

Surround sound *from all angles*

Traditional stereo sound reproduction relies on two loudspeakers to create a sound image. Better understanding of the way in which the brain interprets and locates sounds – along with advances in digital signal processing – make it possible to recreate ‘surround sound’ without a large array of loudspeakers. The resulting technology is now used in innovative audio products incorporated into computer displays, video games, personal audio systems and mobile communications.

Conventional stereophonic sound reproduction puts two loudspeakers in front of a listener who perceives the original source as being between the sound sources. Changing the relative amplitudes of the signals sent to the loudspeakers ‘moves’ the apparent position of the source. This simple technique for generating a ‘virtual image’ of the original sound source has been the basis of two-channel stereophonic sound reproduction for over 70 years.

With conventional stereo, the sounds reaching the listener’s ears are simply an addition of the signals radiated by the loudspeakers. It relies on the fact that the human auditory system locates a sound source from the differences in both the timing and level, or amplitude, of the sound reaching the ears. In traditional stereophony, these differences combine in a subtle and complicated way to generate an illusion of the existence of a virtual sound source.

Recent advances in chip design make it possible to manipulate the signals sent to the two loudspeakers in a much more sophisticated way than by simply changing their amplitude. If we know the signals that the original sound source would produce at the listener’s ears, we can digitally process the original signal before sending it to the loudspeakers to

