General Metatheory of Auditory Localisation

[4PS2.01]
Preprint 3306

GERZON, Michael A.;
Technical Consultant, Oxford, United Kingdom

Presented at the 92nd Convention 1992 March 24–27 Vienna

AN AUDIO ENGINEERING SOCIETY PREPRINT

General Metatheory of Auditory Localisation

Michael A. Gerzon

Technical Consultant, 57 Juxon St., Oxford OX2 6DJ, U.K.

Abstract

This paper presents a general metatheory (theory of theories) of directional sound localisation suited to the design of directional sound reproduction systems using loudspeakers. It is shown that any theory of localisation can be expressed as a composite of "primitive" component theories based on three hierarchies: degree of nonlinearity, order of spherical-harmonic directionality, and degree of use of head movement. These component primitive theories are mathematically tractable for design purposes. An appendix illustrates applications of the metatheory to previous models for stereo localisation.

0. Author's Note

This paper on the basic theory of auditory localisation aimed at designing directional loudspeaker reproduction systems was originally written in 1976/7, under the cumbersome title "The Rational Systematic Design of Surround Sound Recording and Reproduction Systems. Part I. General Theory of Directional Psychoacoustics and Applications".

This paper was circulated (with slight editing of the diametric decoder theorem, which was the subject of U.K. Patent 2,073,556 filed in 1980) at that time to the FCC and EBU as part of technical submissions by the British N.R.D.C. on technical standardisation proceedings on "quadraphonic" systems. Although copies have also been circulated to individuals, the paper has never been available in the open literature.

The original paper was the first part of a massive three-part paper which described the detailed design procedure of Ambisonic encoding and decoding systems, and many of the results of this paper have been applied to commercial ambisonic decoders. Although some of the detailed technical references in the paper may be somewhat "dated", the contents of the paper remain relevant to the design of directional sound reproduction systems, and not merely to those that are "ambisonic". References [41] to [43] are recent examples of technical papers that have made use of the results of this paper, and it is its continuing relevance to contemporary work that has prompted the formal publication of this work.

Apart from changing the ending of the "Conclusions" section to make it relevant to current work, the paper is exactly as written in the 1970's, since I feel that attempts to "update" the paper would not significantly improve on the original presentation.

1. INTRODUCTION

Hitherto, the design of surround sound systems has been a "black art" rather than a systematic procedure. Generally (and the author does not exclude his own early work), the procedure has been to juggle with the mathematical patterns of speaker feeds until they satisfy some pattern considered nice or plausible by the designer, and possibly to back up this by poorly-defined "psychoacoustic" rules-of-thumb that had little general predictive value, and often based on either little hard experimental evidence, or on an extreme extrapolation of results obtained under a limited set of test conditions. Of these heuristic system designs, possibly the most successful [1] (at least in surround reproduction performance) was the UMX system of Cooper and Shiga [2], which was based on requiring a rotational invarience of mathematical properties, and backed up by Makita's localization criterion which sometimes (but not always) gives moderately good predictions.

A number of papers have appeared, of various degrees of sophistication [3],[4],[5],[6],[33] attempting to explain various aspects of directional psychoacoustics as it pertains to surround sound reproduction. The present paper lays the foundations required for a systematic design of complete surround sound systems. Since the aim of such systems is (or should be) to produce a reliable and convincing illusion to domestic listeners of the intended encoded directional effect [7], it is essential to begin such designs with a reliable and mathematically tractable theory of human directional psychoacoustics. It is evident that until one knows what information needs to be presented at the listener's ears, no rational system design can proceed. (Historically, it is interesting that modern 2-speaker stereo was designed backwards from the ears of the listener by Blumlein in 1931 [8]).

This paper presents for this purpose a novel approach to directional psychoacoustics. Much of the philosophy and theory was described in nonmathematical form in a previous paper [6] of the author, and the mathematically less sophisticated reader will find this an ideal preparation for this paper. Essentially, the novelty of this work lies in three important features:

(i) It is assumed that the ears have no single method of localizing sounds, but that many different methods are used. In the case that not all methods give the same results, it is supposed that the ear takes some sort of "majority decision", except when a complete conflict of cues is heard when the localization will also be confused.

(ii) A "metatheory" (i.e. theory of theories) of directional psychoacoustics is developed that in principle allows arbitrary complex methods of sound localization to be expressed in terms of a heirarchy of relatively simple "primitive" theories, rather as in applied mathematics one might approximate a complicated function by a sum of much simpler polynomial terms. It is not suggested that the "primitive" theories themselves necessarily describe the way or ways the ears localize sound, but that if the various requirements for accurate localization demanded by "primitive" theories are satisfied, then the likelihood of correct localization by the ears' actual localization mechanisms is greatly increased. In general, the more "primitive" requirements are satisfied, the more reliable will be the localization heard, and the degree to which various "primitive" theories fail to be satisfied may be used to describe various qualities of faulty localization.

(iii) A great simplification in the theory is obtained by treating the ears as a "black box" responding to an incident sound field, and by not attempting to describe the internal mechanisms in the ears and brain

responsible for the black box behaving in the way it does. For engineering purposes, it is clearly sufficient to know how the ears respond, and not necessary to know why. In fact, a number of the "low degree low order" primitive models described in this paper may be filled out with great detail involving computing acoustic waveform arrivals at the ears, as is done by Clark Dutton and Vanderlyn [9], Makita [10] and Tager [11], among others [5],[12],[13]. However, a comparison of the computations in [5],[9]-[11] with the corresponding black box theory of this paper shows that the same results can be deduced with very much simpler mathematics by the process of ignoring the inner workings of the ears. While the resultant theory looks much more "abstract", it is much more amenable to complicated design calculations and more accessible to intuition because of its fundamental simplicity. Some physical idea of how this black box approach works is described in [6] with especial reference to its figure 1.

We lay emphasis on theory being "simple" and "tractable" (which is not the same as "elementary" or even "easy"), since our aim is more ambitious than merely to be able to analyse an already-designed encode/decode system. The design of encoders and decoders involves tens or hundreds of parameters, all of which can be varied. There is clearly little chance of ending up with a near-optimal design (especially if the number of criteria of goodness used is also large) if one relied on guesswork in choosing designs. Even with large-scale computing facilities, it is virtually impossible to evaluate and optimise systems with more than about 7 or 8 free parameters. Therefore, to do better than this, the mathematics of the psychoacoustic theory must be of such a form that there are general mathematical results (Theorems) that permit a drastic reduction in the number of parameters that need to be

considered. The "diametric decoder theorem" later in this paper is an example of a result that reduces the number of parameters to be considered. (That theorem describes mathematical relationships between speaker feed signals that ensure automatically that more than one localization criterion is satisfied). A more extreme example of reduction of parameters is discussed in part III of this series of papers, whereby we shall show how to optimise the whole encode/decode system by reducing the problem to a 4-parameter problem.

Thus, precisely because the design of decoders would otherwise be too hard a problem to solve, the style of these papers is mathematical. Readers who only wish to understand the physical ideas behind this mathematical approach are referred to [6]. However, we do make some attempt at not burdening the reader with more formalism than he needs for engineering design tasks. An exception is the section below dealing with the heirarchical metatheory of directional psychoacoustics. This is because we owe the reader some general background to explain what would otherwise seem an arbitrary choice of models of directional hearing.

No claim is made that theory described in this paper is good for all possible applications. Without considerable extension or modification, the theory is unsuitable for dealing with ambient sounds and with the effects of inter-speaker time delays. Such aspects will be considered in part in subsequent papers.

It is an essential feature of the psychoacoustic metatheory of this paper that it includes as special cases the theories of auditory localization pursued by a large number of previous authors, such as [3]-[5],[8]-[16]. In Appendix I, we describe briefly how previous theories may be incorporated into the present one, and discuss applications to 2-speaker stereo localization.

The theory in this paper is presented in a form that may be applied either to horizontal-only or with-height reproduction. Although the presentation would have been somewhat simplified by not including the with-height case, we believe that the methods described in this paper and in subsequent parts makes with-height (periphonic [17]) reproduction of sound an entirely practical proposition, and there seemed to be a strong case for advancing the sound reproduction art further by making this information available.

Some difficulties of reading this paper arise because its conceptual framework is not entirely conventional (so that explanations of concepts require careful thought of the reader), and yet has to be expressed in somewhat complex mathematical form in order to do computations. The notations used have not been made so abstract that the statement and proof of results becomes very short but difficult to 'see through' physically, but involve sufficient abstraction not to make all the results totally unwieldy. Some use (explained in the text) has been made of simple tensor notations where this is of some assistance, notably in the so-called "cross-bispectral" models.

The first-time reader is advised to skim through to get the 'gist' rather than to get bogged down in details that he might not need for his purposes. In particular, of all the classes of models considered, the velocity and energy models are the most important, and others may be omitted on first reading. The writing-out of the various equations for familiar decoders (e.g. BMX and TMX decoders [2]) is a useful aid to understanding the significance of the mathematics given, and is recommended.

2. CONVENTIONS

It is convenient to describe some general aspects of notation and some general assumptions adopted in this paper, for ease of reference.

We set up (x,y,z) axes in space that are rectangular cartesian coordinates centred upon the listener, with the x-axis pointing forward, the y-axis leftward and the z-axis upward. Thus, for example, the horizontal plane is the x-y plane, and a loudspeaker placed at azimuth ϕ measured anticlockwise from due front (the x-axis) at a distance d from the listener has coordinates $(x,y) = (d\cos\phi, d\sin\phi)$. All azimuth angles are measured anticlockwise from due front, and symbols involving ϕ will represent the azimuths of reproducing speakers, whereas those involving θ will represent the apparent azimuths of recorded sounds.

It is assumed in this paper (Part I only) that all loudspeakers are placed at an identical distance d from the listener, so that identical sounds emitted from all speakers reach the listener at the same time and with the same amplitude. Unless otherwise specifically stated, it is assumed that the distance d is large, so that the wavefronts from the loudspeakers arrive at the listener in the form of a plane wave. Although such a restriction is not fundamental, it simplifies the theory to consider a sound field as resulting solely from a finite number of infinitely distant point sources. As we shall see, it is possible to remove this restriction by subsequent simple modifications, and this order of doing things is much easier than the opposite one of starting with the greatest generality and then restricting to special cases.

Speaker feed signals, and signals from which they are matrixed, are indicated by sans-serif letters, which are usually capitals, and the corresponding signal gains for individual encoded sounds are

indicated by the same letters in ordinary type-face. Thus LB,LF,RF,RB represent signals fed to left-back, left-front, right-front and right-back speakers, and LB,LF,RF,RB represent the corresponding (generally complex) signal gains. The symbol $j=\sqrt{-1}$ is used to indicate a $+90^0$ relative phase shift (Hilbert transform) when applied to signals and to indicate $\sqrt{-1}$ when applied to signal gains, as usual. The letters X,Y,Z or x,y,z are used to indicate signals associated with the components of sound field velocity in the respective directions of the x,y and z-axes, and in particular we let x,y,z indicate signals with sounds encoded with respective gains x,y,z where $x^2+y^2+z^2=1$ and (x,y,z) is a vector pointing towards the intended direction of the encoded sound. Thus a sound from azimuth θ in the horizontal plane has $x=\cos\theta$, $y=\sin\theta$, $z=0$, and a sound from azimuth θ and elevation η above horizontal has $x=\cos\theta\cos\eta$, $y=\sin\theta\cos\eta$, $z=\sin\eta$. (x,y,z) is termed the direction cosines of the direction of the vector. The symbol 1 is used to indicate a signal with all sounds encoded with the uniform gain 1, and the letter W is used in connection with signals representing reproduced sound field pressure. The letter P is used to indicate general speaker feed signals.

The letter t is used to indicate time, c the speed of sound, F frequency and $\omega=2\pi F$ angular frequency. For complex numbers u+jv, we use the symbols Re, Im and * as follows:

$$\mathrm{Re}(u+jv) = u \quad , \quad \mathrm{Im}(u+jv) = v \quad , \quad (u+jv)^* = u-jv \quad .$$

In some of the paper, we find it convenient to write vectors in the form $x^p = (x^1,x^2,x^3)$ rather than in the form (x,y,z). The superscript here (which is written as ^{p} or ^{q} when any coordinate is considered, where p,q=1,2,3) should not be confused with a power or exponent, but is just a coordinate index. Capital letter subscripts are invariably not intended to stand for numbers, and we occasionally

(and with warning) use the Einstein summation convention whereby repeated superscripts or subscript in a product are intended to be added over all possible values of the repeated superscript or subscript.

