Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Sound and Vibration 297 (2006) 251–266
JOURNAL OF SOUND AND VIBRATION
www.elsevier.com/locate/jsvi

Crosstalk cancellation in virtual acoustic imaging systems for multiple listeners

Y. Kim[a,*], O. Deille[b], P.A. Nelson[b]

[a] Samsung Advanced Institute of Technology, Nongseo-dong, Giheung-gu, Yongin-si, Gyeonggi-do, 446-712, South Korea
[b] Institute of Sound and Vibration Research, University of Southampton, Southampton, SO17 1BJ, UK

Received 30 November 2004; received in revised form 24 March 2006; accepted 27 March 2006
Available online 5 June 2006

Abstract

The perception of a virtual sound source is achieved by ensuring that the sound pressures at the ears of the listener are equivalent to those produced by the source at the virtual position. Theoretically, with only two loudspeakers for a single listener, virtual sources positioned anywhere in space can be presented provided that "crosstalk cancellation" can be achieved. The crosstalk cancellation problem is central to the problem of sound reproduction since an efficient crosstalk canceller gives one complete control over the sound field at a number of target positions. However, all crosstalk cancellation systems implemented so far have in practice produced virtual sources for only a single listener at a time. The design of crosstalk cancellers for multiple listeners involves a detailed study of the relative orientation of both sources and listeners. It is vital in any multiple listener system to first establish the conditioning of the potential geometrical arrangements of transducers and listeners by using simple free field models of the electro-acoustic transfer functions between transducers and ears. This gives an important link between the conditioning of the electro-acoustic transfer function matrix and the inverse filters for crosstalk cancellation. Optimal transducer arrangements for the efficient crosstalk canceller have been identified for the case of two listeners and these are evaluated here with time domain simulations.

1. Introduction

A virtual acoustic imaging system attempts to generate an illusion in a listener of being in a virtual acoustic environment that is entirely different from that of the space in which the listener is actually located. In order to render the virtual environment to a listener, binaural technology [1–3] is often used. The principle of this technology is to control the sound field at the listener's ears so that the reproduced sound field coincides with that produced when the listener is in the desired real sound field. One way of achieving this is to use a pair of loudspeakers at different positions in a listening space with the help of signal processing to ensure that appropriate binaural signals, which contain all the spatial information, are obtained at the listener's ears [4–7]. Of course, headphones have also been shown to be useful for recreating 3D sound scenes despite the feeling of

*Corresponding author. Tel.: +82 31 2808192; fax: +82 31 2808327.
E-mail address: yt65.kim@samsung.com (Y. Kim).

0022-460X/$ - see front matter
doi:10.1016/j.jsv.2006.03.042

intra-cranial sound and front–back confusion problems [1,2]. However, this paper focuses on loudspeaker systems. The technical developments undertaken to date for a single listener are described in full in references on digital filter inversion [5,7–10] associated with crosstalk canceller design, optimal positions of sources and receivers [11] and robustness to head movements [12]. Meanwhile, in many practical applications such as video games, simulators, and television, virtual acoustic imaging systems are required multiple users. In such cases, the systems have to be able to create virtual environments for multiple listeners at the same time. Therefore, it is clearly necessary that research be initiated to examine techniques for the delivery of high quality virtual acoustic images to multiple listeners.

In this paper, the central aspect of the research for multiple listeners involves a detailed study of the basic inversion processes in designing the crosstalk canceller that enables the reproduction of the desired signals. The singular value decomposition (SVD) [13] is used to explain physically the system inversion processes and to identify the particular frequencies where the inversion problem becomes well-conditioned. Even though this work is undertaken in a relatively straightforward fashion by using simple free field models of the electro-acoustic transfer functions between the transducers (such as loudspeakers) and the listeners' ears, this is to first establish the conditioning of the potential geometrical arrangements of transducers and multiple listeners. Also, the work reveals an important link between the conditioning of the electro-acoustic transfer function matrix and the inverse filters for the crosstalk cancellation. Therefore, the crosstalk cancellation solutions to be presented for multiple listeners will be represented by a range of transducers in different locations with each transducer handling a specific frequency bandwidth. In this paper, optimal transducer arrangements associated with the frequency bands will be suggested. Their responses in time domain will also be evaluated, since there appears to be a link between the time domain response of virtual acoustic imaging systems and their robustness to listener head movement [7,10,12,14,15].

2. Matrix inversion for crosstalk cancellation and the condition number

As outlined above, the virtual acoustic imaging systems developed to date have been implemented solely in order to generate high-quality images for a single listener. One of the objectives of the system is to feed to each ear of a listener independently the binaural signals containing the spatial information. However, when two loudspeakers are used, each loudspeaker delivers its signal to both ears through acoustic transmission paths between the loudspeakers and the listener's ears. These paths can be expressed as a matrix of transfer functions that is also referred to as the plant matrix. However, independent control of the signals (the binaural signals) at two receivers (the ears of a listener) can be achieved with two transducers by filtering the input signals to the transducers.

For example, with reference to the block diagram of Fig. 1, assume that the objective of the free field sound reproduction system for a 2-source/2-receiver system is to produce acoustic pressure signals defined by

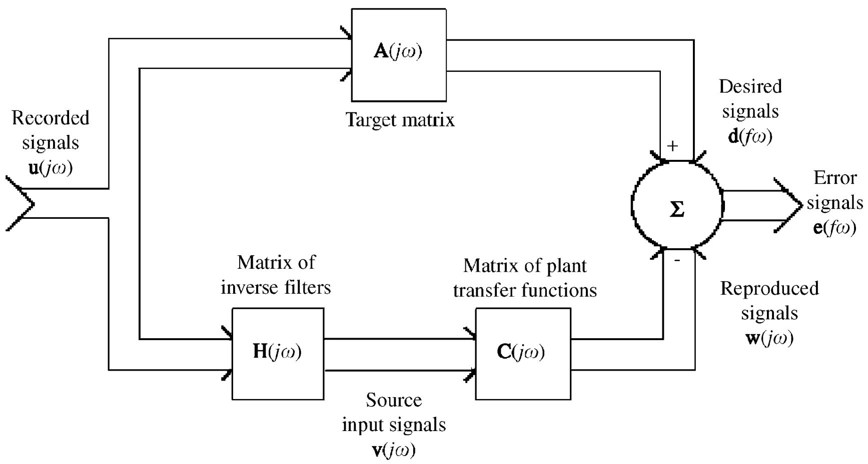

Fig. 1. Block diagram of the free field sound reproduction problem.

$\mathbf{w} = [w_1(j\omega), w_2(j\omega)]^T$. Also when assuming that these signals are expressed in terms of the signals $\mathbf{u} = [u_1(j\omega), u_2(j\omega)]^T$, thus $\mathbf{w} = \mathbf{Cv}$ where $\mathbf{C}$ is the plant matrix and the complex vector $\mathbf{v}$ denotes the transducers' output signals. The signals $\mathbf{u}$ are input to a filter matrix $\mathbf{H}$ designed to produce the requisite source volume acceleration vector $\mathbf{v}$ and thus [11]

$$\mathbf{w} = \mathbf{CHu}. \tag{1}$$

Meanwhile, the desired signals $\mathbf{d}$ $(= [d_1(j\omega), d_2(j\omega)]^T)$ which would produce a desired virtual auditory sensation (for example the binaural signals for crosstalk cancellation) can be obtained by recording sound source signals $\mathbf{u}$ with a dummy head or by filtering the signals $\mathbf{u}$ with the synthesised binaural filters $\mathbf{A}$. In such cases, the matrix $\mathbf{A}$ becomes the identity matrix $\mathbf{I}$. In other words, the desired signals $\mathbf{d}$ in Fig. 1 are the acoustic pressure signals which would have been produced by the closer sound alone whose values are either $d_1(j\omega)$or $d_1(j\omega)$ without disturbance (i.e., crosstalk) due to the other source. This is given by

