Systems for Virtual Sound Imaging

P.A. Nelson, M. Shin, F.M. Fazi, T. Takeuchi and M.Park.
Institute of Sound and Vibration Research (ISVR),
University of Southampton, Highfield, Southampton SO17 1BJ, England.

J. Seo and K.O. Kang
Electronics and Telecommunications Research Institute (ETRI),
Yuseong-gu, Daejon, 305-700, Republic of Korea

Abstract*—This paper will describe recent developments in the science and technology of sound reproduction with an emphasis on the application of new methods for the generation of virtual (or "3D") sound images. The aim of the paper will be to evaluate the potential for application of these new technologies in modern communication systems. A brief summary will first be presented of the factors governing the human perception of sound source location through reference to recent computational models of binaural hearing. Conventional methods of "stereophonic reproduction" will first be reviewed and the limitations discussed of such two-channel techniques and their multi-channel extensions. The problem of binaural reproduction via loudspeakers will be described within the framework of the simple linear algebra associated with a two-input two-output system whose inversion enables the optimal design of cross-talk cancellation filters. The correct implementation of such filters enables the accurate delivery of acoustic signals to the ears of a listener. The influence of ill-conditioning of this system will be described together with the natural consequences for the distribution of acoustic sources as a function of frequency that ensures robust reproduction. A description will be presented that illustrates the remarkable potential offered by a strategy that involves the frequency dependent spatial distribution of acoustic source strength. The extension of these techniques to the generation of robust virtual images for multiple listeners will be discussed briefly. Alternative approaches to the reproduction of sound for multiple listeners will then be described, most of which rely on the reproduction of an acoustic field in its entirety over a defined spatial region, either through a knowledge of the values of the acoustic variables on the boundary of the region or through a knowledge of the natural basis functions used to describe the field within the region. The difficulty of ensuring reproduction with a sparse or non-uniform distribution of acoustic source strength will be outlined and recent work will be described that aims to overcome such problems by seeking to reproduce alternative acoustic field variables. (Abstract)

Keywords- sound; virtual; 3D; binaural; stereophony; multi-channel;ambisonics; wavefield.

I. INTRODUCTION

Methods are described here for producing "3D sound", or the perception by a listener of sound that appears to come from sources located at prescribed positions in the three dimensional space surrounding the listener. Generally, the acoustic signals generated at the listener's ears are manipulated to ensure that they replicate those signals that would be produced by a "virtual source" in the spatial position required. The methods used to accomplish this are reviewed briefly here. They include the use of headphones to deliver the ear signals, the use of a conventional stereo pair of loudspeakers, more recent methods using a number of loudspeaker pairs with suitably processed input signals and the use of arrays of loudspeakers enveloping the listener. These discussions are preceded by a brief introduction to the mechanism of binaural hearing, an understanding of which is essential to the effective design of systems for the production of virtual acoustic images. Some new work is also presented on the design of sparse and irregular loudspeaker arrays for reproducing sound fields over extended spatial regions.

II. CHARACTERISTICS OF BINAURAL HEARING

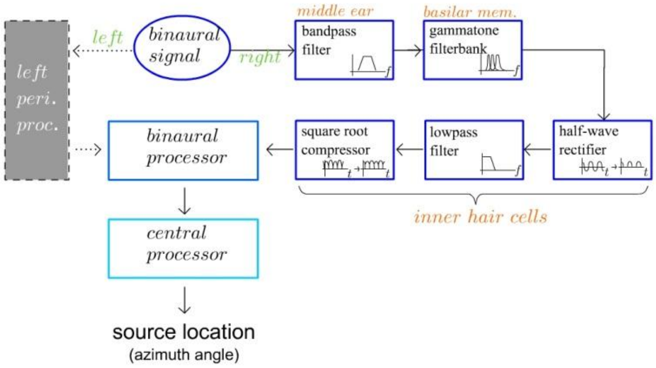


Figure 1. Block diagram of simple signal processing model of the auditory periphery [7]

It is helpful to first outline the factors that influence the human perception of the location of a sound source. Many of these, particularly those associated with the human auditory periphery, can be described quite readily in signal processing terms. The transfer function that characterises the relationship between the signal emitted by an acoustic source at a given position in space and the signal produced at the eardrum is known as the Head Related Transfer Function (HRTF). This transfer function is both linear and time invariant for fixed positions of source and listener's head and can be expressed, for example, by a series of FIR filters describing the transfer functions from the source to both ears of listener. A number of databases of such transfer functions have been measured and are available (see [1] for example). The inner ear, through which signals arriving at the eardrum are transmitted to the cochlea, can broadly be characterised as a band-pass filter that

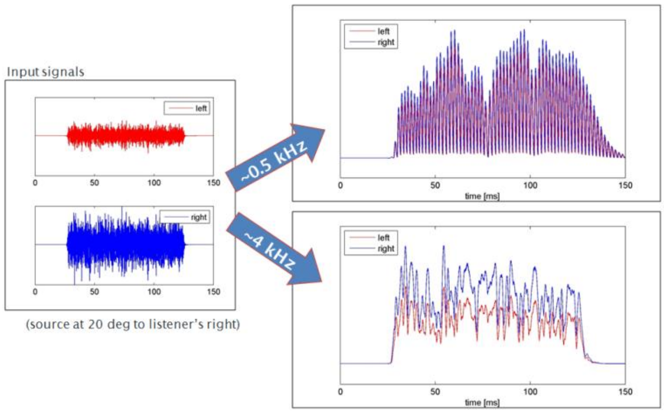


Figure 2. Illustration of the output signals from the auditory periphery (outputs from the "inner hair cells" shown in Figure 1.)

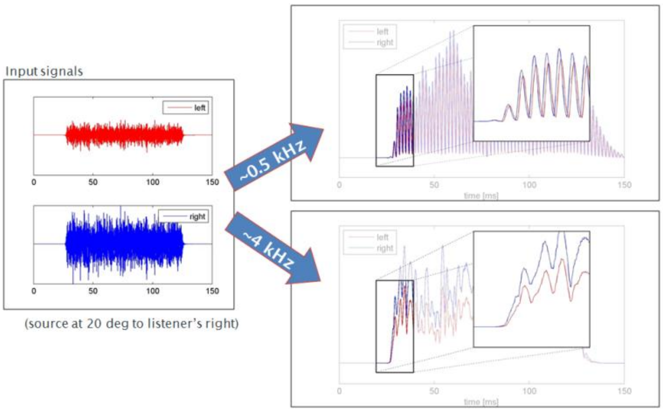


Figure 3. Details of parts of the signals illustrated in Figure 2

has the inverted form of the equal loudness contour [2]. The filtering action of the basilar membrane within the cochlea is often characterised by a series of band-pass filters [3], a popular representation being a series of gammatone filters (see [4] for example). Finally, the neural transduction undertaken by the organ of Corti can to some extent be modeled by a half wave rectifier and low-pass filter to represent the generation of neural impulses, and a square root compressor to approximate the input-output non-linearity of this process [5,6].

A block diagram of this signal processing scheme is shown in Fig. 1 which is based on the model described in [7]. Some typical outputs of this system are illustrated in Fig 2. which shows the model of the signals generated by the hair cells that are subsequently transmitted to the binaural processor. Fig 3 also shows an expanded version of sections of these signals, illustrating the extent to which the relative phase of the ear input signals is preserved at low frequencies, but also how the amplitude difference between the signals is more apparent at high frequencies. Importantly, this figure also illustrates how the time-differences between the envelopes of the ear input signals are also preserved at high frequencies. These factors are all known to be important in the human localization of acoustic sources.

