Proceedings of the Institute of Acoustics

THE IMPLEMENTATION AND APPLICATION OF VIRTUAL ENVIRONMENTS IN TELECONFERENCING

I Drumm University of Salford, The School of Acoustics and Electronic Engineering, U.K.
P Phillipson University of Salford, The School of Acoustics and Electronic Engineering, U.K.

1. INTRODUCTION

The paper describes the implementation and example use of an interactive, real-time computer based application for the construction of virtual environments. The application allows the user to easily construct architectural spaces and so add in furniture and multiple sound sources. Hence, he or she can 'walk through' the virtual environment experiencing simultaneous visual and acoustic simulation. The case discussed below is the generation of binaural virtual environments where speech communication between several different talkers at different room positions can be achieved. This demonstrates the advantages in using a binaural 2 channel communication system and how such a software application can be used to aid acoustics research and teaching.

One advantage in using some form of spatial positioning and auralisation is in producing a more "realistic" and comfortable meeting place where the participants can be assigned a virtual position within a virtual room of a specified acoustic design. Each participant will receive a different set of spatial cues depending on their own virtual position with respect to the others, hence assisting their interpretation of which words come from which person (especially important when all the participants are strangers). However the most significant advantage to be discussed in this paper is the improved speech intelligibility when the signals are spatially processed.

There are numerous situations where a listener is presented with multiple speakers over a single channel such as a telephone meeting line, video-conference, or in the military or emergency services. The listener hence needs to be able to isolate and clearly perceive one voice among many.

It has been known for several decades that binaural listening offers improved signal detection over monaural listening in the presence of masking noise and this is commonly referred at as the "cocktail party effect" [4]. Many investigations have been conducted into this effect and it has often been quantified in terms of the closely related Binaural Masking Level Difference (BMLD); the difference in masking thresholds when the same signals are presented monaurally and binaurally.

If a sound source is positioned to one side of the listener (as shown by Source A in Figure 1) then this produces interaural differences, as the signal reaches one ear before the other and depending on the frequency, may be attenuated before reaching the contralateral ear. These interaural time and level differences (ITD and ILD) in combination with the filtering effects of the outer ear and head (Head Related Transfer Function - HRTF) are all used as the cues in localising a sound source. However, the ITD appears to be the most important in detecting signals within noise [17, 19] and as the ears interpret this as a phase difference at low frequencies, the BMLD is best at frequencies below 1 kHz. The BMLD is also dependant on the azimuth position of the source, as the further this moves closer to a value of 90° the more significant the interaural differences and hence the BMLD is highest. Previous works have shown that for lower frequency sources at the side of the head a typical figure of around 10 dB can be achieved [16, 19].

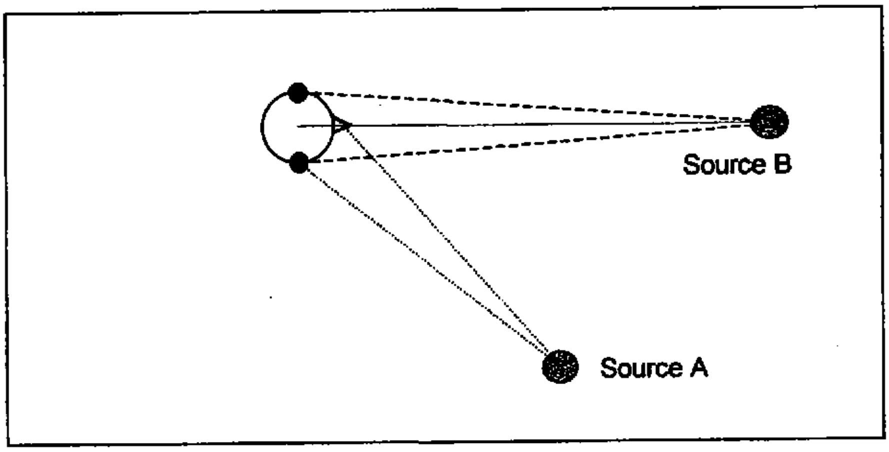

Figure 1: Inter-aural differences in level, time and frequency spectrum (Source A) help us localise sound sources and improve our signal detection in the presence of masking noise.

The software application implements up to 16 virtual sound sources in a room in real time processing by applying the relevant spatial cues. The importance of various aspects of aural and visual localisation on a listener's ability to pick out and understand individual speakers is worthy of investigation giving insights into auditory perception and justification for virtual reality in teleconferencing. The initial work discussed here investigates the advantages of this personal computer based system for use with several speech sources.