$$\mathbf{d} = \mathbf{Iu}. \tag{2}$$

Therefore, in order to ensure the reproduction of the binaural signals at the receivers, it is required that $\mathbf{d} = \mathbf{w}$ which leads to

$$\mathbf{CH} = \mathbf{I}, \tag{3}$$

where the matrix $\mathbf{CH}$ represents the control performance of the system. The elements of the filter matrix $\mathbf{H}$ for crosstalk cancellation can be obtained from the exact inverse of the transfer function matrix of the plant $\mathbf{C}$:

$$\mathbf{H} = \mathbf{C}^{-1}. \tag{4}$$

However, in practice the system inversion involved gives rise to a number of problems [11], for example, the amplification required by the system inversion and cancellation of sound at listener's ears results in loss of dynamic range, low signal to noise ratio, more severe nonlinear distortion and fatigue of transducers. Also, errors contained in the plant, such as individual differences in head related transfer functions or misalignment of loudspeakers and the listener's head, may result in deterioration of control performance and coloration of sound. Therefore, deducing the solution of such problems is not straightforward due to the practical difficulty caused by their inherent behaviour. In order to overcome this difficulty, the problem is replaced by a problem whose solution approximates the required solution. This mathematical treatment is based on least squares estimation. This approach has been fully described in Refs. [16–18] and the least squares estimate of the inverse filter $\mathbf{H}$ is given by

$$\mathbf{H} = \mathbf{C}^{+}, \tag{5}$$

where the matrix $\mathbf{C}^{+} = [\mathbf{C}^{H}\mathbf{C}]^{-1}\mathbf{C}^{H}$ is the 'pseudo inverse' [13] of the plant matrix $\mathbf{C}$ where the superscript H denotes Hermitian transpose.

In dealing with the practical difficulty associated with the system inversion described in the above, good use can be made of the SVD [16–18]. The usefulness of SVD stems from the fact that the matrix $\mathbf{C}$ can be decomposed into the following product of the three matrices [13], thus $\mathbf{C} = \mathbf{U\Sigma V}$ where the matrix $\mathbf{U}$ is a matrix of left singular vectors of the matrix $\mathbf{C}$, and the matrix $\mathbf{V}$ is a matrix of right singular vectors of the matrix $\mathbf{C}$. Both matrices $\mathbf{U}$ and $\mathbf{V}$ are unitary and have the properties $\mathbf{U^H} = \mathbf{U}^{-1}$ and $\mathbf{V^H} = \mathbf{V}^{-1}$. The matrix Σ has only diagonal elements σ_i that comprise the singular values of the matrix $\mathbf{C}$. By substituting this relationship into Eq. (5), and by using the orthonormal properties of the unitary matrices $\mathbf{U}$ and $\mathbf{V}$, the least squares estimation of the inverse filter matrix $\mathbf{H}$ can be written as [16–18]

$$\mathbf{H} = \mathbf{V}\Sigma^{+}\mathbf{U}^{H}, \tag{6}$$

where the matrix Σ^{+} is the pseudo inverse of the matrix Σ and is defined as $\Sigma^{+} = \mathrm{diag}\,(1/\sigma_i)$. It is well known [13] that in any problem involving the inversion of a matrix the conditioning of the matrix associated with the behaviour of the small singular values dictates the sensitivity to errors of the resulting solution. The condition number $\kappa(\mathbf{C})$ of the plant matrix $\mathbf{C}$ is simply defined as a ratio between the largest and smallest non-zero singular values of the matrix $\mathbf{C}$, and is given by [13]

$$\kappa(\mathbf{C}) = ||\mathbf{C}||\,||\mathbf{C}^{+}||, \tag{7}$$

where $\| \ \|$ denotes the two-norm of the matrix. For example, when the condition number is large, the system inversion is very sensitive to small errors in the assumed plant matrix **C** (which is often measured and thus the small errors are inevitable). Therefore, the control performance of the filters for crosstalk cancellation will be crucially determined by the conditioning of the plant matrix **C**.

3. Analysis with a free field model for multiple listeners

3.1. A free field model for a single listener

In the above, the importance of the condition number has been shown as the most important attribute of the plant matrix **C** relating the control performance and robustness in order to achieve crosstalk cancellation. For example, a simple case for a single listener as depicted in Fig. 2 involves the control of two receivers such as the ears of a listener (w_1 and w_2 where the distance between two ears is 0.18 m) with two transducers (which are assumed to be two point monopole sources; v_1 and v_2) under free field conditions. l_1 and l_2 are path lengths between the sources and the field points with $l_2 > l_1$. The case is a symmetric example with the inter-source axis parallel to the inter-receiver axis and the plant matrix can be modelled as

$$\mathbf{C} = \frac{\rho_o}{4\pi}\begin{bmatrix} \mathrm{e}^{-\mathrm{j}kl_1}/l_1 & \mathrm{e}^{-\mathrm{j}kl_2}/l_2 \\ \mathrm{e}^{-\mathrm{j}kl_2}/l_2 & \mathrm{e}^{-\mathrm{j}kl_1}/l_1 \end{bmatrix}, \tag{8}$$

where an $\mathrm{e}^{\mathrm{j}\omega t}$ time dependence is assumed with $k = \omega/c_0$, and where ρ_0 and c_0 are the density and sound speed. For this case depicted in Fig. 2 for example, the condition number in dB ($= 10\log_{10}\kappa(\mathbf{C})$) is shown as a function of frequency in Fig. 3. In this case, the system will be very sensitive to small errors in the matrix **C** around frequencies where $\kappa(\mathbf{C})$ is large. Thus the designed inverse filter matrix **H** is likely to contain large errors due to the small errors in the matrix **C** and results in large errors in the reproduced signal **w** at the receivers. This is because such errors are magnified by the inverse filters but remain uncancelled in the plant. On the contrary, at the frequencies where $\kappa(\mathbf{C})$ is small, the system is robust, and as shown in Fig. 3 the frequency range of the robust inversion becomes wider as the source span becomes smaller. The so-called Stereo Dipole [7,20] describes a virtual imaging system that comprises two closely spaced loudspeakers corresponding to a source span 2θ of 10°. The significance of this angle is that a particular frequency range is best conditioned when this choice of source span is made, and therefore the particular loudspeaker arrangement enables crosstalk cancellation to be readily achieved over an important frequency range. The so-called "optimal source distribution" [11] on the other hand makes use of number of pairs of loudspeakers with different angular spans in order to ensure a well-conditioned inversion problem over a wide range of frequency.

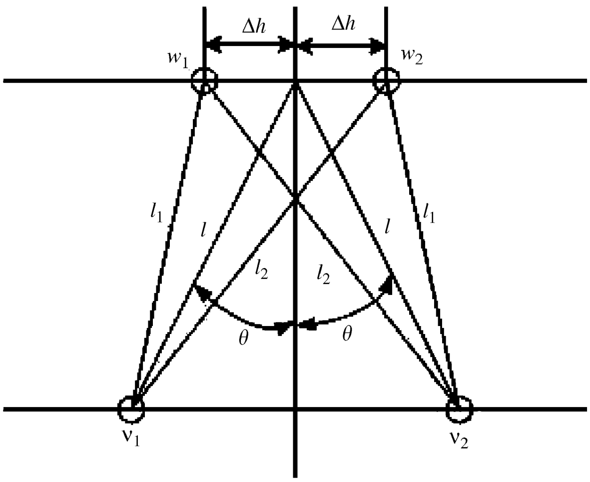

Fig. 2. Geometry of a 2-source (point monopoles) and 2-receiver system.