3. HEIRARCHIES OF MODELS

The models of directional localization that we shall consider are graded in order of complexity in 3 different ways. The first parameter describing the heirarchy of models describes the degree to which the model is nonlinear. Thus a linear model is "first degree", a quadratic or correlation model is "second degree", a cubic or bispectral model is "third degree", etc. The second parameter describing the place of a model in the heirarchy is the order of directionality, i.e. the order of the spherical harmonics in direction to which the model reponds. Thus a zeroth order model is non-directional, a first order model responds to vector aspects of directionality, etc. The third parameter describes the degree to which head-movement is taken into account by the model, i.e. whether the model supposes the head to be stationary, in an arbitrary orientation, or some intermediate situation. We now describe each of these heirarchies in more detail, and then put them together.

(i) Nonlinear Heirarchy.

Under certain conditions, a nonlinear operator N acting on n input signals $f_i(t)$ (i=1,2,...,n) will produce an output signal that may be expressed in the general form of a Volterra Series [18],[19] via

$$N\{\underset{\sim}{f}(t)\} = \sum_{D=0}^{\infty} \int\cdots\int_{D\text{-fold}} k_{i_1\cdots i_D}(t-t_1,\ldots,t-t_D) f_{i_1}(t_1)\cdots f_{i_D}(t_D)\, dt_1\cdots dt_D$$

where we use the Einstein summation convention of summing over all possible values of the indices $i_1,\ldots,i_D$, and where for each set $i_1\cdots i_D$ of indices, $k_{i_1\cdots i_D}$ is a function of D time variables known as the Volterra kernel. In the special case D=1, $k_i(t)$ is the ordinary convolution kernel of a linear system, and in general, the D'th term of the Volterra series describes a nonlinearity of D-th degree in the input signals.

By taking each term separately, we may consider a model of D-th degree nonlinearity. For example, if the ears respond via

$$N\{\underset{\sim}{f}(t)\} = \int\{f_1(t)^2+\ldots f_n(t)^2\}dt,$$

which is the total energy of the signal, the model would be quadratic, i.e. of degree D=2. Similarly, a model with D=3 is termed cubic. Most models considered in the literature are either essentially linear ([4]-[6],[8]-[13],[15],[16]) or essentially quadratic ([3],[6],[14],[20]).

(ii) Directional Heirarchy

The ears as a system may be considered as responding to a function of direction. (This function may be sound waveform amplitude in a linear theory, sound energy in a quadratic theory, and more complicated quantities in a D-th degree theory). The degree to which the polar diagrams describing the reception of the sound information are directional determines the order of the theory.

A function of the direction may [17] be described as a function on the surface of a sphere, and expressed uniquely as a sum of spherical harmonics. The order of the theory is the order of spherical harmonics used by the model. Thus expressing the incoming sound information as a function f(x,y,z) of the direction cosines (x,y,z), the order of a theory is the order of the spherical harmonic components [17] of f to which it responds. Thus a first order theory responds to functions only of the form $1+\alpha x+\beta y+\gamma z$ (α,β,γ constants), whereas a 2nd order theory responds also to functions involving terms quadratic in the direction cosines.

(iii) Moving Head Heirarchy

Given a model of known degree and order, the extent to which use is made of information obtained by moving the head is still not determined. The moving head heirarchy of models is obtained by choosing

which of the parameters in a given model are physically significant. For example, with the head absolutely fixed, vector components of sound arriving in a direction 90^0 from the axis of the ears are not used. In a second class of models (e.g. see [12],[20]), only information that varies to first order in head movement is used, whereas at the other extreme, a model may use information in a totally direction-independent fashion, i.e. without the model having any preferred spatial orientation.

4. SOME "PRIMITIVE" MODELS

Based on the heirarchies described above, it is possible to describe what the lowest models in these heirarchies look like. Here we list some basic "primitive" models, describing them in terms of the information received by the ear/brain system when responding to a number n of equally distant sound signals P_i placed in the direction with direction cosines (x_i, y_i, z_i). We also discuss briefly the approximate physical significance of the various parameters occuring in these models.

(1) First Degree First Order Models (Velocity Models)

Consider the signals (representing pressure and velocity)

$$w'_V = \sum_{i=1}^{n} P_i$$

$$x'_V = \sum_{i=1}^{n} x_i P_i$$

$$y'_V = \sum_{i=1}^{n} y_i P_i$$

$$z'_V = \sum_{i=1}^{n} z_i P_i$$

and for a single encoded sound with associated complex gains w'_V, x'_V, y'_V, z'_V, and write

$$x_V = x'_V / w'_V$$

$$y_V = y'_V / w'_V$$

$$z_V = z'_V / w'_V$$

so as to eliminate the effect of the overall signal level w'_V from the direction-determining aspects of our models.

Then write

$$r_V \hat{x}_V = \text{Re } x_V$$

$$r_V \hat{y}_V = \text{Re } y_V$$

$$r_V \hat{z}_V = \text{Re } z_V$$

where $r_V = \{(\text{Re } x_V)^2 + (\text{Re } y_V)^2 + (\text{Re } z_V)^2\}^{\frac{1}{2}}$ and $\hat{x}_V^2 + \hat{y}_V^2 + \hat{z}_V^2 = 1$, and Re stands for "real part of". Thus $(\hat{x}_V, \hat{y}_V, \hat{z}_V)$ are direction cosines.

The direction $(\hat{x}_V, \hat{y}_V, \hat{z}_V)$ is the apparent direction of the sound according to Makita's theory of sound localization (used in [2],[5],[10]) and is thus called the Makita localization. The quantity r_V equals 1 for a single sound source (as a calculation quickly shows) and ideally should be as close to 1 as possible for a reproduced sound. r_V is called the velocity magnitude of the sound. All low frequency interaural phase theories of sound localization (which ignore any effect of amplitude differences between the ears) assume that the only quantities relevant to localization are the Makita localization $(\hat{x}_V, \hat{y}_V, \hat{z}_V)$ and the velocity magnitude r_V. Such theories apply at audio frequencies somewhat below 700 Hz.

For horizontal sound sources, we may rewrite the above by putting (for speakers with azimuths ϕ_i)

$$x'_V = \sum_{i=1}^{n} P_i \cos\phi_i$$

$$y'_V = \sum_{i=1}^{n} P_i \sin\phi_i$$

and computing x_V, y_V as above, and finally putting

$$r_V \cos\theta_V = \text{Re } x_V$$

$$r_V \sin\theta_V = \text{Re } y_V$$

where θ_V is the Makita azimuthal localization and $r_V > 0$ is again the velocity

magnitude.

The remaining three real parameters

$$\text{Im } \hat{x}_V \quad , \quad \text{Im } \hat{y}_V \quad , \quad \text{Im } \hat{z}_V$$

describing localization in this class of models are termed the phasiness in (respectively) the directions of the x,y and z-axes. As discussed in Appendix II of [21], the phasiness describes departures from the ideal low frequency theories, as well as unpleasant qualities of localization discussed experimentally in [22]. Phasiness describes quite well the overall degree of blurring and unpleasantness caused by the use of phase-shifting circuitry in decoding equipment, and is probably apt at frequencies above about 300 Hz but below about 1500 Hz. These figures are guesstimates based on a mixture of theory and experience. Experimental evidence [22] suggests that the magnitude of the vector component of phasiness in the direction of the ear-axis should not exceed about 0.21 in order to be practically inaudible, although experience suggests that sensitivity to phasiness grows with experience.

For a forward-facing listener (with horizontal head), the component ($\text{Im } \hat{y}_V$) of phasiness parallel to his ear-axis is thought to be subjectively the most important. Ideally, the phasiness should be zero, and certainly should be less than 1.

For sound sources (i.e. loudspeakers) at a finite distance d, the above model must be slightly changed. The quantity w'_V describing sound field pressure is given as above, but the formulae for the velocity components x'_V , y'_V , z'_V at the listener should be changed to:

$$x'_V = \sum_{i=1}^{n} x_i P_i \left(1 - \frac{jc}{\omega d}\right)$$

$$y'_V = \sum_{i=1}^{n} y_i P_i \left(1 - \frac{jc}{\omega d}\right)$$

$$z'_V = \sum_{i=1}^{n} z_i P_i (1 - \frac{jc}{\omega d})$$

where the distance d is in metres, c is the speed of sound in metres/sec (334 m/s), and where ω is the angular frequency of the sound in sec^{-1}. The rest of the calculation of the localization parameters is as before. Essentially, the factor $(1-jc/\omega d)$ is the familiar 'bass boost' of velocity for a source at a finite distance, and is caused by the curvature of the sound field for such a source. Alternatively, it is possible to deduce this modification of the sound localization by complicated calculations on the waves arriving at the two ears in models in which such aspects are calculated . It will be seen that the bass boost affects mainly low frequencies, and that its main effect will be to convert phasiness into an alteration of the Makita azimuth.

We comment that the localization parameters described here may properly be regarded as a function of sound frequency, and it is possible for some designs of equipment that the various parameters might actually be designed to vary with frequency.

(2) Second Degree First Order Models (Energy Models)

The model considered here is similar to the first degree model, except that the sound amplitude gain P_i from each speaker is replaced by its energy gain $|P_i|^2$.

$$w'_E = \sum_{i=1}^{n} |P_i|^2$$

$$x'_E = \sum_{i=1}^{n} x_i |P_i|^2$$

$$y'_E = \sum_{i=1}^{n} y_i |P_i|^2$$

$$z'_E = \sum_{i=1}^{n} z_i |P_i|^2 \quad .$$

As before, we write

$$x_E = x'_E/w'_E$$

$$y_E = y'_E/w'_E$$

$$z_E = z'_E/w'_E$$

and further write

$$r_E\hat{x}_E = x_E$$

$$r_E\hat{y}_E = y_E$$

$$r_E\hat{z}_E = z_E$$

where $r_E = \{(x_E)^2+(y_E)^2+(z_E)^2\}^{\frac{1}{2}}$, and where $(\hat{x}_E)^2+(\hat{y}_E)^2+(\hat{z}_E)^2 = 1$. Then the direction having direction cosines $(\hat{x}_E,\hat{y}_E,\hat{z}_E)$ is called the energy vector localization and the quantity r_E is termed the energy vector magnitude. These are the two quantities describing localization in the model now considered, which is thought to apply to some degree in the frequency region 500 - 5000 Hz. For a single sound source, an easy calculation shows that $r_E=1$, so that ideally this value should be attained for all reproduced sounds. However, we can prove the following theorem:

Theorem 1 If two or more distinct sound sources at a large and equal distance from the listener are fed with a sound with non-zero gains, then the associated energy vector magnitude r_E is strictly less than 1.

Proof This is shown by observing that r_E is the length of an average of unit length vectors (x_i,y_i,z_i) with positive weights $|p_i|^2 / \sum_{j=1}^{n}|p_j|^2$; for r_E to equal 1 it would be necessary for the length of this sum of vectors to equal the sum of the lengths of the vectors, which in turn would require all vectors (and hence speakers) to lie in the same direction, contrary to the assumption of the theorem. This completes the proof.

As in the case of the previous models, the energy vector magnitude and energy vector localization may be computed separately either for each

frequency, or for each of a band of frequencies, in the case that the signal gains P_i vary with frequency.

(3) Third Degree First Order Models (Bispectral Models)

The theory of third degree models is somewhat more complex, and relies on somewhat deeper theory than it is practical to give in the present paper, so that the form given below has, to some extent, to be taken on trust. Although the detailed theory will be published elsewhere, we assume that third degree aspects of the ears act as a bispectral analyser, where for a signal f(t), the bispectrum is defined as that function of pairs of frequencies F_1, F_2 (with $0<F_2<F_1$) that is the Fourier transform of the triple correlation

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} f(t)f(t+t_1)f(t+t_2)dt \ .$$

The theory of the bispectrum is given (very mathematically) in [23],[24], [25], and we comment that the bispectrum measures the degree of mutual correlation between three frequencies F_1, F_2 and F_1+F_2, and is also a measure of waveform asymmetry. Some elementary discussion of the bispectral theory of hearing is given in [26].

