Journal of Sound and Vibration 539 (2022) 117259

Contents lists available at ScienceDirect

Journal of Sound and Vibration

journal homepage: www.elsevier.com/locate/jsvi

Sound field control for multiple listener virtual imaging

P.A. Nelson *, T. Takeuchi, P. Couturier, X. Zhou

Institute of Sound and Vibration Research, University of Southampton, Highfield, Southampton, SO17 1BJ, United Kingdom

ARTICLE INFO

Keywords:
Sound field control
Virtual sound imaging

ABSTRACT

Advances in technology have made it possible to use loudspeakers to generate the perception in a listener of a "virtual" image of a source of sound that appears to be at a position other than that of the loudspeakers. This paper deals with the challenge of simultaneously generating the perception of the same virtual sound image at the ears of multiple listeners. The basis of the analysis is the so-called Optimal Source Distribution (OSD) that enables well-conditioned crosstalk cancellation at the ears of a single listener. This consists of a hypothetically continuous distribution of monopole acoustic strength which has a remarkable frequency independent radiation pattern. This directivity pattern enables crosstalk cancellation at a series of defined listener positions in the far field. However, in any practical application using current technology, the sources will be discrete loudspeaker elements. Previous work already shows the promise of this technique and the work described here aims to evaluate a number of methods for determining the strength of such discrete sources while still ensuring crosstalk cancellation at a number of listener positions.

1. Introduction

Developments in digital signal processing and the associated microelectronic technology has made the active control of acoustic fields an increasingly practical proposition, a possibility recognised at an early stage by Ffowcs-Williams [1]. Whilst much attention has been focused on the reduction of unwanted noise by such electronic means [2,3], it has also been found possible to manipulate acoustic fields in order to generate signals at the ears of listeners that enhance the experience of listening to sound. This can be achieved by controlling the interference between a number of sources to generate the requisite signals at the listener's ears. The method described here builds on the idea of cancelling the crosstalk from two transducers at the ears of a listener, that arguably began with work in the 1960's by Bauer [4] and Atal and Schroeder [5] and that was developed thereafter by number of workers (see, for example, the papers by Damaske [6], Ando et al. [7], Sakamoto et al. [8], Hamada [9], Bauck and Cooper [10], and Kirkeby et al. [11]). A comprehensive review of progress in the application of this technique can be found in the recent PhD thesis by Hamdan [12].

A further development [13] was the introduction of the Optimal Source Distribution (OSD) for achieving binaural sound reproduction for a single listener. This method has been shown to yield excellent subjective results [14] and has since been implemented in a number of products for virtual sound imaging. A remarkable property of the OSD is that the crosstalk cancellation produced at a single centrally placed listener is replicated at all frequencies at a number of other locations in the radiated sound field. This property has since been further investigated experimentally [15,16], and analytically by Yairi et al. [17]. The experiments reported show that once a discrete approximation to the hypothetically continuous OSD is introduced, the effectiveness of the crosstalk cancellation is achieved at many, but not all, frequencies at the non-central positions in the radiated sound field.

* Corresponding author.
E-mail address: p.a.nelson@soton.ac.uk (P.A. Nelson).

https://doi.org/10.1016/j.jsv.2022.117259
Received 22 December 2021; Received in revised form 1 August 2022; Accepted 22 August 2022
Available online 30 August 2022

The work presented here investigates a number of approaches that can be used to generate virtual sound images for multiple listeners. Initial attempts to tackle this problem have been presented by Kim et al. [18] who emphasised the importance of the condition number of the matrix relating the outputs of multiple sources to the pressures at the ears of multiple listeners. Further work by Masiero and Qiu [19] introduced the scattering by the listener's heads into the analysis. Other contributions include those of Mannerheim and Nelson [20], Kurabayashi et al. [21] and Simon-Galvez et al. [22] who use image processing for tracking listener head position in binaural reproduction, and of House et al. [23] who aim to present binaural sound to multiple listeners in an in-car environment. More recently, promising work has been presented by Holleborn et al. [24] using linear arrays with a loudspeaker dependent regularisation in order to deliver binaural sound to three listeners. The aim of the work presented here is to enable the exploitation of the fundamental property of the OSD with the objective of producing exact cross-talk cancellation for multiple listener positions over the maximum possible frequency range. The focus of the work is on the acoustical consequences of different optimisation strategies rather than the design of the associated inverse filters that has been dealt with previously by Kirkeby et al. [11] and Masiero and Vorlander [25].

First, the background to the problem will be introduced through a brief analysis of the original OSD [13] and its three channel extension [26]. Then a number of approaches to the discretisation problem will be developed. These include the linearly constrained least squares solution and the approach minimising the $L2$ norm of the source strength vector that were first presented in Ref. [27]. New solutions based on the minimisation of the $L1$ norm of the source strength vector will also be presented. In particular it is demonstrated numerically that the OSD provides the natural solution to the $L1$ norm minimisation problem. The OSD thus appears to guide the search for sparse solutions to this discretisation problem.

2. The Optimal Source Distribution

Fig. 1(a) shows the geometry of the two-loudspeaker/single listener problem and Fig. 1(b) shows the equivalent block diagram, both figures replicating the notation used by Takeuchi and Nelson [13] and Yairi et al. [17]. The desired signals for reproduction, the source signals and the reproduced signals are respectively defined by

$$\mathbf{d}^{\mathrm{T}} = \begin{bmatrix} d_R & d_L \end{bmatrix}, \ \mathbf{v}^{\mathrm{T}} = \begin{bmatrix} v_R & v_L \end{bmatrix}, \ \mathbf{w}^{\mathrm{T}} = \begin{bmatrix} w_R & w_L \end{bmatrix} \tag{1}$$

whilst the inverse filter matrix and transmission path matrix are defined by

$$\mathbf{H} = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}, \ \mathbf{C} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}, \tag{2}$$

where a harmonic time dependence of $\mathrm{e}^{\mathrm{j}\omega t}$ is assumed and all variables are in the frequency domain. Note that the vector $\mathbf{d}$ of desired signals defines the signals input to the filter matrix and these are the signals that are to be produced at the listener's ears through the process of crosstalk cancellation. Thus $\mathbf{v} = \mathbf{Hd}$ and $\mathbf{w} = \mathbf{CHd}$ and therefore it appears that the vector $\mathbf{w}$ of reproduced signals can be made equal to the desired signals $\mathbf{d}$ if the matrix $\mathbf{H}$ is made equal to the inverse of the matrix $\mathbf{C}$. However if the matrix of filters is to be realisable, the elements of $\mathbf{H}$ must each have a causal impulse response. As shown in what follows, this is accomplished by defining the signals to be reproduced to be delayed versions of the desired signals.

Now assume that the sources are free field point monopoles with volume accelerations respectively given by v_R and v_L. The transmission path matrix $\mathbf{C}$ then takes the form

$$\mathbf{C} = \frac{\rho_0}{4\pi} \begin{bmatrix} \frac{\mathrm{e}^{-\mathrm{j}kl_1}}{l_1} & \frac{\mathrm{e}^{-\mathrm{j}kl_2}}{l_2} \\ \frac{\mathrm{e}^{-\mathrm{j}kl_2}}{l_2} & \frac{\mathrm{e}^{-\mathrm{j}kl_1}}{l_1} \end{bmatrix}, \tag{3}$$

The distances between the assumed point sources and the ears of the listener are as shown in Fig. 1(a) and $k = \omega/c_0$ and ρ_0, c_0 are respectively the density and sound speed. This matrix can be written as

$$\mathbf{C} = \frac{\rho_0 \mathrm{e}^{-\mathrm{j}kl_1}}{4\pi l_1} \begin{bmatrix} 1 & g\mathrm{e}^{-\mathrm{j}k\Delta l} \\ g\mathrm{e}^{-\mathrm{j}k\Delta l} & 1 \end{bmatrix} = \frac{\rho_0 \mathrm{e}^{-\mathrm{j}kl_1}}{4\pi l_1} \mathbf{C}_{\mathrm{N}} \tag{4}$$

where $g = l_1/l_2$ and $\Delta l = l_2 - l_1$. The normalised matrix $\mathbf{C}_{\mathrm{N}}$ can be further simplified by assuming that $l \gg \Delta r$ where Δr is the distance between the ears of the listener (see Fig. 1(a)), a notation that is consistent with that used previously [13,14]. This assumption enables the approximation $\Delta l = \Delta r \sin\theta$ so that the matrix $\mathbf{C}_{\mathrm{N}}$ can be written as

$$\mathbf{C}_{\mathrm{N}} = \begin{bmatrix} 1 & g\mathrm{e}^{-\mathrm{j}k\Delta r\sin\theta} \\ g\mathrm{e}^{-\mathrm{j}k\Delta r\sin\theta} & 1 \end{bmatrix} \tag{5}$$

Now note that the relationship $\mathbf{w} = \mathbf{CHd}$ can be written as

$$\mathbf{w} = \frac{\rho_0 \mathrm{e}^{-\mathrm{j}kl_1}}{4\pi l_1} \mathbf{C}_{\mathrm{N}} \mathbf{Hd} \tag{6}$$

Thus if the target values of the reproduced signals are defined as

$$\hat{\mathbf{w}} = \frac{\rho_0 \mathrm{e}^{-\mathrm{j}kl_1}}{4\pi l_1} \mathbf{d} \tag{7}$$

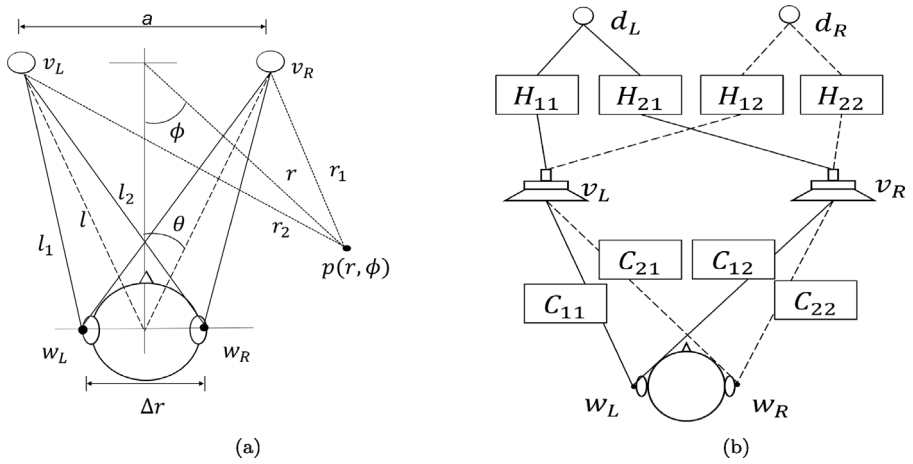

Fig. 1. (a) A two source single listener geometry and (b) the equivalent block diagram.

