Efforts to improve upon the realism of sound reproduction obtainable from stereo records and broadcasting led up a blind alley with "quadraphonics" but appear to have succeeded with "ambisonics", which relies heavily on electronics to give the listener a convincing sensation of being surrounded by sound. From small beginnings ten years ago, the technology is being increasingly adopted by the audio industry, with ambisonic material being broadcast by the BBC and in Europe, and with several companies issuing ambisonic records.

Ambisonics

by P. B. Fellgett, professor of cybernetics of the University of Reading.

Ever since the days of Alexander Graham Bell's telephone and Thomas Edison's cylinder recording, the audio industry has tried to give its listeners a greater sense of naturalness, intelligibility and realism. For a long time there was a preoccupation with improving frequency-response and reducing non-linear distortion, and it seldom seemed to occur to anybody that part of the naturalness of live sound lay in its directionality. An early exception to this generalisation occurred at the Paris Electrical Exposition of 1881, when sound was relayed from a number of carbon microphones on the stage of the Opéra to ear pieces in the exhibition hall. In those days before the development of electronic amplification there was no alternative to feeding each ear-piece from a separate microphone, and some visitors noticed that they got directional effects when they held one ear-piece to each ear.

This seems to be the earliest recognition of what we should now call stereo, but it was not until around 1930 that the basic principles of stereo reproduction were systematically investigated, and not until a third of a century later did stereo records and broadcasts become a regular part of the audio market.

Stereo gives directional information essentially over a front sector bounded by the two loudspeakers. This is very artificial, because of course live sounds come to us from all around. Moreover the two loudspeakers must not be more than 60° apart as seen by the listener, and indeed many listeners prefer to keep the angle much smaller. We now know that this is because stereo presents the ear with conflicting directional clues, the discrepancy between which is negligible for small angles but rapidly becomes larger as the angle is increased. Indeed it appears that a learning process, which most of us have forgotten, is needed before our ears fuse the sounds from the two speakers into a stereo image. Some people are apparently unable to do this; this need not make them feel inferior, on the contrary it means that their ears are particularly sensitive to the artificialities of stereo.

It is an obvious idea to try to extend the sensation of direction to all around the listener. Unfortunately the earlier attempts to do this, usually known by the hybrid name of "quadraphonic", were insufficiently well conceived from the start, and have proved to be a blind alley. Although not necessarily formulated in such an explicit way, the initial idea seems to have been to feed separate and independent signals to four loudspeakers disposed in a square around the listener, and this gives rise to the name "discrete quadraphony". However it was soon realised that this was no more than the four-channel analogue of the kind of pseudo-stereo of some early records which was really dual-mono, and gave rise to the description "ping pong".

The entertainment possibilities of "ping-pang-pung-pong" reproduction were clearly too limited to be of commercial interest. So the next idea was to use the speakers as four stereo pairs surrounding the listener. In the front sector, this means 90° stereo which is well known to give a hole-in-the-middle sensation and cause the images to disappear into one speaker or the other with slight head movements or changes of balance. By standing in a suitable position in front of a stereo pair of loudspeakers and turning around, it is easy to verify that the effect is worse when the loudspeakers are behind the listener, and hardly works at all when they are at the side. Thus four-wall stereo simply does not work well enough.

This "quadraphonic" episode has left behind some unfortunate legacies. Because there was no way in which it could be designed to work properly, different manufacturers made arbitrary and mutually incompatible choices. This combined with poor performance, made it an expensive failure in the marketplace, and this has left behind a prejudice against all surround reproduction among those who do not realise that "quadraphony" is not synonymous with surround reproduction but merely one way of attempting to realise it, and a poor way at that.

These earlier attempts can now be seen as premature, because there was not at the time the knowledge available to solve the essential problems. In particular, suppose that a number of loudspeakers are disposed about a listener. What signals should be fed to each of the loudspeakers so as to give the listener the illusion of sound coming from a particular direction and from a particular distance?

The solution is trivial for a sound intended to come from the direction and distance of any one of the loudspeakers; simply feed the signal to this loudspeaker alone. In all other cases, the problem is more difficult, especially since (as we have seen) ordinary stereo blending is not sufficient. To solve the problem, we need to know more about human hearing.