The neural firing patterns from the left and right ears are combined in a binaural processor, the classical representation

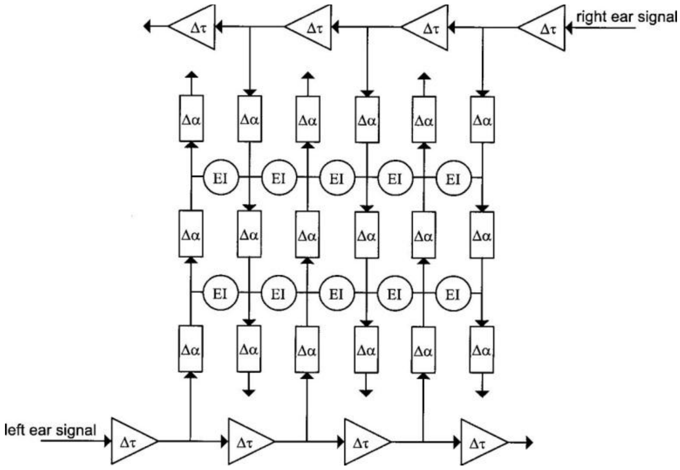


Figure 4. Schematic representation of the array of "EI cells" used to model the binaural processor [7]

of which [8] relies on the computation of inter-aural cross-correlation in order to determine the inter-aural time differences between the signals arriving at the ears, although such approaches do not account for inter-aural level differences, and the use of such models is still debated [9]. A more recent approach [10] to the representation of the binaural processor makes use of an elegant method for dealing with both inter-aural time and level differences. This relies on so-called equalisation-cancellation (EC) networks in which both "delay lines" and "attenuation lines" are represented. Fig. 4 illustrates such a network. The right and left neural signals in a given frequency band are both delayed and attenuated by a series of prescribed amounts and subtracted from one another in so-called excitation-inhibition (EI) cells, the output of which represents the difference between the input signals. The EI cells are arranged in two-dimensional array and the cells with the minimum output identify the most probable inter-aural time and level difference in the given frequency band. Overall, the output of the array of the EI cells can be thought of as defining an "EI pattern" that characterises the relative differences between the left and right ear signals in a given frequency band.

Again, the neural mechanisms used to interpret the binaural information provided by such EI patterns (assuming they exist at the higher levels of the auditory processing system) are far from understood. However, recent work [7] has demonstrated the success of a simple pattern matching procedure that compares, via cross-correlation, the EI patterns generated by an acoustic source in a given location and a series of template EI patterns generated by sources in a series of pre-determined locations. The output of this process is a probability function that represents the similarity between the target and template EI patterns in a given frequency band as a function of azimuthal direction. Simply put, the EI template providing the "best fit" to a given EI pattern is used to determine the location of the source.

III. Conventional Stereophony

Sound reproduction using two-channel stereophonic systems classically relies on the simple procedure of adjusting the relative gain of the identical input signals applied to a pair of loudspeakers positioned to the front of a centrally located

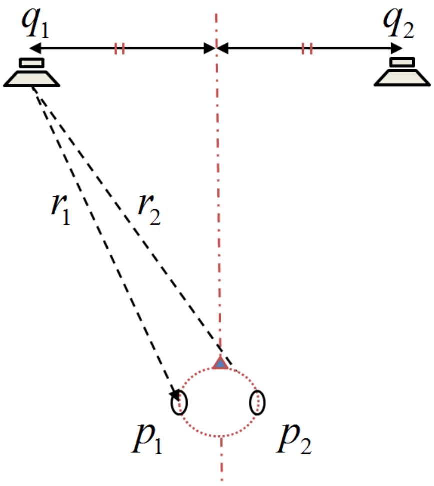


Figure 5. Symmetrical arrangement of two sources producing acoustic pressures p1 and p2 at the listener's ears (neglecting the scattering of the listener's head).

listener and subtending an angle of typically sixty degrees. An early history of the technique is given in [11], where it is also demonstrated that, at low frequencies in a harmonic sound field, the sound fields from the two loudspeakers interfere to give a phase difference at the ears of the listener. This phase difference results in the perception by the listener of a virtual (or phantom) source at an angular location between the two loudspeakers. The amplitude difference between the signals applied to the two loudspeakers can be adjusted to change the phase difference at the ears of the listener and thus the perceived location of the virtual source. This can be understood by following generally the analysis presented in [11]. Thus it is assumed that the sources and listener are arranged symmetrically as depicted in Fig 5. The relationship between the ear pressures p_1 and p_2 and the strengths (volume velocities) of the point monopole sources q_1 and q_2 can be written as

$$\begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = G \begin{bmatrix} 1 & ge^{-j\omega\tau} \\ ge^{-j\omega\tau} & 1 \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} \quad (1)$$

where ω is the angular frequency of time harmonic fluctuations, $G = j\omega\rho_0 e^{-j\omega r_1/c_0} / 4\pi r_1$, and ρ_0, c_0 denote the density and sound speed respectively. The term $g = r_1 / r_2$ is the ratio of distances of the "direct" and "cross-talk" paths from the sources to the ears and $\tau = (r_2 - r_1)/c_0$ is the difference in acoustic travel time between the two paths. If it is assumed that the two sources are in phase and differ only in amplitude, such that $q_2 = Kq_1$, where K is the gain, and furthermore, that the ratio g can be assumed to be unity, then it follows that

$$\frac{p_2}{p_1} = \frac{1 + Ke^{-j\omega\tau}}{e^{-j\omega\tau} + K} \quad (2)$$

The phase difference between the pressures at the two ears is given by in the inverse tangent of the ratio of the imaginary and real parts of this function and can be written as

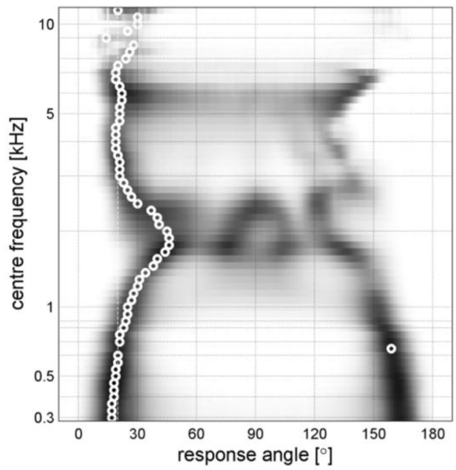


Figure 6. Output of the model of the binaural processor for a stereophonic image at 20 degrees from the normal to the front of the listener [7]

$$\varphi_{12} = \tan^{-1}\left(\frac{\sin\omega\tau - K^2 \sin\omega\tau}{2K + \cos\omega\tau + K^2 \cos\omega\tau} \right) \quad (3)$$

In the low frequency limit, such that $\omega\tau \to 0$, then it follows that

$$\varphi_{12} \approx \tan^{-1}\left(\frac{1-K}{1+K}\omega\tau \right) \approx \frac{1-K}{1+K}\omega\tau \quad (4)$$

Thus under the assumptions given, the simple act of changing the relative amplitude of the two sources fed with the same (in-phase) signal, produces a phase shift between the signals at the listener's ears. Such "amplitude panning" is generally assumed to be effective only at low frequencies, typically below about 700Hz, since at frequencies above this the inter-aural phase difference becomes ambiguous. However, at higher frequencies, the shadowing effect of the listener's head results in inter-aural level differences (ILDs) between the two ears that result in the perception of a virtual source [12, 13].

The model of binaural hearing described in Section II above has been applied to the evaluation of the stereophonic sound field and found to give an excellent representation of listener perception of source location [14, 15]. Simulation results [7] (see Fig 6) confirm the accurate location of virtual sources at low frequencies, with increasing deviations from the target image position as frequency increases, and the prediction of an overestimation of virtual source angular location between 1kHz and 3kHz. Above about 3kHz, the influence of the ILD reduces the bias in the estimation of the angular location of the virtual source, although the position of the virtual source image again becomes ambiguous at about 6kHz.

The predictions of this model have been broadly verified by a series of listening tests as reported in [14, 15]. The human subjects in the experiments used an electromagnetic tracking device to report the perceived position of virtual source images. The stimuli used were 1/3 octave bands of noise centred at 7 frequencies from 0.5 kHz to 6 kHz. The results of the tests shown in Fig. 4 are compared with the predictions of the model of binaural hearing and shown to be in good agreement.

