Optimal source distribution for binaural synthesis over loudspeakers[a)]

Takashi Takeuchi[b)] and Philip A. Nelson
Institute of Sound and Vibration Research, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom

(Received 25 July 2001; revised 11 April 2002; accepted 26 July 2002)

When binaural sound signals are presented with loudspeakers, the system inversion involved gives rise to a number of problems such as a loss of dynamic range and a lack of robustness to small errors and room reflections. The amplification required by the system inversion results in loss of dynamic range. The control performance of such a system deteriorates severely due to small errors resulting from, e.g., misalignment of the system and individual differences in the head related transfer functions at certain frequencies. The required large sound radiation results in severe reflection which also reduces the control performance. A method of overcoming these fundamental problems is proposed in this paper. A conceptual monopole transducer is introduced whose position varies continuously as frequency varies. This gives a minimum processing requirement of the binaural signals for the control to be achieved and all the above problems either disappear or are minimized. The inverse filters have flat amplitude response and the reproduced sound is not colored even outside the relatively large ''sweet area.'' A number of practical solutions are suggested for the realization of such optimally distributed transducers. One of them is a discretization that enables the use of conventional transducer units.
[DOI: 10.1121/1.1513363]

PACS numbers: 43.38.Md, 43.60.Pt, 43.66.Pn [SLE]

I. INTRODUCTION

A. Introduction

Binaural technology[1–3] is often used to present a virtual acoustic environment to a listener. The principle of this technology is to control the sound field at the listener's ears so that the reproduced sound field coincides with what would be produced when he is in the desired real sound field. One way of achieving this is to use a pair of loudspeakers (electroacoustic transducers) 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] Then the listener would experience an extremely realistic three dimensional sound environment.

One of the objectives of this study is to investigate a number of problems that arise from the multichannel system inversion involved in such binaural synthesis over loudspeakers. A basic analysis with a free field transfer function model illustrates the fundamental difficulties that such systems can have. The singular value decomposition helps to understand the role of the inverse filters more intuitively. The amplification required by the system inversion results in loss of dynamic range. The inverse filters obtained are likely to contain large errors around ill-conditioned frequencies. Regularization is often used to design practical filters but this also results in poor control performance around those frequencies. Sound radiation by transducers in directions other than that of the listener can be very large and this results in severe reflection and degrades control performance. Further analysis with a more realistic plant matrix, where the sound signals are controlled at a listener's ears in the presence of the listener's body (pinnae, head, etc.), demonstrates that this is still the case. Such problems are often noted as noise, distortion, fatigue of transducers, loss of directional and spatial perception, and coloration.

The investigation has resulted in the proposal of a system concept that we refer to as the Optimal Source Distribution (OSD). The OSD system overcomes these fundamental problems by means of a conceptual pair of monopole transducers whose span varies continuously as a function of frequency. The underlying theoretical principle is described in detail. The significance is that all of the above problems that are associated with the multichannel system inversion are solved by using this principle. The limitations with this principle are also made clear in terms of the operational frequency range. Several examples of practical solutions that can realize a variable transducer span are also described. One of them is the discretization and it enables the use of conventional transducer units and crossover filter networks with only a little decrease in performance from theoretical limit. The practical system realized has a very good performance over a wide frequency range (e.g., over the whole audible frequency range).

[a)]Portions of this work were presented in ''Optimal source distribution for virtual acoustic imaging,'' presented at the 140th Meeting of the Acoustical Society of America and Noise-con 2000, 3–8 December 2000, Newport Beach, California, and published in ''Optimal source distribution for binaural synthesis over loudspeakers,'' Acoustic Research Letters Online **2**, 7–12 (2001).

[b)]Present address: Kajima Technical Research Institute, 2-19-1 Tobitakyu, Chofu-shi, Tokyo 182-0036, Japan. Electronic mail: tt@kajima.com

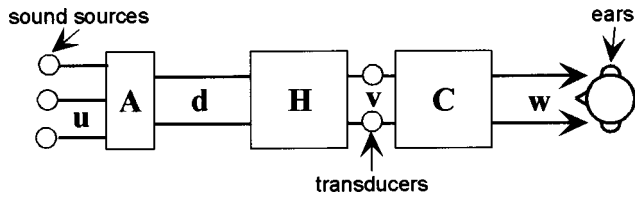

FIG. 1. Block diagram for multichannel sound control with system inversion.