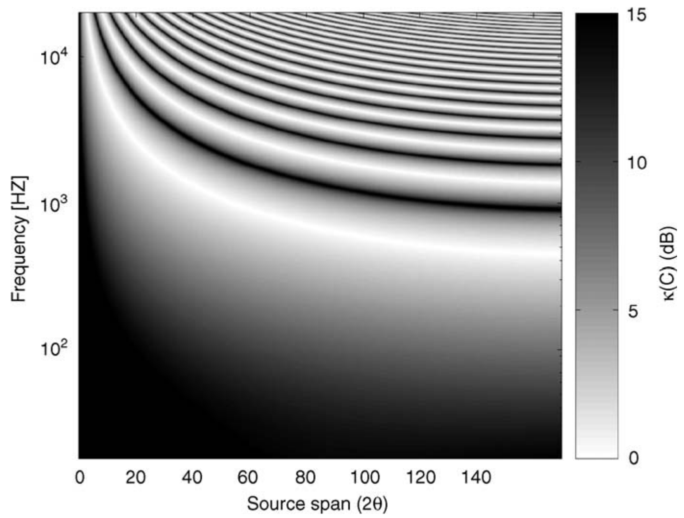

Fig. 3. Condition number $\kappa(\mathbf{C})$ (in dB) of a free field plant matrix **C** as a function of source span and frequency.

3.2. Conditioning of the 4-source/4-receiver system with aligned sources

The observations described in the previous subsection can hold for any symmetrical arrangement of two sources and two field points. Now, the first step is thus to study and find the optimal spatial arrangement of sources and receivers for multiple listeners in order to obtain the smallest condition number of the plant matrix **C**. This paper deals with a virtual imaging system having four sources and four receivers (or equivalently two listeners). However, since this is an initial observation of the dependence of the condition number on geometrical factors for multiple listeners, it will be more helpful to keep matters simple in the first instance. This will enable practical guidelines to be proposed for determining an arrangement of aligned sources in front of the receivers that has the smallest condition number of the plant matrix **C**.

In general, the acoustic pressure field produced by a free field point monopole can be expressed by

$$p(l) = \frac{\rho_0 v}{4\pi l} \mathrm{e}^{-\mathrm{j}kl}, \tag{9}$$

where v is the source volume acceleration and l is the distance between the source and the field data. The general form of the plant matrix **C** for multiple sources and listeners can be written as

$$\mathbf{C} = \begin{bmatrix} C_{11}(\mathrm{j}\omega) & C_{12}(\mathrm{j}\omega) & \bullet & \bullet & C_{1S}(\mathrm{j}\omega) \\ C_{21}(\mathrm{j}\omega) & C_{22}(\mathrm{j}\omega) & \bullet & \bullet & C_{2S}(\mathrm{j}\omega) \\ \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet \\ C_{R1}(\mathrm{j}\omega) & C_{R2}(\mathrm{j}\omega) & \bullet & \bullet & C_{RS}(\mathrm{j}\omega) \end{bmatrix}, \tag{10}$$

where R and S denote the number of the receiver (the ear of the listener) and the source, respectively. In Eq. (10), the element of the matrix **C**, $C_{ij}(\mathrm{j}\omega)$ between the receiver i and the source j is given by

$$C_{ij}(\mathrm{j}\omega =) \frac{\rho_o}{4\pi l_{ij}} \mathrm{e}^{-\mathrm{j}kl_{ij}}, \tag{11}$$

where l_{ij} is the distance between the receiver i and the source j. For example, in the, the 4-source/4-receiver system depicted in Fig. 4, the distance between the closest ears of two listeners is assumed to be fixed to 0.52 m and two ears of single listener is separated by 0.18 m. Also, four sources (v_1, v_2, v_3 and v_4) are aligned in front of the listeners at the distance, 1 m. The purpose of the study is to evaluate the condition number of the plant

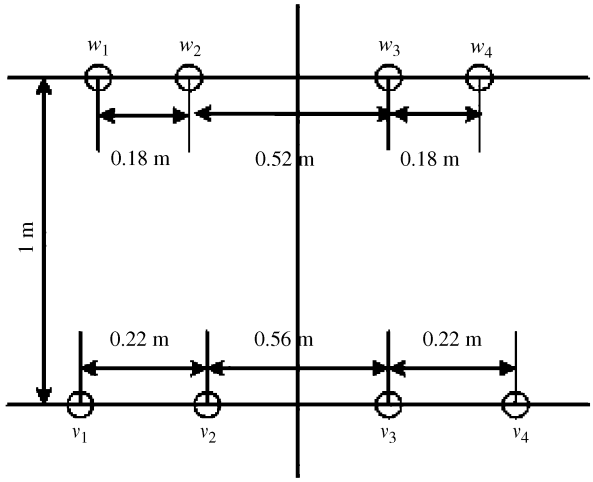

Fig. 4. An example of a configuration for the 4 source/4-receiver system.

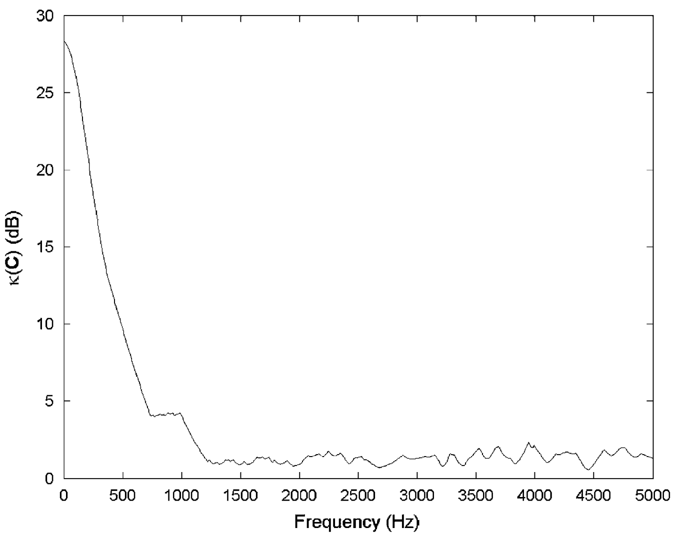

Fig. 5. The smallest condition number $\kappa(\mathbf{C})$ (in dB) obtained with the possible locations of four aligned sources depicted in Fig. 4.

matrix for all possible positions of the four sources with the aligned receivers as depicted in Fig. 4. This evaluation is undertaken for each frequency and the combination of source positions is identified that produces the smallest condition number at each frequency. In undertaking this exercise, the possible locations of the sources are assumed to be at intervals of 0.05 m along the line depicted in Fig. 4 line within the range from −0.6 to 0.6 m. The evaluation of condition number is undertaken for each discrete frequency in the range from 0 to 5000 Hz at intervals of 5 Hz.