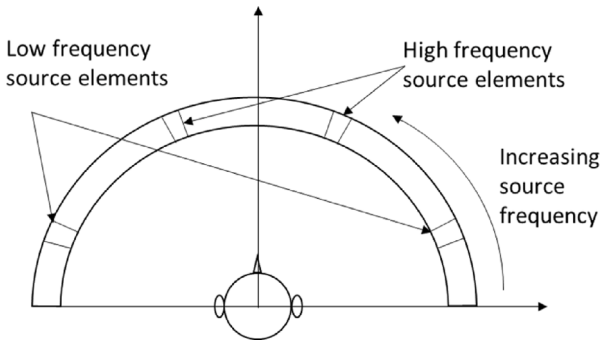

Fig. 2. The Optimal Source Distribution (OSD) consisting of a continuous distribution of monopole source strength whose frequency varies with angular span to ensure that the path length difference from any source element to the two ears of a listener is always one quarter of an acoustic wavelength.

where the term e^{-jkl_1} is equivalent to a pure delay of l_1/c_0 in the time domain and ensures that the target signals for reproduction are delayed versions of the desired signals input to the matrix **H**. Thus to ensure that the reproduced signals $\mathbf{w} = \hat{\mathbf{w}}$, then the matrix of inverse filters is simply given by the inverse of the matrix $\mathbf{C}_N$ where $\mathbf{H} = \mathbf{C}_N^{-1}$. The inverse filter matrix is thus given by

$$\mathbf{H} = \frac{1}{1 - g^2 e^{-2jkr\sin\theta}} \begin{bmatrix} 1 & -g e^{-jk\Delta r\sin\theta} \\ -g e^{-jk\Delta r\sin\theta} & 1 \end{bmatrix} \tag{8}$$

Thus expressing the signals to be reproduced in terms of the sound field produced at the listener's ears by one of the sources [13] ensures the causality of the inverse filter matrix **H**.

The conditioning of the matrix **C** is of particular importance in what follows. The conditioning of the matrix can be defined by its condition number, given by the ratio of the maximum to minimum singular values of the matrix. A high condition number implies poor conditioning and a great sensitivity to errors in the solution involving the inverse matrix (see for example [28] for a full discussion). The singular value decomposition of the matrix **C** (and thus of the matrix **H**) has also been investigated previously [13] and it has been shown that the two singular values of the matrix are equal when

$$k\Delta r\sin\theta = \frac{n\pi}{2} \tag{9}$$

where n is an odd integer (1, 3, 5...). Equality of the two singular values results in a condition number of unity. This in turn implies that the inversion problem is intrinsically well-conditioned when the above criterion is met so that reproduction can be accomplished with minimal error. This condition is equivalent to the difference in path lengths Δl being equal to odd integer multiples of one quarter of an acoustic wavelength λ. The angle 2θ is equal to the angular span of the sources (see Fig. 1), so that this condition implies that the source span should vary continuously with frequency to preserve the $\lambda/4$ path length difference. The OSD is therefore a continuous distribution of pairs of sources, each radiating a different frequency, with sources radiating high frequencies placed close together and those radiating lower frequencies placed further apart. This is illustrated in Fig. 2. Provided the sources are distributed to preserve the path length difference condition $\Delta l = n\lambda/4$ where n is an odd integer, then the inverse filter matrix reduces to

$$\mathbf{H} = \frac{1}{1 + g^2} \begin{bmatrix} 1 & -jg \\ -jg & 1 \end{bmatrix} \tag{10}$$

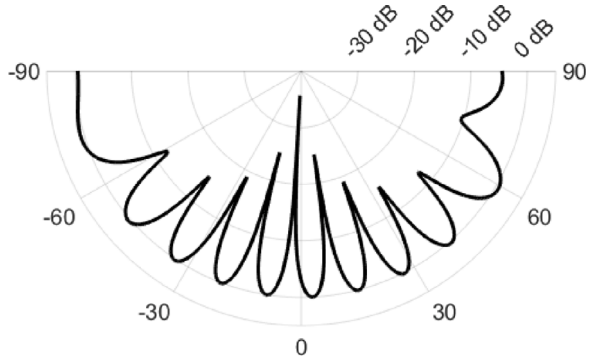

Fig. 3. Far field sound pressure level as a function of the angle ϕ produced by the two channel Optimal Source Distribution.

The term $-\mathrm{j}g$, involves simple inversion, 90-degree phase shift, and a small amplitude adjustment of the input signals. The inverse filter matrix thus takes a particularly simple form.

3. Radiation properties of the Optimal Source Distribution

This idealised source distribution has some attractive radiation properties [13,17]. This can be demonstrated by computing the far field pressure $p(r,\phi)$ radiated by a pair of sources with inputs determined by the above optimal inverse filter matrix. If it is also assumed that the desired signals for reproduction are given by $\mathbf{d}^{\mathrm{T}} = [1 \;\; 0]$, the condition for crosstalk cancellation, then it follows that the source signals are given by

$$\mathbf{v} = \mathbf{H}\mathbf{d} = \frac{1}{1+g^2}\begin{bmatrix} 1 & -\mathrm{j}g \\ -\mathrm{j}g & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{1+g^2}\begin{bmatrix} 1 \\ -\mathrm{j}g \end{bmatrix} \tag{11}$$

The pressure field is then given by the sum of the two source contributions such that

$$p(r,\phi) = \frac{\rho_0}{4\pi(1+g^2)}\begin{bmatrix} \frac{\mathrm{e}^{-\mathrm{j}kr_1}}{r_1} & -jg\frac{\mathrm{e}^{-\mathrm{j}kr_2}}{r_2} \end{bmatrix} \tag{12}$$

Writing $h = r_1/r_2$, shows that this can further be reduced to

$$p(r,\phi) = \frac{\rho_0 \mathrm{e}^{-\mathrm{j}kr_1}}{4\pi r_1(1+g^2)}\begin{bmatrix} 1 & -\mathrm{j}gh\mathrm{e}^{-\mathrm{j}k(r_2-r_1)} \end{bmatrix} \tag{13}$$

In the far field, where $r_1, r_2 \gg a$, the distance between the sources, using the geometry of Fig. 1(a), shows that

$$r_1 = r - \frac{a}{2}\sin\phi, \quad r_2 = r + \frac{a}{2}\sin\phi \tag{14}$$

Taking the squared modulus of the term in square brackets above shows that

$$\left|1 - \mathrm{j}gh\mathrm{e}^{-\mathrm{j}ka\sin\phi}\right|^2 = 1 + (gh)^2 - 2ghsin(ka\sin\phi) \tag{15}$$

and thus the modulus squared of the pressure field is given by

$$\left|p(r,\phi)\right|^2 = \left(\frac{\rho_0}{4\pi r_1}\right)^2 \frac{1 + (gh)^2 - 2gh\sin(ka\sin\phi)}{1+g^2} \tag{16}$$

As r increases, making the approximation $r \approx r_1 \approx r_2$ leads to

$$\left|p(r,\phi)\right|^2 = \left(\frac{\rho_0}{4\pi r_1}\right)^2 (1 - \sin(ka\sin\phi)) \tag{17}$$

Maxima and minima are produced in the sound field when $ka\sin\phi = n\pi/2$ where n is odd. The term $(1 - \sin(ka\sin\phi))$ zero at $n = 1, 5, 9...$ etc. and is equal to two when $n = 3, 7, 11...$etc. The squared modulus of the sound pressure as a function of the angle ϕ is illustrated in Fig. 3. Most importantly, this directivity pattern of the far field pressure is the same at all frequencies, provided the two sources are arranged to always ensure that the path length difference to the listener's ears at a given frequency satisfies the condition of an integer number of quarter wavelengths. It then follows that, as shown above, the value of $k\Delta r\sin\theta$ is an odd integer multiple of $\pi/2$, and since from the geometry of Fig. 1, $\sin\theta = a/2l$ then $ka = n\pi l/2\Delta r$. Thus ka and therefore $\omega a/c_0$ takes a constant value (assuming $n = 1$). The radiation pattern is thus determined solely by the geometrical arrangement of the two on-axis points chosen initially for cross-talk cancellation. The directivity pattern illustrated in Fig. 3 demonstrates that an intrinsic property of the OSD is the production of cross-talk cancellation at multiple angular positions in the sound field. This will be further illustrated below using plots of the sound pressure distribution along an arc in the sound field (see Fig. 6), showing that there are multiple potential positions of a listener's head at which crosstalk cancellation can be produced.

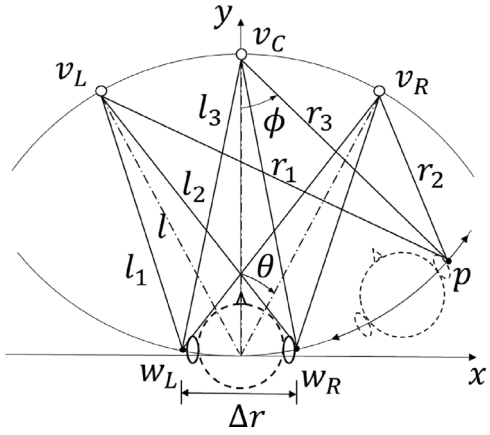

Fig. 4. Geometry of the three channel OSD with a single listener.

4. The three-channel Optimal Source Distribution

It has also been demonstrated [26] that there are some distinct advantages to including an additional centrally located source between the two sources of the OSD. Such an arrangement of sources is illustrated in Fig. 4, where the sources are assumed to be located on a circular arc centred on the head of the listener. This geometry proves helpful to the analysis that follows. In this case the vector of source strengths is defined as

$$\mathbf{v}^{\mathrm{T}} = \begin{bmatrix} v_R & v_C & v_L \end{bmatrix}, \tag{18}$$

and the inverse filter matrix and transmission path matrix are respectively defined by

$$\mathbf{H} = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \\ H_{31} & H_{32} \end{bmatrix}, \ \mathbf{C} = \begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \end{bmatrix}, \tag{19}$$

where in this case the vector of three source strengths is given by $\mathbf{v} = \mathbf{Hd}$ where the vector of two desired signals is again given by $\mathbf{d}$ and the vector of two reproduced signals is given by $\mathbf{w} = \mathbf{CHd}$. This is an underdetermined system of equations but a solution can be found that minimises the $L2$-norm of the vector $\mathbf{v}$ whilst generating delayed versions of the desired signals at the listener's ears. Again assuming the sources are point monopoles radiating into a free field, with volume accelerations respectively given by v_R, v_C and v_L, the transmission path matrix $\mathbf{C}$ can be written as

$$\mathbf{C} = \frac{\rho_0}{4\pi} \begin{bmatrix} \frac{\mathrm{e}^{-\mathrm{j}kl_1}}{l_1} & s\frac{\mathrm{e}^{-\mathrm{j}kl_3}}{l_3} & \frac{\mathrm{e}^{-\mathrm{j}kl_2}}{l_2} \\ \frac{\mathrm{e}^{-\mathrm{j}kl_2}}{l_2} & s\frac{\mathrm{e}^{-\mathrm{j}kl_3}}{l_3} & \frac{\mathrm{e}^{-\mathrm{j}kl_1}}{l_1} \end{bmatrix}, \tag{20}$$

where the distances l_1, l_2 and l_3 are defined in Fig. 4. Note that an additional factor s has been used to multiply the transfer function between the centre source and the listener's ears. This term enables adjustment of the strength of the centre source by the additional factor s. As demonstrated below, the condition number of the matrix $\mathbf{C}$ can in turn be modified through the choice of s. The conditioning of the matrix can again be deduced by examination of its singular values. The singular value decomposition can be undertaken analytically by first writing the matrix in the form

$$\mathbf{C} = \frac{\rho_0 e^{-jkl_1}}{4\pi l_1} \begin{bmatrix} 1 & sg_3\mathrm{e}^{-\mathrm{j}k(l_3-l_1)} & g_2\mathrm{e}^{-\mathrm{j}k(l_2-l_1)} \\ g_2\mathrm{e}^{-\mathrm{j}k(l_2-l_1)} & sg_3\mathrm{e}^{-\mathrm{j}k(l_3-l_1)} & 1 \end{bmatrix}, \tag{21}$$

where $g_2 = l_1/l_2$ and $g_3 = l_1/l_3$. When the listener is in the far field of the sources, $g_2 \approx g_3 = 1$. Furthermore, it follows from the geometry of shown in Fig. 4 that when Δr, the width of the listener's head, is much smaller than the distance l to the listener from the centre source, we may make the approximation $l_2 - l_1 \approx \Delta r \sin\theta$. In addition, to the same order of approximation, $l_3 \approx l$ and thus $l_3 - l_1 \approx (\Delta r \sin\theta)/2$. Now defining the path length difference $\Delta l = l_2 - l_1$ enables the transmission path matrix to be written as

$$\mathbf{C} = \frac{\rho_0 \mathrm{e}^{-\mathrm{j}kl_1}}{4\pi l_1} \begin{bmatrix} 1 & s\mathrm{e}^{-\mathrm{j}k\Delta l/2} & \mathrm{e}^{-\mathrm{j}k\Delta l} \\ \mathrm{e}^{-\mathrm{j}k\Delta l} & s\mathrm{e}^{-\mathrm{j}k\Delta l/2} & 1 \end{bmatrix}, \tag{22}$$