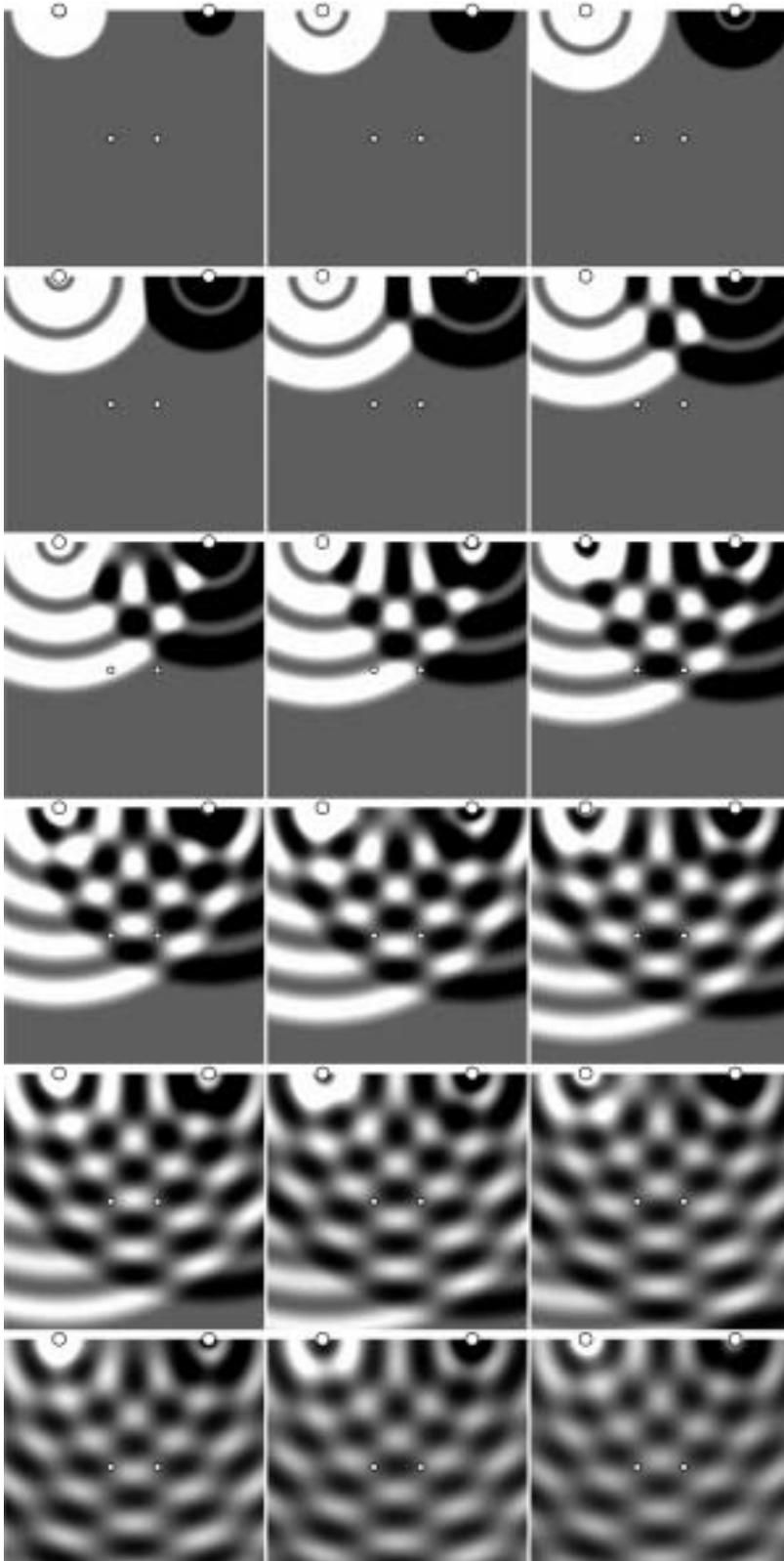


Figure 1 Crosstalk cancellation – a representation of the two loudspeakers as ‘point’ sources

reproduce an almost exact replica of these signals at each instance in time. This gives the listener more accurate and repeatable cues on the timing and level of the sound.

Thus with digital signal processing (DSP) there is much more scope for generating accurate virtual images of the original sound source. For example, the position of the virtual image can be made to vary over a wide range. The sound source does not have to be between the two loudspeakers – it can be above or even behind the listener.

A mathematical analysis soon shows that two loudspeakers can produce any two desired signals at two ears. It is important, however, to understand the physical process involved. This is best explained by first assuming that the two signals that we wish to reproduce are simply a sound pulse of short duration at the left ear of the listener and a zero signal at the listener's right ear.

For simplicity, we can also assume that the listener's head is transparent to sound and that we can represent the two loudspeakers as 'point' sources that radiate spherical sound waves. An exact analysis of the physics involved (Figure 1) shows that we need to cancel sounds from the left-hand speaker so that they do not reach the right ear, and vice versa.

It turns out that this process – known as crosstalk cancellation – becomes difficult at the frequency where the difference in the path length between the two loudspeakers and one of the listener's ears is half the acoustic wavelength. At this frequency, the system also becomes much more prone to error – in the listener's head position, for example. We call this the 'ringing frequency'.

When loudspeakers subtend an angle at the listener of 60 degrees, the ringing frequency is typically in the region of 2 kHz. However, move the loudspeakers closer together and the angle between them reduces, which increases this troublesome frequency. At 10 degrees, the ringing frequency rises to about 11 kHz.

Crosstalk cancellation (see Figure 1)

Digital signal processing can cancel cross-talk between loudspeakers.

The sequence of pictures depicts the sound field generated in reproducing the desired signals.

First, the left loudspeaker emits the sound pulse (white) that is desired at the left ear of the listener. This pulse produces a positive sound pressure that is above atmospheric pressure (grey).

This positive pulse arrives at the left ear after traveling at the speed of sound, and with diminishing amplitude as the sound spreads outwards. The objective is thus achieved as far as the left ear is concerned, but the pulse continues to spread outwards and soon arrives at the listener's right ear.

We desire a zero signal at the right ear. This is achieved by cancelling the first pulse with a negative pulse (black) from the left loudspeaker. That is, we cancel the 'crosstalk' between the left loudspeaker and right ear. Interference between the positive and negative pulses maintains zero sound pressure at the right ear of the listener.

Now there is another problem. The negative (black) pulse continues to spread outwards and arrives at the left ear a short time later. This pulse must therefore be cancelled at the left ear by a second positive (white) pulse from the left loudspeaker. This pulse is, in turn, cancelled at the right ear by a second positive (black) pulse from the right loudspeaker, and so on.