B. Principles of binaural synthesis over loudspeakers

System inversion is often used for multichannel sound control including binaural reproduction over loudspeakers. The principle of such systems is described below with 2-channel case as an example and is illustrated in Fig. 1. The objective of the system is to feed to each ear of the listener independently the binaural signals that contain spatial information of sound as well as the signals associated with the sources in a virtual sound environment. However, when loudspeakers are used for this purpose, each loudspeaker feeds its signal to both ears. There is a matrix of acoustic paths between the loudspeakers and the listener's ears, and this can be expressed as a matrix of transfer functions (plant matrix). Independent control of two signals (such as the binaural sound signals) at two receivers (such as the ears of a listener) can be achieved with two electroacoustic transducers (such as loudspeakers), by filtering the input signals to the transducers with the inverse of the transfer function matrix of the plant. This process is also referred as system inversion. The signals and transfer functions involved are defined as follows. Two monopole transducers produce source strengths (volume accelerations) defined by the elements of the complex vector $\mathbf{v}=[\nu_1(j\omega)\ \ \nu_2(j\omega)]^{\mathrm{T}}$. The resulting acoustic pressure signals are given by the elements of the vector $\mathbf{w}=[w_1(j\omega)w_2(j\omega)]^{\mathrm{T}}$. This is given by

$$\mathbf{w}=\mathbf{C}\mathbf{v}, \tag{1}$$

where $\mathbf{C}$ is the plant matrix (a matrix of transfer functions between sources and receivers). The two signals to be synthesised at the receivers are defined by the elements of the complex vector $\mathbf{d}=[d_1(j\omega)\ \ d_2(j\omega)]^{\mathrm{T}}$. In the case of audio applications, these signals are usually the signals that would produce a desired virtual auditory sensation when fed to the two ears independently. They can be obtained, for example, by recording sound source signals $\mathbf{u}$ with a recording head (e.g., a dummy head) or by filtering the signals $\mathbf{u}$ by matrix of synthesised binaural filters $\mathbf{A}$.

Therefore, a filter matrix $\mathbf{H}$ which contains inverse filters is introduced (the inverse filter matrix) so that $\mathbf{v}=\mathbf{H}\mathbf{d}$, where

$$\mathbf{H}=\begin{bmatrix} H_{11}(j\omega) & H_{12}(j\omega) \\ H_{21}(j\omega) & H_{22}(j\omega) \end{bmatrix} \tag{2}$$

and thus

$$\mathbf{w}=\mathbf{C}\mathbf{H}\mathbf{d}. \tag{3}$$

For convenience in later analysis, we also define the control performance matrix $\mathbf{R}$ given by

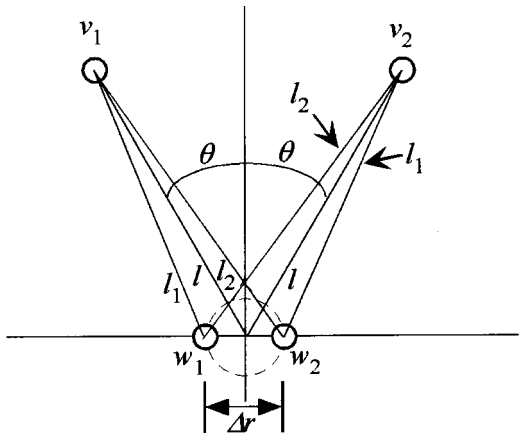

FIG. 2. Geometry of a 2-source 2-receiver system under investigation.

$$\mathbf{R}=\mathbf{C}\mathbf{H}. \tag{4}$$

The inverse filter matrix $\mathbf{H}$ can be designed so that the vector $\mathbf{w}$ is a good approximation to the vector $\mathbf{d}$ with a certain delay.[8,9] When the independent control at two receivers is perfect, the matrix $\mathbf{R}$ becomes the identity matrix $\mathbf{I}$. However, various factors introduce errors into the process. When errors are present, the diagonal term of the matrix $\mathbf{R}$ shows the portion of desired signal transmission and the other terms are the cross-talk components.