Fig. 5 shows the smallest the condition number $\kappa(\mathbf{C})$ of the plant matrix $\mathbf{C}$, for the possible locations of sources on the line in the front of the receivers depicted in Fig. 4. As shown in this figure, except for very low frequencies, the condition number can be made to have a small magnitude at any frequency by correct choice of source position. These results can be compared, for example, with the results illustrated in Fig. 6(b) that are computed for the fixed 4-source/4-receiver systems depicted in Fig. 6(a). It is evident that optimal positions of the four sources exist at each frequency, which ensures the smallest condition number. The optimal positions of the sources are illustrated in Fig. 7. However, it is clearly difficult with current technology to use all the

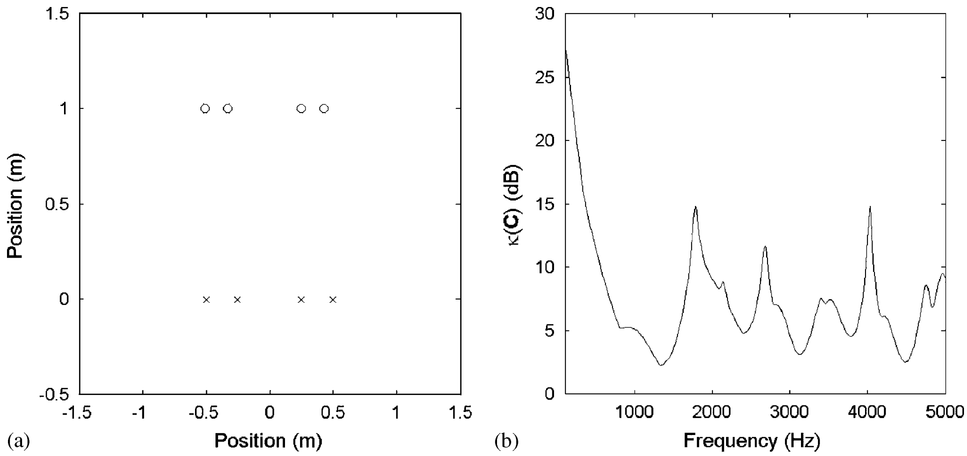

Fig. 6. A 4 source/4-receiver system and its condition number variation as a function of frequency: (a) sources (cross) and receivers (circle) positions; (b) condition number $\kappa(\mathbf{C})$.

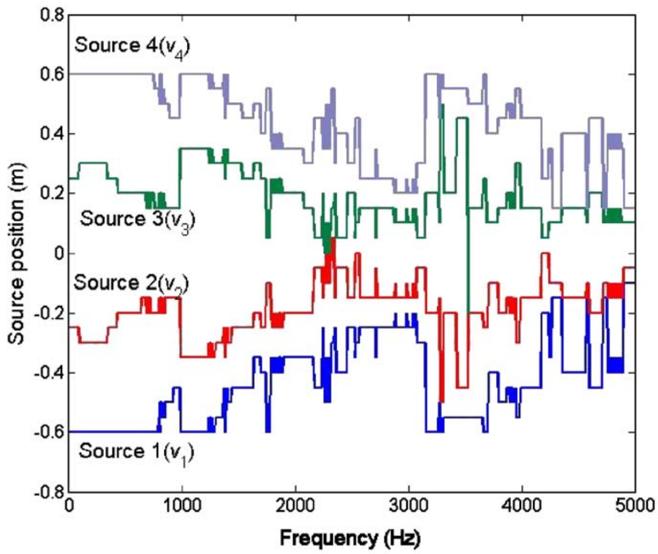

Fig. 7. Positions of the four sources having the smallest condition number at each frequency.

geometrical arrangements of the suggested source positions having the smallest condition numbers. Such an approach would require a finely discretised array of sources each allocated to transmit a narrow band of frequencies. An alternative approach is to divide the frequency range from 0 to 5000 Hz into several narrow frequency bands associated with a single arrangement of four sources. The suggested six frequency bands and the associated source positions are summarised in Table 1 (also, see Fig. 8), and Fig. 8(b) shows the condition number of the associated plant matrix $\mathbf{C}$ at each frequency band. The results of this more practical alternative are illustrated in Fig. 8 which shows the particular choice of arrangement and frequency bandwidth. It should be borne in mind however that a potentially much larger number of the geometrical arrangements of the aligned four sources for the associated the frequency bands can lead to a smaller condition number at all frequencies.

Table 1
Positions of the aligned sources for different frequency bands based on the results shown in Fig. 8

Frequency band (Hz)	Source position (m)			
	V_1	V_2	V_3	v_4
0–800 ($f_0 = 500$ Hz)	−0.6	−0.2	0.2	0.6
801–1750 ($f_0 = 1275$ Hz)	−0.5	−0.28	0.28	0.5
1751–2400 ($f_0 = 2075$ Hz)	−0.35	−0.2	0.2	0.35
2401–3050 ($f_0 = 2500$ Hz)	−0.25	−0.15	0.15	0.25
3051–3900 ($f_0 = 3400$ Hz)	−0.55	−0.2	0.2	0.55
3901–5000 ($f_0 = 4400$ Hz)	−0.42	−0.15	0.15	0.42

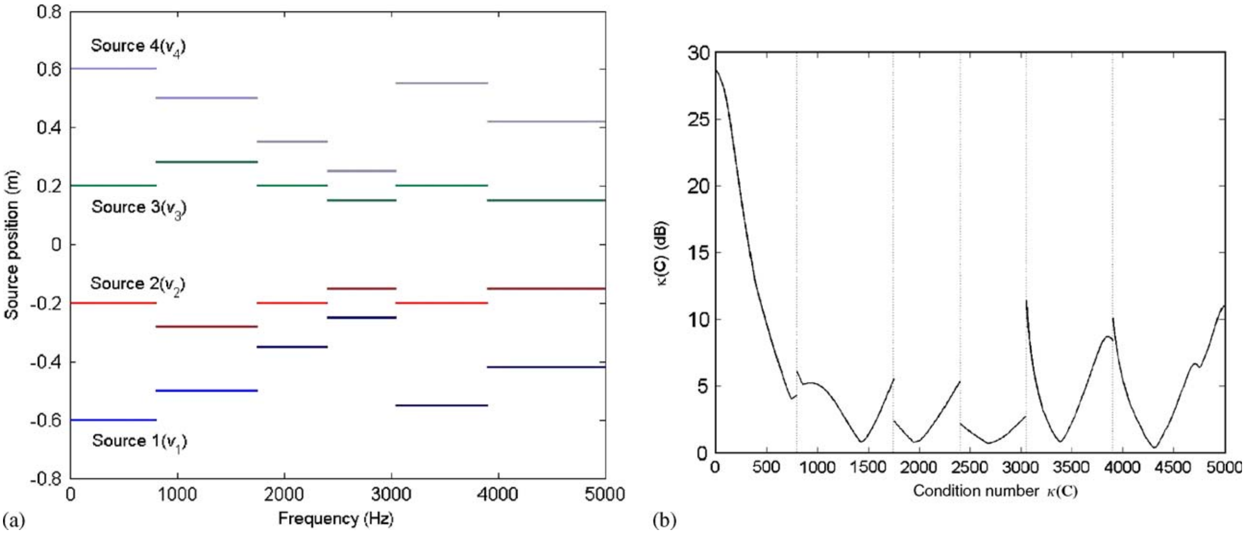

Fig. 8. Six frequency bands where single arrangement of four sources can be represented: (a) the source positions summarised in Table 1 for six frequency bands; (b) condition number $\kappa(\mathbf{C})$.

4. Time domain simulation of crosstalk cancellation

4.1. Introduction

In the previous section, several practical arrangements of four aligned sources associated with the frequency bands have been proposed. Now, it is important to confirm their accuracy with the solutions in the time domain. In general, a time domain simulation typically requires several orders of magnitude more computational time than a frequency domain simulation. Nevertheless, it is still valuable to calculate the full time history of the radiated field, since the solution in the time domain illustrates the form of the reproduced field when attempting to produce a short duration pulse at one ear of each of the listeners. The results of this time domain simulation will show the sound wave interference pattern in the horizontal plane containing the sources and the receivers. The pattern can give an indication of the precision and quality of the virtual acoustic scene produced, especially for example, with regard to the robustness with respect to the listener's head movement and undesirable time domain responses associated with frequencies at which the condition number is large. This corresponds to the 'ringing frequency [21]' problem in the two-channel case. In the case of a single listener illustrated in Refs. [14,21], time domain simulations have shown that crosstalk cancellation can be produced with a well defined wave-field of relatively short duration provided that the frequency content of the pulse to be reproduced lies within the range of frequencies over which the inversion problem is well-conditioned. The short duration of source outputs makes the system less sensitive to the listener's head movement and may have psychoacoustical benefits. The purpose of the time domain simulations presented here is to demonstrate that these observations still hold for multiple listeners in the case of the 4-source/4-receiver system depicted in Fig. 4.