The singular values of this matrix can be calculated by finding the roots of the determinant of the matrix $\lambda\mathbf{I} - \mathbf{C}\mathbf{C}^{\mathrm{H}}$ where λ is an eigenvalue and $\mathbf{I}$ is the identity matrix (see, for example, Hamdan and Fazi [29]). After some lengthy, but straightforward, algebra it follows that the two singular values (the square roots of the eigenvalues) are given by

$$\sigma_+ = \sqrt{[2 + 2s^2 + 2\cos k\Delta l]}, \quad \sigma_- = \sqrt{[2 - 2\cos k\Delta l]} \tag{23}$$

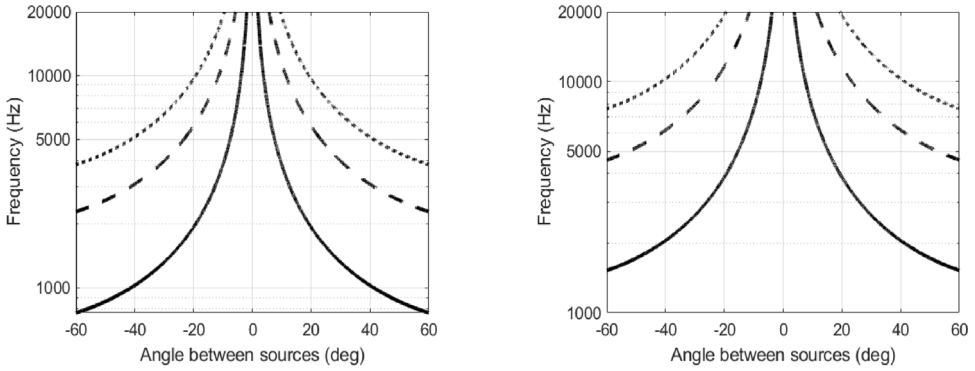

Fig. 5. The dependence of angular position θ of the sources as a function of frequency for (a) the two channel and (b) the three channel OSD for $n = \pm 1$ (solid line) $n = \pm 3$ (dashed line) and $n = \pm 5$ (dotted line).

First note that when $s = 0$, these results reduce to those derived previously [13] for the two channel case when the far field approximation is made. Secondly, as also recognised previously [14], and as demonstrated by Yairi et al. [17], choosing the parameter $s = \sqrt{2}$ results in a minimum of the condition number of the matrix $\mathbf{C}\mathbf{C}^{\mathrm{H}}$ i.e. when the two singular values become equal. Making the choice $s = \sqrt{2}$ and equating the singular values soon shows that this condition is met at frequencies where the path length difference $\Delta l = n\lambda/2$ where n is an odd integer. This is in contrast to the criterion of $\Delta l = n\lambda/4$ for optimal conditioning of the two-channel OSD. The range of values of the optimal angular position of the sources in for the two and three channel OSD are shown as a function of frequency and the value of n in Fig. 5.

Also note that a normalised matrix $\mathbf{C}_{\mathrm{N}}$ can be defined analogously to Eq. (4) and the desired signals for reproduction can be defined in accordance with Eq. (5). The solution for the inverse filters that minimises the $L2$ norm of the source strength vector is then given by $\mathbf{H} = \mathbf{C}_{\mathrm{N}}^{\mathrm{H}}[\mathbf{C}_{\mathrm{N}}\mathbf{C}_{\mathrm{N}}^{\mathrm{H}}]^{-1}$. Using this solution when $\Delta l = \lambda/2$, and assuming the desired signals are again given by $\mathbf{d}^{\mathrm{T}} = [1\ 0]$, then shows that the inverse filter matrix and optimal source strength vector are respectively given by [17]

$$\mathbf{H} = \frac{1}{4}\begin{bmatrix} 1 & -1 \\ \sqrt{2}\mathrm{j} & \sqrt{2}\mathrm{j} \\ -1 & 1 \end{bmatrix}, \quad \mathbf{v}_{opt} = \frac{1}{4}\begin{bmatrix} 1 \\ \sqrt{2}\mathrm{j} \\ -1 \end{bmatrix} \tag{24}$$

Fig. 6(a) and (b) show the radiation patterns of the two- and three-channel OSD systems with the optimal choice of the parameter $s = \sqrt{2}$ used in the three channel system. Note that in contrast to the two-channel system, the three channel OSD produces an interference pattern that has radial dependence, although clear zones of cross talk cancellation are produced on the circular arc at the assumed listener radius of l =1.5 m (shown by the black line in Fig. 6(a) and (b)). Fig. 6(c) and (d) show the resulting sound pressures as a function of frequency at locations along this same listener arc. Clearly, in both cases, cross-talk cancellation is produced very effectively over a wide frequency range, although the spatial zones over which cross-talk is cancelled are markedly larger in the case of the three-channel OSD. However, the lowest frequency at which crosstalk cancellation can be delivered with good conditioning of the inversion problem is higher (1244 Hz) for the three channel OSD than for the two-channel OSD (622 Hz). There are clearly subjective trade-offs to be made in designing such systems and these await a full evaluation.

5. Discretisation of the Optimal Source Distribution

In any real application, any approximation to the OSD must be realised by a discrete number of loudspeakers. The work presented here examines a number of approaches to determining the signals input to a discrete number of sources in order to achieve the objective of producing cross-talk cancellation at multiple listener positions. Now consider the case illustrated in Fig. 7. The strength of a distributed array of acoustic sources is defined by a vector $\mathbf{v}$ of order M, and the pressure is defined at a number of points in the sound field by the vector $\mathbf{w}$ of order L. The curved geometry chosen here replicates that used in Fig. 4. In general $\mathbf{w} = \mathbf{C}\mathbf{v}$ where $\mathbf{C}$ is an $L \times M$ matrix. However, it will be helpful to partition this matrix $\mathbf{C}$ into two other matrices $\mathbf{A}$ and $\mathbf{B}$ and also partition the vector $\mathbf{w}$ of reproduced signals so that

$$\mathbf{w} = \begin{bmatrix} \mathbf{w}_A \\ \mathbf{w}_B \end{bmatrix} = \begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix} \mathbf{v} = \mathbf{C}\mathbf{v} \tag{25}$$

The vector $\mathbf{w}_B$ is of order P and defines the reproduced signals at a number of pairs of points in the sound field at which cross-talk cancellation is sought. Thus $\mathbf{w}_{\mathrm{B}} = \mathbf{B}\mathbf{v}$ where $\mathbf{B}$ defines the $P \times M$ transmission path matrix relating the strength of the M sources to these reproduced signals. The vector $\mathbf{w}_A$ is of order N and defines the reproduced signals sampled at the remaining points in the sound field. Thus $N = L - P$ and the reproduced field at these remaining points can be written as $\mathbf{w}_A = \mathbf{A}\mathbf{v}$ where $\mathbf{A}$ is the $N \times M$ transmission path matrix between the sources and these points. Now note that the desired pressure at the P points in the sound field

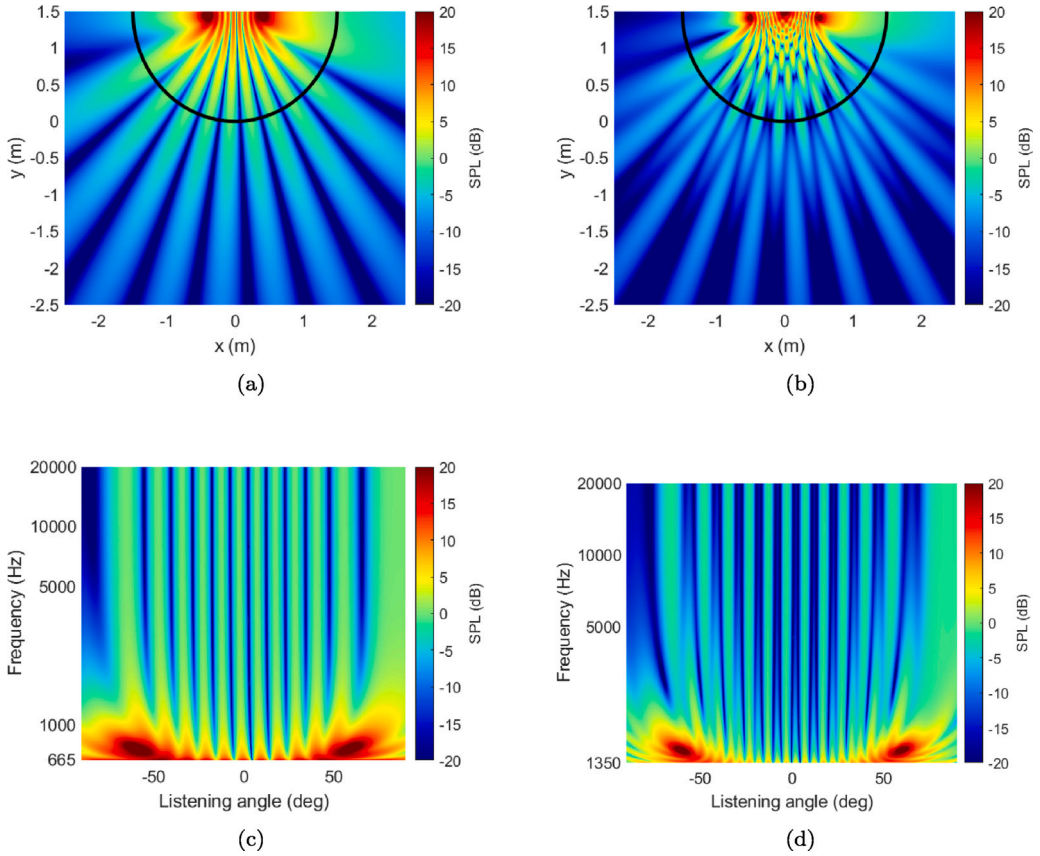

Fig. 6. The radiation pattern of (a) the two channel OSD for a pair of sources at 2058 Hz, and (b) the three channel OSD for the same pair of sources at 3206 Hz, where the black line defines the listener arc. The distribution of sound pressure level on the listener arc as a function of frequency and the listener angle is shown for (c) the two channel OSD and (d) the three channel OSD.