II. ANALYSIS WITH A FREE FIELD MODEL AND THE SINGULAR VALUE DECOMPOSITION

A simple case involving the control of two monopole receivers with two monopole transducers (sources) under free field conditions is first considered here in order to understand the physics underlying binaural synthesis over loudspeakers. The fundamental problems with regard to system inversion can be illustrated in this simple case where the effect of path length difference dominates the problem. The singular value decomposition helps to understand the role of the inverse filter matrix more intuitively, since it decomposes the matrix into multiple parts and each of these has a simple physical role. A matrix of Head Related Transfer Functions (HRTFs) is also analyzed in the later section as an example of a more realistic plant matrix. In such a case, the acoustic response of the human body (pinnae, head, torso, and so on) also comes to affect the problem. However, the fundamental difficulties inherit to such systems can still be seen predominantly.

A. Inverse filter matrix

A symmetric case with the intersource axis parallel to the inter-receiver axis is considered for an examination of the basic properties of the system. The geometry is illustrated in Fig. 2. In the free field case, the plant transfer function matrix can be modelled as

$$\mathbf{C}=\frac{\rho_0}{4\pi}\begin{bmatrix} \mathrm{e}^{-jkl_1}/l_1 & \mathrm{e}^{-jkl_2}/l_2 \\ \mathrm{e}^{-jkl_2}/l_2 & \mathrm{e}^{-jkl_1}/l_1 \end{bmatrix}, \tag{5}$$

where an $\mathrm{e}^{j\omega t}$ time dependence is assumed with $k=\omega/c_0$, and where ρ_0 and c_0 are the density and sound speed. When the ratio of and the difference between the path lengths con-

necting one source and two receivers are defined as $g = l_1/l_2$ and $\Delta l = l_2 - l_1$,

$$\mathbf{C} = \frac{\rho_0 e^{-jkl_1}}{4\pi l_1}\begin{bmatrix} 1 & g e^{-jk\Delta l} \\ g e^{-jk\Delta l} & 1 \end{bmatrix}. \tag{6}$$

Now consider the case

$$\mathbf{d} = \frac{\rho_0 e^{-jkl_1}}{4\pi l_1}\begin{bmatrix} D_1(j\omega) \\ D_2(j\omega) \end{bmatrix}, \tag{7}$$

i.e., the desired signals are the acoustic pressure signals which would have been produced by the closer sound source alone whose values are either $D_1(j\omega)$ or $D_2(j\omega)$ without disturbance due to the other source (cross-talk). This normalization enables a description of the effect of system inversion as well as ensuring a causal solution. The elements of **H** can be obtained from the exact inverse of **C** and can be written as

$$\mathbf{H} = \mathbf{C}^{-1} = \frac{1}{1 - g^2 e^{-2jk\Delta l}}\begin{bmatrix} 1 & -g e^{-jk\Delta l} \\ -g e^{-jk\Delta l} & 1 \end{bmatrix}. \tag{8}$$

When $l \gg \Delta r$, we have the approximation $\Delta l \approx \Delta r \sin\theta$ where $\Theta = 2\theta$ is the source span (hence $0 < \Theta \leqslant \pi$) and under these conditions,

$$\mathbf{H} = \frac{1}{1 - g^2 e^{-2jk\Delta r \sin\theta}}\begin{bmatrix} 1 & -g e^{-jk\Delta r \sin\theta} \\ -g e^{-jk\Delta r \sin\theta} & 1 \end{bmatrix}. \tag{9}$$

B. Singular value decomposition

The singular value decomposition helps to understand the role of the inverse filter matrix **H** more intuitively. As described in the Appendix, the inverse filter matrix **H** can be expressed as

$$\mathbf{H} = \mathbf{U}\boldsymbol{\Sigma}^{-1}\mathbf{V}^{\mathrm{H}} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}\begin{bmatrix} \sigma_i & 0 \\ 0 & \sigma_o \end{bmatrix} \times \frac{1}{\sqrt{2}}\begin{bmatrix} \sqrt{\frac{1+g e^{jk\Delta l}}{1+g e^{-jk\Delta l}}} & \sqrt{\frac{1+g e^{jk\Delta l}}{1+g e^{-jk\Delta l}}} \\ \sqrt{\frac{1-g e^{jk\Delta l}}{1-g e^{-jk\Delta l}}} & -\sqrt{\frac{1-g e^{jk\Delta l}}{1-g e^{-jk\Delta l}}} \end{bmatrix},$$

where

$$\sigma_i = \frac{1}{\sqrt{(1+g e^{-jk\Delta r \sin\theta})(1+g e^{jk\Delta r \sin\theta})}}$$

and

$$\sigma_o = \frac{1}{\sqrt{(1-g e^{-jk\Delta r \sin\theta})(1-g e^{jk\Delta r \sin\theta})}}. \tag{10}$$