The approach taken is to consider the source outputs necessary to achieve crosstalk cancellation for a pulse of both limited time duration and limited bandwidth. The pulse chosen to illustrate the system response is assumed to be the Gaussian pulse having the time history given by [14]

$$d_{\text{pulse}}(t) = \mathrm{e}^{-\pi(at)^2}\cos(\omega_0 t), \tag{12}$$

that has the Fourier transform

$$D_{\text{pulse}}(\mathrm{j}\omega) = \frac{1}{2|a|}\left[\mathrm{e}^{-(\omega-\omega_0)^2/4\pi a^2} + \mathrm{e}^{-(\omega+\omega_0)^2/4\pi a^2}\right], \tag{13}$$

where ω_0 is the carrier frequency and a is a constant that determines the duration and bandwidth of the pulse. The sampling frequency $F_s = 20\,480$ Hz which is chosen to cover a wide enough frequency band. The inverse filter matrix **H** for crosstalk cancellation has been calculated using Eq. (5) or (6) and the inverse filter **H** has 2048 coefficients.

For time domain simulations with multiple sources and listeners, the geometrical arrangement of four receivers depicted in Fig. 4 has been used, i.e., the listeners face the point monopole sources and two ears of a single listener is separated by 0.18 m where the receivers w_1 and w_3 are their right ears, and w_2 and w_4 are their left ears The four sources (v_1, v_2, v_3 and v_4) are aligned in front of the listeners at 1 m and their positions for each frequency band were summarised in Table 1. In the plots presented, the ears of the listeners are represented by white circles, and are positioned at the top of the area to be calculated. The sources are represented by black squares and are generally at the bottom of the area. The total sound field is represented by nine 'snapshots' or frames which are listed sequentially in a reading sequence from top left to bottom right; top left is the earliest time and bottom right is the latest time. The results demonstrate that perfect crosstalk cancellation produces silence at the right ears (w_1 and w_3) and the Gaussian pulse at the left ears (w_2 and w_4). The time increment between each frame Δt is 1.22×10^{-3} s, which is equivalent to the time it takes the sound to travel approximately 0.3 m. In each frame, the sound field is calculated at 131×131 discrete points over the area of $1.3\,\text{m} \times 1.3\,\text{m}$ ($-0.65\,\text{m} < x < 0.65\,\text{m}$ and $-0.15\,\text{m} < y < 1.15\,\text{m}$). For each point, the transfer function

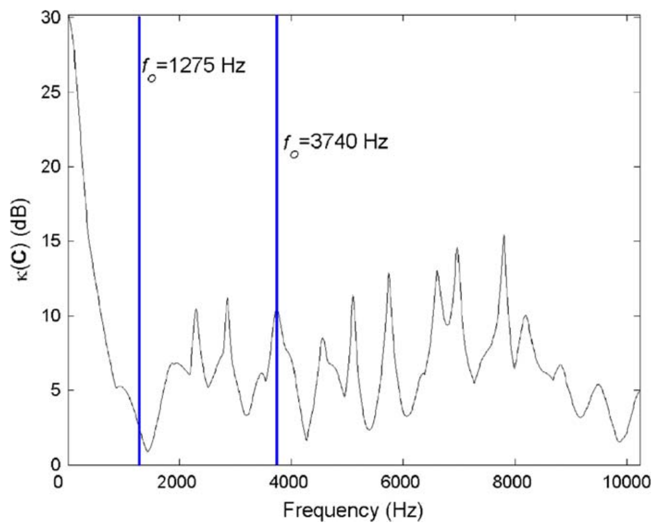

Fig. 9. Condition numbers $\kappa(\mathbf{C})$(in dB) of the plant matrix for the configuration of the four aligned sources for 801–1750 Hz frequency band.

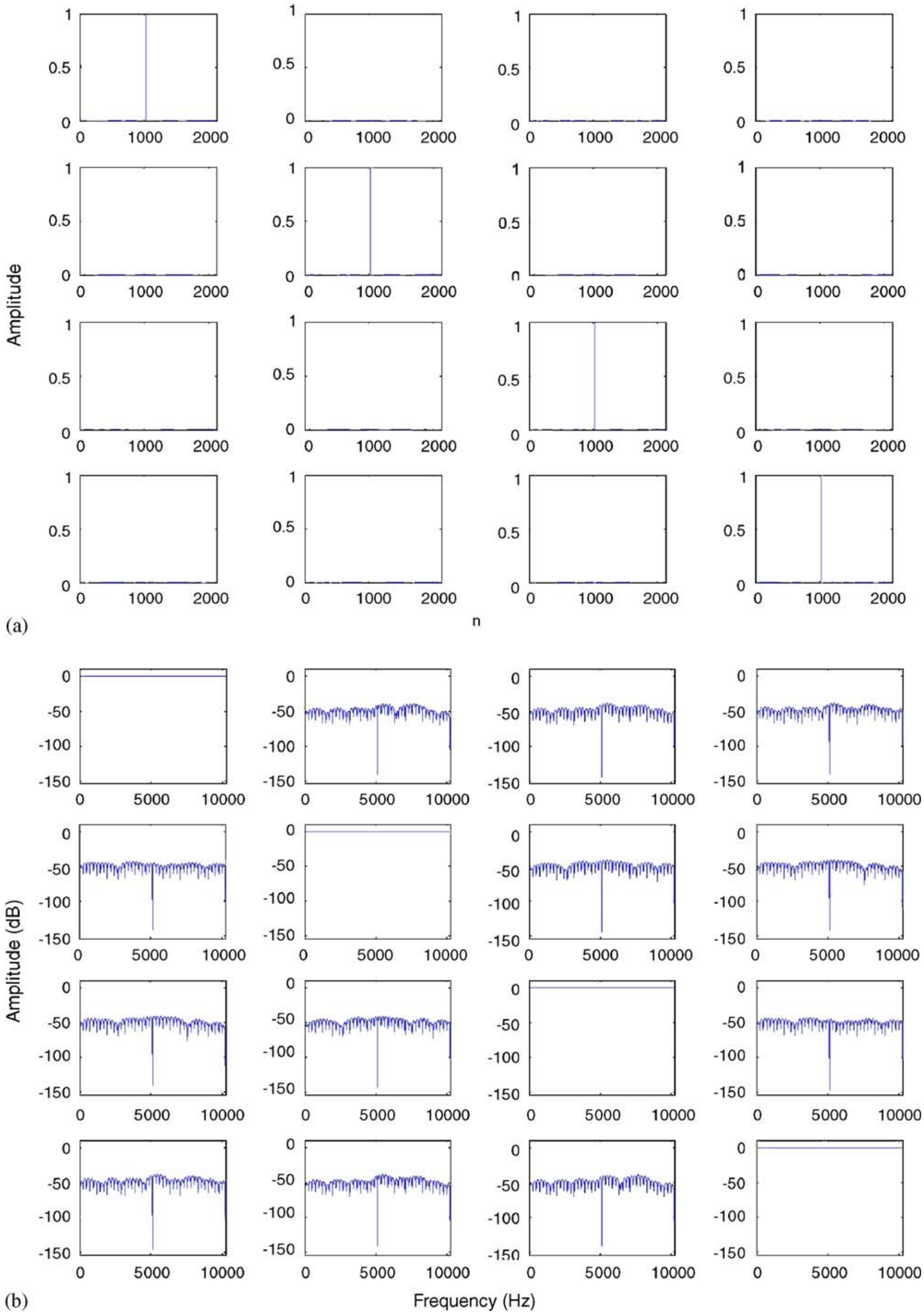

Fig. 10. Time and frequency characteristics of the control performance **CH** for the well-conditioned configuration of the four aligned sources for 801–1750 Hz frequency band: (a) impulse response, CH(n), of the elements of the **CH**; (b) frequency response, CH(f), of the elements of the **CH**.

between each source and the position is calculated by using Eq. (11) and then convolved with the source signal to obtain the acoustic pressure at the position as a function of time. In the frames, values greater than 1 are plotted as white, values smaller than -1 as black, and values between -1 and 1 are shaded appropriately from white to black.