2. IMPLEMENTATION

2.1 Virtual Environment Application

The application was developed with C++ to run on Windows based PCs. In recent years popular 'Application Programming Interfaces' (APIs) such as OpenGL [1], A3D [3] and DirectX [2] have enabled software developers to easily construct applications featuring virtual environments with 3D graphics and 3D room acoustics / auralisation. The A3D API implements Head Related Transfer Functions (HRTF) together with simple first order reflections and a geometry based reverberation algorithm to render sound from multiple sources within a virtual environment. Similar functionality is becoming available via Creative Laboratory's EAX API and Microsoft's DirectX. Dedicated cards for 3D graphics and 3D sound have become popular and affordable, their development largely sponsored by a lucrative entertainment's industry. Common APIs provide a standard programming interface to such hardware acceleration enabling smooth real-time operation.

The author has utilised these APIs together with in-house room acoustic rendering techniques such as the 'adaptive beam tracing' to develop a flexible application for the construction and navigation of virtual environments.

The user specifies planes of different materials each with it's own visual texture, sound absorption and transmission coefficients. Planes can be positioned to construct the main architectural space or the furnishings within. Previously constructed objects can be imported and positioned from the user interface. OpenGL visually renders the data in real-time relative the viewer. The user can also position multiple sound sources within the space; each sound source having an assigned audio file playing at a specified gain. Hence A3D models the propagation of sound within the space so as to apply appropriate reflections/reverberation and HRTF's relative to the listener.

The application implements other room acoustics modelling techniques including the image method, ray tracing and triangular and adaptive beam tracing [5]. The latter has been developed by the author to provide a fast and accurate profile of propagated sound reaching the listener. Subsequently measurable acoustic parameters such as Reverberation Time, Early Decay Time, Clarity Index and Early Lateral Energy Fraction can be predicted.

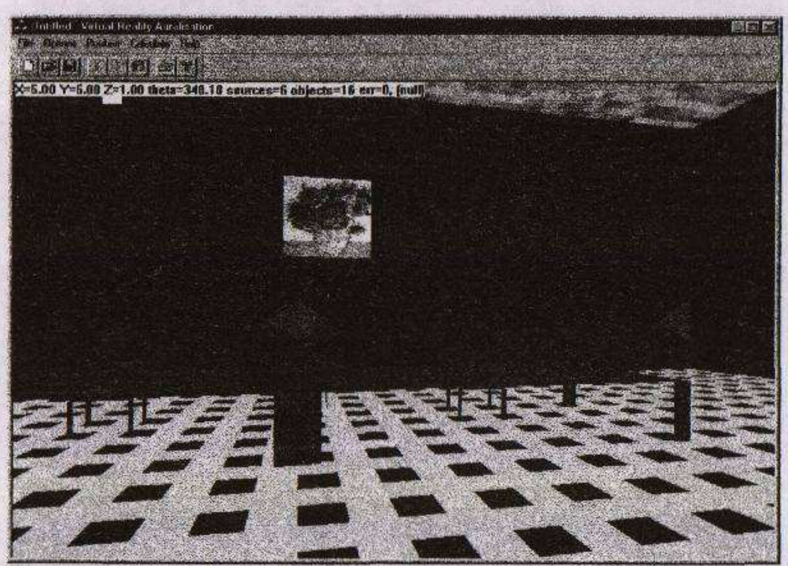

Figure 2: The 'Virtual Environment Application' used to evaluate speech intelligibility with a virtual conference environment.

2.2 Speech Intelligibility

The experiments given are a precursor to a more detailed investigation into the potential benefits of virtual reality in teleconferencing. The aim of the initial test was to determine the speech intelligibility of a binaurally processed virtual voice source when also presented with a further six binaurally processed but diotic virtual speakers. It was decided to use six different human voices as a potential masking source as this was considered to be more relevant to the specific application than using speech band noise.

All speakers voices were recorded in an isolated listening room using a vocal microphone, analogue to digital input channel and hard disk recorder. For reasons of compatibility a sample rate of 44.1 k Hz was used and the audio files were stored in a 16 bit .wav format.

The speaker whose speech was to be identified and understood was recorded reading from lists of phonemically balanced words [18]. Only a male voice was used in this case. This speech was edited to place the words in different random orders and to insert a 6 second pause between them. The 6 "babble" speakers (2 female, 4 male) whose voices would contribute to the masking signal were recorded reading passages from either books or magazines. All speakers were asked to speak at a constant level throughout the recording. Their speech was edited to remove any pauses of significant length and changes in level that would reduce the overall masking level. An appropriate gain factor was applied to all the individual speech files to ensure that the individual speech levels were approximately equal.