Writing the gains of the sound emerging from the i'th loudspeaker at frequency F as $P_i(F)$, the bispectral theory computes for each pair F_1, F_2 of frequencies (with $0<F_2<F_1$) the quantities

$$w'_B = \sum_{i=1}^{n} P_i(F_1)P_i(F_2)\{P_i(F_1+F_2)\}^*$$

$$x'_B = \sum_{i=1}^{n} x_i P_i(F_1)P_i(F_2)\{P_i(F_1+F_2)\}^*$$

$$y'_B = \sum_{i=1}^{n} y_i P_i(F_1)P_i(F_2)\{P_i(F_1+F_2)\}^*$$

$$z'_B = \sum_{i=1}^{n} z_i P_i(F_1)P_i(F_2)\{P_i(F_1+F_2)\}^*$$

where * indicates complex conjugation. Note the curious asymmetric way complex conjugation occurs, whereas in the 2nd degree case we had $|P_i(F)|^2 = P_i(F)\{P_i(F)\}^*$ in our expressions. Then as in the first degree case, we compute the complex quantities

$$x_B = x'_B/w'_B$$

$$y_B = y'_B/w'_B$$

$$z_B = z'_B/w'_B$$

and from these in turn compute r_B and $(\hat{x}_B,\hat{y}_B,\hat{z}_B)$ via the equations

$$r_B\hat{x}_B = \text{Re } x_B$$

$$r_B\hat{y}_B = \text{Re } y_B$$

$$r_B\hat{z}_B = \text{Re } z_B$$

where $r_B = \{(\text{Re } x_B)^2+(\text{Re } y_B)^2+(\text{Re } z_B)^2\}^{\frac{1}{2}}$ and where it follows that $\hat{x}_B^2+\hat{y}_B^2+\hat{z}_B^2 = 1$. The direction $(\hat{x}_B,\hat{y}_B,\hat{z}_B)$ is the <u>bispectral vector localization</u>, and r_B is termed the <u>bispectral vector magnitude</u>. For a single sound source, r_B=1 and its localization is the bispectral vector localization. Additional quantities produced by this localization theory are the <u>bispectral phasiness</u> components

$$\text{Im } x_B \ , \ \text{Im } y_B \ , \ \text{Im } z_B$$

in the directions respectively of the x, y and z axes. Ideally these should be zero.

We reiterate that if a frequency-dependent decoder is used, so that $P_i(F)$ varies with frequency F, the bispectral localization is a function of pairs of frequencies F_1 and F_2 (with $0<F_2<F_1$). This can make computations tedious unless the decoder is frequency-independent.

The bispectrum of a signal describes its timbre or sound-colour, and the frequencies F_1 and F_2 are the frequencies of the formants

of the sound, i.e. frequencies at which the sound has been subject to broad resonances [27]. Although sufficient data is not yet available, the bispectral theories are thought to apply for frequencies F_1, F_2 and F_1+F_2 in the range 300 Hz - 5 kHz approximately.

(4) First Degree Second Order Models

In order to illustrate the way higher orders of directionality can enter into a theory, we consider the case of a 2nd order 1st degree theory; such a theory is probably apt in the frequency region 400 - 1000 Hz in which the 1st order theory starts failing. In the 2nd order theory, the available information includes 9 signals with complex gains corresponding to 9 independent zero first and 2nd spherical harmonics. If $f(x,y,z)$ is a spherical harmonic (i.e. a polynomial function on the sphere of direction cosines orthogonal to all polynomials of lower degree [17]) of 0, 1st or 2nd order, then the signal S'_f with gain

$$S'_f = \sum_{i=1}^{n} f(x_i,y_i,z_i)\, P_i$$

conveys information about sound directionality. Denoting by 1 the special function on the sphere that equals 1 everywhere,

$$S'_1 = \sum_{i=1}^{n} P_i$$

and we may thus consider localization as depending on the 8 complex parameters

$$S_f = S'_f/S'_1$$

for 8 independent 1st and 2nd order spherical harmonics [17]; if we only consider horizontal directions, then we need only consider the 4 complex parameters given by putting

$$f(\cos\theta,\sin\theta,0) = e^{-j\theta}, e^{j\theta}, e^{-2j\theta}, e^{2j\theta}$$

such as considered in [2].

Determining how apparent direction depends on the 16 real parameters $\mathrm{Re}\, S_f$ and $\mathrm{Im}\, S_f$ is clearly a much more complicated task than when we

considered only 6 real patameters as in the first order first degree models. It really requires more experimental data than is presently available to refine the interpretation of the 16 parameters at each frequency, and the present status of 2nd order models is largely to give a general idea of how 1st order models start to go wrong as the frequency is raised, rather than to give detailed predictions.

(5) A 3rd Degree 3rd Order Model (Cross-Bispectral Models)

We here describe an example of a high degree high order model in order to show that although the general class of such models is too complex to handle in any detail, particular models may well be simple enough to give useful detailed predictions. The model we shall use will be called the "cross-bispectral" model, and is a 3rd degree 3rd order model; like the "bispectral" 3rd degree first order model considered earlier it envisages an impression of directionality due to correlations between the frequency components of signals at frequencies F_1, F_2 and $F_3 = F_1+F_2$. The theory given here differs from the bispectral model considered earlier in that it envisages that correlations between output signals from sources in different directions may influence apparent sound direction; for example, a sound emerging at frequencies F_1 and F_2 from one speaker and at frequency $F_3 = F_1+F_2$ from a second speaker may produce an image situated at neither speaker. The present model describes localization in terms of cubic functions of the direction cosines of loudspeakers, and so is a 3rd order model.

Using the notations described earlier, consider the complex quantities at frequencies F_1, F_2:

$$w'_{CB} = \Sigma^i\Sigma^j\Sigma^k P_i(F_1)P_j(F_2)\{P_k(F_1+F_2)\}^* = (\Sigma^i P_i(F_1))(\Sigma^j P_j(F_2))(\Sigma^k P_k(F_3))^*$$

$$x'^{p}_{CB} = \frac{1}{3}\{x^{q}_{i}x^{q}_{j}x^{p}_{k} + x^{q}_{i}x^{p}_{j}x^{q}_{k} + x^{p}_{i}x^{q}_{j}x^{q}_{k}\}P_{i}(F_1)P_{j}(F_2)\{P_{k}(F_3)\}^{*} \quad .$$

where the indices i,j,k run from 1 to n, Σ^{i}, Σ^{j}, Σ^{k} mean summation over all values of these indices, where p,q=1,2,3 represent the coordinates of vectors in respectively the x,y and z directions (so that $x^{1}_{i} = x_{i}$, $x^{2}_{i} = y_{i}$ and $x^{3}_{i} = z_{i}$), and where we use the Einstein summation convention that repetition of an index in any product of terms means that we must sum over the repeated indices (so that $x^{q}_{i}x^{q}_{j} = x^{1}_{i}x^{1}_{j}+x^{2}_{i}x^{2}_{j}+x^{3}_{i}x^{3}_{j}$ and $x_{i}P_{i} = x_{1}P_{1}+\ldots+x_{n}P_{n}$). We use F_3 to denote F_1+F_2. We remark that the vector quantity x'^{p}_{CB} may be regarded as a function $x'^{p}_{CB}u^{p}$ on the sphere $(u^{1})^{2}+(u^{2})^{2}+(u^{3})^{2} = 1$, and that this function is $\frac{5}{3}$ times the 1st spherical harmonic component of the 3rd degree function $a^{pqr}u^{p}u^{q}u^{r}$ on the sphere, where

$$a^{pqr} = x^{p}_{i}x^{q}_{j}x^{r}_{k}\,P_{i}(F_1)P_{j}(F_2)\{P_{k}(F_3)\}^{*} \quad .$$

Given the scalar w'_{CB} and complex vectors x'^{p}_{CB}, as is now familiar, we form a cross-bispectral vector

$$x^{p}_{CB} = x'^{p}_{CB} / w'_{CB}$$

for p=1,2,3 , and form the vector $\hat{x}^{p}_{CB}$ of unit length such that

$$r_{CB}\,\hat{x}^{p}_{CB} = \text{Re}\; x^{p}_{CB}$$

where $r_{CB} > 0$ equals $\{\sum_{p=1}^{3}(\text{Re}\; x^{p}_{CB})^{2}\}^{\frac{1}{2}}$. $\hat{x}^{p}_{CB}$ is the cross-bispectral vector localization and r_{CB} the cross-bispectral vector magnitude. The vector $\text{Im}\; x^{p}_{CB}$ is the cross-bispectral phasiness vector.

Unlike the earlier bispectral model, the cross-bispectral model applies only to very accurately positioned listeners who are precisely equidistant from the sources i=1,...,n. We shall see later that the cross-bispectral model gives the same predicted localization as velocity models under a wide range of special conditions, although in general (e.g. when the different frequency components emerge from different speakers) it gives very different predictions

5. INTERPRETATION OF MODELS

The effect of head movement has not been discussed explicitly in the above models, but it is useful to give some general indication of how head movement may be described explicitly. In the velocity, energy, bispectral and cross-bispectral models, we have a direction cosine vector $(\hat{x},\hat{y},\hat{z})$ and a vector length r given by the models. $(\hat{x},\hat{y},\hat{z})$ may be considered as the direction of the sound as determined by someone orienting their head until the sound seems straight in front of them. In the velocity model, $(\hat{x}_V,\hat{y}_V,\hat{z}_V)$ is called the Makita localization [2],[10],[5]. Alternatively, taking the realistic view that real-world sound fields will be degraded by the presence of random room reflections and the like, it is not to be expected that the effective vector length r "heard" by the ear will be as great as 1 even for single sound sources, so that one presumes that the ears are equipped with means of allowing for this; such means must involve movements of the head, but not necessarily movements so drastic as to make the listener face the source. Leakey [12] was led to the same localization as Makita just by considering information deduced by looking at the change of interaural localization for small head movements.

However, there will be cases when head movements do not occur, or when sounds change too rapidly for the effect heard at different head orientations to be compared, or where changes are small in relation to the complexity of total sound information reaching the ears. In these cases, a fixed head model is apt. While room acoustics may degrade the effective value of r, there will be a short period of time between the arrival of a direct sound and its reflections during which the room has no effect on the value of r. Thus we may expect that in some circumstances fixed-head theories may also have some predictive value. (It has been found by experience that reflections from room walls of loudspeaker

signals may be treated as part of the direct sound provided that the delay of the reflections are less than 10ms, i.e. provided that the speakers are closer than 1½m to the nearest walls. With larger distances from room walls, the ears appear to treat early reflections not as a part of the initial transient, but erroneously as a part of the subsequent period during which room acoustics do not yet affect r. The results are that surround localization is poor for such away-from-wall layouts).

The fixed-head localization is obtained by taking the vector $(r\hat{x},r\hat{y},r\hat{z})$, taking its component in the direction of the unit vector (u^1,u^2,u^3) along the ear axis (pointing towards, say, the left ear), and the resultant quantity (if less than 1) is the cosine of the apparent sound direction's angle from the ear-axis. Thus put

$$\theta = \arccos(r\hat{x}u^1 + r\hat{y}u^2 + r\hat{z}u^3)$$

and the sound appears to arrive at an angle θ from the leftward axis of the ears (see figure 1). It will be seen that:

(1) if $r = 1$, the fixed-head localization is the same as the vector localization $(\hat{x},\hat{y},\hat{z})$, although it suffers from an ambiguity due to the fact (shown in figure 2) that there is a cone of directions at an angle θ to the ear-axis. The well-known front-back ambiguity [16][30] is a well-known example of this ambiguity.

(2) if $r < 1$, then the apparent fixed-head direction of the sound is further from the ear-axis (and closer to plane of symmetry of the head) than the vector localization. This causes a 'narrowing of images'.

(3) if $r > 1$ and $|r\hat{x}u^1 + r\hat{y}u^2 + r\hat{z}u^3| < 1$, then the apparent fixed-head direction is closer to the ear-axis (and further from the symmetry plane) than the vector localization, giving a 'wider image'.

(4) if the projection onto the ear-axis has length > 1, i.e. $|r\hat{x}u^1 + r\hat{y}u^2 + r\hat{z}u^3| > 1$, the quality of localization is unlike any

encountered for real sounds. In such a case, the localization quality may be to the side of the head but disturbing, or may have no clear localization.

We mention the theories of Strutt [16], Clark Dutton Vanderlyn [9], Bauer [15], de Boer [14] and Damaske and Ando [3] as examples of fixed-head theories.

Some of the theories considered (velocity, bispectral and cross-bispectral models) also have a 'phasiness vector' affecting localization, which is zero for real-world direct sounds. The effect of such phasiness quantities on localization is not easy to predict in detail. Not only does phasiness affect the quality of the localized sound (producing listener fatigue and poor localization quality, as well as affecting tone color), but it may also alter the actual localization perceived from that given above to some degree, which may vary with frequency (or with pairs of frequencies in bispectral and cross-bispectral models). The precise way in which this happens varies from theory to theory, but in some theories, the localization may be approximately predicted by taking the vector

$$(r\hat{x}+ap^1,\ r\hat{y}+ap^2,\ r\hat{z}+ap^3)$$

where (p^1,p^2,p^3) is the phasiness vector, and where a is a numerical constant (often equal to about 0.3), and treating this as the vector from which fixed and moving-head localizations may be deduced.

Implicit in the description of the models was two assumptions:

(i) that sounds only occur one at a time, i.e. that only one monophonic sound is fed to the n loudspeakers at any time. In practice many different monophonic sounds with different complex gains for each speaker will occur in actual program material.