The unitary matrix $\mathbf{V}^{\mathrm{H}}$ extracts the in-phase and out-of phase components out of the binaural signals. (Superscript H denotes the Hermitian transpose of a matrix; that is the complex conjugate of the transpose.) It also introduces phase rotation according to the property of the plant but does not change their amplitude. The two singular values are denoted by σ_i and σ_o, and correspond to orthogonal components of the inverse filters. The singular value σ_i corresponds to the amplification factor of the in-phase component of the binaural signals and the other singular value σ_o corresponds to the amplification factor of the out-of-phase component of the binaural signals. The unitary matrix **U** distributes the suitably amplified in-phase and out-of phase components into the pair of transducers.

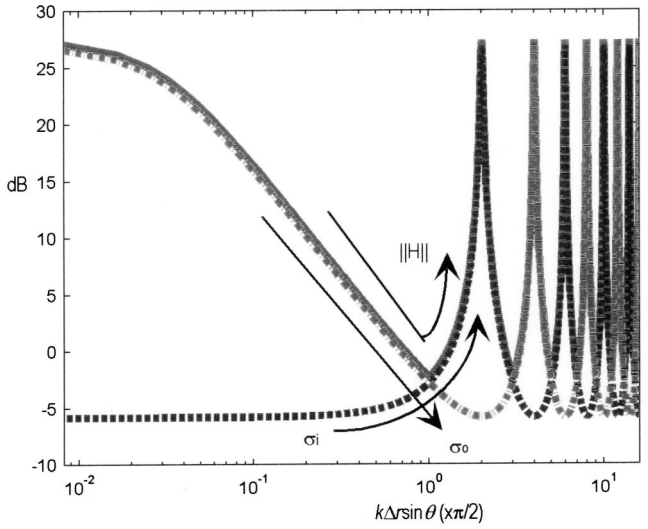

FIG. 3. Norm and singular values of the inverse filter matrix **H** as a function of $k\Delta r \sin\theta$.

The net effect of the inverse filter matrix **H** depends largely on the content of the input signals **d**, i.e., the characteristics of the sound source signal contents and the auditory virtual space being created. However, the maximum amplification of the source strengths required for the arbitrary binaural signal input at each frequency can be found from the 2-norm of **H** ($\|\mathbf{H}\|$). Since $\|\mathbf{U}\| = \|\mathbf{V}\| = 1$, this is equal to the largest of the singular values. Thus

$$\|\mathbf{H}\| = \|\boldsymbol{\Sigma}\| = \max(\sigma_i, \sigma_o). \tag{11}$$

Plots of σ_i, σ_o, and $\|\mathbf{H}\|$ with respect to $k\Delta r \sin\theta$ are illustrated in Fig. 3. The examples throughout this paper use typical values of the distance between the adult human ears for Δr. As seen in Eq. (10) and Fig. 3, the singular values σ_i and σ_o interchange their amplitude as a function of frequency and source span, periodically giving peaks of $\|\mathbf{H}\|$ where k and θ satisfy the following relationship with even values of the integer number n:

$$k\Delta r \sin\theta = \frac{n\pi}{2}. \tag{12}$$

This is where a term in the denominator of the singular values becomes very small (nearly zero) so the singular values become very large. The singular value σ_i has peaks at $n = 2,6,10,\ldots$ where the system is required to use large effort to reproduce the in-phase component of the desired signals. The singular value σ_o has peaks at $n = 0,4,8,\ldots$ where the system is required to use large effort to reproduce the out-of-phase component. Around these frequencies, sound signals from control sources interfere destructively with each other, leaving little response left. In other words, the signals

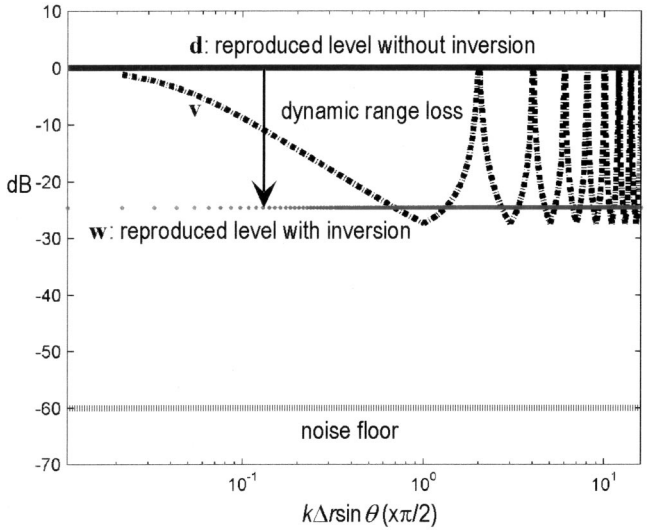

FIG. 4. Dynamic range loss due to system inversion.