The listening test was conducted in an isolated listening room where the speech signals were binaurally presented to simulate a virtual conference environment, within which all "babble" speakers were positioned at a position 3 meters in front of the listener. As the main speaker read the word list the subjects were asked to record which words they believed they heard. After each run of fifty words the gain and position of the main speaker were varied with respect to the other speakers hence a threshold of intelligibility could be determined. The reference condition was chosen where the main speaker was also positioned directly in front of the listener hence providing a diotic (or monaural) signal for the listener.

The aim of the test was investigate the degree of improvement in speech intelligibility using this particular Virtual Environment Application and also to compare two different azimuth angles of 45° and 90°. It was decided after initial listening by the authors to only conduct test cases close to the threshold of intelligibility to limit the number of listening tests to accommodate the subjects. Five subjects were used in the test and nine conditions were considered.

Condition	Azimuth angle (°) (0° = central)	Relative gain of main speaker (dB)
1	0	-2
2	0	-4
3	0	-6
4	45	-10
5	45	-8
6	45	-6
7	90	-10
8	90	-8
9	90	-6

Table 1. Listening test conditions.

2.3 Word Score Results

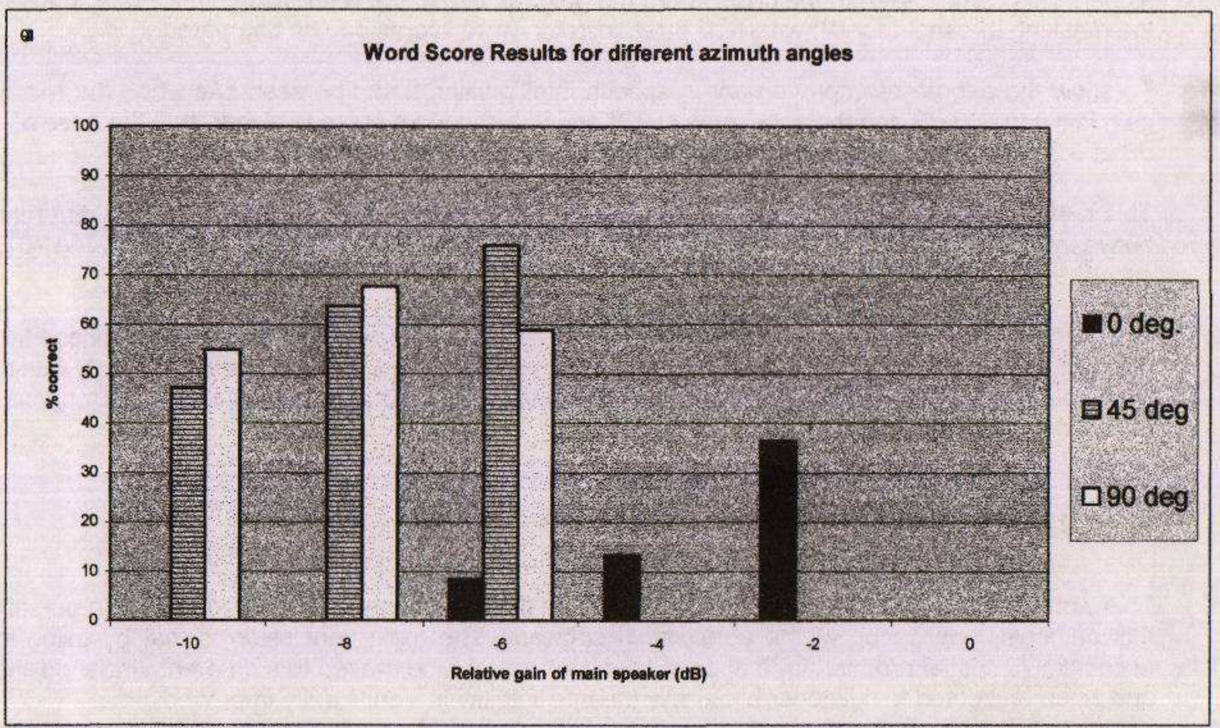

Figure 3. Average percentage word score representation of speech intelligibility with respect to speaker gain and position.