The human ear becomes less sensitive at these higher frequencies. Many sounds of practical interest are concentrated at much lower frequencies. Moving the loudspeakers closer together can push the ringing frequency right out of the audio frequency range, that is above 20 kHz. With infinitesimal spacing between the sources, the ideal sound field for producing the desired signals comes from the combination of a 'monopole' and 'dipole' source.

The sources produce the desired result – of a pulse at the left ear and zero at the right ear – with a single wavefront that contains a zero pressure that coincides with the right ear. This appears very attractive. However, the bad news is that when they are pushed infinitesimally close together, the cones of the two loudspeakers are required to have infinite acceleration.

The 'stereo dipole' offers a compromise. It uses loudspeakers that are close enough to have a high ringing frequency but not so close together to require unreasonably large outputs. As the ringing frequency increases, it also becomes more difficult to achieve crosstalk

cancellation at lower frequencies. However, we have found that that a 10-degree angle between loudspeakers gives a very useful frequency bandwidth, with highly effective cancellation of crosstalk. This arrangement meets the objective very well, with the ringing frequency only just detectable in the sound field.

Virtual images

The stereo dipole is extremely effective in producing the illusion of a virtual source of sound (Figure 2). A comprehensive study of the psychoacoustical properties of the stereo dipole, comparing it with other virtual imaging systems that rely on using two loudspeakers with a much wider spacing, showed that it has many advantages. These include a greater robustness to movements in the listener's head and a larger 'sweet spot', or range of head positions which maintain the illusion. The stereo dipole is especially effective in generating the correct time difference between the ear signals, a cue that is particularly important to the human localisation mechanism.