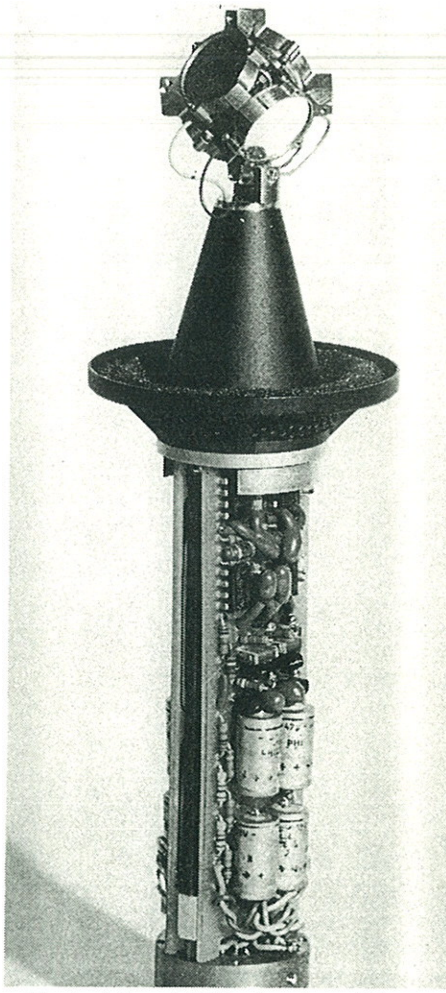

Fig. 1: A Sound Field microphone (Photo: courtesy of Calrec Audio Ltd.)

The literature of perception psychology includes many references to the properties of hearing, including the perception of direction, but the wish to make experiments scientifically controlled has often led to them being conducted under conditions that are too artificial or specialised to be directly helpful to our problem. Moreover the systematic relationship of one experimental result to another may be obscure, and we need a general theoretical framework into which to fit the individual observations. It is as if we possessed a large number of leaves, but did not know how to make a tree.

A crucial step in solving this problem was taken by the Oxford mathematician Michael Gerzon. He turned the question around and instead of asking what information the ear does use, posed the devastatingly simple question of what information the ear could in principle use. Putting the question this way round immediately frees us from having to consider detailed mechanisms within the ear and brain, and

▶ 44

81 MAY

38 ◀
allows us to concentrate on what clues are physically available.

Fortunately there are well-established techniques for describing the information that could be gleaned from the soundfield. These involve describing the available information in terms of a mathematical series, the first term of which would give a first approximation, addition of the second term would improve upon this, and so on. In fact, a relevant expansion is in a double series involving spherical harmonics of the directionality of the soundfield, and the type of processing that the ear could apply to it.

The terms of a Fourier series used to describe a repetitive waveform can appropriately be called circular harmonics. Spherical harmonics are a generalisation from the one-dimensional space of repetitive time to the two dimensions of directionality expressed as the two angles of polar coordinates, for example the altitude and azimuth of the navigator

Summation

The directionality of any soundfield can be expressed with increasingly good approximation by summation of successive spherical harmonics. The spherical harmonic of 0^{th} order corresponds to an omni-directional response. There are three spherical harmonics of 1^{st} order, and they may be pictured as three figure-of-eight responses pointing in the three possible orthogonal directions. There are a further five spherical harmonics, of 2^{nd} order.

It will be seen that so far the total number of spherical harmonics up to and including order n is $(n + 1)^2$, namely 1, 4, 9 respectively for $n = 0, 1, 2$, and indeed this relationship holds generally. The polar responses of all ordinary microphones, omni-directional, figure-of-eight, cardioid, and hyper-cardioid, can be represented by suitably chosen proportions of the spherical harmonics of orders 0 and 1.

The simplest kind of processing the ear could perform may be described as linear. This in fact corresponds to the laws of hearing formulated by the pioneers Ohm, Helmholtz and Lord Rayleigh. In this description the relative phases of different components of a signal are unimportant, in accordance with what is called Ohm's Law of hearing. The next simplest kind of processing is quadratic, and here relative phase does become important. This may be expressed by saying that the ear has access to bispectral information as well as to the more familiar spectral information of linear processing.

This classification of the information available to the ear provides the necessary meta-theory, that is to say, a theory of theories, which can relate the previously isolated experimental results and the more limited theoretical descriptions previously derived from them. In this way, the theory can be expressed in a form which is directly useful for engineering design. This has been a crucial factor in the development of Ambisonics.