(ii) that the listener is precisely equidistant from all loudspeakers.

These two assumptions have some inter-relation and may often be weakened substantially. First we note that in the velocity and energy models, it does not matter that different sounds occur at different frequencies, since the models assume that each frequency or frequency band may be handled independently. Similarly in the bispectral and cross-bispectral models, sounds with different 'bispectral frequency pairs F_1,F_2' (i.e. different formant frequencies) may be handled as if occuring separately even if they actually occur together; the mathematical proof of this lies in the 'stochastic independence' of signals bispectra [23].

However, even for sounds with the same frequencies or 'bifréquencies F_1,F_2', there is a possibility that the ears can simultaneously distinguish among more than one sound independently sounding from different apparent directions. We cannot here go into the mathematical theory of why this should be, except to point out to the interested reader that the spectrum of a multichannel signal is a 'covarience matrix' or complex 'tensor of 2nd rank' and that the bispectrum is a complex 3rd rank tensor [23]. As a result, the number of independent spectral and bispectral variables available to the ear is larger than the number of vector variables used for localization. Indeed, in principle, the bispectral models are capable of localizing up to 4 sounds sharing the same bispectrum and sounding at the same time from different directions [25].

The velocity and cross-bispectral models demand that the listener be precisely the same distance from all speakers for them to give valid predictions, but the energy and bispectral models do not demand this. This is because the latter models assume each speaker is an independent source, and only time-averaged quantities from them (e.g. the energy or bispectrum) add up at the listeners ears. As a result, it is expected that the energy and bispectral models will give useful predictions also

for non-central listeners, provided that the change in intensity of sound due to speaker distance and speaker direction changes are taken into account. We hope to describe in another publication applications of a random version of the velocity models to non-central listeners. However, all these models ultimately give predictions based on the time-averaged properties of signals reaching the ears, and it is possible and likely that the different localizations occuring immediately after transients when the sound of only some speakers has arrived at the listener's ears will modify the overall directional impression (e.g. the Haas precedence effect [28][29]).

6. THEOREMS ABOUT MAKITA LOCALIZATION

We have seen that the Makita localization is only one of many localizations predicted by the various primitive models of sound localization. If the various methods of sound localization do not agree, then the Makita localization is not a reliable way of predicting where sounds come from. However, as we shall see, there is a range of conditions that automatically ensure that various methods of localization do agree with the Makita prediction. Indeed, given that we can design decoding apparatus to give correct Makita localization, we may use the following theorems to design decoders to give correct localization according to other criteria as well.

Thus, in the subsequent work in this and following papers, we shall lay great emphasis on getting correct Makita localization. This is not because we consider Makita's theory to be "correct" or "reliable" (we do not), but because the art of designing good decoders may be reduced to getting correct Makita localization as a first step, and then to using results in this and later papers to get other localization criteria right as well.

The fundamental design theory of decoders is based largely on velocity and energy models. Other models enter largely as an aid to refining designs. Our basic approach will be to demand that decoders should, as a very minimum requirement, give localization that is identical according to both the Makita and energy vector localization criteria. The reason for this is as follows:

We know that the Makita localization is one of the things we wish to get right at low frequencies well below 700 Hz. We can also get the velocity vector magnitude r_V right if in an initial design it does not equal the desired value of one by changing the gain of the sum of

the speaker signals until r_V does equal 1. It is likely, however that such a change of design will give a less-than-optimal value of the energy vector magnitude r_E. There is a rather indeterminate band of frequencies (say 250 Hz - 1500 Hz) where it is not clear which of velocity and energy models applies.

The designer of a decoder will wish to optimise localization both at low and high frequencies. The best way of doing this is to design a decoder which takes the form of one matrix at low frequencies and another at high frequencies. It is necessary that the transition between these two matrices also satisfies relevant localization criteria. Given that it is probably not possible to design a decoder to be simultaneously optimal according to both velocity and energy models, we seek to satisfy both Makita and energy vector localization in the intermediate band of frequencies, and to get some compromise among the other criteria of localization. The following theorems ensure that it is possible to make Makita and energy vector localizations the same, and also tell us how to do this.

Consider four loudspeakers placed in a rectangle (see fig. 3) with speakers handling respective signals LB, LF, RF, RB placed at azimuths $180^0-\phi, \phi, -\phi$ and $-180^0+\phi$ respectively, i.e. with respective direction cosines $(-x,y,0)$, $(x,y,0)$, $(x,-y,0)$ and $(-x,-y,0)$ where $x=\cos\phi$, $y=\sin\phi$. (It is convenient from now on to suppress the third 0 z-coordinate as irrelevant). Then we shall prove:

Theorem 2 (Rectangle Decoder Theorem)

The Makita and energy vector localization of a rectangle speaker layout coincide if the signal

$$Q = \tfrac{1}{2}(-LB+LF-RF+RB)$$

is either zero or bears for all sounds a 90^0 phase relation to X, Y,

where the 3 signals W,X,Y are defined by:

$$W = \tfrac{1}{2}(LB+LF+RF+RB)$$

$$X = \tfrac{1}{2}(-LB+LF+RF-RB)$$

$$Y = \tfrac{1}{2}(LB+LF-RF-RB) \ .$$

In the case Q bears a 90^0 phase relation to W,X,Y the latter 2 signals must bear a real phase relation to one another, and the Makita and energy vector localizations and velocity vector magnitude r_V are not changed by replacing Q by a zero signal, but the energy vector magnitude r_E is increased when Q is replaced by zero. In all these cases, for speaker azimuths $180^0-\phi, \phi, -\phi$ and $-180^0+\phi$ respectively for LB,LF,RF,RB we have that the apparent Makita azimuth (and hence energy vector azimuth) θ_V is given by the proportional equation

$$\cos\theta_V \ : \ \sin\theta_V \ = \ (\cos\phi)\mathrm{Re}(X/W) \ : \ (\sin\phi)\mathrm{Re}(Y/W)$$

or by the equivalent equation

$$\cos\theta_V \ : \ \sin\theta_V \ = \ \cos\phi \ \mathrm{Re}(XW^*) \ : \ \sin\phi \ \mathrm{Re}(YW^*) \ .$$

Proof Note that in terms of the signals W,X,Y,Q defined in the theorem that

$$LB = \tfrac{1}{2}(-X+W+Y-Q)$$

$$LF = \tfrac{1}{2}(X+W+Y+Q)$$

$$RF = \tfrac{1}{2}(X+W-Y-Q)$$

$$RB = \tfrac{1}{2}(-X+W-Y+Q)$$

as an easy algebraic manipulation will verify. It may also be checked that the total energy gain

$$|LB|^2+|LF|^2+|RF|^2+|RB|^2$$

equals

$$|X|^2+|W|^2+|Y|^2+|Q|^2,$$

using the fact that $|\alpha+\beta|^2 = |\alpha|^2+|\beta|^2+2\mathrm{Re}(\alpha\beta^*)$ for any complex numbers α,β in the calculation.

Using the notations of the velocity and energy models described earlier, we have that

$$w'_V = LB+LF+RF+RB = 2W$$
$$x'_V = \cos\phi\ (-LB+LF+RF-RB) = 2X\cos\phi$$
$$y'_V = \sin\phi\ (LB+LF-RF-RB) = 2Y\sin\phi$$

so that the Makita azimuth θ_V is given by

$$r_V\cos\theta_V = \mathrm{Re}(x'_V/w'_V) = \cos\phi\ \mathrm{Re}(X/W)$$
$$r_V\sin\theta_V = \mathrm{Re}(y'_V/w'_V) = \sin\phi\ \mathrm{Re}(Y/W)$$

as required in the statement of the theorem. Clearly, no prediction of the velocity models of localization is changed by any choice of the signal Q, since it does not enter the formulae for w'_V, x'_V or y'_V, so that r_V and θ_V are not affected by replacing Q by zero.

Note for any complex numbers α,β that $\alpha/\beta = (\alpha\beta^*)/(\beta\beta^*) = |\beta|^{-2}(\alpha\beta^*)$, so that $\mathrm{Re}(\alpha/\beta) = |\beta|^{-2}\mathrm{Re}(\alpha\beta^*)$, which proves the last equation of theorem 2.

Calculating the energy model parameters shows

$$w'_E = |LB|^2+|LF|^2+|RF|^2+|RB|^2 = |X|^2+|W|^2+|Y|^2+|Q|^2$$
$$x'_E = \cos\phi(-|LB|^2+|LF|^2+|RF|^2-|RB|^2) = \cos\phi(2\mathrm{Re}XW^*+2\mathrm{Re}YQ^*)$$
$$y'_E = \sin\phi(|LB|^2+|LF|^2-|RF|^2-|RB|^2) = \sin\phi(2\mathrm{Re}YW^*+2\mathrm{Re}XQ^*)\ .$$

Clearly we have $x_E : y_E = x'_E : y'_E = \cos\phi\ \mathrm{Re}XW^* : \sin\phi\ \mathrm{Re}YW^*$ provided only that $\mathrm{Re}XQ^* = \mathrm{Re}YQ^* = 0$, so that Q being in 90^0 phase relation to X and Y is sufficient to ensure that the Makita and energy vector localizations coincide. Finally, we note that replacing a Q

having a 90^0 phase relation to both X and Y by zero leaves θ_V, r_V and $\theta_E = \theta_V$ unaltered, but that r_E is then multiplied by $\{|X|^2+|W|^2+|Y|^2+|Q|^2\} / \{|X|^2+|W|^2+|Y|^2\} > 1$. Thus removal of the Q signal increases the energy vector magnitude r_E when Q has 90^0 phase relation to X and Y. This completes the proof of theorem 2.

Theorem 2 tends to show that the following remarkable fact holds: better results for non-speaker directions of sound will be obtained for a central listener to a rectangle of loudspeakers if only 3 independent channels of information (W,X,Y) are used to feed them; the presence of a fourth channel Q can only degrade the results. This is especially true for signals all having a real phase relation to one another, where it will be seen from the above method of proof that the following corollary holds.

Corollary 2A

If the signals W,X,Y,Q have a real phase relation to one another, then the Makita and energy vector localization for sounds not precisely in one of the four loudspeaker directions will not coincide unless $Q = 0$.

Proof $x'_E : y'_E = \cos\phi(XW+YQ) : \sin\phi(YW+XQ)$ for real W,X,Y,Q, and if this coincides with the Makita localization, it equals $\cos\phi(XW) : \sin\phi(YW)$, so that if $Q = 0$ $(\cos\phi)Y : (\sin\phi)X = (\cos\phi)X : (\sin\phi)Y$ i.e. $Y^2 : X^2 = 1 : 1$, i.e. $\cos\theta_V : \sin\theta_V = \pm\cos\phi : \pm\sin\phi$. Thus $\theta_E = \theta_V$ and $Q = 0$ implies that $\theta_V = 180^0-\phi, \phi, -\phi$ or $-180^0+\phi$ as required to prove corollary 2A.

Using the velocity models, an easy calculation of the vector $(\text{Im } x'_V/w'_V , \text{Im } y'_V/w'_V)$ shows that the first sentence of the next corrolary holds.

Corollary 2B

For a rectangle speaker layout, phasiness according to the velocity model of hearing is avoided if and only if the signals W,X,Y (defined in theorem 2) have a real phase relationship. For non-speaker directions, the Makita and energy vector localizations for such signals coincide only if Q has 90^0 phase relation to all of W,X,Y , and the energy vector magnitude is maximised by putting Q = 0.

Proof It remains only to show that Q must bear a 90^0 phase relation to the signals W,X,Y whose gains may be presumed real. $\theta_E = \theta_V$ only if

$$\cos\phi\ Y\ (\mathrm{Re}Q) : \sin\phi\ X\ (\mathrm{Re}Q) = \cos\phi\ X : \sin\phi\ Y .$$

As in the proof of corollary 2A, this implies ReQ = 0 except for sounds from the 4 speaker directions. This proves the corollary 2B.

The above results show that, for the most common loudspeaker layouts, there is a definite disadvantage in having a 4th channel Q to feed the 4 loudspeakers, i.e. 3 channels is best for rectangle speaker layouts. We now consider some results concerning decoding 3 channels through regular polygon loudspeaker layouts.

Given 3 signals W,X,Y with an intended sound localization θ_I given by

$$\cos\theta_I : \sin\theta_I = (\mathrm{Re}\ X/W) : (\mathrm{Re}\ Y/W) ,$$

the problem arises of how to decode such signals through an arbitrary speaker layout. We have already solved part of this problem for a rectangle speaker layout.