Further details of the results are discussed in [7], but broadly one concludes that the model is a reasonable representation of the key elements of the auditory processes underlying the binaural localisation of sound.

It has proved tempting to extend the amplitude panning scheme associated with conventional two channel stereophony by surrounding the listener with multiple loudspeakers and simply activating the pair of loudspeakers between which one wishes to generate a virtual source. However, the amplitude panning approach, particularly for loudspeakers placed to one side of a listener, does not result in the same low frequency phase differences that are produced by a frontal pair of loudspeakers. Thus for a pair of loudspeakers placed symmetrically to one side of a listener, the difference in path lengths from one of the listener's ears to the two loudspeakers becomes very much less than is the case for a pair of loudspeakers to the front of the listener. Hence amplitude differences between the loudspeaker input signals are not so readily converted into phase differences between the ears. It has been shown conclusively in recent listening tests [15] that the image position may not be controlled by amplitude-panning, and subjects in the localisation experiments simply reported the position of the louder transducer as the location of the acoustic image.

IV. Binaural Reproduction

The binaural reproduction of sound is achieved by accurate replication of the signals at the ears of a listener that would have been produced at that listener's ears by an acoustic source at a prescribed spatial position. The implementation of such an approach improves greatly upon conventional stereophony and can in principle result in the perception by the listener of virtual sound images throughout the entirety of the surrounding three-dimensional space. One might expect that reproduction of the requisite ear signals by using a pair of headphones would be the solution of choice, especially since the headphones might be expected to provide a good environment for the accurate control of the desired ear signals. However, it has long been known (and is matter of common experience) that presentation of acoustic signals to the listener's ears via headphones or (in-ear transducers) generally results in the perception by the listener of the auditory image being "inside the head".

Many explanations for this phenomenon were investigated in early studies [16] although it is now thought to be due simply to the fact that, without any pre-processing of the input signals, headphones are not generally successful in replicating the signals at the listener's eardrums that would be produced by a source under free field listening conditions. It has been established [17-20] that if care is taken in presenting accurately the waveforms to the listener's ear-drums that would be generated under free field conditions, then the correct perception of an "externalised" image is produced. However, the HRTF is known to be highly variable from individual to individual, this variability mostly being due to the effect of diffraction by the outer ear (pinna). A knowledge of the HRTF appropriate to a given individual is therefore required if the headphone or earphone is to be fed with input signals that generate the correct ear-drum signals. Whilst it has been established that this can be deduced, for example, by laser scanning an individual pinna to establish the geometry, and that computational methods can then be used to establish the HRTF [21, 22], a fast and practical method for providing individual data has yet to be developed. Approaches that assume a certain HRTF (associated for example with that of a standard "dummy" head) have found to be partially successful in providing a degree of externalisation [23]. It is also well known [24, 25] that the addition of artificial room reverberation can enhance the degree of externalisation perceived.

V. Binaural Reproduction using Loudspeakers

The problem of producing externalised auditory images does not generally occur when the binaural signals are presented to the listener by a pair of loudspeakers to the front of a listener. However, there is the fundamental difficulty produced by the "cross-talk" signals generated at the left ear by the right loudspeaker and at the right ear by the left loudspeaker. The two by two matrix of transfer functions relating the loudspeaker input signals to the listener ear signals can be inverted by pre-processing the loudspeaker input signals by two by two matrix of "cross-talk cancellation" filters. Considerable work has been undertaken on the design of systems based on this approach [26-35] and the principle issues can readily be described with reference to the free field model of two monopole sources used above. Thus if $\mathbf{G}$ denotes the "plant" matrix relating the loudspeaker input signals to the listener ear signals, and $\mathbf{H}$ denotes the matrix of cross-talk cancellation filters, then one wishes to design the filters such that $\mathbf{GH} \approx \mathbf{I}e^{-j\omega d}$ where $\mathbf{I}$ is the identity matrix and d is a delay. Using the expression for the plant matrix given in equation (1), it follows that the expression for the filter can be written in terms of the inverse of the plant matrix such that

$$\mathbf{H} \approx \frac{e^{-j\omega d}}{G\left[\left(1-ge^{-j\omega\tau}\right)\left(1+ge^{-j\omega\tau}\right)\right]}\begin{bmatrix} 1 & -ge^{-j\omega\tau} \\ -ge^{-j\omega\tau} & 1 \end{bmatrix} \quad (5)$$

The plant matrix becomes ill-conditioned when the difference in path length between the two loudspeakers and one of the listener's ears is equal to one half of an acoustic wavelength (assuming a symmetrical arrangement of loudspeakers to the front of the listener). This results in the inverse filter matrix having a large gain when $\omega\tau = n\pi$ (i.e. at integer numbers of half wavelength path differences), with the denominator in equation (5) approaching zero. On the other hand, when the path length difference is one quarter of an acoustic wavelength, the matrix is very well conditioned. Full details of the analysis can be found in [36], including a formal consideration of the conditioning of this inversion problem using the singular value decomposition.

An approach to binaural synthesis that ensures optimal conditioning at all frequencies is provided by the Optimal Source Distribution (OSD). Conceptually, this is provided by a pair of continuous distributions of monopole source strength radiating sound at a frequency of the sound that depends upon spatial position. The frequency radiated by each element of the pair of source distributions is determined to ensure that there is always a one-quarter wavelength path length difference between source elements and the ears of the listener. This means that the angular separation of the loudspeakers becomes smaller as frequency becomes higher. This also ensures that n is an odd integer number at all frequencies (except at very low frequencies) and that the singular values of the plant matrix are equal [35]. Under these circumstances, of a one-quarter wavelength path difference, it also follows that $\tau = \pi / 2\omega$, and the expression for the inverse filter matrix simplifies to

$$\mathbf{H} = \frac{e^{-j\omega d}}{G(1+g^2)}\begin{bmatrix} 1 & jg \\ jg & 1 \end{bmatrix} \quad (6)$$

Thus the cross-talk cancellation is produced simply by a 90 degree phase change in the cross-talk path in the inverse filter matrix without any change in amplitude response. The frequency response of the inverse filter is thus the same at all frequencies. Since the sound is always synthesised by constructive interference at all frequencies, there is no dynamic range loss or loss of quality compared to the case without system inversion. Thus the OSD can be thought of as providing "lossless" cross-talk cancellation.

Obviously it is not easy in practise to build a pair of distributed transducers that realise such a continuous distribution of acoustic source strength. However, a suitable discretisation of the distribution into (say) three or four pairs of loudspeakers has been found to give excellent results in practise [37]. Whilst in principle, the details of the HRTF comprising the two-by-two plant matrix are also required to be known, it has also been found in practice that the HRTFs associated with a particular individual can be substituted by generic HRTFs associated with a dummy head [30, 37] whilst still producing convincingly externalised images. This approach to binaural reproduction has a number of other advantages. The sound radiated by the OSD is always smaller in directions other than those corresponding to the listener, and is also smaller than the sound radiated by a single monopole transducer producing the same sound level at the listener's ears. This therefore results in a system that has a good signal to noise ratio, reduced distortion, and which is robust to reflections in a reverberant environment. Furthermore, the radiation pattern becomes constant as a function of frequency and repeats periodically in the listening space. This offers the possibility of the perception of nearly correct binaural signals by multiple listeners [35]. The inverse filters have a flat frequency response so there is no coloration at any location in the listening room. When the listener is far away from the intended listening position, the spatial information perceived may not be ideal. However, the spectrum of the sound signals is not changed by the inverse filters and therefore a listener will continue to perceive correctly reproduced sound. It has also been recognised that the performance of the OSD can be improved still further, especially at low frequencies, with the addition of a third centrally located loudspeaker channel [35].