The Gerzon theory has enabled the small ambisonics team, supported over the past decade by the National Research Development Corporation, to go straight to the building of workable prototypes, and to take large steps in their subsequent development and refinement. The team has thus been able to proceed much faster, and much more cheaply, than was possible for the often much larger teams within audio manufacturing companies which earlier had to proceed largely empirically.

The most immediate application of this psychoacoustic theory is to ambisonic decoders. In accordance with what was said above, the job of a decoder is to provide loudspeaker feeds such that the sounds radiated by the individual speakers will appear to the listener to fuse into the impression of a sound coming from the intended direction. Contrary to "quadraphonic" assumptions, valid loudspeaker feeds cannot exist at any earlier stage in the reproduction chain, because the number and layout of the listener's loudspeakers is unknown. Any reasonable number of loudspeakers could in principle be used, and ambisonic decoders have been built giving a choice of either four or six loudspeakers in various layouts. Distance and layout controls enable the decoder to be matched to the particular layout in use.

The use of more than the minimum of four loudspeakers is not necessarily more expensive, since it may be possible to use smaller and cheaper individual speakers, and may be domestically more convenient. To obtain the best directional illusion in general requires cooperative radiation from every available loudspeaker, again contrary to the false "quadraphonic" goal of "separation" which merely calls attention to individual loudspeakers as sources of sound, instead of giving the desired illusion of reproducing the direction of an original sound no matter how the loudspeakers are disposed.

According to our present understanding, the ear does not use just one mechanism of localisation but a number of different methods. If these different mechanisms can be supplied with clues which agree with each other, the illusion will seem natural and satisfying, while if they differ the result will be poorly localised and fatiguing. Decoder design aims therefore to satisfy simultaneously and consistently as many as possible of the mechanisms of localisation used by the ear.

The extent to which this can be done depends on the number of available loudspeakers and transmission channels, and on other circumstances, in a way which has been described in detail in the Gerzon theory. This theory takes account of non-central as well as central listening positions, and it is found in practice that the satisfactory listening-area is comparable with but larger than that for stereo; indeed in a small domestic room it may cover most of the living area.

One specific result of the theory may serve as an illustration. At low frequencies, the ear uses a combination of mechanisms which collectively are usually known as Makita localisation. At medium-high frequencies, localisation corresponds to what is called the energy-vector theory. The gradual changeover from Makita to energy-vector localisation takes place over the region where the ear-to-ear spacing is comparable with the wave-length of sound, and it is very disturbing if the two localisations conflict.

Gerzon has shown that they will in fact agree if, and only if, the loudspeakers are disposed in pairs symmetrically about a central point, and the sum of the signals fed to the two loudspeakers of any pair must be the same as for any other pair. This redundancy relationship speaker-feeds means that the optimum number of independent signals to feed four loudspeakers disposed in a square or rectangle, is actually three, and it will come as no surprise that this is at variance with the "quadraphonic" assumption.

The decoders mentioned so far have been concerned only with horizontal surround reproduction, but of course in nature sounds can come from a whole sphere of possible directions. Spherical surround reproduction, known as periphony, is also included in the ambisonic technology. It may be shown that the minimum number of loudspeakers that can be used is six, disposed at the corners of a regular octahedron. There are two possible layouts using eight loudspeakers, cuboid or bi-rectangular, each of which has some advantages of its own. Four transmission channels are optimally needed to feed these periphonic arrays.

Generation and transmission

However well designed a decoder may be, it cannot function until fed with suitable input signals which must be generated from the original programme material and then transmitted to the decoder. These signals must be what is called *well decodable*, that is to say, subject to practical constraints they must contain the necessary information in the right form for it to be possible for a linear decoder to generate psychoacoustically accurate loudspeaker-feed signals.

The starting point is the direction of a sound, and all Ambisonic signal formats conform to kernel specifications. This means that the format is defined as a smooth function of the direction that is encoded. This is in contrast to "quadraphonic" formats which often change discontinuously at corner positions, and are specified in terms of ill-defined signals associated with these corners.

The ambisonic format recommended for studio processing and recording is known as B-format. It consists of three signals for horizontal surround, or four for periphony,

▶ 46

44 ◀
representing the four spherical harmonics of orders 0 and 1. The relative amplitudes are chosen so that each signal would carry the same energy in the representation of a homogeneous random sound field; thus the probability of overload is approximately equalised in each channel.