Consider a naive decoder for a regular polygon loudspeaker layout with n speakers at azimuths ϕ differing by $360^0/n$. The naive decoder feeds the speaker at azimuth ϕ with the signal

$$W + X\cos\phi + Y\sin\phi,$$

as one might expect then we have:

Theorem 3 (Regular Polygon Decoder Theorem)

Let a regular polygon of $n > 4$ loudspeakers be fed with 3 signals W,X,Y presented to the loudspeaker at azimuth ϕ in the form

$$W + X\cos\phi + Y\sin\phi \ .$$

Then the Makita and Energy vector localizations coincide, and the energy vector magnitude r_E cannot exceed $1/\sqrt{2}$. $r_E = 1/\sqrt{2}$ if and only if there is a θ such that W,X,Y have real phase relation and $X = 2^{\frac{1}{2}}W\cos\theta$, $Y = 2^{\frac{1}{2}}W\sin\theta$. In general, the Makita and energy vector Azimuths are given by

$$\cos\theta_I : \sin\theta_I = Re(X/W) : Re(Y/W) = Re(XW^*) : Re(YW^*) \ .$$

All velocity and energy vector model localization criteria for given signals W,X,Y are identical for all numbers $n > 4$ of speakers in any regular polygon array, including a continuous circle of loudspeakers.

Proof We prove the last statement first, since the rest of theorem 3 need then be proved in the special case of a circle of loudspeakers only. For a function $f(\phi)$ of angle ϕ, The integral $(2\pi)^{-1}\int_{-\pi}^{\pi} f(\phi)d\phi$ (which we hereafter write $\int f(\phi)d\phi$) is equal to $\frac{1}{n}\sum_{i=1}^{n} f(\phi_i)$ for $\phi_i = \phi_0 + 2\pi i/n$, i.e. the integral may be replaced by the average over a regular polygon of n points, provided only that $f(\phi)$ may be written as a sum of terms of the form $a_m \sin m\phi$ or $b_m \cos m\phi$ with $m < n$. (This is based on the easily proved fact that

$$\frac{1}{n}\sum_{i=1}^{n}\cos\left(\frac{2\pi mi}{n}\right) = 0 = \int\cos m\phi \, d\phi$$

for $0 < m < n$ with n an integer, and a similar relation with "sin" replacing "cos"). Now for the polygon decoder described above in the theorem (with $n > 4$)

$$\frac{1}{n} w'_V = \frac{1}{n} \sum_{i=1}^{n} (W+X\cos\phi_i+Y\sin\phi_i) = \int (W+X\cos\phi+Y\sin\phi)d\phi = W$$

$$\frac{1}{n} x'_V = \frac{1}{n} \sum_{i=1}^{n} (W+X\cos\phi_i+Y\sin\phi_i)\cos\phi_i$$

$$= \int (W\cos\phi+ \frac{1}{2} X+ \frac{1}{2} X\cos 2\phi+ \frac{1}{2} Y\sin 2\phi)d\phi = \frac{1}{2} X$$

and $\frac{1}{n} y'_V = \frac{1}{2} Y$ similarly.

also

$$\frac{1}{n} w'_E = \int |W+X\cos\phi+Y\sin\phi|^2 d\phi = |W|^2+ \frac{1}{2}|X|^2+ \frac{1}{2}|Y|^2$$

$$\frac{1}{n} x'_E = \int |W+X\cos\phi+Y\sin\phi|^2 \cos\phi d\phi = \int 2\mathrm{Re}(XW^*)\cos^2\phi d\phi$$

$$= \mathrm{Re}(XW^*)$$

and $\frac{1}{n} y'_E = \mathrm{Re}(YW^*)$ similarly,

by replacing $\frac{1}{n} \sum_{i=1}^{n}$ by $\int \ldots d\phi$, which is permissible for $n > 4$, as no trigonometric function of $m\phi$ for $m > 3$ occurs here. It is now easy to see that the Makita and energy vector localizations θ_V and θ_E are given by

$$\cos\theta_V : \sin\theta_V = \frac{1}{2} \mathrm{Re}(X/W) : \frac{1}{2} \mathrm{Re}(Y/W) = \mathrm{Re}\ XW^* : \mathrm{Re}\ YW^*$$

and $\cos\theta_E : \sin\theta_E = \mathrm{Re}\ XW^* : \mathrm{Re}\ YW^*$ also, so that $\theta_E = \theta_V$.

Moreover, we easily compute that

$$r_E^2 = \frac{\{\mathrm{Re}(XW^*)^2+ \mathrm{Re}(YW^*)\}^2}{\{|W|^2+ \frac{1}{2}|X|^2+ \frac{1}{2}|Y|^2\}^2} = \frac{(\mathrm{Re}\ \alpha)^2+(\mathrm{Re}\ \beta)^2}{\{1+ \frac{1}{2}|\alpha|^2+ \frac{1}{2}|\beta|^2\}^2}$$

where $\alpha =$ (X/W), $\beta =$ (Y/W).

Thus r_E^2 is maximized for a given value of $|\alpha|^2$ and of $|\beta|^2$ by requiring α,β to be real, and in that case, putting $u^2 = \alpha^2+\beta^2$

$$r_E^2 = \frac{u^2}{(1+\frac{1}{2}u^2)^2}$$

which is easily shown (by differential calculus or otherwise) to be maximized when $u = \sqrt{2}$. Thus we may put $\alpha = \sqrt{2}\cos\theta$ $\beta = \sqrt{2}\sin\theta$ for some θ; the maximum occurs only when this is possible, i.e. when $X = \sqrt{2}\cos\theta\ W$, $Y = \sqrt{2}\sin\theta\ W$. In this case $r_E = \frac{\sqrt{2}}{2} = \frac{1}{\sqrt{2}}$, which proves the theorem 3.

This limitation that $r_E \leqslant \frac{1}{\sqrt{2}}$ described in theorem 3 does not apply for all sounds to non-regular-polygon-decoders, but the <u>average</u> r_E over all azimuths obtained from 3 signals W,X,Y still has to meet this limitation. By way of comparison we mention (without the routine computational detail) that the standard BMX decoder [2] has $r_E = 0.500$, the standard TMX decoder [2] has $r_E = 0.667$, the standard QMX decoder [2] via 5 or more speakers has $r_E = 0.750$, as compared with the maximum r_E consistently obtainable from 3 channels for a regular polygon decoder of $r_E = 0.707$. It seems that the relatively small improvement from $r_E = 0.707$ to $r_E = 0.750$ is not justification enough for adding a fourth channel to a regular polygon decoding system.

The results above for rectangular and polygonal decoders can be extended to rectangular cuboid and regular polyhedron decoders in 3 spatial dimensions. We omit proofs, which are broadly similar but more complex, but state the results here.

<u>Theorem 4 (Cuboid Decoder Theorem)</u>

Let LBD, LBU, LFD, LFU, RFD, RFU, RBD, RBU (L = left, R = right, B = back, F = front, D = down, U = up) be the eight speaker signal gains of eight speakers placed in a cuboid (see fig. 4) at direction cosines respectively equal to

(-x,+y,-z), (-x,+y,+z), (+x,+y,-z), (+x,+y,+z), (+x,-y,-z),(+x,-y,+z), (-x,-y,-z), (-x,-y,+z).

Define eight signals W, X, Y, Z, Q_X, Q_Y, Q_Z, Q_Q via

$$W = \frac{1}{2\sqrt{2}} (LBD+LBU+LFD+LFU+RFD+RFU+RBD+RBU)$$

$$X = \frac{1}{2\sqrt{2}} (-LBD-LBU+LFD+LFU+RFD+RFU-RBD-RBU)$$

$$Y = \frac{1}{2\sqrt{2}} (LBD+LBU+LFD+LFU-RFD-RFU-RBD-RBU)$$

$$Z = \frac{1}{2\sqrt{2}} (-LBD+LBU-LFD+LFU-RFD+RFU-RBD+RBU)$$

$$Q_X = \frac{1}{2\sqrt{2}} (-LBD+LBU-LFD+LFU+RFD-RFU+RBD-RBU)$$

$$Q_Y = \frac{1}{2\sqrt{2}} (-LBD+LBU+LFD-LFU+RFD-RFU-RBD+RBU)$$

$$Q_Z = \frac{1}{2\sqrt{2}} (-LBD-LBU+LFD+LFU-RFD-RFU+RBD+RBU)$$

$$Q_Q = \frac{1}{2\sqrt{2}} (LBD-LBU-LFD+LFU+RFD-RFU-RBD+RBU) .$$

Then for the Makita and energy vector localizations to coincide, it is sufficient either that $Q_X = Q_Y = Q_Z = Q_Q = 0$ or that $Q_Q = 0$ and Q_X, Q_Y, Q_Z are in 90^0 phase relation with X, Y, Z or that $Q_X = Q_Y = Q_Z = 0$. The Makita localization $(\hat{x}_V, \hat{y}_V, \hat{z}_V)$ is given by

$$\hat{x}_V : \hat{y}_V : \hat{z}_V = xRe(X/W) : yRe(Y/W) : zRe(Z/W)$$
$$= xRe\ XW^* : yRe\ YW^* : zRe\ ZW^* .$$

In the case $Q_X = Q_Y = Q_Z = Q_Q = 0$, we have

$$LBD = \frac{1}{2\sqrt{2}} (W - X + Y - Z)$$

$$LBU = \frac{1}{2\sqrt{2}} (W - X + Y + Z)$$

$$LFD = \frac{1}{2\sqrt{2}} (W + X + Y - Z)$$

$$LFU = \frac{1}{2\sqrt{2}} (W + X + Y + Z)$$

$$RFD = \frac{1}{2\sqrt{2}} (W + X - Y - Z)$$

$$RFU = \frac{1}{2\sqrt{2}}(W + X - Y + Z)$$

$$RBD = \frac{1}{2\sqrt{2}}(W - X - Y - Z)$$

$$RBU = \frac{1}{2\sqrt{2}}(W - X - Y + Z) .$$

Theorem 5 (Regular Polyhedron Theorem)

Let four signals W, X, Y, Z be fed to a layout of more than 4 loudspeakers placed on a sphere at all the points of (i) the face-centers, (ii) the edge-centers, or (iii) the vertices of a regular polyhedron (see Stroud [40] for the use of such point-sets in spherical integration) such that the speaker at direction cosines (x,y,z) is fed with

$$W + xX + yY + zZ .$$

Then all velocity and energy model localization parameters are the same as for a continuous sphere of loudspeakers fed as indicated, and the energy vector and Makita localization are the same $(\hat{x}_V, \hat{y}_V, \hat{z}_V)$, where

$$\hat{x}_V : \hat{y}_V : \hat{z}_V = \text{Re } XW^* : \text{Re } YW^* : \text{Re } ZW^* .$$

The maximum possible energy vector magnitude with such a decoding arrangement is $r_E = 1/\sqrt{3}$, and this value is attained if and only if

$$X = \sqrt{3}xW, \quad Y = \sqrt{3}yW, \quad Z = \sqrt{3}zW,$$

where (x,y,z) are real direction cosines of some direction, i.e. $x^2+y^2+z^2 = 1$.

A result of considerable general use applies to arbitrary decoders having loudspeakers arranged in diametrically opposed pairs, i.e. if one speaker is in the direction (x,y,z), another one is in the direction (-x,-y,-z). Such decoders need not be regular.

Theorem 6 (Diametric Decoder Theorem)

Let 2n loudspeakers be arranged equidistant from a listener such that the loudspeakers are placed in n diametrically opposed pairs

of speakers. Suppose further that the sum of the signals emitted by the two speakers in an opposite pair is the same for all pairs, then the Makita and energy vector localizations of the resultant sound are the same.

Proof Let the loudspeakers at direction cosines (x_i, y_i, z_i) and $(-x_i, -y_i, -z_i)$ be fed with signals $W+P_i$ and $W-P_i$ respectively, as required by the theorem. Then a computation of velocity and energy localization parameters gives, for p=1,2,3

$$w'_V = 2nW$$

$$x'^p_V = \sum_{i=1}^{n} \{x_i^p (W+P_i) - x_i^p (W-P_i)\} = 2 \sum_{i=1}^{n} x_i^p P_i$$

and

$$w'_E = \sum_{i=1}^{n} \{|W+P_i|^2 + |W-P_i|^2\} = 2n|W|^2 + 2\sum_{i=1}^{n} |P_i|^2$$

$$x'^p_E = \sum_{i=1}^{n} \{x_i^p |W+P_i|^2 - x_i^p |W-P_i|^2\} = 4 \sum_{i=1}^{n} x_i^p \, \mathrm{Re}\, P_i W^*$$

thus $\mathrm{Re}(x'^p_V / w'_V) = (x'^p_E / w'_E)\{\frac{n|W|^2 + \Sigma^i |P_i|^2}{2n|W|^2}\}$,

for p=1,2,3 , which proves that the Makita and energy vector localizations coincide. This proves theorem 6.