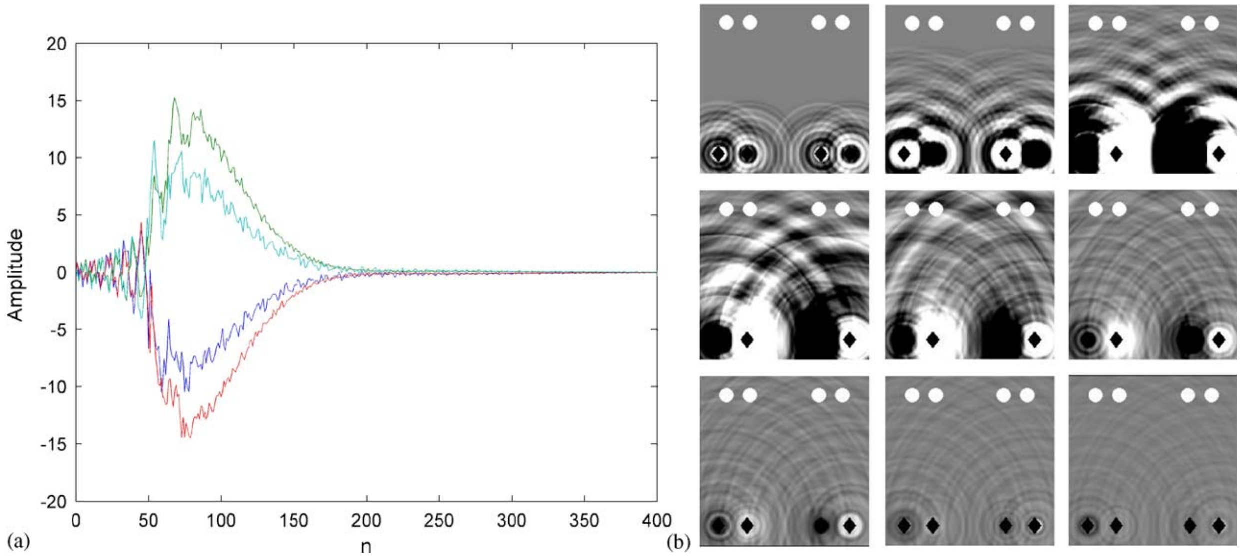

Fig. 11. Sound field reproduced by the aligned sources at the well-conditioned frequency band ($f_0 = 1275$ Hz). The source inputs are designed to achieve crosstalk cancellation at the listeners' right ears: (a) the signals of the four sours achieving crosstalk cancellation with a Gaussian pulse; (b) time domain solutions.

4.2. The suggest configuration of four aligned sources in 801–1750 Hz frequency band

The suggested configuration of the aligned sources for the frequency band 801–1750 Hz is presented in Fig. 8(a) and Table 1, and the condition number in the frequency band is shown in Fig. 9. For this case, time and frequency characteristics of the control performance of the system **CH** defined by Eq. (3) are shown in Fig. 10. It has been shown [19] that perfect crosstalk cancellation requires the diagonal terms of the matrix of impulse responses CH(n) should be a single peak at half of the impulse response length due to the modelling delay. In such cases, the amplitude of the off-diagonal elements must be zero for all time histories. This corresponds that in frequency domain the diagonal elements the matrix of frequency responses CH(f) should have a flat frequency response with 0 dB in amplitude. The other elements should have a great attenuation in amplitude over the whole frequency range. As shown in Fig. 10 for the case considered here, the control performance of the system with the crosstalk cancellation filter matrix **H** seems to be very good since the spectrums of the diagonal terms of CH(f) are perfectly flat with 0 dB and the amplitude of the off-diagonal elements are attenuated by more than 45 dB.

With reference to the suggested configuration of the aligned sources at the well-conditioned frequency band, time domain simulation is presented in Fig. 11 when the central frequency of the Gaussian pulse in Eq. (9), $f_0(=\omega_0/2\pi)$, is 1275 Hz (see the condition number at $f_0 = 1275$ Hz in Fig. 9) and the constant $a = 500$. In this case, the source signals to be emitted as shown in Fig. 11(a) are simple, and thus the solution in the time domain shows a good result as illustrated in Fig. 11(b). In this figure, it can be seen that at the initial stage, some emitted signals with small amplitude near field signals vanish rapidly before reaching the receivers. Also, once even main pulses have gone, the signals decay very quickly and thus the source outputs ring on for only a short duration. This means that pulses centred on well-conditioned frequencies are easily reproduced and crosstalk cancellation is naturally easy within the sound field.

In order to show the effects of attempting to achieve crosstalk cancellation for a signal whose frequency range is outside the band of optimum inversion (i.e., in an ill-conditioned frequency band), the centre frequency of the Gaussian pulse is made equal to $f_0 = 3740$ Hz (see the condition number at $f_0 = 3740$ Hz in Fig. 9). The results of the changes in the source signals are plotted in Fig. 12(a). The source signals have much more complicated and sharp fluctuations compared with the results shown in Fig. 11(a), and thus the source outputs ring on for a long duration as illustrated in Fig. 12(a). In such cases, in order to cancel crosstalk, the

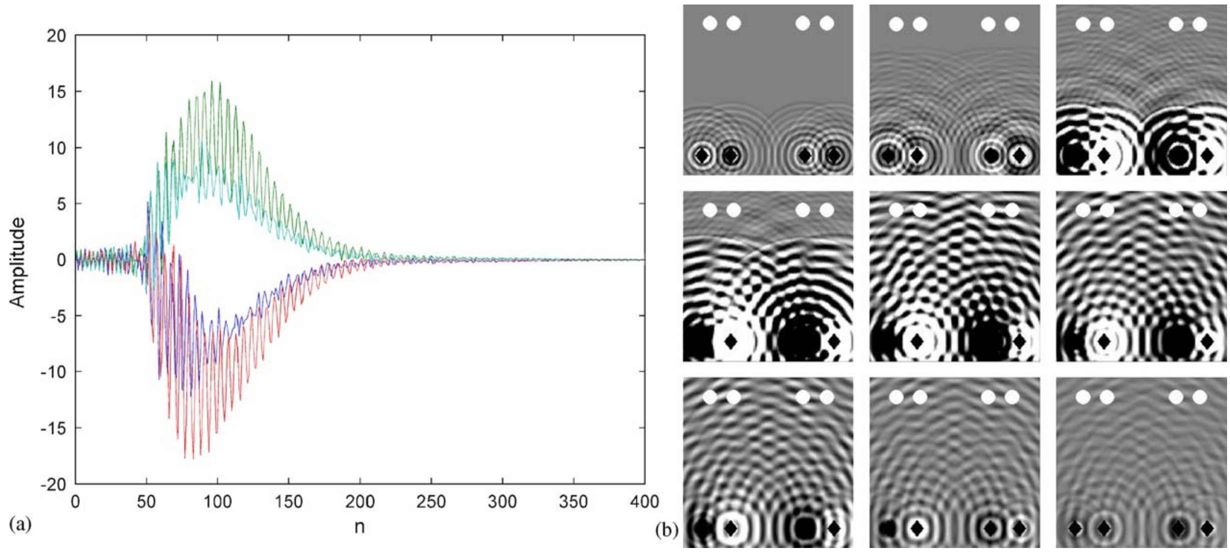

Fig. 12. Sound field reproduced by the aligned sources at the ill-conditioned frequency band ($f_0 = 3740$ Hz). The source inputs are designed to achieve crosstalk cancellation at the listeners' right ears: (a) the signals of the four sours achieving crosstalk cancellation with a Gaussian pulse; (b) time domain solutions.