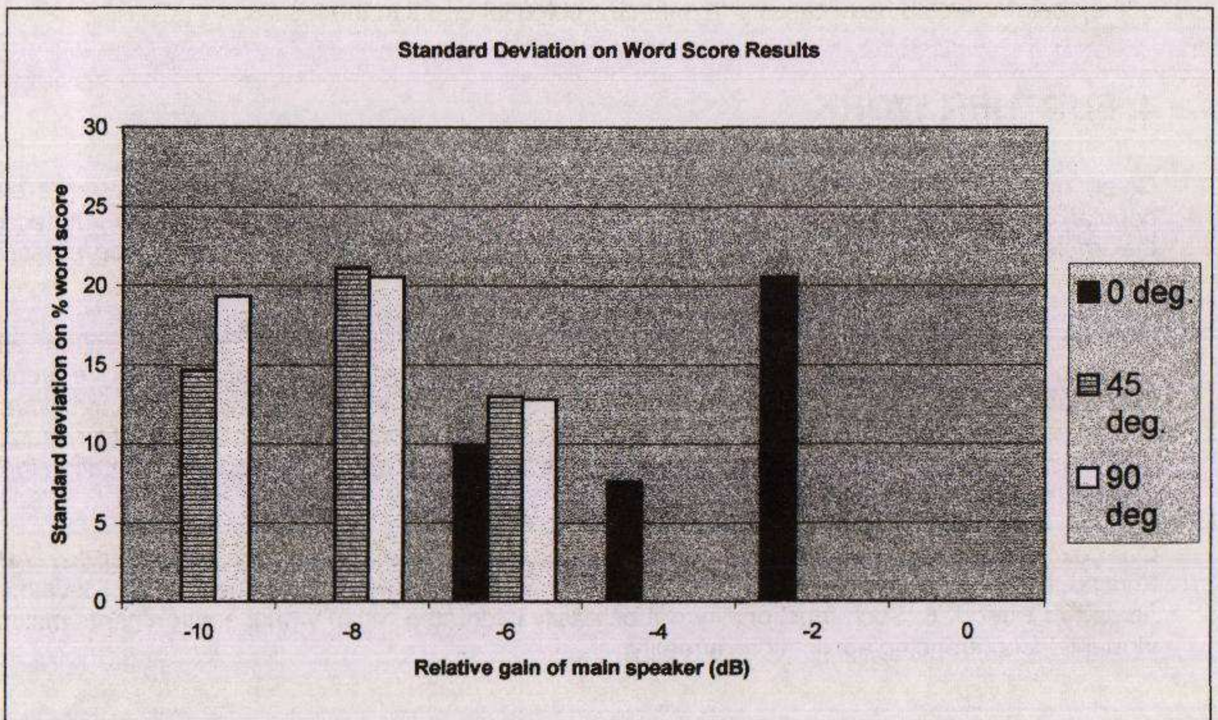

Figure 4. Standard deviation of the five subjects for each condition.

Figure 3 shows the % scores from the word list test. It is clear that when the main speaker is at a virtual position to one side of the listener the number of correctly identified words is significantly higher. This can be seen in the case where the gain relative to the reference gain is –6 dB (close to the masking threshold for 0°) where all three azimuth word scores can be compared.

To show the degree of improvement in speech intelligibility, it can be seen that when the relative gain is set to –10 dB and the speaker is at 45° and 90°, the word score is higher than the case of –2 dB at 0°. This suggests a BLMD of approximately 10 dB.

This result compares well with previous work in this area such as [16,19]. However, the difference between the 45° and 90° word scores in this test is not as significant as expected and in the –6 dB case the most intelligible condition occurred when the source was positioned at 45°.

The standard deviation plots indicate that there were significant differences in the responses from the different subjects during the listening test. This may have been due to varying familiarity with the words but may well have been due to some undetected hearing damage. All the subjects responses followed a common trend.

3. CONCLUSIONS

This work was a crude initial study intended to give the authors some idea of the value in pursuing this particular application of the auralisation software. The main conclusion is that by using the application's spatial processing, a speech sound source can be moved to a different virtual position and when presented to a listener binaurally the intelligibility is significantly improved.

The differences in word scores between the five subjects were significantly high and may have caused some error. Unfortunately time constraints prevented the authors from conducting audiometric tests on the subjects. The use of more test subjects and some assessment of the state of their hearing would be necessary to minimise potential errors in the results.

4. FURTHER WORK

Given developments in ever more powerful and more affordable processors, the internet and popular 3D graphics and 3D acoustic sound cards using standard extensible software interfaces, the implementation of virtual environments for teleconferencing between typical office personal computers is clearly feasible.

The benefits of spatial processing to produce interaural differences for localisation and communication are clear. Furthermore multiple participants could have the ability to move around virtual conference environments as avatars communicating in a virtual world. Hence individuals could move into smaller groups for clearer communication when needed or even into separate rooms. Integration with other audio and video signals generated from multimedia presentation tools could also be considered.