The final general result of this paper relates the results given by the velocity and cross-bispectral models.

Theorem 7

For an arbitrary loudspeaker layout lying equidistant from the listener, and for any signal fed to those speakers with complex gains that are independent of frequency, the cross-bispectral localization

parameters are determined by the velocity model localization parameters. For a decoder for which in addition the velocity phasiness vector is zero, the cross-bispectral vector localization is identical to the Makita localization, and the cross-bispectral vector magnitude r_{CB} equals the cube r_V^3 of the velocity vector magnitude.

Proof Since we have assumed frequency-independent gains, $P_i(F) = P_i$ for all frequencies F. Then using the abbreviated notation that we introduced in connection with cross-bispectral models,

$$w'_V = \Sigma^i P_i$$

$$x'^p_V = x_i^p P_i$$

$$w'_{CB} = (\Sigma^i P_i)^2 (\Sigma^j P_j)^*$$

$$x'^p_{CB} = \frac{1}{3}\{2(x_i^p P_i)(x_j^q P_j)(x_k^q P_k)^* + (x_i^p P_i)^*(x_j^q P_j)^2\}$$

so that putting

$$x^p_{CB} = x'^p_{CB} / w'_{CB} \quad \text{and} \quad x_V^p = x'^p_V / w'_V$$

we have

$$x^p_{CB} = \frac{2}{3} x_V^p x_V^q (x_V^q)^* + \frac{1}{3} (x_V^p)^* x_V^q x_V^q \quad .$$

This proves that the cross-bispectral localization parameters x^p_{CB} are dependent only on the velocity localization parameters.

In the special case that there is no phasiness in the velocity model (i.e. that $(x_V^p)^* = x_V^p$), we have

$$x^p_{CB} = x_V^p (x_V^q x_V^q)$$

so that the two real vectors x^p_{CB} and x_V^p are proportional and hence the Makita and cross-bispectral vector localizations coincide. Moreover,

$$r_{CB}^2 = x_{CB}^p x_{CB}^p$$

and $$r_v^2 = x_V^p x_V^p$$

so that $r_{CB}^2 = (x_V^p r_V^2)(x_V^p r_V^2) = r_V^6$, i.e. $r_{CB} = r_V^3$,

as required to prove theorem 7.

Corollary 7A If all components of the complex velocity model vector x_V^p are either purely real or purely imaginary (in any orthogonal coordinate system), then the Makita and cross-bispectral vector localizations coincide.

Proof Under these assumptions $x_V^q x_V^q$ is real, and so x_{CB}^p is a real linear combination of x_V^p and $(x_V^p)^*$, with coefficients whose sum is greater than zero (or equal to 0). Thus the cross-bispectral vector localiztion in the direction of $\mathrm{Re}\, x_{CB}^p$ is also in the direction of $\mathrm{Re}\, x_V^p = \mathrm{Re}(x_V^p)^*$. This proves the corollary.

7. CONCLUSIONS

Although a metatheory (theory of theories) of sound localization leads to a large number of possible "primitive" models of sound localization, many of these primitive theories can be rendered useful for designing decoders because their mathematical structure permits the proof of theorems that show that a variety of different localization criteria are satisfied simultaneously provided that various easily-arranged relationships between the signals fed to loudspeakers are designed into the decoder. A particular case of one of thses theorems has shown that an optimal decoder feeding four loudspeakers LB, LF, RF, RB in a rectangle array should have $-LB+LF-RF+RB = 0$ for best results over the whole frequency range, and as a result, 3-channel systems for horizontal-only reproduction will actually give better localization than 4-channel systems, except for sounds precisely in the direction of the 4 loudspeakers.

Of the various models of localization considered in this paper, the most important are the velocity models (apt at frequencies below 700 Hz, but possibly having some application up to 1500 Hz) and energy models (apt at frequencies above 1000 Hz, but possibly having some application down to say 400 Hz). Most models of localization considered in the previous literature (other than high frequency interaural delay models and pinna-coloration models [31]) are subsumed in the velocity and energy models as special cases. While most of the design theory concentrated on the velocity and energy models, other high order and degree models have been described for finer investigation of the properties of decoders; such models have proved valuable in practice.

This paper, despite its length, has had to be rather sketchy about some of the foundations of the theory, and also on the detailed proofs of the more complex theorems. It is intended to publish many of

these details elsewhere, notably in a yet unpublished paper [25] originally prepared in 1974/5. That paper has been quite widely circulated to individuals since the 1970's, and it is hoped to formally publish a version of it in the near future.

As illustrated by both the appendix of this paper and by the recent references [41]-[43] on multispeaker stereo and ambisonic systems, the theory of this paper is a very practical tool for designing concrete directional encoding and decoding systems, having a wide range of applications varying from the design of panpots to the design of decoding and transmission systems [44, 45], and we shall publish other applications in the future.

No claim is made that the theory of this paper is complete and exhaustive. In particular, this paper says little about frequencies above around 4 or 5 kHz where pinna colouration cues become dominant, as noted in [6], although ref. [41] described some methods of adapting the theory to this high frequency region to a limited degree. The paper also has not dealt with noncentral listening, although the methods actually extend to that case, as mentioned in [6] and discussed briefly in [41].

However, the generality of the methods of this paper, and the fact that it takes account of many auditory localisation mechanisms, means that designs based on satisfying several "primitive" component theories of hearing tend to have much lower listening fatigue, and tend to be much more robust under conditions of technical or user abuse, than directional sound reproduction systems based on satisfying only one or two sound localisation mechanisms. This ability to design "robust" directional reproduction systems is the main use and strength of the work of this paper.

APPENDIX I. EXISTING LOCALIZATION THEORIES AND STEREO REPRODUCTION

The energy vector and velocity models of this paper reduce in special cases to a number of localization theories in the existing literature. We detail some of these connections, and consider their applications to the perception of 2-speaker stereo sound. This is not only of interest in existing stereo applications, but 2-speaker presentation provides a means of experimentally determining how much of various types of localization fault is subjectively acceptable [22], and it is useful for surround reproduction applications to determine the tolerable associated localization parameters such as phasiness, r_E and r_V.

The Makita azimuth θ_V is the localization considered by Makita [10], Leakey [12], Bernfeld [4], Nishimaki and Hirano[5] and Cooper and Shiga [2]. All except Leakey derive it as the azimuth which the head must face to give zero interaural phase difference at low frequencies. As we showed in the description of velocity models, at very low frequencies in the presence of phasiness, the Makita azimuth computed assuming very large speaker distance is <u>not</u> the same as for when the speaker distance is finite, although this 'infinite distance' assumption is common in the literature.

Another class of low frequency interaural theories considers fixed heads, usually facing straight forward. The component of the velocity vector $(r_V\cos\theta_V, r_V\sin\theta_V)$ along the ear axis is $r_V\sin\theta_V$, and the apparent fixed-head localization θ_F is given by

$$\sin\theta_F = r_V\sin\theta_V \qquad (1)$$

since a fixed head generally has no way of knowing that $r_V \neq 1$. This fixed head theory has been used by Bernfeld [4], Strutt [16], Blumlein [8], Clark Dutton and Vanderlyn [9] and Bauer [15], among others and is conveniently termed the CDV localization theory. (Although we have named theories after

Makita and Clark-Dutton-Vanderlyn, we do not necessarily imply that these were the first to give explicit expression to these theories, only that they were the first to popularize these theories among audio engineers).

Theories based on the energy models have long been popular and include de Boer [14], Damaske and Ando [3] and Gerzon [20][6]. Such theories, however, have hitherto had a mathematically intractible form, and the equivalence of an 'interaural correlation' model (such as [3] and Sayers and Cherry [34]) and energy models of first or higher order is not immediately obvious, since correlations occur in the time domain whereas spectra occur in the Fourier transform of the time domain. Our theory formulated in terms of what happens at each frequency may be shown to be a reformulation of cross-correlation models via the Fourier transformation.

The energy vector azimuth θ_E appears to be new (except for a previous discussion in [6] by the author), since direction-finding by making both ear signals identical does not appear to have been considered in the energy or correlation theory literature. However, the localization given by the energy analog of CDV localization, with azimuth θ_{FE} given by

$$\sin\theta_{FE} = r_E \sin\theta_E \qquad (2)$$

has in effect been considered by de Boer [14] and Damaske and Ando [3], but not in that language. Also the models of [3] and [14] to some extent (especially at higher frequencies) include 2nd and higher order directional sensitivities of the ears.

The higher degree theories involving triple correlations and (in their Fourier formulation) bispectra appear to be new with the author [25], [26], but there is strong evidence from phonetics [27] and the cocktail party effect that the ears must make use of such triple correlations in perceiving sounds; in particular, bispectral theories include the formant

theories of tone-color perception [27] as a special case. We remark that predictions from the bispectral model of this paper have been confirmed experimentally, and we shall give full details elsewhere.

Consider now applications to conventional 2-speaker stereo localization. Consider two loudspeakers situated (initially at a large distance) at azimuths $\pm\phi$ relative to due front, and put $x = \cos\phi$, $y = \sin\phi$ so that the speaker direction cosines are $(x,\pm y)$. (see figure 5).

Initially we consider the simplest case where the signal gains L and R fed to the left and right speaker are real (as in ordinary stereo pan-potting). Then supposing that $L+R > 0$, the localization parameters of the velocity model are real and easily computed to be given by

$$x_V = x'_V/w'_V = (L+R)\cos\phi/(L+R) = \cos\phi$$

$$y_V = y'_V/w'_V = (L-R)\sin\phi/(L+R) = \frac{L-R}{L+R}\sin\phi$$

Thus the Makita azimuth θ_V is given by

$$\tan\theta_V = y_V/x_V = \frac{L-R}{L+R}\tan\phi \qquad (3)$$

which is the stereophonic law of tangents of Leakey [12] and Makita [10]. Using (1), the CDV localization is given by

$$\sin\theta_F = y_V = \frac{L-R}{L+R}\sin\phi \qquad (4)$$

which is the stereophonic law of sines of Bauer [15] and Clark, Dutton, Vanderlyn [9].

The velocity vector magnitude r_V is given by

$$r_V = (x_V^2+y_V^2)^{\frac{1}{2}} = (L^2+R^2+2LR\cos 2\phi)^{\frac{1}{2}}/(L+R) \ . \qquad (5)$$

This equals 1 for $L = 0$ or $R = 0$, and equals $\cos\phi$ for $L = R$.

The energy vector localization θ_E is similarly given by

$$\tan\theta_E = \frac{L^2-R^2}{L^2+R^2}\tan\phi \quad , \tag{6}$$

the fixed-head localization θ_{FE} as in (2) is given by

$$\sin\theta_{FE} = \frac{L^2-R^2}{L^2+R^2}\sin\phi \tag{7}$$

and the energy vector magnitude r_E is given by

$$r_E = (L^4+R^4+2L^2R^2\cos 2\phi)^{\frac{1}{2}}/(L^2+R^2) \quad . \tag{8}$$

(6),(7) and (8) are the same as (3),(4),(5) except that L^2 and R^2 replace L and R. Except for the cases L = 0, R = 0 or L = R, the energy localization θ_E does not equal the Makita localization θ_V, and similarly $\theta_F \neq \theta_{FE}$ except for L = 0, R = 0 or L = R. For bispectral localizations, we replace L and R in (3),(4),(5) by L^3 and R^3.

Using theorem 7 of this paper in the real gain case also gives a cross-bispectral localization

$$\theta_{CB} = \theta_V = \tan^{-1}\left(\frac{L-R}{L+R}\tan\phi\right) \tag{9}$$

and a cross-bispectral vector magnitude:

$$r_{CB} = r_V^3 = (L^2+R^2+2LR\cos 2\phi)^{3/2}/(L+R)^3 \quad . \tag{10}$$

In order to illustrate the theories further, consider the fixed-head localization heard not by a listener facing forward, but by one facing an azimuth ψ. When L and R are both positive (in-phase sounds), then the sound image will tend to be displaced towards the head azimuth since $r_V < 1$, $r_E < 1$ and $r_{CB} < 1$.