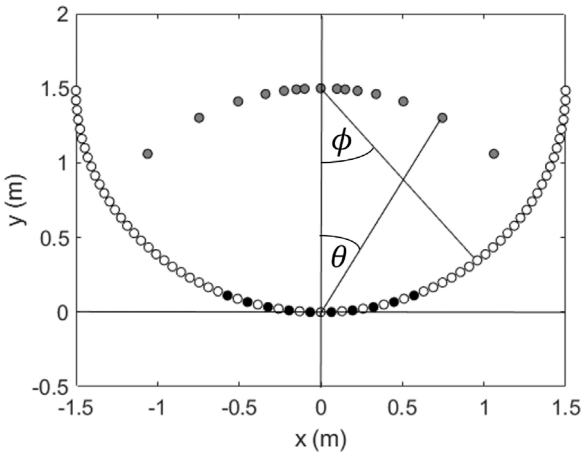

Fig. 7. The arrangement of M point sources (grey symbols) and L points in the sound field (white symbols) at which the complex sound pressure is sampled. Cross-talk cancellation is desired at P points (black symbols) whilst a least squares fit to the OSD sound field is desired at $L - P = N$ points (white symbols). The black symbols denote the ear positions of five listeners chosen to coincide with maxima and minima in the OSD sound field.

at which cross-talk cancellation is required can be written as the vector $\hat{\mathbf{w}}_B$. This vector can in turn be written as $\hat{\mathbf{w}}_B = \mathbf{Dd}$ where the matrix $\mathbf{D}$ defines the reproduced signals required in terms of the desired signals. As a simple example, suppose that cross-talk

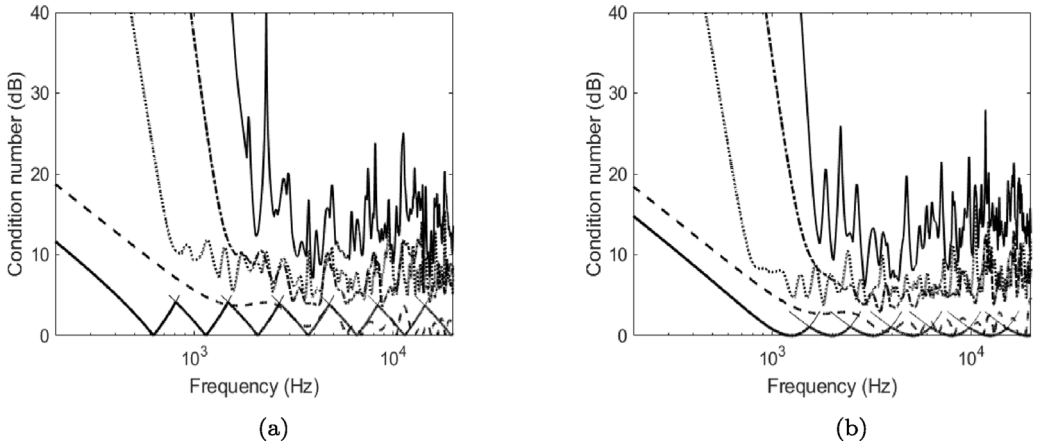

Fig. 8. The condition number of the matrix **C** when (a) 14 sources are used in the two-channel OSD and (b) 15 sources are used in the three-channel OSD, where the solid thick black line shows the condition number associated with each pair of sources over the best conditioned frequency range. Also shown are the condition number of the matrix **B** in each case when there is a single listener (dashed line), three listeners spaced apart (dotted line) three listeners close together (dash–dotted line), and five listeners (solid thin line).

cancellation is desired at two pairs of points in the sound field such that

$$\begin{bmatrix} \hat{w}_{B1} \\ \hat{w}_{B2} \\ \hat{w}_{B3} \\ \hat{w}_{B4} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} d_R \\ d_L \end{bmatrix} \tag{26}$$

The matrix **D** has elements of either zero or unity, is of order $P \times 2$, and may be extended by adding further pairs of rows if cross talk cancellation is required at further pairs of points. Similar to the analysis presented above, we assume that the inputs to the sources are determined by operating on the two desired signals defined by the vector **d** via an $M \times 2$ matrix **H** of inverse filters. The causality of the matrix **H** can again be ensured by including a term such as $e^{-j\omega\Delta}$ on the right side of Eq. (25), where Δ represents the maximum delay necessary to ensure causality.

Some numerical simulations have been undertaken using MATLAB in order to examine a number of approaches to the determination of the optimal strength of the discrete sources used to represent the OSD. The arrangement of sources chosen is illustrated in Fig. 7 where the sources are placed on an arc of radius 1.5 m surrounding a central listener. Fig. 5 suggests that increasing the density of sources as frequency increases will also increase the possibility of ensuring a well-conditioned problem. The positions of the sources on the arc chosen to emulate the optimal source distribution and are thus spaced logarithmically. This is accomplished by first defining the lowest frequency f_{min} at which a path length difference of $\lambda/4$ (or $\lambda/2$ in the three channel case) can be achieved at the listeners ears. For example, if the outermost sources are spaced at angle of $\theta_{max} = \pm 45$ deg then the calculation of the lowest frequency at which the matrix **C** is well conditioned is given by $\Delta r sin\theta_{max} = \lambda/4$.

It is helpful to include the effect of the head related transfer function (HRTF) of the listener by making the approximation $\Delta r = \Delta r_0[1 + (2\theta/\pi)]$ where Δr_0 is the physical distance between the listeners ears. This was first identified in Ref. [13] as providing a good approximation to the HRTF of a typical dummy head. Choosing $\Delta r_0 = 0.13$ m and the speed of sound $c_0 = 343$ m/s then results in a frequency $f_{min} = 622$ Hz. The equivalent calculation for the three channel OSD yields a frequency $f_{min} = 1244$ Hz

The highest frequency f_{max} is then chosen at which a minimum condition number is required. Assuming this to be 20 kHz, then the loudspeaker span giving a $\lambda/4$ path length difference at the listener's ears will be given by $\theta_{min} = \pm 1.89$ deg. The span for a $\lambda/2$ path length difference is similarly given by $\theta_{min} = \pm 3.78$ deg. The angular locations θ_i of the $m = M/2$ pairs of sources (or $m = (M-1)/2$ in the three channel case) are chosen to be logarithmically spaced between θ_{min} and θ_{max}. This results in minima in the condition numbers that are also spaced approximately logarithmically as a function of frequency. If there are m source pairs and the ith source pair is spaced at an angle $\pm\theta_i$ then one can write $\theta_i = u^i\theta_{min}$ where $i = 0...m$ and the factor u can be found from $u = \log^{-1}[(1/(m-1))\log(\theta_{max}/\theta_{min})]$.

Fig. 8 (a) shows the condition number of the matrix **C** relating the strengths of each pair of sources to the sound pressures at the ears of a centrally located listener in a discrete approximation of the two channel OSD consisting of fourteen sources. Note that this condition number plot assumes that each source pair acts only over a specific range of frequencies. The equivalent plot is shown for the three channel OSD in Fig. 8(b). These figures show that choosing each pair of sources to act only over a given frequency range results in a well-conditioned inversion problem. This is in contrast to the case where all 14 or 15 sources are used across the entire frequency range to control the field the ears of multiple listeners, the dotted, dash–dotted and solid lines in Fig. 8(a) and (b) showing how poorly conditioned this inversion problem becomes, especially at low frequencies.

6. Minimising the deviation from the OSD sound field

An obvious approach to the discretisation problem is to simply choose a large number of M of sources and attempt a "best fit" to the OSD sound field at a number N positions where $N \geq M$. Thus for example one might seek to solve the standard least squares problem

$$min\left[\left\|\mathbf{A}\mathbf{v}-\hat{\mathbf{w}}_A\right\|_2^2+\beta\left\|\mathbf{v}\right\|_2^2\right] \tag{27}$$

which has the well-known solution given by

$$\mathbf{v}_{opt}=[\mathbf{A}^{\mathrm{H}}\mathbf{A}+\beta\mathbf{I}]^{-1}\mathbf{A}^{\mathrm{H}}\hat{\mathbf{w}}_A \tag{28}$$

where $\left\|\ \right\|_2$ indicates the $L2$ norm of the vector, H denotes the Hermitian transpose and β is a regularisation parameter.

First note that the problem to be solved is fundamentally ill-conditioned at low frequencies when the acoustic wavelength starts to become significant relative to the distances separating the sources and relevant positions in the sound field. Fig. 8 shows, for both two and three channel OSD systems, that the condition number increases rapidly with reducing frequency below the minimum frequency f_{min}.

The results of computing the solution for the least squares fit to the two channel OSD radiated field given above are shown in Fig. 9. This shows the results for the case of $M = 62$, 30 and 14 sources and in each case a number $N = 95$ points on the arc of radius 1.5 m spanning angles $\phi = \pm 90$ deg. For a large number of sources, the OSD sound field is well replicated at frequencies above f_{min} and the optimal distribution of source strength as a function of frequency is also well replicated. However, when a more practical number of sources such as 14 is used, the replication of the desired field deteriorates markedly and the source strength distribution deviates from that associated with the OSD. The results presented in Fig. 9 have been computed using a regularisation parameter of $\beta = 0.01$. This has the effect of suppressing the excessive source strengths at low frequency, but has little effect on the solution at frequencies above f_{min}. It is also interesting to note the distributions of source strength that are illustrated in Fig. 9(b), (d) and (f). These plots show that the solutions used tend naturally to select the sources at the centre of the distribution to deliver the sound field at high frequencies, whilst at low frequencies the whole range of available sources appears to be recruited to provide the solution. Whilst the results show that it is indeed possible to replicate the OSD sound field given a sufficient number of sources, the challenge remains of finding solutions using a smaller number of sources.