VI. Sound Field Reproduction

A number of other approaches have been taken to the generation of virtual acoustic images for multiple listeners. These generally rely on the use of multiple transducers to generate an interference field that replicates as closely as possible, over a spatial region that is as large as possible, the field generated by a given virtual source. Such an approach is provided by "Wave Field Synthesis" described in detail, for example, in [38-42]. The approach is based on the Kirchhoff-Helmholtz integral equation which describes the sound field in a spatial volume in terms of the pressure and pressure gradient on the surface surrounding the volume. Thus the acoustic pressure field inside the volume V is described in terms of the integral over the bounding surface S such that

$$p(\mathbf{x}) = \int_S \left[g(\mathbf{x}|\mathbf{y})\nabla_y p(\mathbf{y}) - p(\mathbf{y})\nabla_y g(\mathbf{x}|\mathbf{y}) \right] \cdot \mathbf{n} dS \quad (7)$$

where $\mathbf{x}$, $\mathbf{y}$ are position vectors, the operator ∇_y denotes the gradient operator with respect to the y coordinate, $\mathbf{n}$ is the unit vector perpendicular to S at $\mathbf{y}$ and g denotes the free-field Green function (note that $g = G / j\omega\rho_0$) The assumption made in Wave Field Synthesis is that the bounding surface is assumed to be planar, in which case the Kirchhoff-Helmholtz integral reduces to Rayleigh's second integral [39] which allows the pressure within the "volume" (to one side of the planar bounding surface) can be determined from a knowledge of the pressure on the surface. The principle therefore suggests that the measurement or computation of the pressure on the surface allows the determination of the source strength distribution on the surface that will enable reproduction of the field.

Another approach based on the Kirchhoff-Helmholtz integral is that proposed in [43] in which sources outside of the volume V are used to reconstruct the pressure and pressure gradient on the surface S, thereby ensuring correct reproduction inside V. Of course in practise it is impossible to sense both pressure and pressure gradient continuously over the bounding surface and discrete measurement points are necessary to describe both the pressure and the pressure gradient, the latter in principle being measurable by a pair of microphones spaced apart by a suitable fraction of the acoustic wavelength. It has also been shown [44-46] that it is possible to simply reconstruct only the pressure (or indeed the velocity) on the bounding surface S, these parameters describing uniquely the sound field inside a source free volume except at the eigen-frequencies (or resonant frequencies) of that volume (i.e. the Dirichlet eigenvalues in the case of pressure and the Neumann eigenvalues in the case of velocity). The technical feasibility of the approach to reconstructing the acoustic pressure on the bounding surface has been clearly demonstrated in [47].

A further well-known technique is that known as "Ambisonics". This was first proposed in the early 1970's [48,49] and since has been extended to so-called "Higher Order Ambisonics" (HOA) [50-53]. This approach is based on undertaking a spherical harmonic analysis of the field to be reproduced, the spherical harmonics providing a means of describing a three dimensional sound field in terms of natural spatial basis functions. An attempt is made to reproduce these functions by a series of loudspeakers surrounding the region in question. The accuracy of the spatial reproduction generally increases with the order of the spherical harmonics that are reproduced and, broadly speaking, this in turn implies that the number of loudspeakers required also increases.

Another approach, described previously in [54], is simply to find the source strengths (or loudspeaker signal inputs) that provide the best fit of the reproduced sound field to the desired, or target, sound field associated with the virtual source to be simulated. Classical least squares techniques can be used to define the optimal source strengths necessary to minimise a cost function based, for example, on the sum of the squared differences between the desired and reproduced acoustic pressures [55]. This approach provides a numerical approach to the solution of the "inverse problem" of determining the optimal source strengths and is not restricted to particular geometrical arrangements of sources or field points to be controlled. The approach has been studied extensively in connection with the active control of sound and vibration and provides the basis for the discussion that follows. Other features of the sound field reproduction problem have been discussed in [56-60].

VII. Sound Field Reproduction using Sparse and Irregullar Loudspeaker Arrays

The approach taken in [44-47] was to find the source strengths necessary to ensure the reproduction of the acoustic pressure on the surface that bounds the volume in which reproduction is sought. Satisfactory reproduction of the field within the enclosed volume is, of course, strictly only possible at frequencies that do not coincide with the eigen-frequencies of that volume. Whilst it has been demonstrated, both by computer simulation and by experiment, that this is an entirely satisfactory approach when the sources that surround the volume that are used for reproduction are spaced in a regular layout, it has also been recognized that this approach tends to fail if the loudspeakers used are arranged in a sparsely populated or irregular array. In such cases there is a tendency for the source strengths to "blow up" and produce exceptionally large outputs, and whilst the field within the chosen volume can still be reproduced with some accuracy, the field elsewhere can have far from desirable characteristics.

This observation will be illustrated with the results of some numerical simulations presented below. The approach taken is to simulate the reproduction process again by using a discrete number of sources and attempting to reproduce the acoustic pressure at a discrete number of points on the surface that bounds the volume. This approach will be compared with an alternative method that attempts to reproduce the acoustic particle velocity (a vector quantity) at a discrete number of points on the bounding surface. In both cases the conventional "classical" least squares approach will be used. It will be shown that the reproduction of velocity has a number of desirable characteristics.

As illustrated in Figure 7, the source strengths $\mathbf{q}$ and reproduced pressures $\hat{\mathbf{p}}$ can be defined in terms of the complex vectors given by

$$\mathbf{q}^T = \left[q(\mathbf{y}_1) \quad q(\mathbf{y}_2) \quad \cdots \quad q(\mathbf{y}_N) \right], \tag{8}$$

$$\hat{\mathbf{p}}^T = \left[\hat{p}(\mathbf{x}_1) \quad \hat{p}(\mathbf{x}_2) \quad \cdots \quad \hat{p}(\mathbf{x}_M) \right], \tag{9}$$

where N and M define the total number of sources (loudspeakers) and control points respectively. The acoustic pressures induced by the source strengths can be represented by

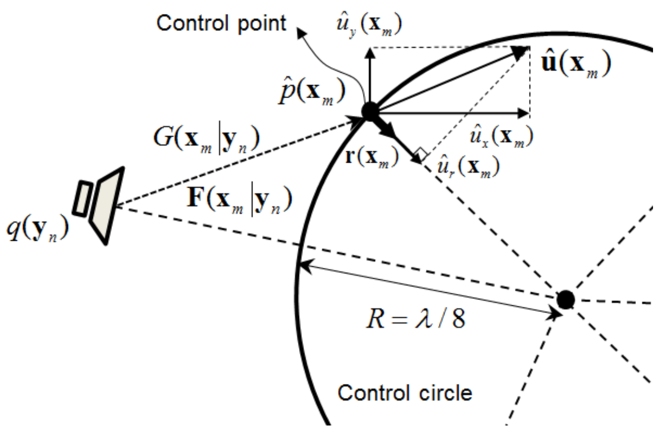


Figure 7. Definition of acoustic parameters in an acoustic field: $q(\mathbf{y}_n)$ is the volume velocity of the source at the loudspeaker location $\mathbf{y}_n$, $\hat{p}(\mathbf{x}_m)$ is the reproduced pressure at the control point location $\mathbf{x}_m$, $\hat{\mathbf{u}}(\mathbf{x}_m)$ is the velocity vector at $\mathbf{x}_m$, $G(\mathbf{x}_m|\mathbf{y}_n)$ and $\mathbf{F}(\mathbf{x}_m|\mathbf{y}_n)$ are the transfer functions of pressure and velocity respectively from the source at $\mathbf{y}_n$ to the control point at $\mathbf{x}_m$, $\hat{u}_x(\mathbf{x}_m)$, $\hat{u}_y(\mathbf{x}_m)$ and $\hat{u}_r(\mathbf{x}_m)$ are x, y and radial components of velocity vector $\hat{\mathbf{u}}(\mathbf{x}_m)$, $\mathbf{r}(\mathbf{x}_m)$ is the radial unit vector at $\mathbf{x}_m$, R is the radius of the control circle and λ is the wavelength of target wave field.