Microphone

The soundfield microphone, invented by Michael Gerzon and Dr. Peter Craven and developed by collaboration between Calrec Audio Ltd. and the Department of Cybernetics of the University of Reading, particularly through the work of Dr. G. Barton, provides a means of transducing live soundfields directly to Ambisonic B-format. It consists of a cluster of individual microphone capsules placed very close together, followed by appropriate mixing and phase-amplitude equalising circuitry, the whole design being based on the theory of sampling on the surface of a sphere.

The minimum number of capsules are never used separately, but only after combination into the B-format signals. These signals can be further combined to give any of the first order microphone responses, from figure-of-eight through hyper-cardioid and cardioid to omni-directional, pointing in any direction horizontally or vertically.

By doing this more than once, it is thus possible to synthesise a truly coincident stereo pair, with any desired polar response and angle of fire. This can be done, if desired, on recorded B-format signals after the sessions are completed. The soundfield microphone has thus opened up new possibilities in stereo recording, in addition to its Ambisonic uses. Other microphone arrays can be used to generate B-format signals with varying degrees of approximation or compromise.

There is clearly no difficulty in generating B-format signals synthetically through an Ambisonic version of the pan-pot, and indeed Ambisonic reverberation, spread, delocalisation and other synthetic facilities are readily available for those who may wish to use them, including some effects which cannot be realised or even defined in less advanced technologies. All these devices can be properly engineered because of the precise specification of Ambisonic signal formats.

For transmission of the information to the eventual listener, either by broadcasting or recordings, Ambisonics provides a set of signal specifications known as UHJ. The set begins with a two-channel version BHJ, this clearly being the minimum number of channels that can give surround reproduction. Two-channel UHJ can be distributed through any of the existing stereo media without the need for extra multiplexing or other complications.

This BHJ format has been chosen so as to be mono- and stereo- compatible within limitations set largely by the mutual incompatibility in definition between mono and stereo themselves. Indeed it turns out that conventional stereo signal formats are not themselves stereo-compatible over anything but a frontal sound-stage! The attempt to design a truly stereo-compatible form of stereo leads indeed to something very close to BHJ, even apart from the possibility of surround decoding. This coincidence of surround and stereo optimisation has been confirmed by a number of excellent reviews of Ambisonic records that have been published, often from reviewers who had not noticed that they were listening to anything but conventional stereo.

The use of only two channels does however impose some compromises, and improved surround decoding can be provided by adding a third channel, giving the THJ member of the UHJ set. This is the recommended "no compromise" form for horizontal surround. If a fourth channel is

▶48

46 ◀
available, it may be used to include the height information for periphonic reproduction, giving the PHJ member of the set.

A useful compromise between two-channel and three-channel working is to use a third channel of restricted amplitude or band-width. Although much easier to implement, for example on carrier-disc or by additional multiplexing in f.m. stereo broadcasting, this is found to give results which are subjectively quite close to full three-channel working. This is the so-called two-and-a-half-channel member of the UHJ set, designated SHJ.

All members of the UHJ set are directly inter-compatible. This means that one goes from one member of the set to another simply by adding or deleting additional signals, without any change to those that remain. This gives great freedom for example to broadcasters to begin immediately with two-channel transmissions, and add further whole or partial channels at a later stage. It also means that Ambisonic technology is able to take full advantage of the information capacity of the various multi-channel digital or other media which are at present vying for a share of the marketplace. Since in addition to fulfilling this inter-compatibility requirement, all members of the set have to be well decodable, and the base-band two-channel signals have also to be mono and stereo compatible, the design of the UHJ set presented some formidable mathematical problems the solution of which is again due to Michael Gerzon.

Fig. 2: A Sound Field control unit (Photo: courtesy of Calrec Audio Ltd.)

To judge by recent engineering publications, the Ambisonic technology has now essentially been adopted by the BBC for surround broadcasting. Broadcasts in Ambisonic format have been made by the IBA, and by NOS in the Netherlands. A number of record companies have begun to issue ambisonic records, and ambisonic decoders are available from more than one manufacturer.

The practical experience of the ambisonic team now goes back for approximately a decade, beginning with extremely primitive experiments and building up to state-of-art recording sessions with distinguished musicians at the present time. In all these ways, much information has been built up about the many problems which in audio go beyond the essential basis of good theory and good engineering implementation. Both from the research point of view, and industrially, the foundations of ambisonics have been well tested, and it seems to be founded on the rock of firm fundamental principles.