7. Minimising the deviation from the OSD sound field with the constraint of crosstalk cancellation

A further approach is to attempt to achieve a "best fit" of the radiated sound field to the known sound field of the OSD whilst also attempting to achieve cross-talk cancellation at multiple pairs of points. Such a problem can be described as a least squares minimisation with a linear equality constraint. The constraint at P points in the sound field can be expressed as $\hat{\mathbf{w}}_B = \mathbf{D}\mathbf{d}$. Thus the desired sound field is that produced by the OSD and one seeks to minimise the sum of squared errors at the remaining $N = L - P$ points of interest, where again it is assumed that $N \geq M$. One therefore seeks the solution of

$$min\left\|\mathbf{A}\mathbf{v}-\hat{\mathbf{w}}_A\right\|_2^2 \quad \text{subject to} \quad \hat{\mathbf{w}}_B=\mathbf{B}\mathbf{v} \tag{29}$$

The solution that is used here is based on that given by Golub and Van Loan [28] but modified to enable application to the complex valued variables. The solution can be written as

$$\mathbf{v}_{opt}=\mathbf{Q}_2\mathbf{A}_2^{\dagger}\hat{\mathbf{w}}_A+(\mathbf{Q}_1\mathbf{R}^{\mathrm{H}-1}-\mathbf{Q}_2\mathbf{A}_2^{\dagger}\mathbf{A}_1\mathbf{R}^{\mathrm{H}-1})\hat{\mathbf{w}}_B \tag{30}$$

As shown in the Appendix, the method relies on a QR factorisation of the matrix $\mathbf{B}^{\mathrm{H}}$ and the terms in the solution above are defined by the relationships

$$\mathbf{B}^{\mathrm{H}}=\mathbf{Q}\begin{bmatrix}\mathbf{R}\\ \mathbf{0}\end{bmatrix} \quad \mathbf{A}\mathbf{Q}=\begin{bmatrix}\mathbf{A}_1 & \mathbf{A}_2\end{bmatrix} \quad \mathbf{Q}=\begin{bmatrix}\mathbf{Q}_1 & \mathbf{Q}_2\end{bmatrix} \tag{31}$$

where the dimensions of the matrices involved are defined in the Appendix and $\mathbf{A}_2^{\dagger}=[\mathbf{A}_2^{\mathrm{H}}\mathbf{A}_2]^{-1}\mathbf{A}_2^{\mathrm{H}}$ is the pseudoinverse of the matrix $\mathbf{A}_2$. It is worth noting that in previous work [27] two further methods of solution were investigated. The first of these made use of Lagrange multipliers to find the solution to the constrained optimisation problem that was previously proposed by Olivieri et al. [30]. Once a minor error in this solution is corrected (see [27] for details) it was found in numerical simulations that this solution gave identical results to that given above. Numerical experiments were also undertaken using the approximate solution given by Golub and Van Loan [28] referred to as the "Method of Weighting" but was found, as anticipated [28] to suffer from numerical difficulties under some conditions. The results of applying this solution to the current problem are shown in Fig. 10.

This shows the sound pressure as a function of frequency and the listener angle when the constraint ensures crosstalk cancellation at the ears of three spaced apart listeners having ear positions corresponding to the central and two outermost pairs of sensors in Fig. 7. Again the problem is poorly conditioned at low frequencies but the application of regularisation to the inversion of the matrix $\mathbf{A}_2^{\dagger}$ succeeds to some extent in controlling excessive source outputs at low frequency. It is likely that the regularisation of the matrix $\mathbf{R}^{\mathrm{H}-1}$ is also necessary to further suppress excessive source outputs. It is also found that when the number of sources used is made larger, as one might expect, the characteristics of the OSD sound field become more closely reproduced, although the results are not shown here.

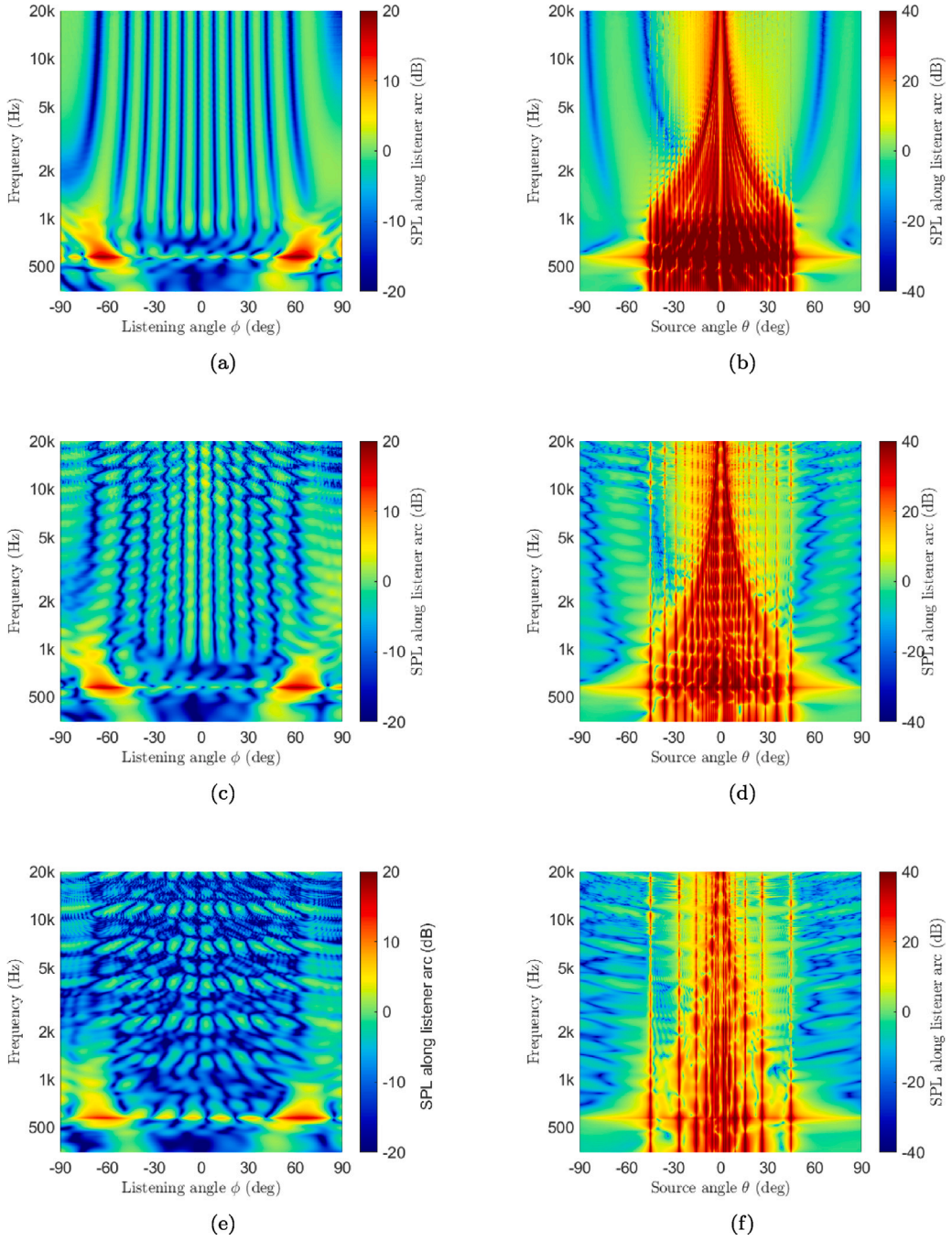

Fig. 9. The results of determining the strengths of a number of M sources to minimise the deviation from the two channel OSD sound field a number N = 95 on the listener arc. Results are shown for the sound pressure level as a function of frequency and listener angle for (a) M = 62 (c) M = 30 (d) M = 14 sources. The corresponding distributions of relative source strength as a function of frequency are shown on a decibel scale in (b), (d) and (f) respectively.

8. Minimising the $L2$ norm of the source strength vector

A further approach is simply to choose the number of sources to be greater than that of the number of points at which crosstalk cancellation is required i.e. $M \geq P$, a method also used recently by Holleborn et al. [24]. One may then seek solutions that minimise the source strengths required to satisfy the constraint of achieving the desired signals at the ears of multiple listeners.

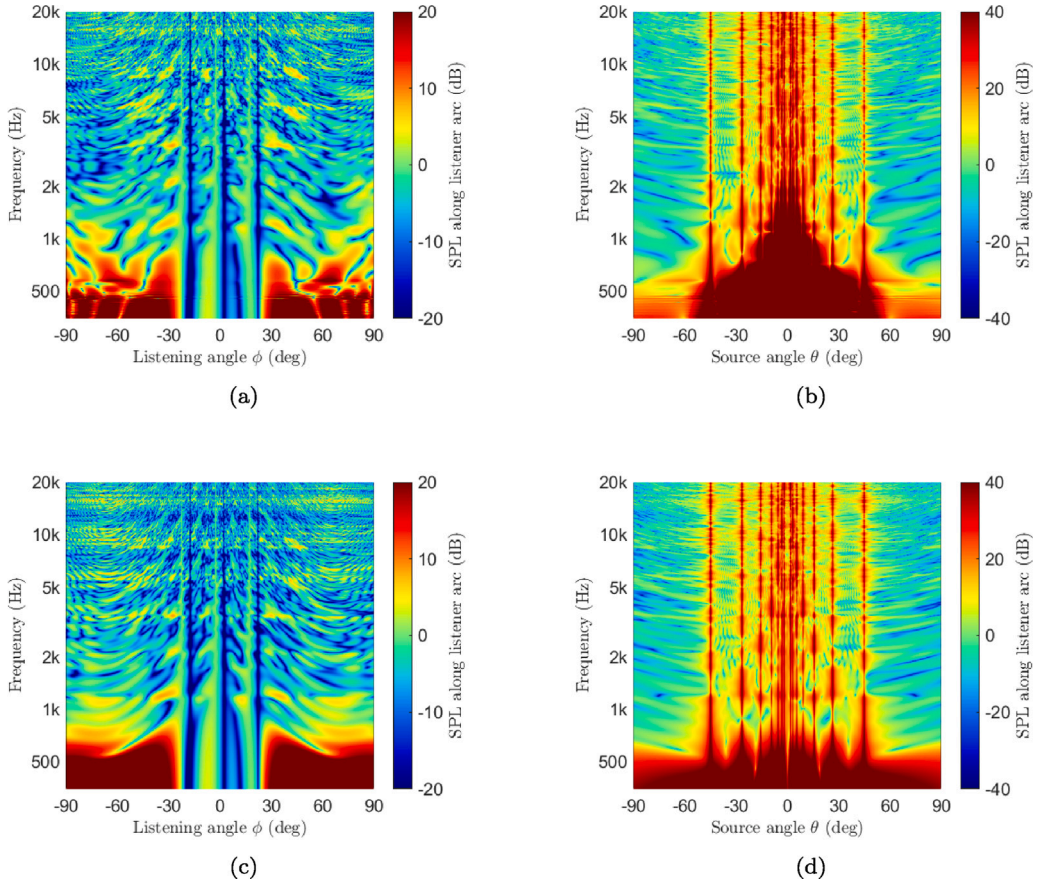

Fig. 10. The results of minimising the deviation from the two channel OSD sound field whilst ensuring crosstalk cancellation at three spaced apart listeners showing sound pressure level as a function of listener angle without (a) and with (c) regularisation and relative source strength on a decibel scale without (b) and with (d) regularisation, with $\beta = 10$. Results are for $M = 14$ sources, $N = 95$ positions in the sound field, and $P = 6$ points at which crosstalk cancellation is ensured.

The optimisation problem can be stated as

$$min\left\|\mathbf{v}\right\|_2^2 \quad \text{subject to} \quad \hat{\mathbf{w}}_B = \mathbf{B}\mathbf{v} \tag{32}$$

This problem has the well-known solution given by [28]

$$\mathbf{v}_{opt} = \mathbf{B}^{\mathrm{H}}[\mathbf{B}\mathbf{B}^{\mathrm{H}}]^{-1}\mathbf{D}\mathbf{d} \tag{33}$$

and this solution can also be regularised by using the parameter β such that

$$\mathbf{v}_{opt} = \mathbf{B}^{\mathrm{H}}[\mathbf{B}\mathbf{B}^{\mathrm{H}} + \beta\mathbf{I}]^{-1}\mathbf{D}\mathbf{d} \tag{34}$$

which, as noted by Masiero and Vorlander [25], can be shown to be equivalent to Eq. (27). This can be demonstrated by using the identities described by Henderson and Searle [31] in their discussion of the matrix inversion lemma (Sherman–Morrison–Woodbury formula).