One exciting development comes with additions to Microsoft's forthcoming DirectPlay API. Functions for multiple 'voice over the net' communication will be available to the software developer as part of DirectX 8. Such functionality will be easily integrated within virtual environments making virtual teleconferencing applications a reality.

Future work will involve additional speech intelligibility and localisation tests to determine the relationship between smaller azimuth angle differences and intelligibility. Additionally, important aspects to be considered are that of intelligibility when all the speech sources are at different azimuth positions, the influence of the room modelling technique on this and the use of visual cues. Other possible audio related avenues are to investigate the performance when the signals are band limited, dynamically compressed or perceptually coded.

A potential problem with using this type of teleconferencing system will be related to the requirement for a headphone / VDU interface. An investigation into using a loudspeaker reproduction system should be considered.

A more robust technique for the word score test [18] will be implemented in future work.

5. REFERENCES

[1] Segal & Akeley, "The Design of the OpenGL Graphics Interface", Silicon Graphics Incorporated, http://www.opengl.org/

[2] Microsoft, "About DirectX", http://www.microsoft.com/directx/

[3] Aureal, "A3D White Paper", Aureal Semiconductor, www.aureal.com and www.a3d.com

[4] Cherry, E.C, " Some Experiments on the Recognition of Speech with One and Two Ears," JASA, vol. 25, pp. 975-979, 1953.

[5] Drumm, I. "The Development and Application of an Adaptive Beam Tracing Algorithm to Predict the Acoustics of Auditoria" - Ph.D. Thesis. University of Salford. 1997.

[6] Drumm, I and Lam,Y.W, "The adaptive beam tracing algorithm," J. Acoust. Soc. Am. **107**, 1405, 2000.

[7] Drumm, I and Lam,Y.W, "'The development of adaptive beam tracing," IOA Proceedings, Auditoria, 1999.

[8] 'Drumm, I," The real-time simulation of the acoustics of virtual environments on personal computers," IOA Proceedings, Reproduced Sound 15, 1999.

[9] M. Volander, "Simulation of Transient and Steady State Sound Propagation in Rooms using a New Combined Ray Tracing / Image Source Algorithm," J. Acoust. Soc. Am., 86 (1), (1989).

[10] D. Van Maercke & J. Martin, "The Prediction of Echograms and Impulse Responses within the Epiduare Software", Applied Acoustics 38, 93-114, (1993).

[11] T. Lewers, "A Combined Beam Tracing and Radiant Exchange Computer Model of Room Acoustics," Applied Acoustics," 38, 161--178, (1993).

[12] Hogdson, Heerema, Halingten, "PlantNoise sysem for the prediction, visulisation and auralization of industrial noise during computer walk-through," JASA Feburary 1999 - Volume 105, Issue 2 p. 989.

[13] Funkhouser, " Interactive acoustic modelling in complex environments," JASA Vol 105, No 2, pp 1357-1358, February 1999.

[14] Kutruff,H. "Auralisation of Impulse Responses Modeled on the Basis of Ray-Tracing Results," J. Audio Eng Vol 41, No 11, 1993 November.

[15] Rainer Feistel, "Ears Auralisation Software,"J. Audio Eng Vol 41, No 11, 1993 November.

[16] Begault, D. R., "Virtual Acoustic Displays for Teleconferencing: Intelligibility Advantage for "Telephone-Grade" Audio," J. Audio Eng, Vol. 47, No. 10, 1999 October, pp 824-828.

[17] Bronkhorst, D.R. and Plomp R.,"The Effect of Head-Induced Interaural Time and Level Differences on Speech Intelligibility in Noise," JASA vol. 83, 1988, pp. 1508-1516.

[18] ISO/TR 4870:1991(E), " Acoustics – The construction and calibration of speech intelligibility tests."

[19] Blauert, J," Spatial Hearing – The Psychophysics of Human Sound Localisation." MIT Press. 1997.

HOW MANY CHANNELS ARE ENOUGH?

F. J. AMPEL MS, MAES, MASA, MInstSCE, MIoA,
Technology Visions s.a. - Overland Park, KS, U.S.A.
FrederickAmpel@Compuserve.com

1.0 INTRODUCTION

Throughout the seven decades in which multi-channel audio has been a presentation format, the debate regarding how many recording/storage and playback/presentation channels are sufficient to 'create' or 're-create' the desired physical and subjective auditory reality has been both consistent and often acrimonious.