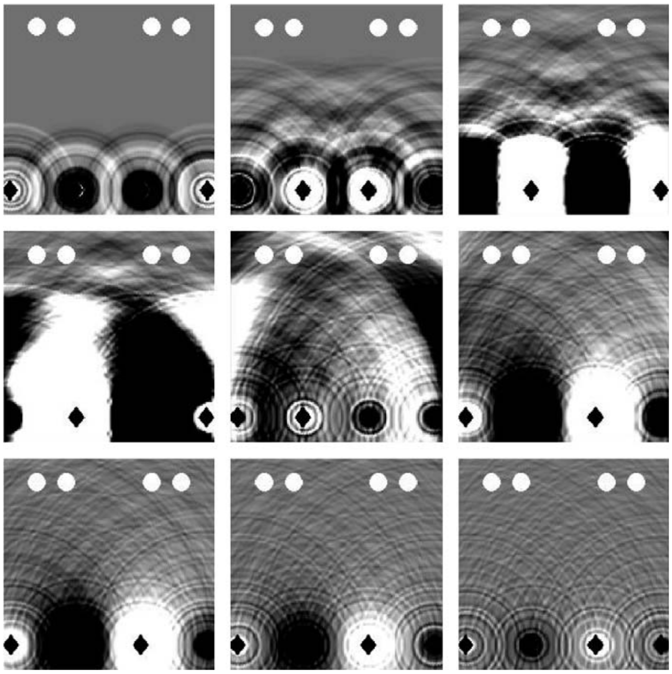

Fig. 13. Sound field reproduced by the aligned sources for 0–800 Hz frequency band ($f_0 = 500$ Hz). The source inputs are designed to achieve crosstalk cancellation at the listeners' right ears.

positions of the listeners are strictly limited due to the very small size of the sweet spots. It is therefore evident that the long duration of source outputs makes the system very sensitive to the listener's head movement. In addition, these results confirm that provided the frequency content of the pulse to be produced lies within the range of frequencies over which the inversion problem is well-conditioned, then crosstalk cancellation can be produced with a well defined wave-field of relatively short duration even for multiple listeners.

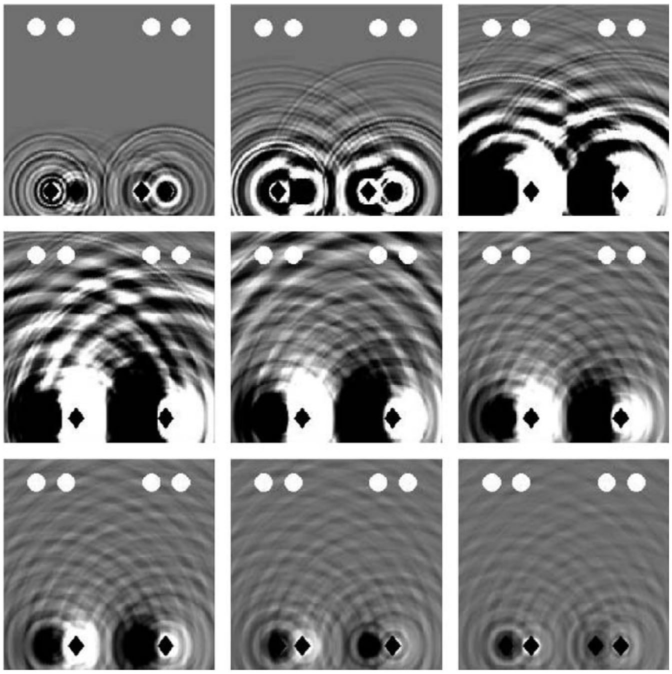

Fig. 14. Sound field reproduced by the aligned sources for 1751–2400 Hz frequency band ($f_0 = 2075$ Hz). The source inputs are designed to achieve crosstalk cancellation at the listeners' right ears.

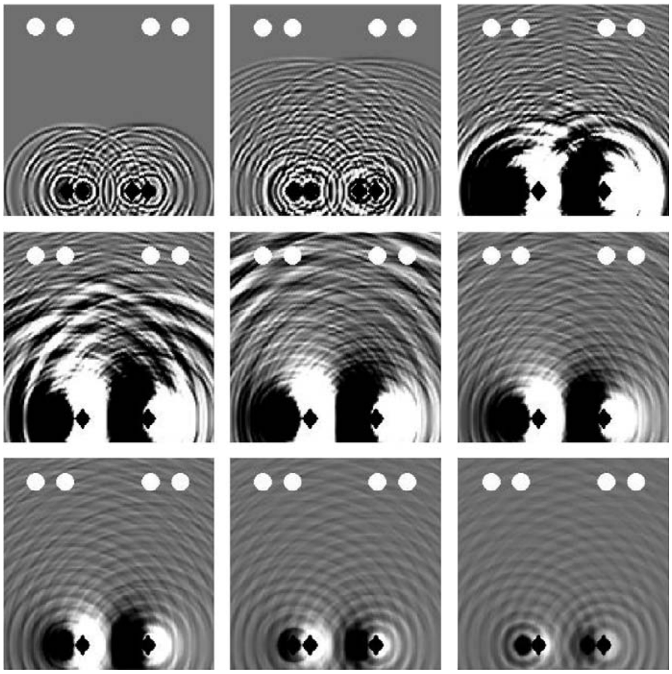

Fig. 15. Sound field reproduced by the aligned sources for 2401–3050 Hz frequency band ($f_0 = 2500$ Hz). The source inputs are designed to achieve crosstalk cancellation at the listeners' right ears.

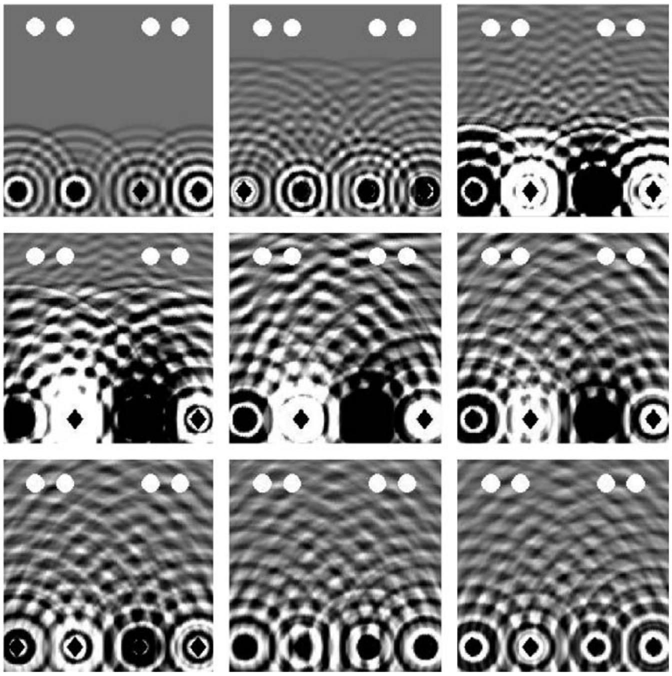

Fig. 16. Sound field reproduced by the aligned sources for 3051–3900 Hz frequency band ($f_0 = 3400$ Hz). The source inputs are designed to achieve crosstalk cancellation at the listeners' right ears.

