SPATIAL AUDIO TELECONFERENCING - WHICH WAY IS BETTER?

Michael J. Evans, Anthony I. Tew and James A.S. Angus
Department of Electronics
University of York
YORK YO1 5DD, United Kingdom
+44 1904 432361
mje100@ohm.york.ac.uk

ABSTRACT

This paper examines the two basic philosophies of spatial audio reproduction, with reference to their application to teleconferencing services. *Sound Field Simulation*, as exemplified by multiple loudspeakers techniques such as Ambisonics, encodes information about a remote or virtual sound field, and allows the reproduction of that field across a listening space. Alternatively, an application might employ *Perceptual Synthesis*, in which measured or simulated sound localisation cues (e.g. Head-Related Transfer Function (HRTF) data) are imposed on the signals reproduced over headphones or a suitably set-up pair of loudspeakers. The relative merits and drawbacks of each approach are discussed in terms of cost, implementation logistics, flexibility, specification and, critically, perceived performance.

INTRODUCTION

In the development of modern telecommunications services, displays and user interfaces there exists the general paradigm of *Telepresence*; the idea that interaction between parties at remote locations can be made more effective and efficient if it makes use of as much perceptual information as possible. Under such circumstances, telecommunications will more closely emulate real interaction between people in the physical world. This philosophy supports the development of large visual displays with increased colour and resolution, 3-D displays, interactive support for head movement and improved audio. In auditory display, *SPATIAL AUDIO* is also considered to be a desirable enhancement to telecommunications interfaces, facilitating the reproduction of the voices of other parties in such a way so they are perceived as emanating from particular locations in space - generally the positions of associated visual stimuli. When communicating with one or more people in person our perception of the scene, both auditory and visual, is inherently spatial and, therefore, spatial auditory displays of this type should have a higher level of telepresence than non-spatial reproduction through their presentation of information in a more natural manner; a style of presentation we each have a lifetime's experience in understanding. Such exploitation of our sound localisation ability also has the potential to yield benefits in discriminating a voice from others, or from vocal or other noise present on the channel.

However, the spatial audio community seems entrenched in two camps, each favouring a different approach to the spatialisation of sound, with different implications upon implementation and performance in teleconference applications. We can consider the first approach to be *Sound Field Simulation*, in which an array of loudspeakers (in practice four or more) are used to reproduce a sound field across a listening space. The reproduced sound field can be sourced by special microphone configurations at a real, but remote, location or might be a virtual sound field, artificially encoded from individual sources. Whatever the origin of the sound field being reconstructed, ideally it should be presented across the listening space in such a way so that listeners within the space are exposed to auditory information identical to that which would be present in the real environment. The listeners should, therefore, perceive the spatial qualities of the sound field accordingly. Ambisonics is a popular and widespread implementation of the sound field simulation technique [1].

The second general method of creating spatial audio is *Perceptual Synthesis*, in which we attempt to ensure that the perceptual cues used by the human auditory system to localise sound as emanating from particular positions are present in the signals reaching the listener's ears. Physical recordings containing such localisation information can be made by means of a microphones placed in the ears of a binaural dummy head or a real listener [2], effectively capturing the sound, including localisation cues, present. Alternatively, artificial binaural recordings can be made by modelling the directionally-dependent features present in the sound reaching a listener's ears, generally in the from of *Head-Related Transfer Functions* (HRTFs), and imposing them on the sound signals that we wish to spatialise [3]. Binaural recordings are inherently two-channel and are naturally suited to reproduction over headphones, although the technique of cancelling left-right crosstalk can be used to enable a pair of loudspeakers to reproduce binaural stimuli [4].

Sound field simulation and perceptual synthesis are two quite different approaches to tackling the same task, reproducing spatial audio. This paper examines the relative merits and drawbacks of implementing spatial audio into a teleconference environment by means of the two techniques (as exemplified by Ambisonics and HRTF-based spatialisation). In teleconferencing we must weigh the improved telepresence that may be achieved by the use of a spatial audio display with financial and logistical concerns and the flexibility of the systems. Also, and critically, we must ensure that any increase in telepresence is not achieved at the expense of a reduction in the effectiveness with which the words themselves are understood by the listener since, in speech telecommunications, the intelligible, unambiguous and comfortable transmission of the spoken word is of primary importance.