Fig. 8(a) and (b) also show the condition numbers of the matrix **B** that relates the source strengths to the sound pressures at the positions at which crosstalk cancellation is sought. This is shown for a single listener and for three and five listeners arranged with ears at the locations illustrated in Fig. 7. These locations were chosen to coincide with those at which the OSD system naturally provides crosstalk cancellation at multiple sound field positions. Again it can be seen that the matrix is poorly conditioned at low frequencies. However, the least poorly conditioned problem is that for which three listeners are spaced apart. Some results of applying the minimum L_2 norm solution for this case are shown in Fig. 11. This shows the degree of cross-talk cancellation at each of the listener positions as a function of frequency for a range of values of the regularisation parameter β.

Fig. 11 shows again that cross-talk cancellation can be achieved at the ears of all three listeners at all frequencies above a lower frequency limit. Below this frequency the problem becomes very poorly conditioned and the source strengths become excessive. Whilst the regularisation can indeed assist in suppressing the low frequency source strengths it does not enable crosstalk cancellation.

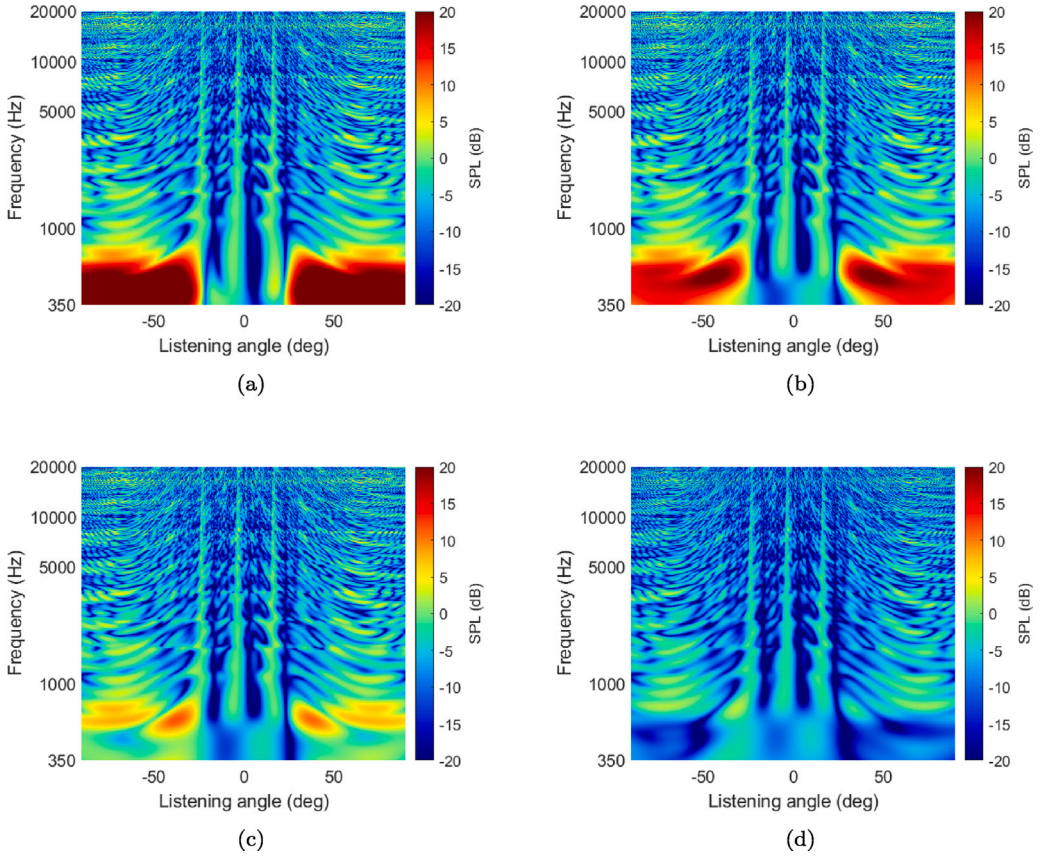

Fig. 11. The effect of regularisation on cross-talk cancellation performance for (a) $\beta = 0$, (b) $\beta = 0.0001$, (c) $\beta = 0.001$, (d) $\beta = 0.01$. Results are for $M = 14$ sources and $N = 6$ (3 spaced apart pairs) of points in the sound field at which crosstalk cancellation is ensured.

Another characteristic of this solution is that all sources are recruited in providing a solution to the problem. An example is given in Fig. 12 of the distribution of source strength as a function of frequency. Fig. 13 shows an example of the radiated field and associated source distribution for the three channel OSD at a frequency of 1995 Hz, the frequency of the second minimum of the condition number curve shown in Fig. 7(b).

9. Minimising the $L1$ norm of the source strength vector

It is well known that in solving optimisation problems of the type considered here, that the minimisation of the $L1$ norm of the appropriate vector will help to find so-called sparse solutions to the problem at hand. That is, the number of sources required to deliver the solutions will be minimised. Since the minimisation of the $L2$ norm described above results in a large number of sources required at a given frequency, it seems logical to attempt to find and evaluate minimum $L1$ norm solutions. The problem can thus be stated as

$$min\left\|\mathbf{v}\right\|_1 \quad \text{subject to} \quad \hat{\mathbf{w}}_B = \mathbf{B}\mathbf{v} \tag{35}$$

This is a convex optimisation problem [32] and solutions can be computed numerically by using the CVX software package [33]. The papers by Gerstoft et al. [34], Gauthier et al. [35] and Franck et al. [36] also provide some examples of using $L1$ norm minimisation to find sparse solutions to acoustical problems.

An example of the solution found by minimising the $L1$ norm of the source strength vector are given in the case of a single listener in Fig. 14(a) and (b). The hoped-for sparsity in the solution is indeed generated. However, there is clearly asymmetry in the solution for the source strengths, most likely caused by the natural asymmetry resulting from the requirement of the desired signals at the listeners' ears to be unity and zero respectively. It is possible to recast the optimisation problem to include a symmetric constraint on the source strengths. This can be written as

$$min\left\|\mathbf{v}\right\|_1 \quad \text{subject to} \quad \hat{\mathbf{w}}_B = \mathbf{B}\mathbf{v} \quad \text{and} \quad v_m = -v_{-m} \tag{36}$$

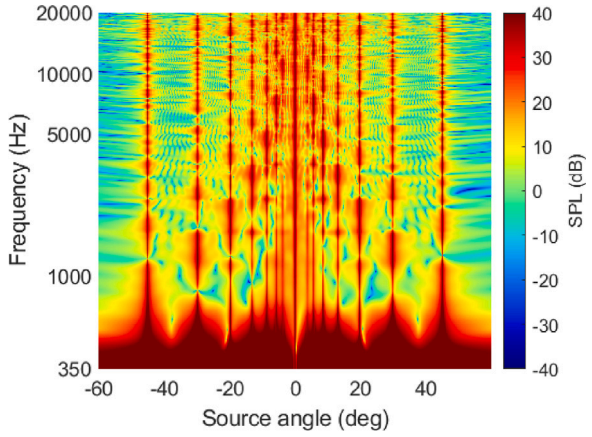

Fig. 12. The distribution of relative source strength on a decibel scale as a function of frequency for the three channel OSD system consisting of fifteen sources when ensuring crosstalk cancellation at the ears of three spaced apart listeners.

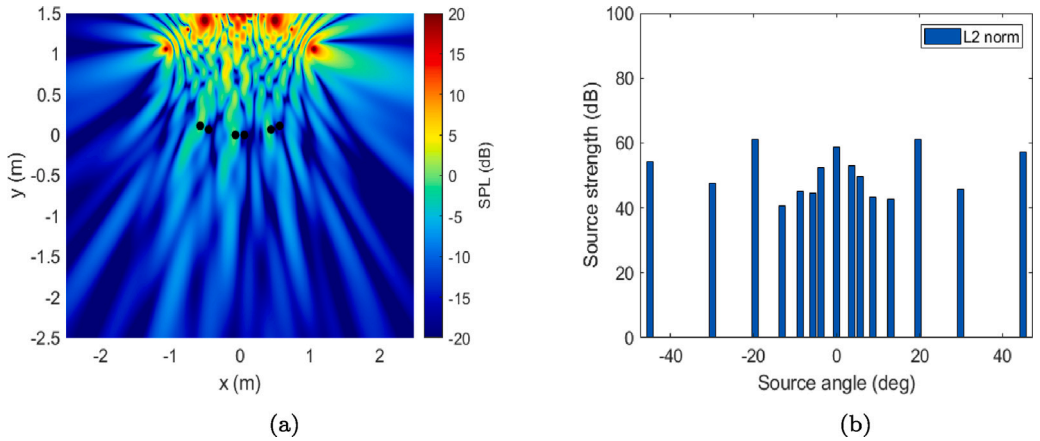

Fig. 13. An example of the radiated field and the relative source strength distribution on a decibel scale at a frequency of 1995 Hz when minimising the $L2$ norm of the source strength vector of fifteen sources used to emulate the three channel OSD.

where the index m is used to denote the source, with positive and negative values of m respectively denoting sources to the right and left of the vertical axis in Fig. 7. Note also that the choice $v_m = -v_{-m}$ is consistent with the opposite phase of the two outer sources of the three channel OSD [17].

The results of solving this problem are shown in Fig. 14(c) and (d) which demonstrates that the solution is both sparse and symmetric. Furthermore, the results in Fig. 15 show the solution to this problem for the entire range of frequencies of interest. Evidently the minimisation of the $L1$ norm seems almost always to select the pair of sources most closely associated with the OSD at a given frequency. However, at some frequencies, the minimisation appears to result in the selection of those pairs of sources corresponding to values of the odd integer n shown in Fig. 5.

Finally, some results of using the $L1$ norm minimisation solution to determine the sound field generated by the three-channel OSD with three spaced apart listeners are shown as a function of frequency in Fig. 16. This shows that the performance of crosstalk cancellation is comparable with that of the $L2$ norm solution, but that the solution is much more selective in the choice of source required. Thus comparison with Fig. 12 shows that the minimum $L1$ norm solution gives a greater concentration of source strength in a smaller number of sources, especially at higher frequencies. In this case the symmetrically constrained and unconstrained solutions show very little difference and only the symmetrically unconstrained solution is shown here.

