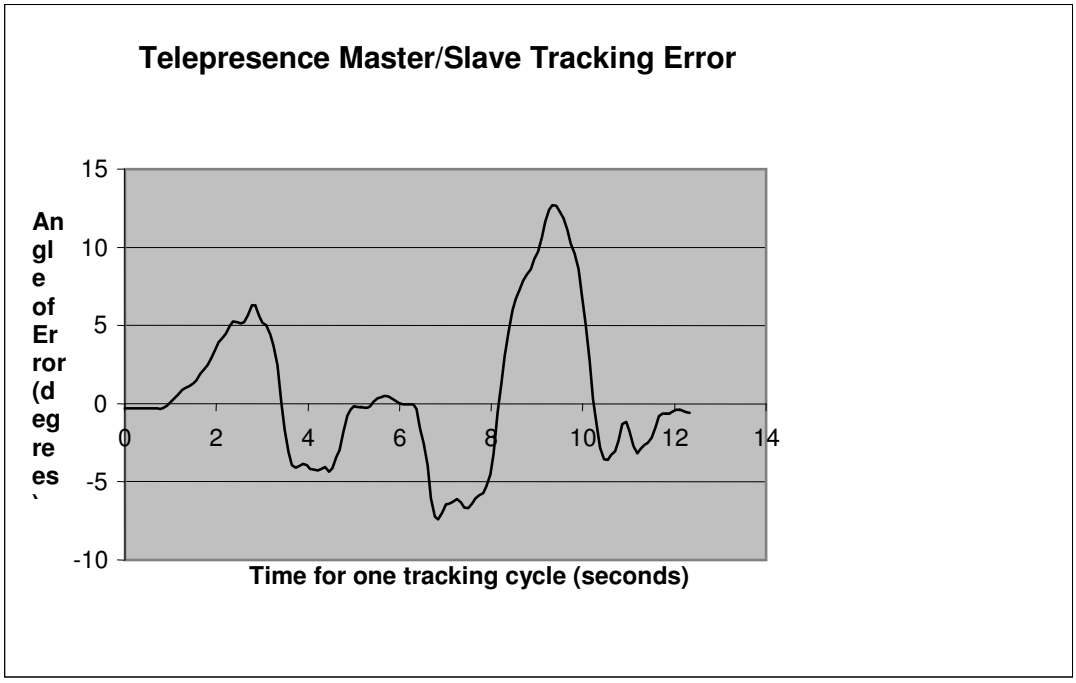

Figure 7 Tracking error for panning axis

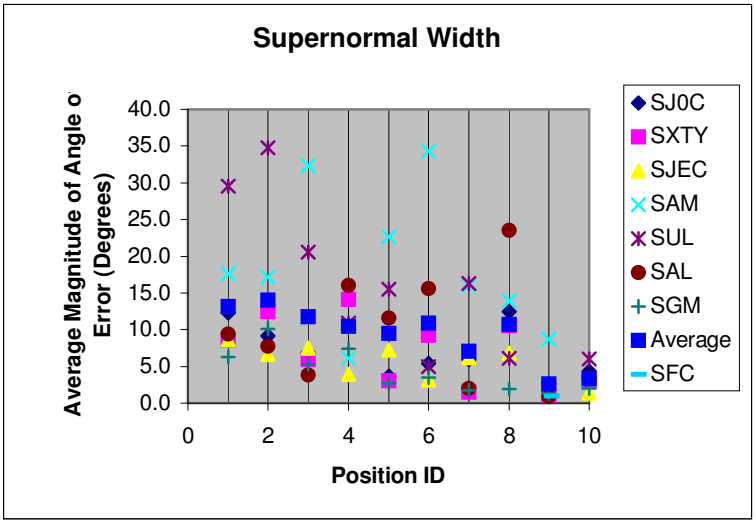

Figure 8 Angle of Error vs position all subjects