TELECONFERENCE APPLICATIONS

Figure 1 shows an artist's impression of a future teleconference system with an extremely high level of telepresence [5]. This application consists of completely unobtrusive cameras and microphones and a 3-D visual display which does not require glasses. The application is also multipoint and multi-user, with each participant experiencing the sense of shared workspace. Reproducing spatial audio in such a way as it supports the natural interaction of such a service rather than compromising it will be essential. Teleconference facilities as advanced as these are still some years beyond the scope of current technology. However, there is a rapidly increasing market for the current generation of commercially-available teleconference systems, produced by companies such as PictureTel in the USA, and BT in Britain, which have exploited the bandwidth provided by the ISDN. The digital network has also allowed the inclusion of ancillary tools to increase telepresence and effectiveness, such as collaborative drawing utilities and data sharing.

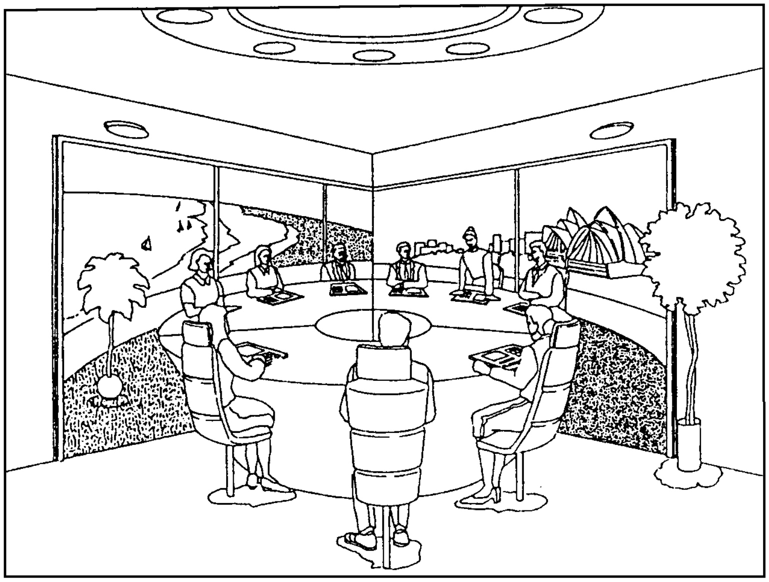

Figure 1: A future teleconference system proposed by NTT, reproduced from [5]

Most ISDN teleconferencing systems can be classified into the following general categories:

- Permanent Videoconference Rooms
- "Rollabout" Teleconference Systems
- PC/Desktop Videotelephones

Although forming the common view of teleconferencing systems, rooms permanently set aside for videoconference use are comparatively absent from the marketplace. Such facilities will generally involve large video-projection screens and allow the inclusion of a large number of participants. Clearly the logistics and costs involved in such large scale applications restrict the attractiveness of this option, but benefits are yielded in the form of high quality, high telepresence presentation of visual images. Permanent videoconference facilities are almost exclusively a niche, custom-built market.

"Rollabout" or group videoconferencing systems provide more practical and flexible access to teleconference facilities whist maintaining many of the benefits of the larger-scale systems, in terms of multiple users and moderately large screens. Products in this category generally house the bulk of the hardware, including monitor, video camera, microphones and loudspeakers, on a movable, and thus to a certain extent portable, trolley. BT's "Rollabout" product is the VS2. PictureTel produce three very similar systems of which the highest specification model, the Concorde 4500, features a 27 inch monitor and a camera which automatically tracks the position of the person speaking, by means of a small beamforming microphone array. Systems of this kind are rapidly becoming the standard product for situations in which groups of parties wish to participate. They have the flexibility and sufficient portability to be moved between different sized rooms, and support is often included for peripherals such as document cameras or digital overhead projection scanners in order to share documents.