10. Discussion

It has been demonstrated that the Optimal Source Distribution (OSD), can deliver crosstalk cancellation at the ears of multiple listeners above a lower limiting frequency. Using this observation as a starting point, solutions have presented for the optimal vector of source strengths when the OSD has discrete source elements. It has been shown that the discrete approximation to the OSD can

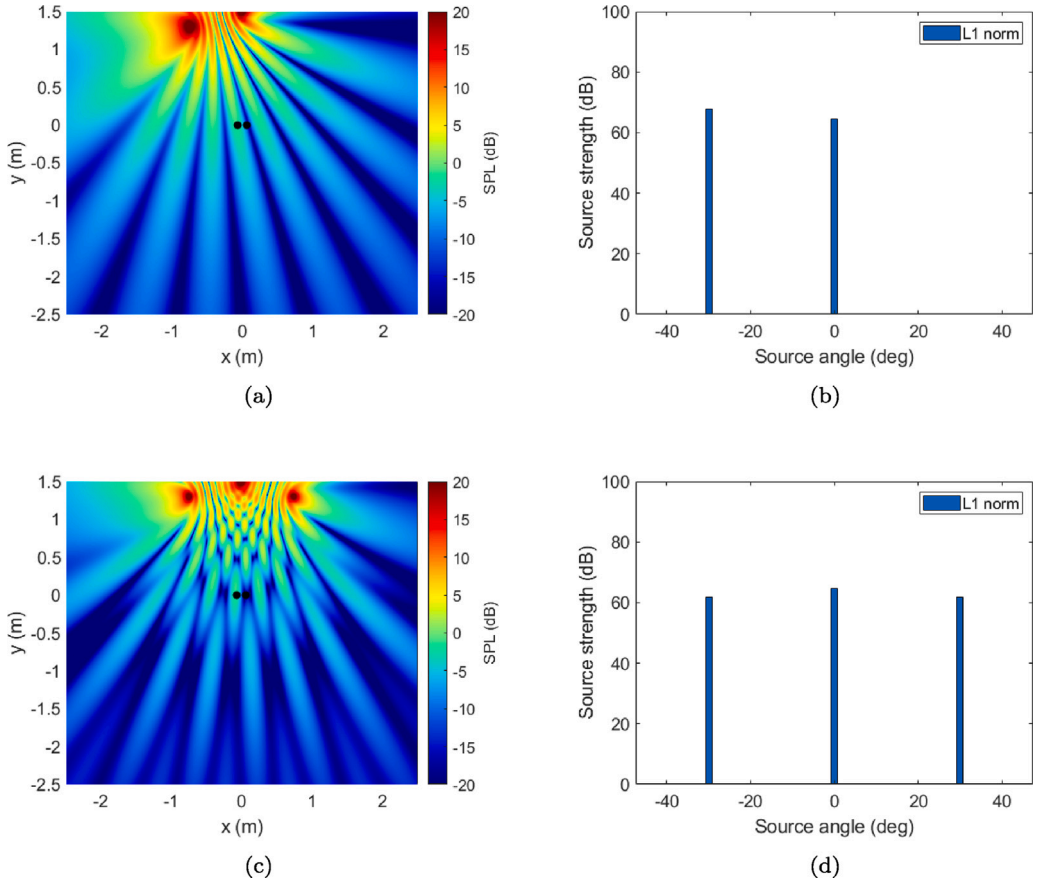

Fig. 14. The results of minimising the $L1$ norm of the source strength vector when using fifteen sources to produce crosstalk cancellation at a single centrally placed listener at a frequency of 1995 Hz showing the sound pressure field (a) without and (c) with the symmetry constraint and the source strength distribution (c) without and (d) with the symmetry constraint.

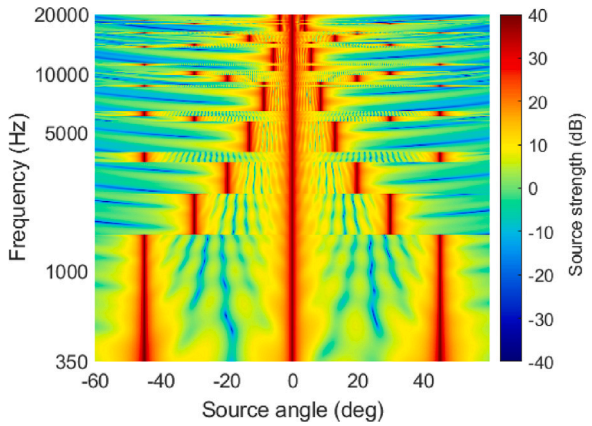

Fig. 15. The result of minimising the $L1$ norm of the symmetrically constrained vector of fifteen source strengths when ensuring crosstalk cancellation at the ears of a single centrally placed listener. The relative source strength distribution is shown on a decibel scale.

replicate the OSD sound field provided a sufficiently large number of sources is used. It has also been shown that linearly constrained least squares solutions can ensure crosstalk cancellation at a prescribed number of positions.

However, when the number of sources available is insufficient to provide replication of the OSD sound field over all frequencies, minimising both $L1$ and $L2$ norms of the source strength vector subject to the constraint of ensuring crosstalk cancellation appear to provide viable solutions. In particular for a single listener at a given frequency, the spacing of the elements of the OSD is shown to

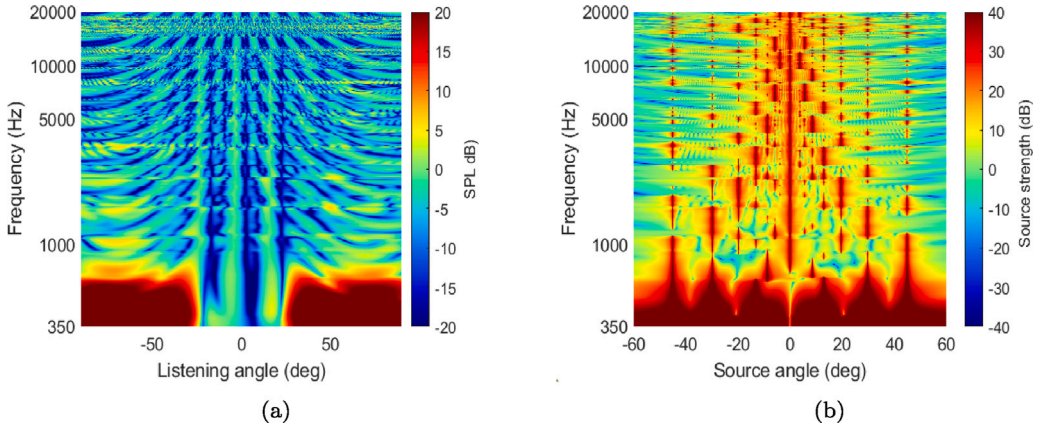

Fig. 16. An example of the radiated field and the relative source strength distribution on a decibel scale when minimising the $L1$ norm of the source strength vector of fifteen sources with the constraint of crosstalk cancellation at three spaced apart listeners.

provide a solution equivalent to minimising the $L1$ norm of the source strength vector with the constraint of crosstalk cancellation. Minimisation of the $L1$ norm also seems to provide sparse solutions that make use of the properties of the OSD through selecting the sources that provide well-conditioned relationships to the listener positions, although further investigations of this are necessary. To summarise, the comparison of the various methods used, it appears from the results generated to date that minimisation of either $L2$ or $L1$ norms of the source strength vector, or some combination of both, shows the most promise for engineering application.

A significant number of questions remain unanswered. In particular, the extent to which any of the strategies examined will yield subjectively acceptable solutions remains unclear. The simple use of each pair of sources comprising the OSD over their best conditioned frequency range in a modest number (typically less than a dozen) alone can produce virtual sound imaging [16,17]. Subjective experiments comparing this approach with the strategies outlined here are required. Further analyses that include factors such as loudspeaker directivity and scattering by the heads of listeners would also be helpful, although the informal listening experiments so far undertaken [16,17] do not indicate that there are any serious problems introduced by these factors. Whilst the emphasis of the work described here has been on the control of the physical properties of the sound field through crosstalk cancellation, which has hitherto yielded good subjective results, it may be that other physical properties may be relevant [37,38].

The design of inverse filters necessary to deliver the appropriate sound field is also a challenge that remains, especially if $L1$ norm solutions are to be exploited. Recent work on inverse filter design using more sophisticated cost functions may also be relevant [39]. Related to that is the time domain response of the systems that might be implemented and their subjective implications. It has been found from previous studies that crosstalk cancellation systems can be associated with highly recursive (and subjectively disruptive) "ringing" in the delivery of acoustic fields in the time domain [40]. Finally, the use of mixed $L2$ and $L1$ norm cost functions such as those used in the LASSO [35,41,42] would also be worthy of further investigation

11. Conclusions

The Optimal Source Distribution for single listener virtual sound imaging provides a useful basis for the delivery of multiple listener virtual sound imaging, although the number of listeners for whom this is possible will depend on the number of sources available. For a practical number of sources (such as fifteen for example), it seems feasible to deliver crosstalk cancellation for three listeners over a wide frequency range above a lower limiting value. Minimisation of either or both $L2$ and $L1$ norms of the source strength vector appear to offer the most promise for system design, although a number of challenges remain before such systems can be fully optimised.

Data availability

The authors do not have permission to share data.

Acknowledgements

This paper has been written in honour of the late Professor J.E. Ffowcs-Williams whose leadership in the field of acoustics was truly inspirational.

The long-standing support of the Kajima Corporation of Japan for research into binaural sound reproduction is gratefully acknowledged.

Appendix

The solution to the linearly constrained least squares problem relies on a QR factorisation of the matrix $\mathbf{B}^{\mathrm{H}}$ where H denotes the Hermitian transpose. This can be written as

$$\mathbf{B}^{\mathrm{H}} = \mathbf{Q}\begin{bmatrix}\mathbf{R}\\ \mathbf{0}\end{bmatrix} \tag{A.1}$$

The matrix $\mathbf{B}^{\mathrm{H}}$ is of order $M \times P$, the matrix $\mathbf{Q}$ is square of order $M \times M$ and has the property $\mathbf{Q}^{\mathrm{H}}\mathbf{Q} = \mathbf{Q}\mathbf{Q}^{\mathrm{H}} = \mathbf{I}$ whilst $\mathbf{R}$ is an upper triangular matrix of order $P \times P$ and the zero matrix is of order $(M - P) \times P$. The vector $\mathbf{v}$ is then expressed in terms of the matrix $\mathbf{Q}$ using

$$\mathbf{v} = \mathbf{Q}\begin{bmatrix}\mathbf{y}\\ \mathbf{z}\end{bmatrix} \tag{A.2}$$

The above definitions can then be used to show that

$$\mathbf{B}\mathbf{v} = \left(\mathbf{Q}\begin{bmatrix}\mathbf{R}\\ \mathbf{0}\end{bmatrix}\right)^{\mathrm{H}}\mathbf{v} = \begin{bmatrix}\mathbf{R}^{\mathrm{H}} & \mathbf{0}\end{bmatrix}\mathbf{Q}^{\mathrm{H}}\mathbf{v} = \begin{bmatrix}\mathbf{R}^{\mathrm{H}} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{y}\\ \mathbf{z}\end{bmatrix} = \mathbf{R}^{\mathrm{H}}\mathbf{y} \tag{A.3}$$

Furthermore, using the property $\mathbf{Q}\mathbf{Q}^{\mathrm{H}} = \mathbf{I}$ and defining the partitioned matrix given by $\mathbf{A}\mathbf{Q} = \begin{bmatrix}\mathbf{A}_1 & \mathbf{A}_2\end{bmatrix}$ also shows that

$$\mathbf{A}\mathbf{v} = \mathbf{A}\mathbf{Q}\mathbf{Q}^{\mathrm{H}}\mathbf{v} = \begin{bmatrix}\mathbf{A}_1 & \mathbf{A}_2\end{bmatrix}\begin{bmatrix}\mathbf{y}\\ \mathbf{z}\end{bmatrix} = \mathbf{A}_1\mathbf{y} + \mathbf{A}_2\mathbf{z} \tag{A.4}$$