4.3. The suggested configurations of four aligned sources in other frequency bands

The suggested positions of aligned sources in other frequency bands have been summarised in Table 1. The centre frequency of the Gaussian pulse defined by Eq. (12) has also been listed in Table 1 and the arbitrary constant $a = 500$ for all frequency bands. The solutions of time domain simulations for these cases are illustrated in Figs. 13–17. Firstly Figs. 13–15 show the sound field produced by the suggested configurations of the aligned sources at the associated well-conditioned frequency bands (0–800 Hz, 1751–2400 Hz and 2401–3050 Hz). Similarly to the results illustrated in Fig. 11, the emitted signals decay very quickly and thus the source outputs ring on for a short duration. However, as shown in Figs. 16 and 17, even though the pulse emitted from the associated sources is in the well-conditioned frequency band, the source outputs ring on for a long duration. This ringing phenomenon in these cases may be caused by the non-minimum phase nature of the crosstalk cancellation filter matrix **H** irrespective of the conditioning of the plant matrix. For example in the case shown in Fig. 16, the source outputs do not decay quickly and thus it can be easily found that the ringing still continues even when the main pulse has passed. Also, in the case illustrated in Fig. 17, even though the source outputs ring on for a relatively short duration compared with that of the case shown in Fig. 16, a lot of ringing is generated before the main pulse is emitted.

5. Conclusions

This paper has focused on the design of crosstalk cancellers for multiple listeners, for example a 4-source/4-receiver system. The source of ill-conditioning in the crosstalk cancellation system with free field model has been shown to be associated with the system inversion. The practical aspects of the conditioning of the plant matrix relating the acoustic pressures at the field points to the source volume accelerations have been detailed for the design of the crosstalk cancellation filters both in the frequency domain and in the time domain. Therefore, as a result of an extensive search with a linear array of sources at fixed locations, the source locations have been identified for which the condition number is smallest in different frequency bands. For

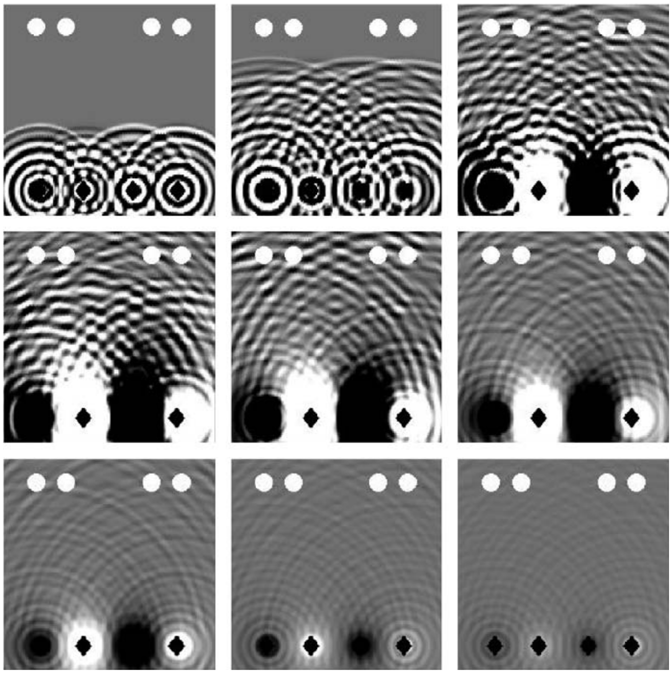

Fig. 17. Sound field reproduced by the aligned sources for 3901–5000 Hz frequency band ($f_0 = 4400$ Hz). The source inputs are designed to achieve crosstalk cancellation at the listeners' right ears.

practical applications this has resulted in an example of an optimal arrangement of the four aligned sources for six different frequency bands from 0 to 5000 Hz.Through time domain simulations, at the ill-conditioned frequencies, the spatial extent of the regions over which crosstalk cancellation is effective are found to be limited, whilst at well-conditioned frequencies this is found to be much greater. In addition, it has been demonstrated that the ill-conditioned frequencies are found to be associated with source outputs that are of long duration in the time domain. At well-conditioned frequencies, the source outputs are shown to be well-contained in time and result in simpler interference field in the time domain. However, at the higher frequencies, even at well-conditioned frequencies, the ringing is found to be of relatively long duration in the time domain.

References

[1] H. Moller, Fundamentals of binaural technology, *Applied Acoustics* 36 (1992) 171–218.

[2] D.R. Begault, *3-D Sound for Virtual Reality and Multimedia*, AP Professional, Cambridge, MA, 1994.

[3] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT Press, Cambridge, MA, 1997.

[4] P. Damaske, Head-related two-channel stereophony with loudspeaker reproduction, *Journal of the Acoustical Society of America* 50 (4) (1971) 1109–1115.

[5] O. Kirkeby, Reproduction of Acoustic Fields, PhD Thesis, University of Southampton, 1995.

[6] J.L. Bauck, D.H. Cooper, Generalized transaural stereo and applications, *Journal of the Audio Engineering Society* 44 (1996) 683–705.

[7] P.A. Nelson, O. Kirkeby, T. Takeuchi, H. Hamada, Sound fields for the production of virtual acoustic images, *Journal of Sound and Vibration* 204 (1997) 386–396.

[8] O. Kirkeby, P.A. Nelson, F. Orduna-Bustamante, H. Hamada, Local sound field reproduction using digital signal processing, *Journal of the Acoustical Society of America* 100 (3) (1996) 1584–1593.

[9] O. Kirkeby, P.A. Nelson, Digital filter design for inversion problems in sound reproduction, *Journal of the Audio Engineering Society* 47 (1999) 583–595.

[10] J. Yang, W.S. Gan, S.E. Tan, Improved sound separation using three loudspeakers, *Acoustics Research Letters Online* 4 (2) (2003) 47–52.

[11] T. Takeuchi, P.A. Nelson, Optimal source distribution for binaural synthesis over loudspeakers, *Journal of the Acoustical Society of America* 112 (6) (2002) 2786–2797.
[12] T. Takeuchi, P.A. Nelson, H. Hamada, Robustness to head misalignment of virtual sound imaging systems, *Journal of the Acoustical Society of America* 109 (3) (2001) 958–971.
[13] S. Barnett, *Matrices: Methods and Applications*, Clarendon Press, Oxford, 1990.
[14] P.A. Nelson, The time domain response of some systems for sound reproduction, *Proceedings of the Institute of Acoustics* 25 (8) (2003) 125–152.
[15] O. Deille, Virtual Acoustic Imaging for Multiple Listeners, MSc dissertation, University of Southampton, 2003.
[16] P.A. Nelson, A review of some inverse problems in acoustics, *International Journal of Acoustics and Vibration* 6 (3) (2001) 118–134.
[17] Y. Kim, Spatial Resolution Limits for Reconstruction of Acoustic Source Distribution by Inverse Techniques, PhD Thesis, University of Southampton, 2002.
[18] Y. Kim, P.A. Nelson, Optimal regularisation for acoustic source reconstruction by inverse methods, *Journal of Sound and Vibration* 275 (2004) 463–487.
[19] D.B. Ward, G.W. Elko, Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation, *IEEE Signal Processing Letters* 6 (5) (1999) 106–108.
[20] O. Kirkeby, P.A. Nelson, The Stereo Dipole—a virtual source imaging system using two closely spaced loudspeakers, *Journal of the Audio Engineering Society* 46 (5) (1998) 387–395.
[21] P.A. Nelson, Active control of virtual acoustics, *Proceedings of ACTIVE 2002*, Vol. 1, Southampton, England, 2002, pp. 67–89.