Videotelephones - small units with visual and audio reproduction generally intended for single users - have not found as active a market as might have been expected by considering what seems to be a logical progression from conventional telephony. ISDN videotelephones are available, such as BT's VC9000, also known as Presence. However, the fact is that, in the workplace at least, a large proportion of potential videotelephone users already have video and audio playback hardware on their desk, in the form of a PC. Thus, a natural transition has been to incorporate the functionality of the videotelephone into the computer. Products such as PictureTel's Live100 or the DVS100 from BT, consist of a PC expansion card which connects to the ISDN, and associated software, generally running under a windows-type environment. A microphone and small monitor videocamera complete the system. The visual image of the other party in the teleconference is shown in a window on the PC desktop and the sound reproduced either over headphones or loudspeakers. Shared applications and additional telepresence tools can run in other windows. Such desktop systems clearly have a lower level of telepresence, primarily because of the small visual image. However, applications of this kind, as well as forming a highly cost-effective teleconferencing solution, are also ideal for day-to-day use, generally by individuals at their desks.

SPATIAL AUDIO - COST, LOGISTICS AND FLEXIBILITY

Channels and Loudspeakers

A central influence upon the cost and logistics of physically realising a spatial audio system is the number of channels used in transmission and in reproduction. To a very great extent, the number of channels is the critical factor in determining the resources that the implementation will require. Ambisonic material encoded into B-format requires four channels of audio for full 3-D spatialisation. If elevation of sources above or below the horizontal plane is not important then this is reduced to three. Sufficient transmission bandwidth must be available to support these multiple channels, comparing unfavourably with binaural systems which consist of, by definition, just two channels.

Similarly, the reproduction of decoded ambisonic stimuli in a listening room requires a minimum of four or six loudspeakers for horizontal plane reproduction, and at least eight if elevated sources are to be spatialised, along with the appropriate number of amplifiers to drive them [1]. Binaural spatialisation requires a pair of headphones or, if crosstalk-cancellation is employed, a single pair of loudspeakers and associated amplifiers.

Additional Hardware and Processing

Ambisonic presentation of spatial audio also requires a decoder which extracts the sound field information stored in the B-format representation and distributes it between the loudspeakers in the array according to their spatial distribution. Ambisonic decoders are generally straightforward hardware devices but must be set-up for the specific listening environment, in terms of number and distribution of loudspeakers. The need for a decoder and the necessity to calibrate it for a particular listening set-up imposes particular cost and logistical overheads upon any teleconference system which makes use of ambisonic reproduction. HRTF spatialisation requires no hardware decoding. However, for presentation over a pair of loudspeakers the crosstalk cancellation process requires the binaural sound to be processed using a combination of the HRTFs associated with the particular relative positions of listener and loudspeakers in the listening room. If this relative position alters, usually because of a change in position of the listener's head, then to maintain spatialisation this processing may have to take account of the changed HRTFs between loudspeakers and ears. This information can be provided by a device which tracks any head movements and indicates the new relative position of the listener's ears, allowing appropriate HRTFs to be substituted [6].

A variety of wireless head tracking devices exist, generally consisting of a fixed source and a sensor attached to the listener, or sometimes vice versa. Tracking sensing can be electromagnetic (Polhemus Fastrak), ultrasonic (Logitiech) or optical (Origin Instruments' infrared Dynasight) and are generally very unobtrusive and non-disruptive to telecommunications. However, the need for such hardware does impose an extra cost on such a perceptual synthesis type of teleconference system. Also, in general, if listener and loudspeakers can take up any position relative to one another then a set of HRTF measurements with notionally infinite resolution is required. This problem can be solved by means of a continuous, functional model of HRTFs; essentially a mathematical formula which, when supplied with a particular azimuth and elevation, can provide an HRTF corresponding to that direction [7]. However, the chain of events formed by detecting movement with the head tracker, obtaining the new HRTFs, and resubstituting these responses into the spatialisation system, comprises a potentially large amount of processing. In such systems sufficient computational resource must be available so that the time taken for this processing to be carried out and the system to respond to a head movement is imperceptible by the listener.