$$\hat{p}(\mathbf{x}_m) = \sum_{n=1}^{N} G\left(\mathbf{x}_m|\mathbf{y}_n\right) q\left(\mathbf{y}_n\right), \tag{10}$$

where $G\left(\mathbf{x}_m|\mathbf{y}_n\right)$ is the transfer function relating the pressure at the control point $\mathbf{x}_m$ to the strength of the source at $\mathbf{y}_n$. The relation between pressure values and source strengths can be expressed in the matrix form given by equation (11) below

$$\hat{\mathbf{p}} = \mathbf{G}\mathbf{q}, \tag{11}$$

where
$$\mathbf{G} = \begin{bmatrix} G(\mathbf{x}_1|\mathbf{y}_1) & G(\mathbf{x}_1|\mathbf{y}_2) & \cdots & G(\mathbf{x}_1|\mathbf{y}_N) \\ G(\mathbf{x}_2|\mathbf{y}_1) & G(\mathbf{x}_2|\mathbf{y}_2) & \cdots & G(\mathbf{x}_2|\mathbf{y}_N) \\ \vdots & \vdots & & \vdots \\ G(\mathbf{x}_M|\mathbf{y}_1) & G(\mathbf{x}_M|\mathbf{y}_2) & \cdots & G(\mathbf{x}_M|\mathbf{y}_N) \end{bmatrix}.$$

In the same way, the velocity vectors in the reproduced field can be defined as

$$\hat{\mathbf{u}}(\mathbf{x}_m) = \sum_{n=1}^{N} \mathbf{F}\left(\mathbf{x}_m|\mathbf{y}_n\right) q\left(\mathbf{y}_n\right) \tag{12}$$

where $\mathbf{F}\left(\mathbf{x}_m|\mathbf{y}_n\right)$ is the transfer function relating the acoustic particle velocity at the control point at $\mathbf{x}_m$ due to the source at $\mathbf{y}_n$. If it is assumed the sources are simple monopoles in a free field, the transfer functions G and $\mathbf{F}$ can be represented respectively by [61],

$$G\left(\mathbf{x}_m|\mathbf{y}_n\right) = j\omega\rho_0 \frac{e^{-jkr}}{4\pi r}, \tag{13}$$

$$\mathbf{F}\left(\mathbf{x}_m \middle| \mathbf{y}_n\right) = \frac{ke^{-jkr}}{4\pi r}\left(j + \frac{1}{kr}\right)\mathbf{a}_{m,n}, \qquad (14)$$

where $r = |\mathbf{x}_m - \mathbf{y}_n|$ and the unit vector $\mathbf{a}_{m,n}$ is defined by

$$\mathbf{a}_{m,n} = \frac{\mathbf{x}_m - \mathbf{y}_n}{|\mathbf{x}_m - \mathbf{y}_n|}. \qquad (15)$$

With the parameters defined above, the cost function to be minimized using the conventional approach of minimizing the sum of squared differences between the reproduced pressures at the control points and the target pressures at the control points defined by the vector of field $\mathbf{p}$ is given by

$$J = \|\mathbf{p} - \hat{\mathbf{p}}\|^2 + \beta_G \|\mathbf{q}\|^2 = \|\mathbf{p} - \mathbf{G}\mathbf{q}\|^2 + \beta_G \|\mathbf{q}\|^2. \qquad (16)$$

where β_G is a regularization parameter. The solution for the source strengths that minimizes this cost function is given by

$$\mathbf{q} = \left(\mathbf{G}^H\mathbf{G} + \beta_G\mathbf{I}\right)^{-1}\mathbf{G}^H\mathbf{p}. \qquad (17)$$

A convenient approach to dealing with the minimization of the velocity cost function is to work with the radial velocity vector normal to the surface bounding the volume in which reproduction is sought. It is also convenient to define the unit radial inward vector $\mathbf{r}$ normal to the surface. The radial velocity component $\hat{u}_r$ of $\hat{\mathbf{u}}$ is related to the source strengths by

$$\hat{u}_r(\mathbf{x}_m) = \sum_{n=1}^{N} F_r\left(\mathbf{x}_m \middle| \mathbf{y}_n\right) q\left(\mathbf{y}_n\right), \qquad (18)$$

where $F_r\left(\mathbf{x}_m \middle| \mathbf{y}_n\right) = \frac{ke^{-jkr}}{4\pi r}\left(j + \frac{1}{kr}\right)\mathbf{a}_{m,n} \cdot \mathbf{r}(\mathbf{x}_m)$.

The relation between radial velocity components and source strengths can be expressed in the matrix form given by

$$\hat{\mathbf{u}}_r = \mathbf{F}_r\mathbf{q}, \qquad (19)$$

where the matrix is defined by

$$\mathbf{F}_r = \begin{bmatrix} F_r(\mathbf{x}_1|\mathbf{y}_1) & F_r(\mathbf{x}_1|\mathbf{y}_2) & \cdots & F_r(\mathbf{x}_1|\mathbf{y}_N) \\ F_r(\mathbf{x}_2|\mathbf{y}_1) & F_r(\mathbf{x}_2|\mathbf{y}_2) & \cdots & F_r(\mathbf{x}_2|\mathbf{y}_N) \\ \vdots & \vdots & & \vdots \\ F_r(\mathbf{x}_M|\mathbf{y}_1) & F_r(\mathbf{x}_M|\mathbf{y}_2) & \cdots & F_r(\mathbf{x}_M|\mathbf{y}_N) \end{bmatrix}$$

The cost function to be minimized for the velocity least square minimization to the target radial velocity $\mathbf{u}_r$ is represented by,

$$J = \|\mathbf{u}_r - \hat{\mathbf{u}}_r\|^2 + \beta_F \|\mathbf{q}\|^2 = \|\mathbf{u}_r - \mathbf{F}_r\mathbf{q}\|^2 + \beta_F \|\mathbf{q}\|^2. \qquad (20)$$

And the optimal solution for the minimization problem is given by

$$\mathbf{q} = \left(\mathbf{F}_r^H\mathbf{F}_r + \beta_F\mathbf{I}\right)^{-1}\mathbf{F}_r^H\mathbf{u}_r. \qquad (21)$$

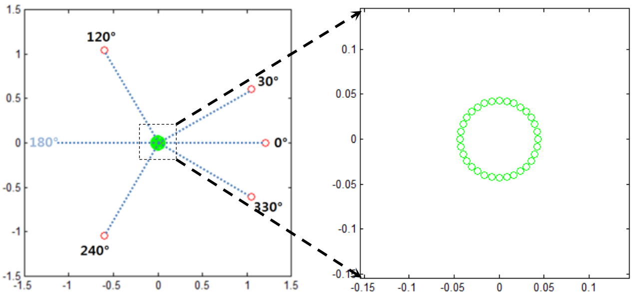


Figure 8. ITU 5.1 channel source distributions depicted with red circles and regularly distributed 32 control points represented with green circles.

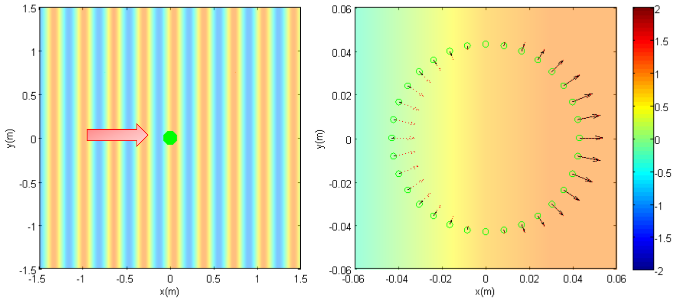
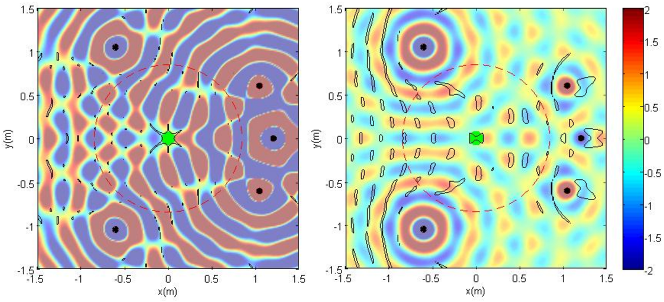


Figure 9. Target pressure field (1000Hz) from the backward (180 degree) direction and the velocities (black solid arrows) and intensities (red dotted arrows) at the control points.