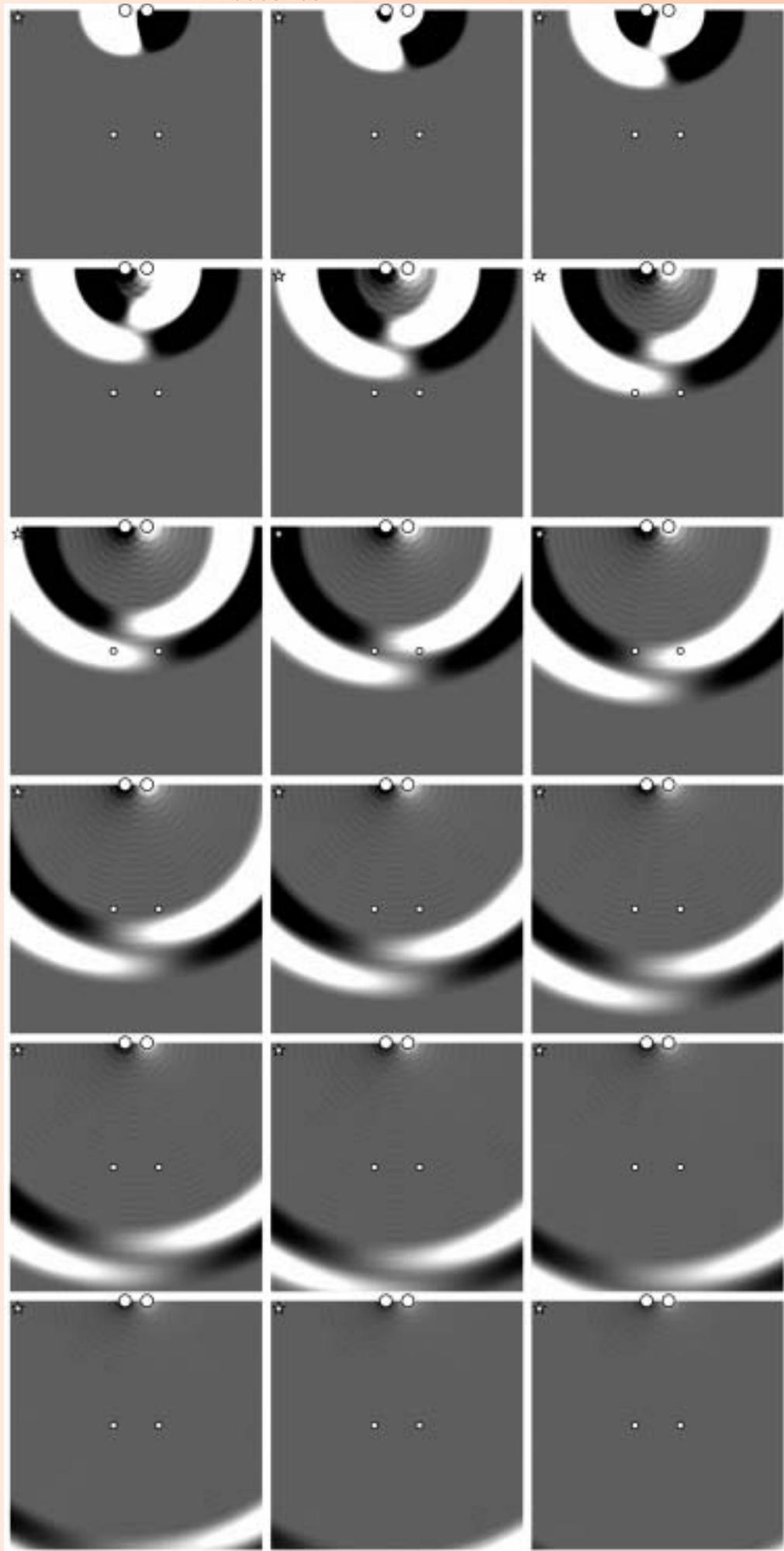
Figure 2 The stereo dipole generates a virtual sound source
 We assume that the desired signals at the listener's ears are from a virtual source in the same (horizontal) plane as the two loudspeakers (illustrated by a star). We also assume that the virtual source emits the signal emitted as a simple positive short sound pulse. To produce the correct illusion of the existence of the virtual source in the position shown, a pulse must first reach the left ear. Then, a short time later, the same pulse (but of slightly lower amplitude) must reach the right ear. The stereo dipole does this by radiating two wavefronts in quick succession. The first pulse produces the desired pulse at the left ear and the second produces the desired pulse at the right ear.

A computational model of how the sound field from a stereo dipole interacts with the listener's head allows us to study the extent to which the arrangement produces the correct level differences (Figure 3). Subjective experiments (Figure 4) show that, if we know the detailed shape of the listener's pinnae, the parts of the outer ears that focus sounds on the inner ear, we can produce virtual sources not just in the front half of the horizontal plane of the listener, but also to the rear. However, we seldom know the details of how a listener's head responds to incoming sound – the listener's 'head related transfer function' – so this has to be estimated for a wide range of listeners. This estimate is then used in designing the signals processing prior to transmission by the loudspeakers.

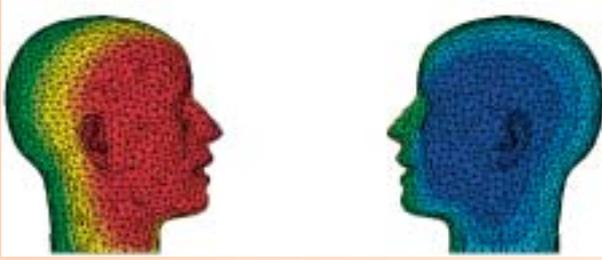
Digital signal processing.

Getting the signals 'just right' at the listener's ears is the secret of success in generating virtual sound images. In practice, this requires very careful processing of the virtual source signal before sending it to the loudspeakers.