More precisely, the leftward ear-axis has direction cosines $(-\sin\psi,\cos\psi)$, so that for the velocity model, the projection of (x_V,y_V) onto this axis is $(-x_V\sin\psi, +y_V\cos\psi)$ and this equals $\sin(\theta_F'-\psi)$ where θ_F'

is the apparent sound azimuth in the fixed-head case. Thus

$$\sin(\theta_F'-\psi) = \sin\phi\,\cos\psi\,\frac{L-R}{L+R} - \sin\psi\,\cos\phi \qquad (11)$$

gives the apparent localization for the listener with head at angle ψ. For example, when he faces the left speaker (so that $\psi = \phi$), we have manipulating (11):

$$\sin(\theta_F'-\phi) = -\frac{R}{L+R}\sin 2\phi\ . \qquad (12)$$

For example, for $L = R$,

$$\theta_F' = \phi - \sin^{-1}(\tfrac{1}{2}\sin 2\phi) > 0 \qquad (13)$$

so that central sounds shift towards the left speaker. For a typical interspeaker angle $2\phi = 60^0$, the shift is given by $\theta_F' = 4.34^0$, whereas for $2\phi = 90^0$, the shift is $\theta_F' = 15^0$, which shows that wide interspeaker angle 2ϕ in stereo leads to images which are unstable under head movement. The most extreme case of (11) when $\psi = 90^0$ (i.e. speaker pair at side of listener) gives $\theta_F' = \pm\phi$, i.e. the supposed central sound is drawn unstably to one speaker or the other (ambiguously), which certainly agrees with experimental results on side-image localization via 2 speakers [1],[3], [35],[36]. We observe that a similar theory replacing L and R by L^2 and R^2 predicts a similar ambiguity in the energy models, so that one presumes, using the philosophy of the introduction to this paper, that the similar ambiguous positioning according to two different models will make such ambiguities likely in practice for pairs of loudspeakers at the side of the listener.

Stereophonic localization for forward-facing listeners when the gains L and R are complex (i.e. with interchannel phase differences) is of particular interest. For this case we find that

$$x_V = \cos\phi$$

$$y_V = \frac{L-R}{L+R}\sin\phi$$

as before, except that now y_V is complex. Thus

$$\text{Re } y_V = \text{Re}\left(\frac{L-R}{L+R}\sin\phi\right) = \sin\phi \frac{|L|^2-|R|^2}{|L|^2+|R|^2+2\text{Re}(LR^*)} .$$

This gives a more complicated localization theory, although certain ways of looking at the localization using the energy sphere model ([21], Appendix II) of 2-channel systems can provide insight into both the localization and phasiness aspects of velocity models. We get the Makita localization θ_V and forward fixed head localizations θ_F as before by:

$$\tan\theta_V = \tan\phi \frac{|L|^2-|R|^2}{|L|^2+|R|^2+2\text{Re}(LR^*)} \tag{14}$$

$$\sin\theta_F = \sin\phi \frac{|L|^2-|R|^2}{|L|^2+|R|^2+2\text{Re}(LR^*)} . \tag{15}$$

In general, for given speaker outputs $|L|^2$ and $|R|^2$, with interspeaker phase ξ we have

$$\text{Re}(LR^*) = |L||R|\cos\xi \tag{16}$$

so that the denominators of (14) and (15) diminish as interspeaker phase increases, so that the sound image widens.

As for phasiness, since only y_V is non-real, the phasiness affects only the ear-axis direction, with magnitude Im y_V, which equals

$$q = \sin\phi \frac{2\ \text{Im}\ LR^*}{|L|^2+|R|^2+2\text{Re}(LR^*)} , \tag{17}$$

which equals

$$q = \frac{2 \sin\phi \sin\xi}{\left|\frac{L}{R}\right| + \left|\frac{R}{L}\right| + 2\cos\xi} \qquad (18)$$

when the left speaker phase leads the right speaker by a phase angle ξ. We remark that the "phasiness" Q introduced in [21] Appendix II and in [37] omitted the factor $\sin\phi$ in (17) and (18), so that the "phasiness" q encountered in this and subsequent parts of this paper is smaller than that discussed in [21] and [37]. In those references, we were only concerned with properties not involving the precise positioning of speakers.

BBC data [22] shows that for $\phi = 30^0$, and central sound images ($|L| = |R|$), an interspeaker phase difference of $|\xi|$ up to 45^0 is "negligable", i.e.

$$q = \frac{2 \cdot \frac{1}{2} \cdot \sin 45^0}{1+1+2\cos 45^0} = \frac{1}{2}(\sqrt{2}-1) = 0.207$$

is the maximum for "negligable" effect. Similarly, $|\xi|$ of up to 90^0 was found to be "acceptable", i.e.

$$|q| = \frac{2 \cdot \frac{1}{2}}{1+1} = 0.500 \quad .$$

Thus we see that, according to these criteria,

$$|q| < 0.207$$

is "negligable" and

$$|q| < 0.500$$

is "acceptable", as reported earlier (in terms of $Q = 2q$) in [21] Appendix II. We caution the reader that interspeaker phase actually has no "minimum audible" value, and in some circumstances, and with suitably experienced listeners, values of $|q| < 0.05$ can be heard, depending on the program material and especially on the loudspeakers used.

Use of (18) will quickly show that a given interspeaker phase

difference ξ gives less phasiness at the edge of the stereo stage (i.e. $|L/R| << 1$ or $|R/L| << 1$) than at the centre. Thus for $\xi = 45^0$ and $|L/R| = 3$, we compute that

$$q = 0.124$$

which is smaller than the centre-stage value $q = 0.207$ for $\xi = 45^0$.

One of the most interesting aspects of interspeaker phase in stereo reproduciton is the effect of having speakers at a finite distance d. Putting the speed of sound = c and the frequency of a sound = F, the speaker proximity modifies the values of x_V and y_V as follows

$$x_V = \cos\phi \ (1 - \frac{jc}{2\pi Fd})$$

$$y_V = \sin\phi \ (\frac{L-R}{L+R})(1 - \frac{jc}{2\pi Fd})$$

due to the bass boost $1 - \frac{jc}{2\pi Fd}$ of velocity components of the sound field due to speaker proximity. For Makita and fixed head localization, we see that Re x_V is unchanged, but that Re y_V has the value

$$\text{Re } y_V = \text{Re } y_V^{\infty} + \frac{c}{2\pi Fd} \text{ Im } y_V^{\infty} \tag{19}$$

where y_V^{∞} is the value for infinite distance. Thus the finite-distance Makita localization is given by

$$\tan\theta_V = \tan\phi \ \{\frac{|L|^2-|R|^2+\frac{c}{\pi Fd}|L||R|\sin\xi}{|L|^2+|R|^2+2|L||R|\cos\xi}\} . \tag{20}$$

For example, if $|L| = |R|$ and $\xi = 90^0$ (i.e. left leads an equal right channel by 90^0, and if $c = 340$m/s, and the speaker distance d is 2m, then

$$\tan\theta_V = \tan\phi \ (\frac{27.06}{F})$$

where F is the frequency in Hz. For $F = 100$ Hz and $2\phi = 60^0$, this gives

a Makita azimuth equal to

$$\theta_V = 8.88^0$$

which gives an image displaced about 0.3 of the way torwards the left loudspeaker at 100 Hz. This image shift, as a proportion of subtended interspeaker angle, diminishes inversely proportional to speaker distance and inversely proportional to frequency. Nevertheless, it will be seen that interspeaker phase difference lead to significant displacements of the bass frequencies of sounds towards the phase-leading speaker. This phenomenon of shift towards phase leading speakers has been noted by Bauer et al. [38], although we do not claim that the proximity effect is the only mechanism involved. (Indeed, in [39] it is shown that similar shifts occur at higher frequencies where other effects must be responsible).

It might be argued that the effect on localization at such low frequencies is "unimportant", but we believe this not to be so insofar as the more things that are made correct the better. Also, the effect is significant below 300 Hz, i.e. over about half of the 700 Hz range over which low frequency localization theory is expected to be apt. Fortunately, it is easy to modify a stereo reproducer to avoid proximity effect by putting the difference L-R signal only (and not the sum signal L+R) through an RC high-pass filter with response

$$1 \; / \; (1 - \frac{jc}{2\pi Fd}) \qquad (21)$$

with -3dB point at 54/d Hz (d in metres). It will be seen that this restores y_V to its ideal y_V^∞ form. In practice, a fixed compensation corresponding say to d = 3m would give useful improvements and a more accurate stereo reproduction for all types of program. Most of the effect of the high-pass filter (21) is due to its effect on phase response rather than to the small effect on amplitude response.

ACKNOWLEGEMENTS

I would like to thank Jerry Bruck for drawing my attention to Tager's paper, Dr. Duane Cooper for correspondence and discussion that clarified many of these ideas, and Professor Peter Fellgett for his assistance and help in the experimental program of the N.R.D.C. ambisonic technology, which has helped refine these ideas from a general abstract scheme into a practical design tool.

REFERENCES

[1] O. Kohsaka, E. Satoh, and T. Nakayama, "Sound-Image Localization in Multichannel Matrix Reproduction", J. Audio Eng. Soc., vol. 20, pp. 542-547 (Sept. 1972)

[2] D.H. Cooper and T. Shiga, "Discrete-Matrix Multichannel Stereo", J. Audio Eng. Soc., vol. 20, pp. 346-360 (June 1972)

[3] P. Damaske and Y. Ando, "Interaural Crosscorrelation for Multichannel Loudspeaker Reproduction", Accustica, vol. 27, pp. 232-238 (1972)

[4] B. Bernfeld, "Simple Equations for Multichannel Stereophonic Sound Localization", J. Audio Eng. Soc., vol. 23, pp. 553-557 (Sept. 1975)

[5] M. Nishimaki and K. Hirano "Localization of Sound Sources in 4-channel Stereo", Characteristics of the Sansui QS Vario-Matrix based on a Psycho-acoustic Study of the Localization of Sound Sources in Four-Channel Stereo, Part 2, QS Regular Matrix System Technical Analyses D3 (distributed by Sansui Electric Co., Ltd).

[6] M.A. Gerzon, "Surround Sound Psychoacoustics", Wireless World, vol. 80, pp. 483-486 (Dec. 1974); originally published in French under the title "Criteres Psychoacoustiques relatif a la Conception des Systemes Matriciels et Discrets en Tetraphonie", Conf. des Journees d'Etudes, Festival International du Son, March 1974, Editions Radio, Paris 1974, pp. 145-155.

[7] M.A. Gerzon, "Criteria for Evaluating Surround Sound Systems", J. Audio Eng. Soc., vol. 25, pp. 400-408 (June 1977)

[8] A.D. Blumlein, U.K. Patent 394,325 (filed Dec. 1931, published 14 June, 1933)

[9] H.A.M. Clark, G.F. Dutton and P.B. Vanderlyn, "The Stereosonic Recording and Reproducing System", I.R.E. Trans. on Audio, vol. pp. 96-111, (1957).

[10] Y. Makita, "On the Directional Localization of Sound in the Stereophonic Sound Field", E.B.U. Review, part A no. 73, pp. 102-108 (1962)

[11] P.G. Tager, "Some Features of the Physical Structure of Acoustic Fields of Stereophonic Systems", J. Soc. MPTE, vol. 76, pp. 105-110 (1967)

[12] D.M. Leakey, "Some Measurements on the Effect of Interchannel Intensity and Time Differences in Two-Channel Sound Systems" J. Acous. Soc. Am., vol. 31, pp. 977-987 (1959)

[13] B. Bernfeld, "Attempts for Better Understanding of the Directional Stereophonic Listening Mechanism", Preprint, Audio Engineering Society 44th Convention, Rotterdam, Feb 1973.

[14] K. de Boer, "Stereophonic Sound Production", Philips Tech. Rev., vol. 5, pp. 107-144 (1940)

[15] B.B. Bauer, "Phasor Analysis of Some Stereophonic Phenomena", J. Acous. Soc. Am., vol. 33, pp. 1536-1539 (1961)

[16] J.W. Strutt (Lord Rayleigh), "On our Perception of Sound Direction", Phil. Mag., vol. 13, pp. 214-232 (1907)

[17] M.A. Gerzon, "Periphony: With-height Sound Reproduction", J. Audio Eng. Soc., vol. 21, pp. 2-10 (Jan. 1973)

[18] R.B. Parente, "Nonlinear Differential Equations and Analytic System Theory", SIAM J. Appl. Math., vol. 18, pp. 41-66 (1970)

[19] J.F. Barrett, "The use of Functionals in the Analysis of Nonlinear Physical Systems", J. Electronics and Control, vol. 15, pp. 567-615 (1963). Reprinted in Nonlinear Systems (ed. A.H. Haddad), Dowden, Hutchinson & Ross, Stroudsburg, Pensylvania, 1975.

[20] M.A. Gerzon, "Recording Techniques for Multichannel Stereo",

Brit. Kinematography, Sound and Telev., vol. 53, pp. 274-279 (July 1971)

[21] M.A. Gerzon, "A Geometric Model for Two-Channel Four-Speaker Matrix Stereo Systems", J. Audio Eng. Soc., vol. 23, pp. 98-106 (Mar. 1975)

[22] J.S. Bower, "The Subjective Effects of Interchannel Phase-shifts on the Stereophonic Image Localization of Wideband Audio Signals", BBC Research Department Report BBC RD 1975/27 (Sept. 1975).