Head Movements, Multiple Listeners and Individual Differences

In principle, a sound field simulation technique such as ambisonics requires no such special purpose hardware in order to tolerate head movements. By definition, the multiple loudspeakers used in ambisonics are used to reproduce sound information across the listening space as a whole, not just at the specific positions of the listeners ears. Of course, in practice complete accuracy in reproducing any general sound field would require a potentially infinite number of channels. However, the sound field approximation afforded by the use of B-format creates a graceful degradation in recreated sound field accuracy away from a ‘sweet spot’, which should permit spatialisation which can tolerate a high degree of listener movement without any need for head tracking. Similarly, ambisonic presentation has the added flexibility that presentation is not merely restricted to a single listener. The reproduction of the sound field across the listening space as a whole means that multiple listeners can be placed within the listening space [8]. Perceptual synthesis with a crosstalk cancelled pair of loudspeakers possess no such flexibility since the whole nature of such systems are predicated on the use of two transducers (the loudspeaker pair) actuating two other transducers (the ears of a single listener). Additional listeners would require the scaling up of the technique to use more than two loudspeakers, and the ensuing increase in complexity of the crosstalk cancellation [9].

Provided that ambisonic sound field simulation is sufficiently accurate the flexibility of such spatialisation also extends to the ease with which different listeners can use the system. HRTFs are highly variable from person to person [10]. With ambisonic reproduction any listener will effectively be perceiving the presented sound field by means of their own, personal HRTF information. However, an HRTF-based system will either have to have access to the HRTFs of any listener making use of it or, more practically, use a set of HRTFs which are an acceptable match to those of the bulk of listeners. Such ‘non-individualised’ HRTF sets can be sourced from dummy heads which make use of average, or some other optimum, anthropomorphic data. Alternatively, a system might make use of the HRTFs of a particular listener who has been found to be a particularly ‘good localiser’ of sound [11].

SPATIAL AUDIO - PERFORMANCE AND PERFORMANCE SPECIFICATION

Spatialisation Performance

Ultimately, the primary consideration in comparing these two possible approaches to spatial audio for teleconferencing is that of their relative performance; ensuring that each creates the desired spatial perceptions and does not compromise the quality with which the words being conveyed are understood. This gives us two areas of performance to examine, the effectiveness with which the sound is spatialised, and the quality with which the speech is perceived. Spatialisation performance was assessed by means of a straightforward set of perceptual tests. 36 seated listeners were each exposed to sentence pairs spatialised to one of 12 azimuths on the horizontal plane (corresponding to a 30° spacing) for each of three types of spatialisation; ambisonics using a square array of loudspeakers, crosstalk-cancelled presentation using a very detailed (17th degree spherical harmonic - see [12]) HRTF model, and crosstalk-cancelled presentation using more computationally simple HRTF model (6th degree spherical harmonic). Listeners were asked to call out - according to the ‘hour of the clock’ - the apparent direction from which each speech stimulus was emanating. Figures 2-4 portray the responses of the subjects. The results are in the form of subjects’ azimuth judgements plotted against intended azimuth. The area of each circle is proportional to the number of judgements for that combination of stimulus and response.

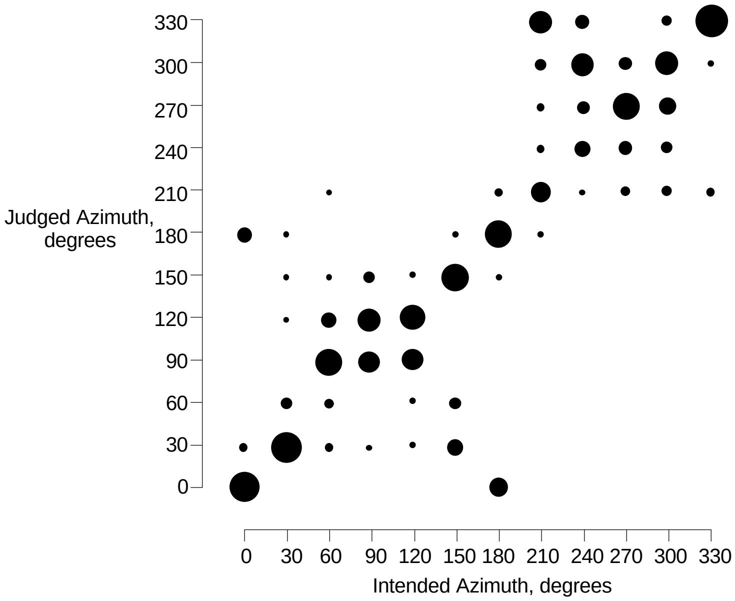

Figure 2: Listener judged azimuths verses the intended source azimuth, for the low-degree binaural stimuli