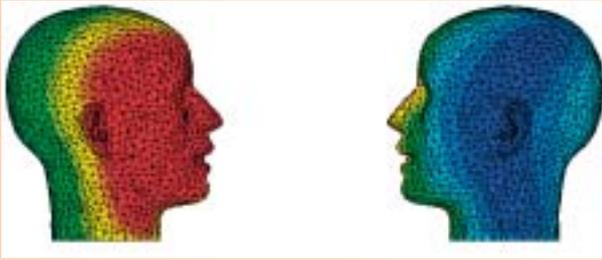
The signals are digitally filtered before transmission. This involves taking the recorded signal and sampling



200 Hz



1 kHz



3 kHz



5 kHz

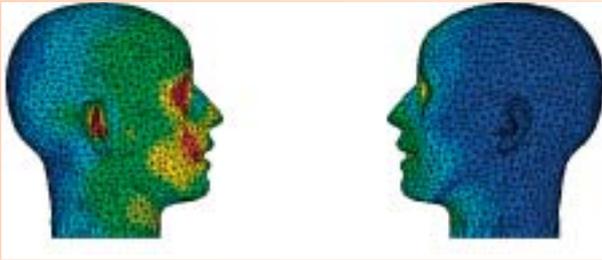


Figure 3 How a stereo dipole interacts with the head of a listener

The sound pressure generated over the head of a listener is illustrated as a function of frequency when the dipole creates a unit value (red) at one ear while producing a zero value (blue) at the other ear. The correct level difference is always produced in this simulation, but the desired unit value of pressure at one ear clearly becomes more localised as the frequency increases.

it, for example, 48,000 times a second to produce a sequence of numbers that represent the source signal. Typically, the first 256 of these numbers are then multiplied by a pre-determined set of 256 'filter coefficients' and the results added to give the first number representing the signal to one of the loudspeakers. The signal transmitted to the other loudspeaker is similarly calculated, but with a different set of 256 filter coefficients.

The next sample representing the loudspeaker signal is then calculated by

omitting the first sample from the source signal and multiplying the next 256 samples – that is sample numbers 2 to 257 – by the filter coefficients and adding the results. The other loudspeaker signal is similarly calculated. This whole process repeats at every sample. A modern DSP chip can accomplish this process in real time. Thus a total of 512 multiplications and additions are undertaken 48,000 times a second to process the signal from one virtual source. This is a remarkable achievement of modern electronics, but



Figure 4 Subjective experiments on sound perception

The cage surrounding the listener is in an anechoic chamber covered with black visually opaque but acoustically transparent cloth. Subjects are asked to judge the location of both real sound sources and virtual sound images generated by the virtual sound imaging system.

one that now allows us to contemplate such approaches in sound reproduction.

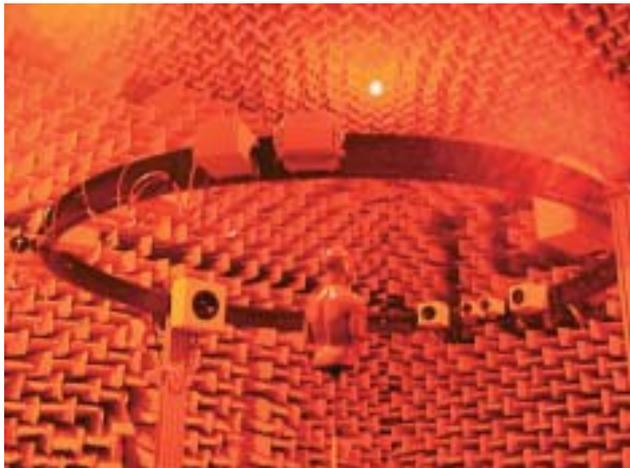
The trick of getting the signals right at the listener's ears relies on the correct choice of filter coefficients. This filter design has enabled the implementation of the system in practice. This is not as straightforward as it might at first appear, but the process is now thoroughly understood as a result of extensive research.

Virtual surround sound

The modern trend in sound reproduction is to use not only two loudspeakers but typically five loudspeakers, three in the front of the listener and two to the rear. This arrangement is often supplemented with a 'sub-woofer' radiating very low frequency sound. Such formats, derived from cinema sound tracks, are finding their way into domestic use. DVDs are likely to increase the use of 'surround-sound' systems. DVDs store vast amounts of both video and audio data and can record the necessary five or six audio tracks.