The problem can now be transformed into a minimisation problem described by $\left\|\mathbf{A}_1\mathbf{y} + \mathbf{A}_2\mathbf{z} - \hat{\mathbf{w}}_A\right\|_2^2$ subject to $\mathbf{R}^{\mathrm{H}}\mathbf{y} = \hat{\mathbf{w}}_B$. Then since $\mathbf{y} = \mathbf{R}^{\mathrm{H}-1}\hat{\mathbf{w}}_B$ the minimisation problem can be written as

$$min\left\|\mathbf{A}_2\mathbf{z} - (\hat{\mathbf{w}}_A - \mathbf{A}_1\mathbf{y})\right\|_2^2 = min\left\|\mathbf{A}_2\mathbf{z} - (\hat{\mathbf{w}}_A - \mathbf{A}_1\mathbf{R}^{\mathrm{H}-1}\hat{\mathbf{w}}_B)\right\|_2^2 \tag{A.5}$$

The solution for the optimal source strength vector can be written in terms of the vector $\mathbf{z}$ where

$$\mathbf{v}_{opt} = \mathbf{Q}\begin{bmatrix}\mathbf{y}_{opt}\\ \mathbf{z}_{opt}\end{bmatrix} \tag{A.6}$$

such that the least squares solution to the minimisation problem can be written as

$$\mathbf{z}_{opt} = [\mathbf{A}_2^{\mathrm{H}}\mathbf{A}_2]^{-1}\mathbf{A}_2^{\mathrm{H}}(\hat{\mathbf{w}}_A - \mathbf{A}_1\mathbf{R}^{\mathrm{H}-1}\hat{\mathbf{w}}_B) \tag{A.7}$$

This solution can be written in terms of the optimal vector of source strengths by partitioning the matrix $\mathbf{Q}$ to give

$$\mathbf{v}_{opt} = \mathbf{Q}_1\mathbf{y}_{opt} + \mathbf{Q}_2\mathbf{z}_{opt} \tag{A.8}$$

Writing the pseudoinverse of matrix $\mathbf{A}_2$ as $[\mathbf{A}_2^{\mathrm{H}}\mathbf{A}_2]^{-1}\mathbf{A}_2^{\mathrm{H}} = \mathbf{A}_2^{\dagger}$ then enables the solution to be written as

$$\mathbf{v}_{opt} = \mathbf{Q}_1\mathbf{R}^{\mathrm{H}-1}\hat{\mathbf{w}}_B + \mathbf{Q}_2\mathbf{A}_2^{\dagger}(\hat{\mathbf{w}}_A - \mathbf{A}_1\mathbf{R}^{\mathrm{H}-1}\hat{\mathbf{w}}_B) \tag{A.9}$$

which can also be written as

$$\mathbf{v}_{opt} = \mathbf{Q}_2\mathbf{A}_2^{\dagger}\hat{\mathbf{w}}_A + (\mathbf{Q}_1\mathbf{R}^{\mathrm{H}-1} - \mathbf{Q}_2\mathbf{A}_2^{\dagger}\mathbf{A}_1\mathbf{R}^{\mathrm{H}-1})\hat{\mathbf{w}}_B \tag{A.10}$$

References

[1] J.E. Ffowcs-Williams, Anti-sound, Proc. R. Soc. Lond. A 395 (1984) 63–88.
[2] P.A. Nelson, S.J. Elliott, Active Control of Sound, Academic Press, 1991.
[3] S.J. Elliott, Signal Processing for Active Control, Academic Press, 2001.
[4] B.B. Bauer, Stereophonic earphones and binaural loudspeakers, J. Audio Eng. Soc. 9 (2) (1961) 148–151.
[5] B.S. Atal, M.R. Schroeder, United States Patent 3,236,949A, Apparent sound source translator, 1966.
[6] P. Damaske, Head-related two-channel stereophony with loudspeaker reproduction, J. Acoust. Soc. Am. 50 (4B) (1971) 1109–1115.
[7] Y. Ando, S. Shidara, Z. Maekawa, K. Kido, Some basic studies on the acoustic design of room by computer, J. Acoust. Soc. Japan 29 (1973).
[8] S. Sakamoto, T. Gotoh, T. Kogure, M. Shimbo, A. Clegg, Controlling sound-image localisation in stereophonic reproduction, J. Audio Eng. Soc. 29 (11) (1981) 794–799.
[9] H. Hamada, Construction of orthosterophonic system for the purposes of quasi-insitu recording and reproduction, J. Acoust. Soc. Am. 39 (5) (1983) 337–348.
[10] J.L. Bauck, D.H. Cooper, Generalised transaural stereo and applications, J. Audio Eng. Soc. 44 (9) (1996) 683–705.
[11] O. Kirkeby, P.A. Nelson, H. Hamada, F. Orduna-Bustamante, Fast deconvolution of multichannel systems using regularisation, IEEE Trans. Speech Audio Process. 6 (2) (1998) 189–194.
[12] E.C. Hamdan, Theoretical Advances in Multichannel Crosstalk Cancellation Systems (Ph.D. thesis), University of Southampton, 2020.
[13] T. Takeuchi, P.A. Nelson, Optimal source distribution for binaural synthesis over loudspeakers, J. Acoust. Soc. Am. 112 (6) (2002) 2786–2797, http://dx.doi.org/10.1121/1.1513363, URL https://www.ncbi.nlm.nih.gov/pubmed/12508999.
[14] T. Takeuchi, M. Teschl, P.A. Nelson, Objective and subjective evaluation of the optimal source distribution for virtual acoustic imaging, J. Audio Eng. Soc. 55 (11) (2007) 981–987.

[15] D.G. Morgan, T. Takeuchi, K.R. Holland, Off-axis cross-talk cancellation evaluation of 2-channel and 3-channel Opsodis soundbars, in: Proceedings of Reproduced Sound 2016, Southampton UK, 2016.
[16] L.A.T. Haines, T. Takeuchi, K.R. Holland, Investigating multiple off-axis listening positions of an Opsodis.
[17] M. Yairi, T. Takeuchi, K.R. Holland, D.G. Morgan, L. Haines, Binaural reproduction capability for multiple off-axis listeners based on the 3-channel optimal source distribution principle, in: Proceedings of the 23rd International Congress on Acoustics, Aachen, Germany, 2019.
[18] Y. Kim, O. Deille, P.A. Nelson, Crosstalk cancellation in virtual acoustic imaging systems for multiple listeners, J. Sound Vib. 297 (1–2) (2006) 251–266, http://dx.doi.org/10.1016/j.jsv.2006.03.042.
[19] B. Masiero, X. Qiu, Two listeners crosstalk cancellation system modelled by four point sources and two rigid spheres, Acta Acust. United Acust. 95 (2) (2009) 379–385.
[20] P. Mannerheim, P.A. Nelson, Virtual sound imaging using visually adaptive loudspeakers, Acta Acust. United Acust. 94 (2008) 1024–1039.
[21] H. Kurabayashi, O. Otani, M. Hashimoto, M. Kayama, Development of dynamic crosstalk cancellation system for multiple-listener binaural reproduction, Acoust. Sci. Technol. 36 (6) (2015) 537–539.
[22] M. Simon-Galvez, T. Takeuchi, F.M. Fazi, Low-complexity, listener's position-adaptive binaural reproduction over a loudspeaker array, Acta Acust. United Acust. 103 (5) (2017) 847–857.
[23] C. House, S. Dennison, D.G. Morgan, N. Rushton, G. White, J. Cheer, S.J. Elliott, Personal spatial audio in cars: Development of a loudspeaker array for multi-listener transaural reproduction in a vehicle, in: Proceedings of the Institute of Acoustics, UK, 2017, p. 39(2).
[24] J. Holleborn, F.M. Fazi, M.F. Simon Galvez, A multiple listener crosstalk cancellation system using loudspeaker-dependent regularization, J. Audio Eng. Soc. 69 (3) (2021) 191–203.
[25] B. Masiero, M. Vorlander, A framework for the calculation of dynamic crosstalk cancellation filters, IEEE/ACM Trans. Audio Speech Lang. Process. 22 (9) (2014) 1345–1354.
[26] T. Takeuchi, P.A. Nelson, Extension of the optimal source distribution for binaural sound reproduction, Acta Acust. United Acust. 94 (6) (2008) 981–987, http://dx.doi.org/10.3813/aaa.918114.
[27] P.A. Nelson, T. Takeuchi, P. Couturier, Optimal source distribution for multiple listener virtual sound imaging, in: Proceedings of the Institute of Acoustics, UK, 2019, pp. 35–48.
[28] G.E. Golub, C.F.V. Loan, Matrix Computations, third ed., The John Hopkins University Press, 1996.
[29] E. Hamdan, F. Fazi, A modal analysis of multichannel crosstalk cancellation systems and their relationship to amplitude panning, J. Sound Vib. 490 (2021) 1–21.
[30] F. Olivieri, F.M. Fazi, P.A. Nelson, S. Fontana, Comparison of strategies for accurate reproduction of a target signal with compact arrays of loudspeakers for the generation of zones of private sound and silence, J. Audio Eng. Soc. 64 (11) (2016) 905–917, http://dx.doi.org/10.17743/jaes.2016.0045.
[31] H.V. Henderson, S.R. Searle, On deriving the inverse of a sum of matrices, SIAM Rev. 23 (1) (1981) 53–60.
[32] S. Boyd, L. Vandenberghe, Convex Optimisation, Cambridge University Press, 2004.
[33] CVX: Matlab software for disciplined convex programming, version 2.0, 2012, http://cvxr.com/cvx.
[34] P. Gerstoft, C.F. Mecklenbrauker, W. Seong, M. Bianco, Introduction to compressive sensing in acoustics, J. Acoust. Soc. Am. 143 (6) (2018) 3731–3736.
[35] P.A. Gauthier, P. Lecomte, A. Berry, Source sparsity control of sound field reproduction using the elastic-net and the lasso minimizers, J. Acoust. Soc. Am. 141 (2018) 2315–2326.
[36] A. Franck, W. Wang, F.M. Fazi, Sparse L1-optimal multiloudspeaker panning and its relation to vector base amplitude panning, IEEE/ACM Trans. Audio Speech Lang. Process. 25 (5) (2017) 996–1010, http://dx.doi.org/10.1109/taslp.2017.2674975.
[37] H. Hacıhabiboglu, E. De Sena, Z. Cvetkovic, J. Johnston, J.O. Smith III, Perceptual spatial audio recording simulation and rendering, IEEE Signal Process. Mag. May (2017) 36–54.
[38] E. De Sena, H. Hacıhabiboğlu, Z. Cvetković, Analysis and design of multichannel systems for perceptual sound field reconstruction, IEEE Trans. Audio Speech Lang. Process. 21 (8) (2013) 1653–1665.
[39] A. Mertins, M. Maass, F. Katzberg, Room impulse response shaping and crosstalk cancellation using convex optimistaion, IEEE Trans. Audio Speech Lang. Process. 29 (2021) 489–502.
[40] P.A. Nelson, O. Kirkeby, T. Takeuchi, H. Hamada, Sound fields for the production of virtual acoustic images, J. Sound Vib. 204 (2) (1997) 386–396, http://dx.doi.org/10.1006/jsvi.1997.0967.
[41] G.N. Lilis, D. Angelosante, G.B. Giannakis, Sound field reproduction using the Lasso, IEEE Trans. Audio Speech Lang. Process. 18 (8) (2010) 1902–1912.
[42] Q. Feng, F. Yang, J. Yang, Time-domain sound field reproduction using the group Lasso, J. Acoust. Soc. Am. 143 (2) (2018) EL55–EL60.