The earliest papers on the topic, produced by Harvey Fletcher and his colleagues at Bell Laboratories during the mid- to late-1920's suggested emphatically, that at least three and as many as five were the minimum needed.

During the decades after those finding were published, the motion picture industry, never moved beyond a single channel, despite both the option to do so as early as 1940, and significant evidence that it would improve the motion picture experience. It was not until the 1950's that any form of multi-channel audio entered the theatrical realm, with the premiere of the Cinerama films in 1957, and both British engineer Raymond Spottiswoodes 3 channel sound for film, and the introduction of the remarkably similar Warnerphonic sound process, in 1953.

The music recording industry also remained firmly monophonic, again ignoring numerous options to move to more channels, until the approval of LP stereo by the Record Industry Association of America (R.I.A.A.) on 25 March 1958, after a very ugly and rancorous war of the technologists. Until the introduction of the two thankfully ill-fated quadraphonic systems, and remarkably for many decades thereafter, two channels were regarded as the holy grail of audio reproduction and many purists still regard that as all that is required.

As we noted earlier, the motion-picture industry has tried 1, 2, 3, 4, 5, 6, 7, 8 and even more experimentally. The renowned (and much missed) Michael Gerzon, said unequivocally that 'all the information needed to re-create a sound field in three dimensions can be transmitted and contained within four audio channels.' The music-recording industry has experimented with 3, 4, and 8+ (Ambisonic) channel formats. The perceptual-encoding experts say they can create as many as we would like, and others now deem 10 as the latest altar upon which to proffer the one true answer. Who's right? Are any of these schemes the precise answer? If not where does it lie?

2.0 WHY MORE THAN ONE CHANNEL ANYWAY?

This whole debate about how many channels are needed, was probably started (albeit unintentionally) at the Exposition Internationale d'Électricité in Paris in 1881.

There, the French engineer and inventor Clément Ader successfully created what is arguably the first acoustic simulacrum (virtual-reality) experience. Lacking any recording technology, or even thoughts of such capability, he made no attempt 'time shifting' - his system worked only for live performances.

Using a device he euphonically named the 'theatrephone' the outputs of two sets of telephone transmitters (microphones) across the front of the stage of the Paris Opéra, were sent 3 km through distribution coils to the Palais d'Industrie, where the populace could listen to the performances through electro-acoustically based stethoscopic-type headphones, with one side connected to the transmitter(s) on the left, and another on the right, of the Opéra's prompt box.

The singers, the orchestra, and the audience were spectacularly audible, and (probably not deliberately) in the spatially correct position.

So, despite the potential demonstrated by Ader, during the formative years of the technology of recording and reproducing sound, no one was considering how many channels were needed to accurately reproduce anything, they were simply trying to develop and perfect the basic technology.

To keep things simple and manageable, they worked with what we would now call one-channel or monophonic reproduction, although it was not viewed through a channel-based lens at the time. After all, getting the whole complex process to work at all was deemed a sufficient goal.

Once the basic technology had been converted from a laboratory curiosity to a viable mechanical system, the now more interesting issues of quality, presentation and 're-creation' began to surface.

In a seminal 1926 paper on the then-brand-new 78 rpm <u>monophonic</u> Western Electric electrical recording system, Joseph P. Maxfield and H.C. Harrison noted that:

> 'perfect reproduction requires that the various components of the reproduced sound reaching the listeners ears should have the same relative intensity and phase relationships as the sound reaching the ears of an imaginary listener to the original performance would have contained.'

Maxfield and Harrison then gave their audience what appears to have been an unintended preview of the highly secretive work then going on at Bell Labs on 'multi-channel' by saying 'It would be difficult if not impossible to attain that goal with a single channel of capability. . .a minimum of two would be required to match the electrical relationships deemed critical when the signal was reproduced.'

The Maxfield and Harrison statement: 'Various components of the reproduced sound. . .should have the same relative intensity and phase relationships' clearly implies that an infinite number of reproduction channels would be the ideal, however impossible that might have been then or even now.

More recent and quite extensive and lengthy research into sound localization by human listeners shows that such precise intensity and phase relationships are *not* needed.

This is because the ear simply cannot in fact localize any sound source more accurately than about ± 3.6Eand that's only for the forward-facing direction (in front of the nose).

At 90E to the side of the head the localization accuracy is severely degraded, to no better than ± 10E. Vertically above the head, localization is accurate to a best case of only ± 22E. To the rear/back of the head, localization capability returns to the limited 10° positional accuracy window (Jens Blauert, *Spatial Hearing, the Psychophysics of Human Sound Localization*, revised edition, MIT Press, 1997, pp. 38-43.)