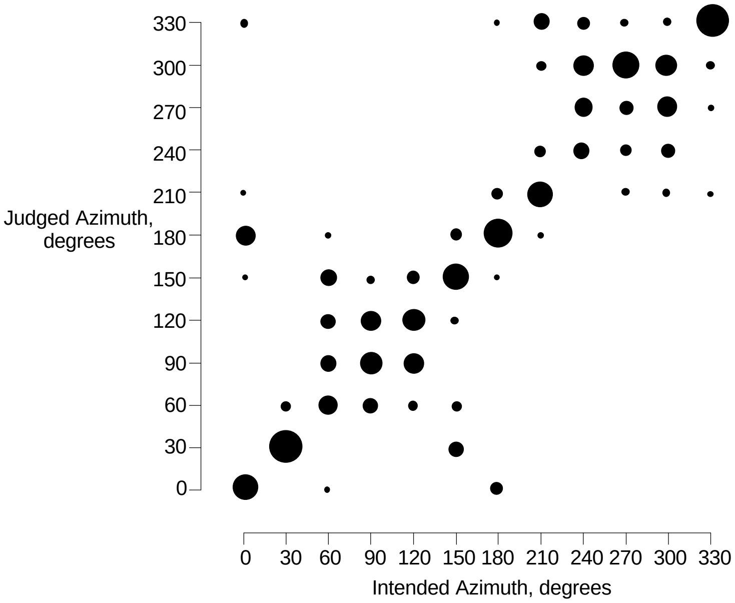

Figure 3: Listener judged azimuths verses the intended source azimuth, for the high-degree binaural stimuli

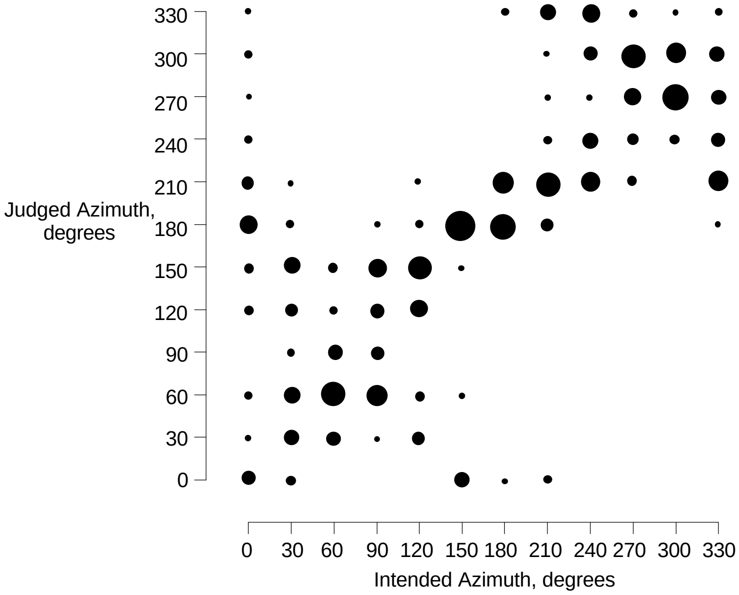

Figure 4: Listener judged azimuths verses the intended source azimuth, for the ambisonic stimuli

On each of these plots, large circles along the leading diagonals and little distribution of responses around this are indicative of highly accurate spatialisation performance. Both binaural techniques seem particularly accurate, with the majority of incorrect localisations differing by just a single ‘hour’, or in the form of a front-back reversal. (By reflecting in the transverse plane, judgements that are subject to front-back reversal can lead to a disproportionately large error in terms of angular separation from the correct azimuth.) The results for the ambisonic stimuli, although still relatively good, seem less accurate, with often quite large spreads of judgements either side of the correct azimuth, although, proportionally, front-back reversals seem less pronounced.