[23] D.R. Brillinger, "An Introduction to Polyspectra", Ann. Math. Stat., vol. 36, pp. 1351-1374 (1965)

[24] P.J. Huber, B. Kleiner, T. Gasser and G. Dumermuth, "Statistical Methods for Investigating Phase Relations in Stationary Stochastic Processes", IEEE Trans. on Audio and Electroacoustics, vol. AU-19, pp. 78-86 (1971)

[25] M.A. Gerzon, "Nonlinear Models for Auditory Perception", (to be published)

[26] M.A. Gerzon, "Psychoacoustics - The Criteria of Hearing and Microphone Techniques", in: Sound From Microphone to Ear, British Kinematograph, Sound and Television Society Education & Training Committee, London, 1976.

[27] F. Winckel, "Music, Sound and Sensation", Dover, New York, 1967, pp. 12-23, 112-119.

[28] H. Wallach, E.B. Newman and M.R. Rosenzweig, "The Precedence Effect in Sound Localization", J. Audio Eng. Soc., vol. 21, pp. 817-826 (Dec. 1973). Reprinted from Am. J. Psychology, vol. 62, pp. 315-336 (1949).

[29] H. Haas, "The Influence of a Single Echo on the Audibility of Speech", J. Audio Eng. Soc., vol. 20, pp. 145-159 (Mar. 1972)

[30] S.S. Stevens and E.B. Newman, "Localization of Actual Sources of Sound", Am. J. Pshychol., vol. 48, pp. 297-306 (1936)

[31] A.W. Mills, "Auditory Localization", Chapter 8 of: Foundations of Modern Auditory Theory (ed. J.V. Tobias) vol II, Academic Press, New York, 1972.

[32] M.A. Gerzon, "Ambisonics, Part II. Studio Techniques", Studio Sound, vol. 17 no. 8, pp. 24,26,28-30 (Aug. 1975). Correction: ibid, vol. 17 no. 10, p. 60 (Oct. 1975)

[33] K. Nakabayashi, "A Method of Analyzing the Quadraphonic Sound Field", J. Audio Eng. Soc., vol. 23, pp. 187-193 (Apr. 1975)

[34] B. McA. Sayers and E.C. Cherry, "Mechanisms of Binaural Fusion in the Hearing of Speech", J. Acous. Soc. Am., vol. 29, pp. 973-987 (1957)

[35] P.A. Ratliff, "Properties of Hearing Related to Quadraphonic Reproduction", BBC Research Department Report, BBC RD 1974/38 (Nov. 1974)

[36] G. Thiele and G. Plenge, "Localization of Lateral Phantom Sources", J. Audio Eng. Soc., vol. 25, pp. 196-200 (Apr. 1977)

[37] M.A. Gerzon, "Compatible 2-channel Encoding of Surround Sound", Electronics Letters, vol. 11, pp. 615-617 (11 Dec. 1975)

[38] B.B. Bauer, D.W. Gravereaux and A.J. Gust, "A Compatible Stereo-Quadraphonic (SQ) Record System", J. Audio Eng. Soc., vol. 19, pp. 638-646 (Sept. 1971)

[39] J.S. Bower, "The Subjective Effects of Interchannel Phase-Shifts on the Stereophonic Image Localisation of Narrowband Audio Signals", BBC Research Department Report, BBC RD 1975/28 (Sept. 1975).

[40] A.H. Stroud, "Approximate Calculation of Multiple Integrals" Prentice Hall, Englewood Cliffs N.J., 1972, p. 302.

[41] M.A. Gerzon, "Optimal Reproduction Matrices for Multispeaker Stereo", Preprint 3180 of the 91st Audio Engineering Society Convention, New York (1991 Oct.)

[42] M.A. Gerzon, "Panpot Laws for Multispeaker Stereo", Preprint presented at the 92nd Audio Engineering Society Convention, Vienna, March 1992

[43] M.A. Gerzon, "Ambisonic Decoders for HDTV", to be presented at the 92nd Audio Engineering Society Convention, Vienna, March 1992

[44] M.A. Gerzon, "Hierarchical Transmission System for Multispeaker Stereo", Preprint 3199 of the 91st Audio Engineering Society Convention, New York (1991 Oct.)

[45] M.A. Gerzon, "Hierarchical System of Surround Sound Transmission for HDTV", to be presented at the 92nd Audio Engineering Society Convention, Vienna, March 1992

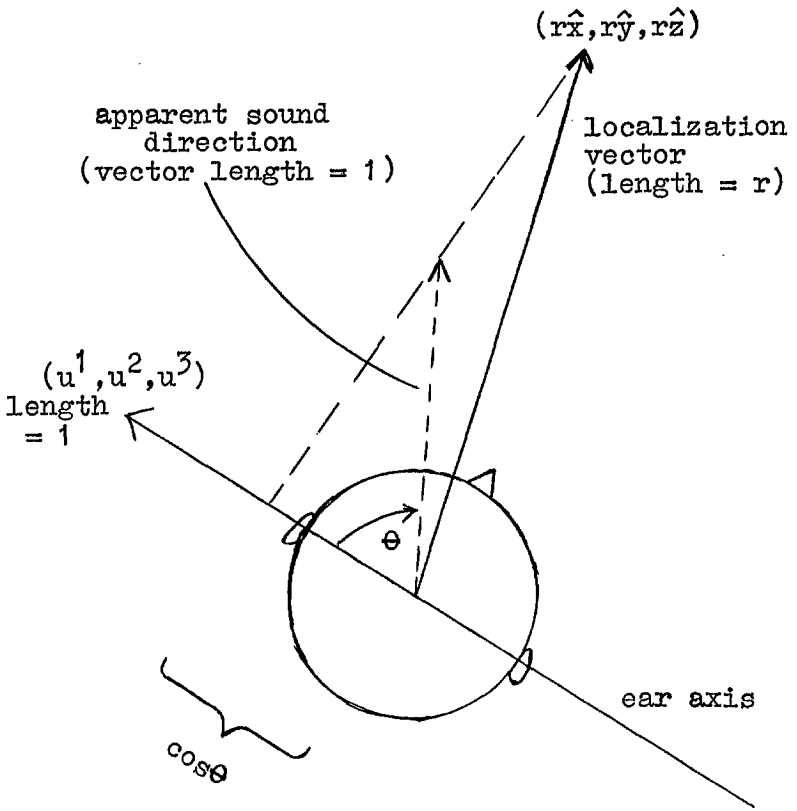

Figure 1. Fixed head sound localization θ .

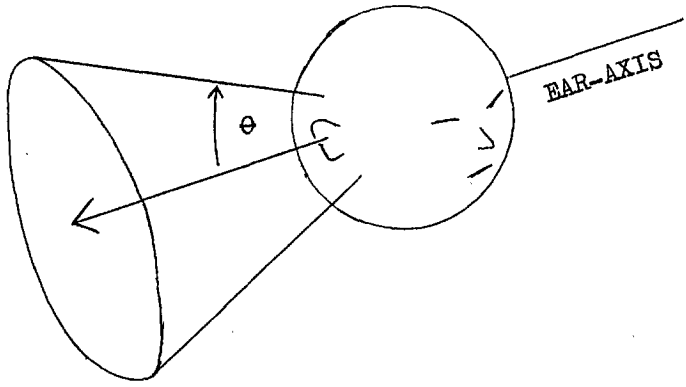

Figure 2. Ambiguity cone of directions at angle θ to ear axis. (See [31] for a further discussion.)

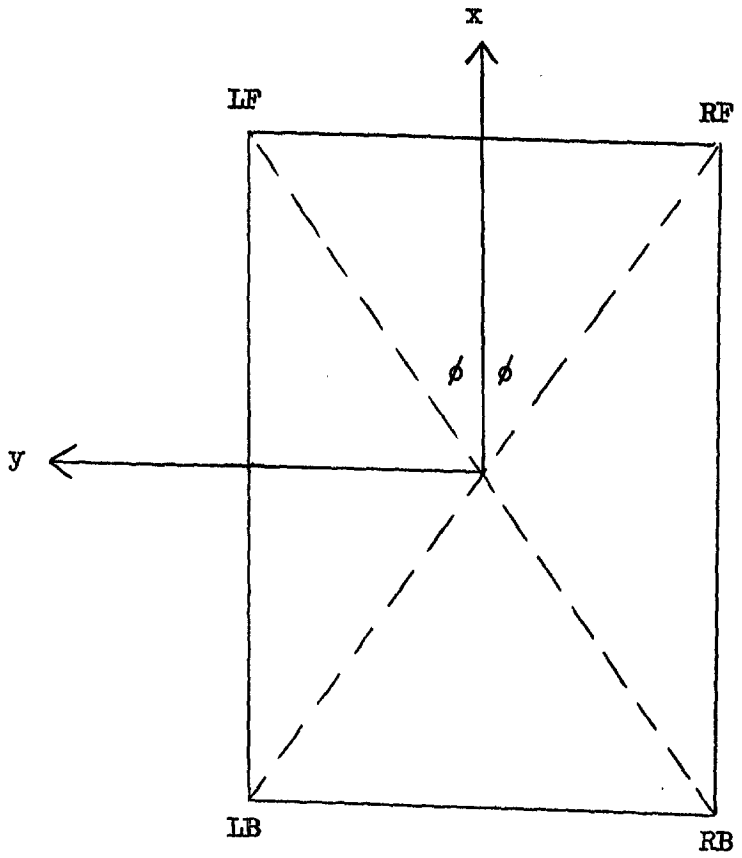

Figure 3. Rectangle layout used in this paper.

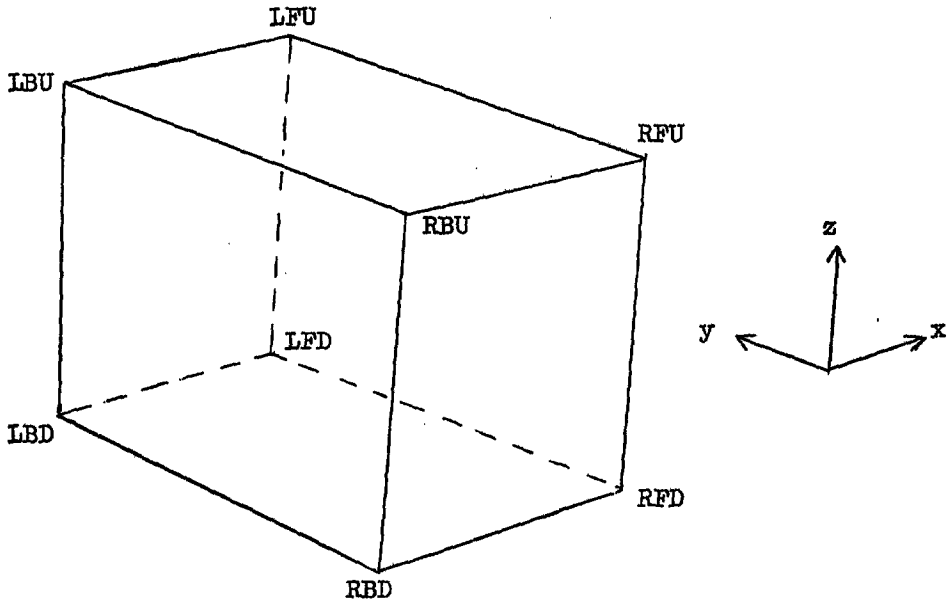

Figure 4. Rectangular cuboid loudspeaker layout, showing (x,y,z) axis directions.

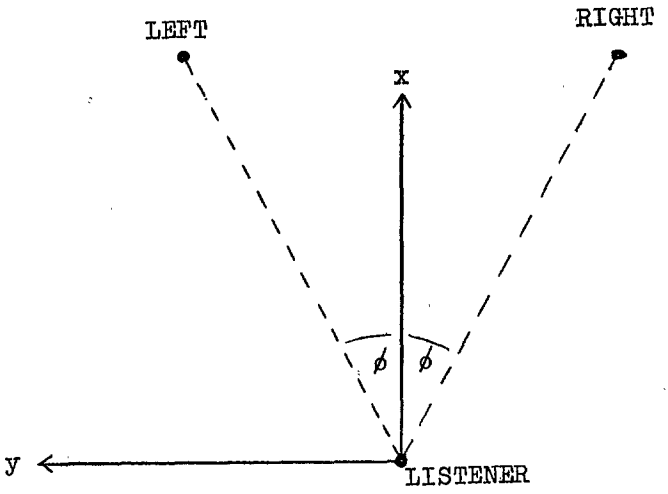

Figure 5. 2-speaker stereo sound reproduction, showing x- and y-axes.