It is vitally important to recognize that the goal being sought by Maxfield and Harrison, and slightly later by Harvey Fletcher and his colleagues, was the viable recreation of a stage based presentation of music, actors, etc., to a seated listener in a large theatrical type space — simply because it was the presentation format that was dominant at the time and with which they and the potential audience were most familiar.

In that case, precise localization and spatial depth were not issues (even if they had been understood), simply re-creating the frontal sonic perspective and having lateral positional accuracy was more than adequate .

That same year in a separate paper, Fletcher and his co-authors would note, recognizing both Ader's demonstration and Maxfield's technical points: 'The use of two ears, that is, two-channel listening, gives the listener a sense of lateral direction for each of the various sources of sound to which at a given moment he may be listening, and, therefore, he apprehends them in their relative distribution in space.'

While this statement may have been a basic recognition of the need for a more complex, three dimensional soundfield to produce more acceptance of the presentation, it was not until technology provided the opportunities that this knowledge would again become truly useful.

3.0 SO. . .WHAT IS THE GOAL?

To logically determine an answer that effectively takes into account current technological realities, we first must ask what was and is the goal being sought.

In its simplest form the goal of most multi-channel capable audio reproduction systems is the creation of the best feasible auditory reality. Underlying that basic purpose is the more complex intent of producing in the mind of the listener a very complete and thoroughly willing suspension of auditory disbelief. Those were the essential objectives for Ader, Maxfield, and Fletcher, and all who have followed in their footsteps.

But it is also crucial to recognize, even if the early experimenters did not, that within these two core concepts resides yet another objective: the manipulation of *emotional response*, to move the listener beyond actuality, reality or accepted belief.

The now celebrated 27 April 1933 experiment conducted by Bell Labs achieved that third concept, perhaps more effectively than anyone at the time realized. Using three microphones in the Academy of Music in Philadelphia, one at each side of the orchestra and one in the center, each about 20 ft (6 m) back and 10 ft (3 m) above the lip of the orchestra platform, a live performance of the Philadelphia Orchestra was transmitted to Constitution Hall in Washington, D.C., where the public was permitted to hear the live transmission through loudspeakers.

The reaction was essentially identical to that of Ader's audiences more than a half-century earlier — stunned amazement and disbelief!

Why three channels? The Bell Labs researchers had calculated and observed in their preliminary work on the subject, the apparent angles of localization for two through five channels.

They found two channels to have serious spatial distortions, while three channels were the minimum which would permit an almost linear localization transfer function.

Fletcher explained it this way: 'Reproduction systems do not consist of two, three, or any other fixed number of channels. There must be sufficient of these to give a good illusion of an infinite number.', .With this statement it seems as if he was clearly pointing us back to the work of Maxfield and Harrison, just five years earlier.

In addition he was he was essentially re-stating the core goal: the creation of a convincing audio simulacrum, which requires the suspension of disbelief, and to work effectively must make the listeners believe what they are hearing.

4.0 ACHIEVING THE GOAL

To make people in fact believe that what they are hearing is authentic, we must successfully produce the psychological phenomenon known as *recruitment*. Simply put, the audio presentation must provide sufficient reality so as to 'recruit' the listener into believing that what they are hearing is an accurate representation (as they would understand it,) of the source being portrayed.

Recruitment is a very complex process and in large part depends on the prior knowledge and experience of the intended audience. Inherent in the process is the depth of understanding regarding the particular sound, picture or image which the listeners (or viewers) bring to the presentation.

For example, in the very early days of film making technicians at Edison's labs photographed waves breaking on the New Jersey shoreline. Unfortunately they positioned themselves a bit too close to the water and one larger-than-normal breaker inundated the camera.

The film they ended up with, once it had been dried out and developed, when projected for audiences new to the whole cinematic concept, made people leap from their seats and run for the exits as that large wave accelerated towards them from the screen. Those audiences had little to no recruitment to the idea of images on film being just that, images not reality, and thus accepted what they saw as being very real indeed.

To our 21st century minds this seems ridiculous. Our ears, eyes and brains have been 'recruited' to a much higher level of sophistication, but to the inexperienced viewers of the turn of the previous century this was very real indeed. To make things realistic and convincing in the year 2000 requires a much higher degree of performance and detail, to compensate for the much higher recruitment levels in the intended audiences, but the underlying psychology remains the same.