Audio Perceived Performance

Successful creation of spatial audio in a teleconferencing environment has the potential to enhance the effectiveness of the communication by increased immersion and telepresence. However, it is important that we verify that this is not achieved at the expense of a reduction in the perceived performance of the speech being conveyed, as tested using the formal techniques laid down by the International Telecommunications Union (ITU) and expressed using their established subjective metrics. A formal set of tests investigating the perceived performance of spatial audio in a teleconference environment is described in detail in [13]. Once again these tests collected listeners’ opinions of speech stimuli spatialised using the four loudspeaker ambisonic approach, the high-accuracy crosstalk-cancelled HRTF model, and the HRTF model based on less computation. Also, equivalent non-spatialised (mono) material was included as a reference. 24 subjects were tested in each of two tests.

The *Babble Resilience Test* established listening effort scores for spatialised and non-spatialised speech in the presence of different levels of peripheral vocal babble. Listening effort is defined as *the effort required to understand the meaning of sentences*, according to the following scale:

5	*Complete relaxation possible; no effort required.*
4	*Attention necessary; no appreciable effort required.*
3	*Moderate effort required.*
2	*Considerable effort required.*
1	*No meaning understood with any feasible effort.*

Average listening effort scores awarded by subjects to stimuli spoken by a female talker for each babble level and spatialisation technique are summarised in Table I.

	No Babble	**Low Babble**	**High Babble**
Non-spatial audio	5.00	2.29	1.17
Crosstalk-cancelled binaural (low degree model)	4.96	2.92	1.54
Crosstalk-cancelled binaural (high degree model)	5.00	3.42	1.88
Ambisonic (square loudspeaker array)	4.96	3.58	2.00

Table I: Average listening effort scores from the Babble Resilience Test

The second formal test of perceived performance, *Multiple Voice Discrimination*, used the Comparison Category Rating (CCR) technique to quantify listeners' preference for the reproduction of three simultaneous voices when those voices were spatialised to three distinct positions across the front half of the horizontal plane, as opposed to when they were reproduced non-spatially. The three spatialisation techniques were compared to equivalent mono reference stimuli, with four distributions of talkers across the three positions - FFM (female-female-male), FMF, MFM and MMF - and responses expressed by subjects using the following scale:

The second compared to the first is

+3	*Much Better*
+2	*Better*
+1	*Slightly Better*
0	*About the Same*
-1	*Slightly Worse*
-2	*Worse*
-3	*Much Worse*

Table II gives the mean CCR scores awarded to stimulus pairs consisting of the non-spatial reference followed by the spatialised speech, giving preferences for the spatialised material.

	FFM	**FMF**	**MFM**	**MMF**	**Average**
Crosstalk-cancelled binaural (low degree model)	0.33	0.50	-0.13	0.42	0.28
Crosstalk-cancelled binaural (high degree model)	1.54	1.33	1.50	1.25	1.41
Ambisonic (square loudspeaker array)	1.63	1.71	0.79	1.25	1.35

Table II: Average CCR scores from the Multiple Voice Discrimination Test

Discussion

Any results derived from subjective tests must be treated with care since they will, in general, contain a large amount of variance arising from differences between test subjects as well as from the variables which we wish to investigate. An extensive Analysis of Variance (ANOVA) carried out in [13] has allowed us to draw reliable conclusions from the results of the Babble Resilience and Multiple Voice Discrimination Tests. With reference to Table I, there are no significant differences between the performance of any of the reproduction techniques when no babble is present. Increasing the babble level always incurs a reduction in the mean listening effort score. However, the low degree binaural spatialisation is significantly better than mono when babble is present, and the high degree binaural and ambisonic stimuli are significantly better still. Spatial audio has yielded a clear improvement in perceived performance when peripheral vocal babble is present. This improvement is particularly marked for the ambisonic and high degree crosstalk cancelled techniques.

Similarly, examining the results in Table II for the Multiple Voice Discrimination Test: Although the low degree binaural spatialisation yields no significant preferences over mono, the high degree and ambisonic stimuli are awarded a significant level of preference for virtually every distribution of talkers in space.