Virtual sound imaging also has a role to play in this trend. It can treat the five signals intended for the surround-sound loudspeakers as if they were virtual-source signals. Thus, digitally filtering the signals appropriately can generate

Figure 5
An array of six loudspeakers, under test in the anechoic chamber of the Institute of Sound and Vibration Research, with each pair handling different sound frequencies. Such an arrangement comes close to the holy grail of recreating the acoustics of a concert hall at the listener's ears.



the illusion for the listener of the existence of five virtual loudspeakers. A single compact unit containing just two loudspeakers, placed conveniently to the front of the listener, can generate the impression of a real surround-sound system.

Of course, the real-time signal processing becomes even more demanding with a wider sound stage. It requires 2560 multiplications and additions 48,000 times a second. However, modern DSP chips can accomplish this and a listener can enjoy the illusion of almost complete surround sound without the collection of wires and loudspeakers associated with a real surround-sound system. However, the approach is limited to a single listener, although research is investigating ways to lift this restriction.

Further developments

Recent work demonstrates that the stereo dipole can enhance the effective frequency range of a virtual sound imaging system. As we have already explained, the low-frequency output of the stereo dipole is limited by the need to drive the loudspeakers with high amplitudes. The system, therefore, operates at a limited, albeit very useful, range between the low frequency end of the spectrum and the ringing frequency. This limited frequency range can be overcome with a number of pairs of loudspeakers making different angles with the head of the listener.

To replicate the desired signals at the listener's ears most effectively, loudspeakers need to be continuously distributed at an angle relative to the listener. Loudspeakers generating the highest frequencies should be very close together to give a small angle relative to the listener. Conversely, loudspeakers generating the lowest frequency sound should be at the maximum angle relative to the listener. The latter is, of course, 180 degrees, with loudspeakers on either side of the listener.

Ideally, between these two extremes, there should be a continuous distribution of acoustic sources, with frequency content varying continuously from high to low as the angle subtended at the listener increases. Until we can devise such an ideal 'optimal source distribution', recent approaches use of three pairs of transducers. One pair is at a 6-degree angle, a pair at 30 degrees and a pair at 180 degrees (Figure 5). This system has proved spectacularly successful in generating 'virtual acoustic reality' and has come extremely close to realising the long-sought ability to recreate the sound of the concert hall at a listener's ears.

One of the main factors influencing the success of the system is the replication of the correct time difference between the signals produced at the ears of a listener by a source of sound in a given position. The technique can be combined with modern multi-channel approaches to sound

reproduction to generate 'virtual surround sound'. With multiple pairs of loudspeakers, we can extend the principles of the stereo dipole to ensure accurate reproduction of signals at the ears of a listener over a very broad range of frequencies. Other applications of these principles are sure to follow. ■

Notes and acknowledgements

Research into virtual acoustics at the Institute of Sound and Vibration Research and Tokyo Denki University has been sponsored by Yamaha Corporation, Hitachi Ltd, Alpine Electronics, Nittobo Acoustic Engineering, Kajima Corporation and the UK Engineering and Physical Sciences Research Council.

The virtual sound imaging technology described in this paper is the subject of a number of granted international patents and patent applications.

*<http://www.isvr.soton.ac.uk>
<http://www.dimagic.co.jp>*

Philip Nelson is Professor of Acoustics and Director of the Institute of Sound and Vibration Research at the University of Southampton. He has particular research interests in the active control of sound and vibration, multi-channel signal processing, inverse problems in acoustics and the generation and control of aerodynamic noise. He has been awarded the Tyndall and Rayleigh medals of the Institute of Acoustics and was elected a Fellow of the Royal Academy of Engineering in 2002. He is also a Fellow of the Institution of Mechanical Engineers, the Institute of Acoustics and the Acoustical Society of America.



Hareo Hamada is Professor of Electrical Engineering at Tokyo Denki University and is CEO of DiMagic Co. Ltd, Tokyo, Japan.