A sound reproduction system with multiple loudspeakers arranged in a two-dimensional array has been investigated with respect to the efficiency of both pressure and radial velocity control methods. The simulation results give several clear indications that the proposed velocity control method provides benefits when the loudspeakers are irregularly arranged. Multiple loudspeakers in almost all standards are arranged irregularly and here the ITU 5.1 channel configuration has been chosen for illustration, especially since it is one of the best known standards realizing sound reproduction with five loudspeakers. Figure 8 shows the location of the sources (depicted with red circles) and control points (green circles) on the “control circle” surrounding the volume of to be controlled. A total of 32 control points are regularly distributed on the control circle. An important point to make is that the control circle is chosen to have a radius that is made *frequency dependent*. It has been shown previously [47] that using a frequency dependent control volume results in source strengths whose outputs do not show rapid changes as a function of frequency and this results in highly desirable characteristics of the filters used to process the source input signals. The radius of the control volume studied here is thus chosen to be one eighth of an acoustic wavelength (λ/8) at all frequencies.

The results of the simulations are illustrated here by using a plane wave target pressure field at a single frequency of 1000Hz. The plane wave is presented from the backward direction (i.e. from the direction of the pair of loudspeakers to the rear of the listener). For this target pressure field, the radial components of velocities (black solid arrows) and acoustic intensities (red dotted arrows) at all of the control points are represented in Figure 9.



(a) (b)

Figure 10. Reproduced pressure field controlled by pressure (a) and velocity (b) control methods.

Figure 10 shows the reproduced sound fields generated by the pressure and velocity control methods respectively using the regularization factors β_G=7.3×10^{-5} and β_F= 1.0×10^{-7}. The optimal source strengths associated with the pressure control method shows the excessive power problems which make the system give undesirable results. The optimal source strengths obtained by the velocity control method solves the excessive power problem and also generates stable and reasonable pressure field with much less energy.

The source strengths with all of the five loudspeaker channels and the total energy with respect to target pressure field angles over the full range of incidence angles have been computed at the same frequency. The resulting panning functions are represented in Figure 11. The magnitude and phase of each of the channels are overlaid with the L-2 norm of all of the sources, $\|\mathbf{q}\|$ the latter giving an indication of the total energy used in reproduction. The latter is depicted with black dashed lines over the 0°~360 ° ranges of target field incidence angles.

Based on the panning functions of the pressure control method, it is obvious that the optimal source strengths show excessive power whenever the target plane is arriving from a direction for which the source array is sparsely populated, such as 30°~120°, 120°~240° and 240°~330°. If the target plane wave comes from a densely populated region area -30°~30°, or from the angle of source locations such as 120° and 240°, the optimal solutions of the conventional pressure control method are stable without excessive powers. However, the panning function associated with the velocity control method produces stable energies and reasonable source distributions throughout the entire range of incidence angles.

Compared with the conventional pressure control method, the proposed velocity control method has no excessive power problem and produces less and evenly distributed stable energy as an optimal solution. In addition, changes of phase angle in the source strength associated with the velocity control method are less pronounced than for those arising from the pressure control method.

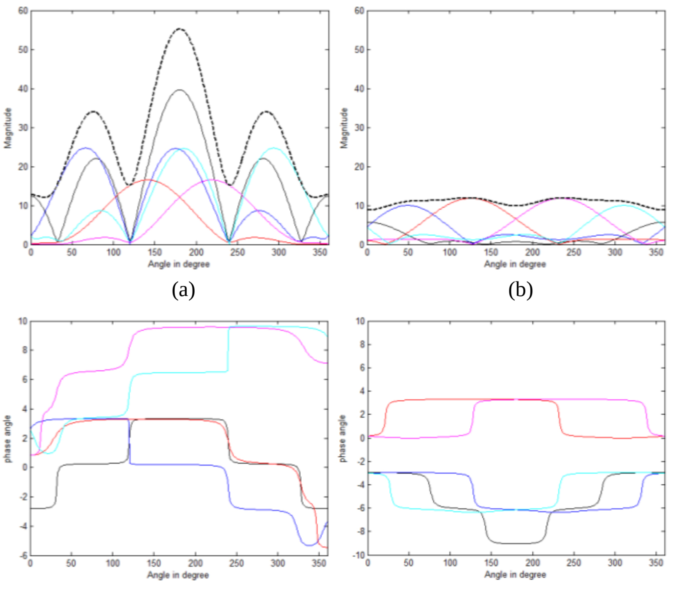
Based on the reproduced sound pressure and the particle velocity calculated from the pressure difference method [62], the sound intensity flow diagram can be obtained as shown in Figure 12. In this figure, only the direction of the intensity flow is illustrated since at each point the magnitude has been normalised. The intensity flow resulting from the pressure control method deviates considerably from the target intensity flow. The intensity flow error between target and reproduced fields may cause a deterioration of the perceptual localization performance. The proposed velocity control method appears to produce a better result in terms of intensity flow, but a better quantitative evaluation of the respective performances of the two methods is given by the Intensity Flow Error (IFE) defined as

$$IFE(\%) = \frac{\mathrm{mod}\left(\left|\theta_t - \theta_r\right|, \pi\right)}{\pi} \times 100, \quad (22)$$

where θ_t and θ_r are incident angles of target and reproduced intensity vectors and mod(A,B) is the modulus after division A/B. The IFEs associated with the pressure and velocity control methods are shown in Figure 13. Based on the IFE plots, it is clear that the proposed velocity control method gives better intensity flow than the conventional pressure control method. A “region of interest” can be defined that is within 0.7 of the distance between loudspeakers and the center of the system. This is depicted with red dashed lines in Figure 13. (Note that this region is much larger than the control circle upon which reproduction is sought). A “sweet area” can then be defined in percentage terms as the ratio between the area having less than 20% IFE and the total area within the region of interest.

The exact sweet area for Figure 13 (a) is 5% and for (b) is 42.2%. In order to identify the performance enhancement by the velocity control method, the changes of sweet areas with respect to the target incidence angle have been represented in Figure 14 for both pressure and velocity control methods. Based on the sweet area curves, it is obvious that the excessive power problem greatly reduces the sweet areas in the case of pressure control method. The velocity control method has much wider sweet areas, especially when the target plane wave is presented from directions that are sparsely populated with loudspeakers. Consequently, based on the intensity flow analyses, the proposed velocity control method appears to have some advantages over the conventional pressure control method, especially in terms of IFE and when multiple loudspeakers are arranged irregularly.

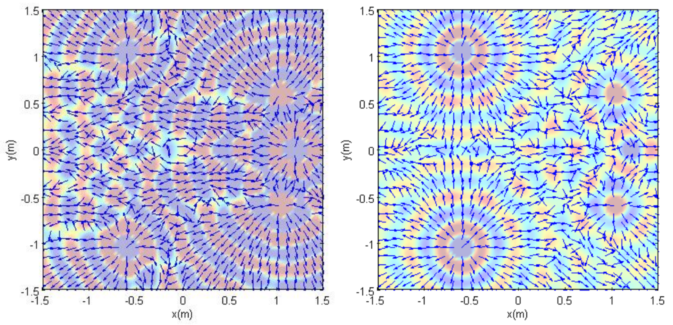
The frequency depenednce of the two techniques have been investigated by computing the filters that would be neecsseary to process the input signal associated with a target plane wave in order to deduce the source strength signals. The filters in Figure 15 within the frequency range from 0 to 3000 Hz have been obtained for both the pressure and velocity control methods when the target pressure field comes from the backward (180°) direction. Note that these filters have been equalized by a $j\omega\rho_0$ factor. The total energies (L-2 norm) depicted with black dotted lines of both control method are also represented. The filters generated by the pressure control method suffer from the excessive power problem within the crucial frequency range around 1000 Hz and above. However, filters generated by the proposed velocity control method are much flatter than the conventional method even if they show some roll-off in the very low frequency range around 10 Hz.