These subjective test results illustrate that spatial audio reproduction in teleconference conditions does not compromise the perceived performance of the conveyed speech, and can yield a performance improvement in the presence of occluding material. Of particular interest to our comparative assessment is the fact that despite the fact that crosstalk-cancelled binaural presentation demonstrated better spatialisation performance that ambisonics, this was not reflected in the perceived performance. It seems that, in teleconferencing where the task is the understanding of speech, highly accurate localisation is less important than keeping the speech and occluding material as spatially distinct as possible. Therefore, the lower directional accuracy of ambisonics is countered by the comparatively higher occurrence of front-back reversals in the crosstalk-cancelled stimuli, particular in those prepared using the low degree HRTF model. These reversals may have the effect of 'folding' peripheral babble present in the rear half of the horizontal plane onto the front half, where its occluding of the main speech will be exaggerated.

Performance Specification

Formal metrics and testing techniques allow the meaningful, reliable and comparative assessment of performance. If spatial audio is to become a widely used component of telecommunications applications then such techniques must be developed and agreed upon. If not, developers will be forced, as we have when discussing spatialisation performance in this paper, to use their own, specific techniques and metrics, permitting little or no comparisons with other results or systems. Standard metrics might take the from of an estimate of the range of directions from which a sound is thought to emanate centred around a subject’s best guess, giving a measure of localisation ‘focus’. Perhaps even more importantly, we require the means to evaluate and express the performance of teleconferencing systems in terms of the level of perceptual realism or telepresence that they supply to users. Developing such meaningful measures of systems which, in general, create a combined sensory experience using sound and vision is a process which will require higher level psychological and possibly cognitive considerations based on the integration of information from ear and eye [14]. However, such measures will be essential in the assessment of perceptually realistic telecommunications: If researchers only have access to tools for investigating the performance of each sensory modality in isolation then compromises in telepresence arising from mismatches in auditory and visual information, and effects where information in one modality influences perceptions in the other, will not be able to be quantified.

CONCLUSIONS AND THE FUTURE

As might be expected with two such different solutions to what is essentially the same problem, sound field simulation and perceptual synthesis each have strengths and weaknesses in delivering spatial audio in teleconference services. Ambisonics’ ability to present spatial information to multiple listeners spaced or moving within the listening space makes this form of presentation particularly well suited to dedicated, permanent videoconference rooms, for which a B-format decoder can be calibrated. Binaural spatial audio, by means of perceptual synthesis, lags well behind in terms of such flexibility. However, the requirement of ambisonic systems to have a regularly distributed loudspeaker array and, particularly, loudspeakers to the rear of listeners means that it seems highly unlikely that the technique is desirable for use in other kinds of teleconference application. The essence of the popularity of ‘rollabout’ systems is their convenience and portability, with all the required apparatus situated on a single trolley which can be wheeled from room to room. Having to also set up rear loudspeakers and, potentially, recalibrate a B-format decoder seems highly impractical. Similarly, it is difficult to envisage desktop videoconferencing easily incorporating rear loudspeakers, since they would almost certainly be an obstruction in a day-to-day working environment.

The direct reproduction of sound spatialised by means of HRTFs or other binaural cues is naturally suited to headphones. The discomfort and isolation from the physical auditory environment which may be incurred by wearing headphones makes this type of spatialisation generally unsuitable for teleconference use. (Exceptions may possibly lie in any application that make use of a head-mounted, virtual environment style of visual display, or some brief use of a desktop system.) However, we have demonstrated the usefulness of crosstalk-cancelled reproduction to reproduce binaural sound over a pair of loudspeakers. Spatialisation of this type lends itself very well to desktop teleconferencing, where a pair of loudspeakers can easily be placed at either side of the screen and the layout of the screen and loudspeakers implicitly restricts the user’s position and movement. In principle, such spatialisation could also be incorporated into rollabout or room-based teleconference services, since the pair of loudspeakers could be mounted on the trolley or at either side of the screen. However, the relative positions of listener and loudspeakers are clearly not as fixed when such systems are used. Also, two loudspeaker crosstalk cancellation can, by definition, only spatialise to a single listener. Rollabout and room-based systems are not generally intended for single use but for presentation to a room, possibly filled with participants. Even if head tracking were supported crosstalk-cancelled perceptual synthesis would still be unsuitable for this form of teleconference.