5.0 WHAT DO WE WANT TO CONVINCE OUR EARS OF?

For any reproduced audio source to be convincing it must produce the following responses in the listener:

1. It must be 'real' or sound like what the listener expects that source to sound like, in spite of any enhanced recruitment that may exist.

2. The sound source must have an appropriate vector; that is, direction, distance, moving or stationary.

We might aptly state that our visual sense determines the position of objects in space while our auditory sense determines the position of events in time.

It is this spatialization capability and the human brain's need for details that often makes it far more difficult to create a convincing auditory reality.

It is considerably easier to get an audience to believe they are on an intergalactic starship (they have at best a very limited recruitment level since none of them have ever been on such a craft) than to make them believe they are in a live performance of any type when that performance is delivered from a recording.

Relatively recent research by psychologist Neil Todd at the University of Manchester (U.K.) has shown that we begin to hear in the womb. For the final four or five months of our gestation our ears are functioning.

Suspended in the amniotic fluid, we are surrounded by the pulsating low frequencies of our mother's heartbeat, the ELF of the gastro-intestinal tract, the 100-300 Hz fundamentals of the mother's spoken voice.

This constant stimulus energizes a relatively primitive organ of the inner ear called the sacculus, which is connected not to the higher brain but to the hypothalamus, controlling emotionally triggered hormones, and basic instinctive responses.

This pre-training (or pre-recruitment) of the auditory system means that sound can also be used to produce mood changes. A classic illustration of this is the use of underscore music in motion pictures for dramatic effect or to elicit those emotional responses. Perhaps the best example of this process is Bruce's theme (Bruce was the shark) from *Jaws*, where the use of pulsating low-frequency sound by composer John Williams produced an adrenalin shot for the audiences, heightening the tension and fear within the movie substantially.

Another more recent example is the dramatic low frequency thuds, and the visuals of the shaking water in the glass produced for the climatic approach of the T-Rex in *Jurassic Park.*

Knowing about and using these stimulus-response pathways, which was clearly the intent of these modern filmmakers, allows those seeking to successfully create auditory reality to build in cues to trigger the needed reactions — greatly assisting the process of disbelief suspension.

6.0 TIME CHANGES EVERYTHING

The things that those using any form of reality creation (visual or auditory) want their audiences to believe change with time. Fletcher wanted his listeners to believe they were listening to a live performance on a stage in a large hall. The creators of the first cyber-reality movie (*Tron*), wanted their listeners and viewers to believe they were inside a computer, while the producers of *The Matrix* wanted us to believe that our whole reality was artificial.

Our ability to produce the needed stimuli changes as well, from Edison's wax cylinders to today's computer based VR systems, and multi-channel audio data streams.

The technology itself evolves, often quite rapidly over time. In Fletcher's world the room acoustics were a fixed physical reality. Using the powerful DSP engines now available we can manipulate perceived room acoustics to a substantial degree. This capability has significantly aided the movement of motion picture 'reality creation' from the large picture palaces of seventy-five years ago to the private home theaters of today.

7.0 SO. . .HOW MANY CHANNELS ARE ENOUGH??

Given the current state of the art, and the physical requirements determined by the need to use loudspeakers, it would seem that somewhere between six and eight data streams, properly decoded and manipulated will produce an extremely convincing auditory reality.

While in some specific applications more physical locations (loudspeakers) may well be needed, the actual signals being sent to them can be produced from a basic six-to-eight-channel data set. This conclusion is based on our current capabilities to sample reality using microphones, and reproduce it over loudspeakers.

As technology opens up new options for every step in the process, it is highly likely that many more 'channels' will be available, and therefore will be used. The old adage that things expand to fill the available space, and to use the available computing power, applies here as well.

When 'transducers' are produced that can be installed in places where loudspeakers can't or won't work, they will be employed. If this means we can establish a finer level of detail in our auditory re-creation process, it will absolutely enhance the level of reality that can be produced.

On the other hand, simply adding more channels to the current state of the art is not likely to produce an increasing benefit level, unless it provides a way to significantly expand the creation of needed details.

The critical importance of the auditory experience, especially when presented in surround formats, is well described by Stephen Handel (*Listening, an Introduction to the Perception of Auditory Events*, MIT Press, 1991, p. xi):

> Listening puts me in the world. Listening gives me a sense of emotion, a sense of movement, and a sense of being there that is missing when I am looking. I am more frightened by thunder than by lightning, even though I know that thunder is harmless and lightning is deadly. I feel far more isolation living with ear plugs than living with blinders. Listening is centripetal; it pulls you into the world. Looking is centrifugal; it separates you from the world.