(a) (b)

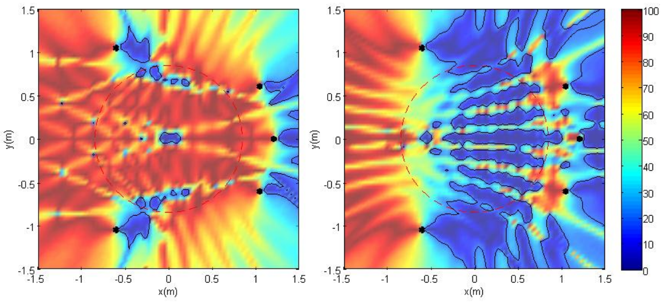
(c) (d)

Figure 11. Panning functions and corresponding phase angles obtained by pressure (a, c) and velocity (b, d) control methods (black dotted lines: L^2 norm, solid lines: channel 1 (black), channel 2 (blue), channel 3 (red), channel 4 (magenta) and channel 5 (cyan)).



(a) (b)

Figure 12. Intensity flows based on the pressure fields obtained by pressure (a) and velocity (b) control methods.



(a) (b)

Figure 13. Intensity flow error by pressure (a) and velocity (b) control methods.

The excessive power problem with the pressure control method with an irregular loudspeaker layout can to some extent be improved by using a change of regularization which must therefore become an angular dependent parameter, with a larger value of β used for target fields from sparsely populated directions, and a smaller β chosen for more to densely populated directions [63]. However, the proposed velocity control method could be made to function with a regularization factor that was independent of target incidence angle.

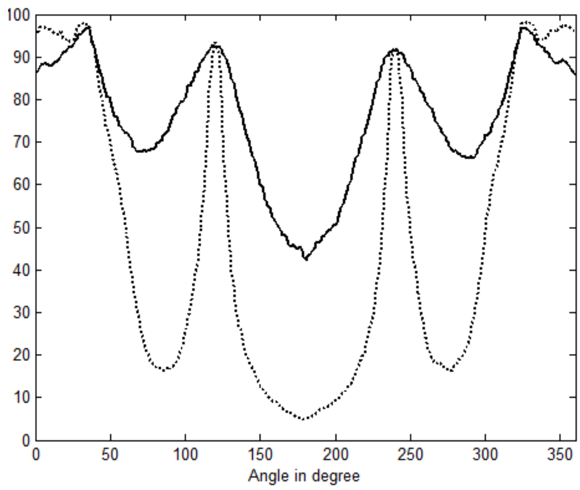
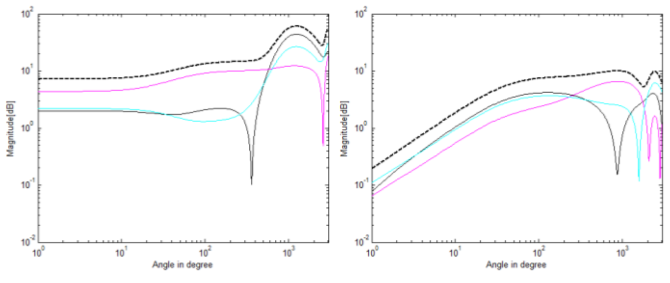


Figure 14. Change of sweet area by pressure (dotted lines) and velocity (solid line) control method.



(a) (b)

Figure 15. Filters obtained by pressure (a) and velocity (b) control methods when the target pressure field from the backward (black dotted lines: L^2 norm, solid lines: channel 1 (black), channel 2 (blue), channel 3 (red), channel 4 (magenta) and channel 5 (cyan)).

VIII. Conclusions

Methods for producing virtual sound images have been reviewed briefly, starting with conventional stereo methods, discussing more recently developed techniques based on binaural reproduction, and finishing with approaches based on multichannel loudspeaker arrays. Some new work has been presented that illustrates the difficulties in reproducing sound over spatial volumes when sparse and irregular loudspeaker arrays are used to control the pressure field on the surface surrounding a reproduction volume. Results suggest that many of these difficulties can be overcome by controlling the acoustic particle velocity on the surface of the volume.

REFERENCES

[1] H. Moller, "Fundamentals of Binaural Technology", Applied Acoustics, 36, 171-218 (1992).

[2] B.C.J. Moore, B.R. Glasberg, and T. Baer, "A model for the prediction of thresholds, loudness, and partial loudness", Journal of the Audio Engineering Society, 45, 224-240 (1997).

[3] R.M. Stern, and C. Trahiotis, "Models of binaural perception", in Binaural and spatial hearing in real and virtual environments, Gilkey, R.H., and Anderson, T.R., Eds., Lawrence Erlbaum Associates, New Jersey, 499-531 (1997).

[4] M. Slaney, "An efficient implementation of the Patterson-Holdsworth auditory filter bank", in Apple Computer Technical Report, Apple Computer, Inc. (1993).

[5] R.D. Patterson, M.H. Allerhand, and C. Giguere, "Time-Domain Modeling of Peripheral Auditory Processing - a Modular Architecture and a Software Platform", Journal of the Acoustical Society of America, 98, 1890-1894 (1995).

[6] B.C.J. Moore, An introduction to the psychology of hearing, Academic Press (2003).

[7] M.Park, P.A. Nelson and K.O.Kang, "A model of sound localisation applied to the evaluation of systems for stereophony", Acta Acustica united with Acustica, 94, 825-839 (2008).

[8] L.A. Jeffress, "A place theory of sound localization", J. Comp. Physiol. Psychol., 41, 35-39 (1948).

[9] N.S. Harper, and D. McAlpine, "Optimal neural population coding of an auditory spatial cue", Nature, 430, 682-686 (2004).

[10] J. Breebaart, S. van de Par, and A. Kohlrausch, "Binaural processing model based on contralateral inhibition. I. Model structure", Journal of the Acoustical Society of America, 110, 1074-1088 (2001).

[11] H. Clark, G. Dutton, and P. Vanderlyn, "The stereophonic recording and reproducing system", IRE. Trans. Audio, 5, 96-111 (1957).

[12] F. Rumsey, Spatial Audio, Focal Press, London (2001).

[13] J. Blauert, Spatial Hearing: The Psychophysics of Human Sound Localization, MIT Press, London (2001).

[14] M. Park, P.A. Nelson, and K.O. Kang, "Evaluation of stereophonic images with listening tests and model simulations", in the 124th AES Convention, Amsterdam (2008).

[15] M. Park, P.A. Nelson, and K.O. Kang, "Application of an auditory process model for the evaluation of stereophonic images", in IOA Spring Conference, Reading (2008).

[16] F.E. Toole, "In-head localisation of acoustic images", Journal of the Acoustical Society of America, 48, 943-949 (1970)

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

[18] F.L. Wightman and D.J. Kistler, "Headphone simulation of free-field listening II. Psychophysical validation", 85, 869-878 (1989).

[19] W.M. Hartmann and A. Wittennberg, "On the externalisation of sound images", Journal of the Acoustical Society of America, 99, 3678-3688 (1996).

[20] S.M. Kim and W. Choi, "On the externalisation of virtual sound images in headphone reproduction: A Wiener filter approach", Journal of the Acoustical Society of America 117, 3657-3665 (2005).