In summary, therefore, it is clear that sound field simulation seems to be appropriate only for those teleconference applications which can permanently make use of a suitably rigged listening room. Such applications are and will probably remain a custom-supplied, niche market. Perceptual synthesis shows great potential for integrating spatial audio into desktop teleconference systems for single users; a type of application which seems sure to gain an enormous amount of popularity in the near future, thanks to the ISDN. In their current form, neither spatial audio philosophy seems ready to tackle ‘rollabout’ videoconference systems, which seem almost certain to form a large proportion of the teleconferencing market, due such applications being predicated on flexibility and portability. In order to integrate spatialised sound into these and other services it is very much hoped that the two opposing camps in spatial audio can make use of each others’ expertise; perhaps yielding binaural systems which can more readily support movements and multiple listeners, or sound field synthesis which is made more effective by using knowledge of sound localisation cues - new forms of spatial audio in which the whole is more than the sum of their parts.

ACKNOWLEDGEMENT

This research was funded by an EPSRC CASE award partly funded by BT. Contact: Phil Sheppard, MLB 3 7, BT Laboratories, Martlesham Heath, IPSWICH IP5 7RE, United Kingdom

REFERENCES

[1] Malham, D. G., Myatt, A. 3-D sound spatialisation using ambisonic techniques, Computer Music Journal, vol. 19, no. 4 (1995), pp. 58-70

[2] West, J. E., Blauert, J., MacLean, D. J., Teleconferencing system using head-related signals, Applied Acoustics, vol. 36 (1992), pp. 327-333

[3] Wightman, F. L., Kistler, D. J., Headphone simulation of free-field listening II: Psychophysical validation, Journal of the Acoustical Society of America, vol. 85 (1989), pp. 868-878

[4] Cooper D. H., Bauck, J. L., Prospects for transaural recording, Journal of the Audio Engineering Society, vol. 37 (1989), pp. 3-19

[5] Miyoshi M., Koizumi N., Research on acoustics for future telecommunication services, Applied Acoustics, vol. 36 (1992), pp. 307-326

[6] Begault, D. R., 3-D Sound for Virtual Reality and Multimedia, AP Professional, Cambridge, MA, 1994

[7] Evans, M. J., Angus, J. A. S., Tew, A. I., A spherical harmonic analysis of head-related transfer function measurements, Proceedings of the Institute of Acoustics, vol. 18, no. 8 (1996), pp. 191-200

[8] Burraston, D. M., Hollier, M. P. Hawksford, M. O., Limitations of dynamically controlling the listening position in a 3-D ambisonic environment, Audio Engineering Society preprint 4460 (1997)

[9] Bauck J., Cooper, D. H., Generalized transaural stereo and applications, Journal of the Audio Engineering Society, vol. 44 (1996), pp. 683-704

[10] Wightman, F. L., Kistler, D. J., Headphone simulation of free-field listening I: Stimulus synthesis, Journal of the Acoustical Society of America, vol. 85 (1989), pp. 858-867

[11] Wenzel, E. M., Arruda, M., Kistler, D. J., Wightman, F. L., Localization using nonindividualized head-related transfer functions, Journal of the Acoustical Society of America, vol. 94 (1993), pp. 111-123

[12] Evans, M. J., Tew, A. I., Angus, J. A. S., The perceived performance of speech spatialized using a spherical harmonic model of head-related transfer functions, Audio Engineering Society preprint 4406 (1997)

[13] Evans M. J., The Perceived Performance of Spatial Audio for Teleconferencing, DPhil Thesis (1997), University of York, York, UK

[14] Knudsen, E. I., Brainard, M. S., Creating a unified representation of visual and auditory space in the brain, Annual Review of Neuroscience, vol. 18 (1995), pp. 19-43

NOTE

After 30 September 1997 Michael J. Evans should be contacted at the following address:

Department of Engineering
The University of Reading
PO Box 225
Whiteknights
READING RG6 6AY
United Kingdom
Tel: +44 118 931 8567
Fax: +44 118 931 3327