[21] Y. Kahana and P.A. Nelson, "Numerical modelling of the spatial response of the human pinna", Journal of Sound and Vibration, 292, 148-178 (2006)

[22] Y. Kahana and P.A. Nelson, "Boundary element simulation of the transfer function of the human head and baffled pinnae using simple and complex geometric models", Journal of Sound and Vibration, 300, 552-579 (2007).

[23] E.M. Wenzel, M. Arruda, D.J. Kistler and F.L. Wightman, "Localisation using nonindividualised head related transfer functions" Journal of the Acoustical Society of America, 94, 111-123 (1993).

[24] N. Sakamoto, T. Gotoh and Y. Kimura, "On out-of-head localisation in headphone listening", Journal of the Audio Engineering Society, 24, 710-715 (1976).

[25] T. Choi, Y.C. Park and D.H. Youn, "Efficeient out of head localisation system for mobile applications" Audio Engineering Society Convention Paper 6758 Presented at the 120th Convention, Paris, France (2006).

[26] M. R. Schroeder, B. S. Atal, "Computer Simulation of Sound Transmission in Rooms," IEEE Intercon. Rec. Pt7, 150-155 (1963).

[27] P. Damaske, "Head-related Two-channel Stereophony with Reproduction," J. Acoust. Soc. Am. 50, 1109-1115 (1971).

[28] H. Hamada, N. Ikeshoji, Y. Ogura and T. Miura, "Relation between Physical Characteristics of Orthostereophonic System and Horizontal Plane Localisation," Journal of the Acoustical Society of Japan, (E) 6, 143-154, (1985).

[29] J. L. Bauck and D. H. Cooper, "Generalized Transaural Stereo and Applications," J. Acoust. Soc. Am. 44 (9), 683-705 (1996).

[30] P. A. Nelson, O. Kirkeby, T. Takeuchi, and H. Hamada, "Sound fields for the production of virtual acoustic images," J. Sound. Vib. 204 (2), 386-396 (1997).

[31] O. Kirkeby, P.A. Nelson and H. Hamada, "The 'Stereo Dipole'- a virtual source imaging system using two closely spaced loudspeakers", Journal of the Audio Engineering Society 46, 387-395 (1998).

[32] T. Takeuchi and P. A. Nelson, "Optimal source distribution for binaural synthesis over loudspeakers", Acoustic Research Letters Online, 2(1), 7-12 (2001).

[33] J. Bauck, "A Simple Loudspeaker Array and Associated Crosstalk Canceller for Improved 3D Audio," Journal of the Audio Engineering Society, 49, 3-13 (2001).

[34] T. Takeuchi and P. A. Nelson, "Optimal source distribution for binaural synthesis over loudspeakers", Journal of the Acoustical Society of America, 112, 2786-2797 (2002).

[35] T. Takeuchi and P.A. Nelson, "Extension of the Optimal Source Distribution for Binaural Reproduction," Proc. Inst. Acoustics Vol 30 Pt 2. (2008).

[36] P.A. Nelson and J.F.W. Rose, "The time domain response of some systems for sound reproduction", Journal of Sound and Vibration 296, 461-293 (2006).

[37] T. Takeuchi, M. Teschl and P.A. Nelson, "Objective and subjective evaluation of the optimal source distribution for virtual acoustic imaging" Journal of the Audio Engineering Society 55, 981-987 (2007).

[38] A.J. Berkhout, "A holographic approach to acoustic control." Journal of the Audio Engineering Society, vol. 36, pp. 977–995, (1988).

[39] A.J. Berkhout, D. de Vries and P. Vogel, "Acoustic control by wavefield synthesis", Journal of the Acoustical Society of America 93, 2764-2778 (1993)

[40] D. de Vries, E.W. Start, V.G. Valstar, "The Wave Field Synthesis concept applied to sound reinforcement: Restrictions and solutions," in 96th Convention of the Audio Engineering Society, (1994).

[41] S. Spors, H. Buchner, R. Rabenstein, "Efficient active listening room compensation for Wave Field Synthesis," in 116th Convention of the Audio Engineering Society (2004).

[42] P. Gauthier, A. Berry, "Adaptive wave field synthesis with independent radiation mode control for active sound field reproduction: Theory", Journal of the Acoustical Society of America, Vol. 119, No. 5, pp. 2721-2737 (2006)

[43] S. Ise, "A Principle of Sound Field Control Based on the Kirchhoff–Helmholtz Integral Equation and the Theory of Inverse Systems," Acustica—Acta Acustica, vol. 85, pp. 78–87 (1999).

[44] F.M. Fazi and P.A. Nelson "A theoretical study of sound field reconstruction techniques" 19th International Congress on Acoustics (2007)

[45] F.M. Fazi and P.A. Nelson, "Application of functional analysis to the sound field reconstruction", Proceedings of the Institute of Acoustics, 23rd Conference on Reproduced Sound (2007).

[46] F.M. Fazi "Sound field reproduction" PhD Thesis, University of Southampton (2010)

[47] F.M. Fazi, P.A. Nelson, J.E.N. Christensen and J. Seo, "Surround system based on three-dimensional sound field reconstruction. In 125th International Convention of the Audio Engineering Society, San Francisco, USA, (2008).

[48] M.A. Gerzon “Periphony : With-height sound reproduction.” Journal of the Audio Engineering Society 21 (1), 2-10 (1973).

[49] J. S. Bamford, J. Vanderkooy, “Ambisonic Sound for Us,” in 99th Convention of the Audio Engineering Society, (1995).

[50] J. Daniel “Acoustic field representation, application to the transmission and the reproduction of complex sound environments in the multimedia context.” PhD. Paris: Université Paris 6 (2000).

[51] J. Daniel, R. Nicol, S. Moureau, “Further Investigations of High-Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging,” in 114th Convention of the Audio Engineering Society, (2003).

[52] D. B. Ward, T. D. Abhayapala “Reproduction of a Plane-Wave Sound Field Using an Array of Loudspeakers.” IEEE Transaction on Speech and Audio Processing, 9, 697-707 (2001).

[53] M. Poletti, “Three-Dimensional Surround System Based on Spherical Harmonics”. Journal of the Audio Engineering Society, Vol. 53, No. 11 (2005).

[54] O. Kirkeby and P.A. Nelson “Reproduction of plane wave sound fields.” Journal of the Acoustical Society of America 94, 2992-3000 (1993).

[55] P.A. Nelson and S.J. Elliott, “Active Control of Sound”. Academic, London (1992).

[56] B. Støfringsdal, U. P. Svensson, “Conversion of discretely sampled sound field data to auralization formats,” Journal of the Audio Engineering Society, Vol. 54., No. 5, pp 380-400 (2006).

[57] V. Pulkki, “Virtual sound source positioning using vector base amplitude panning,” Journal of the Audio Engineering Society, vol. 45, no. 6, pp. 456–466, (1997).

[58] A. Farina et al. “Ambiophonic Principles for the Recording and Reproduction of Surround Sound for Music”, in 19th Convention of the Audio Engineering Society (2001).

[59] T. Betlehem, T. Abhayapala, “Theory and Design of Sound Field Reproduction in Reverberant Rooms,” Journal of the Acoustical Society of America, vol. 117, pt. 1, pp. 2100–2111 (2005).

[60] J.Ahrens and S. Spors, “Reproduction of a plane-wave sound field using planar and linear arrays of loudspeakers.”, IEEE Int. Symposium on Communications Control and Signal Processing (ISCCSP) (2008)

[61] E. Williams, “Fourier acoustics. Sound Radiation and Nearfield Acoustical Holography”. New York: Academic (1999).

[62] F. J. Fahy, “Sound Intensity”, Elsevier Science Prublishers (1989).

[63] M. Poletti, “Robust Two-Dimensional Surround Sound Reproduction for Nonuniform Loudspeaker Layouts”. Journal of the Audio Engineering Society, Vol. 55, No. 7/8 (2007).