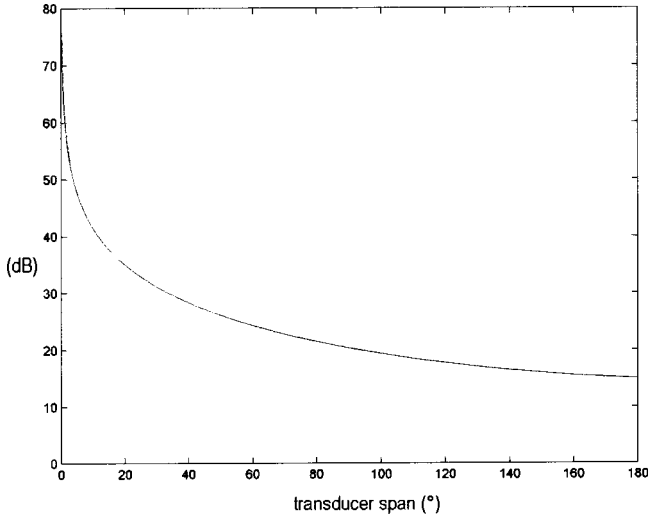

FIG. 5. Dynamic range loss as a function of source span.

cancel each other. Therefore, the solution for the inverse, i.e., the amplification required to produce the desired sound pressure at each receiver, becomes substantially large. The low frequency boost as a consequence of the peak at $n=0$ has often been addressed in several papers[3] but the other features, especially the peak for the in-phase component, has drawn less attention.

III. FUNDAMENTAL PROBLEMS OF EXISTING SYSTEMS

A. Loss of dynamic range

In practice, since the maximum source output is given by $\|\mathbf{H}\|_{\max}$, this must be within the range of the system in order to avoid clipping of the signals. The required amplification by the inverse filter matrix $\mathbf{H}$ results directly in the loss of dynamic range illustrated in Fig. 4. The level of the output source signals ($\mathbf{d}$: without system inversion, $\mathbf{v}$: with inversion) and the resulting level of the acoustic pressure at listener's ears ($\mathbf{d}$: without inversion, $\mathbf{w}$: with inversion) are plotted assuming that the maximum output levels and dynamic range of the systems are the same. Where $\|\mathbf{H}\|$ is large, each transducer is emitting very large sound most of which is cancelled by the sound from the other transducers. As a result, the levels of synthesized binaural signals at the listener's ears are significantly smaller than that those without cancellation. The given dynamic range is distributed into the system inversion and the remaining dynamic range that is to be used by the binaural auditory space synthesis, and also most importantly, by the sound source signal itself. Thus the signal-to-noise ratio of the signals $\mathbf{w}$ becomes low. Since the transducers are working much harder than usual use to produce usual sound level at the ears, nonlinear distortion becomes more significant and is often audible. For the same reason, fatigue of the transducers is more severe. Conventional driver units are not designed to be used in this manner and they can be easily destroyed by fatigue.

The dynamic range loss is defined by the difference between the signal level at the receiver with one monopole source and the signal level reproduced by two sources having the same maximum source strength when the system is inverted. The frequency of the peaks of $\|\mathbf{H}\|$ do not affect the amount of dynamic range loss but the magnitude of the peaks do. Since $\|\mathbf{H}\|$ here is normalized by the case without system inversion by Eq. (7), the dynamic range loss Γ is given by $\|\mathbf{H}\|_{\max}$. For example, when $n=2,6,10,\ldots,$ $k\Delta r\sin\theta=\pi,3\pi,5\pi,\ldots,$ therefore, $e^{-jk\Delta r\sin\theta}=e^{-jk\Delta r\sin\theta}=-1$. Thus

$$\Gamma=\|\mathbf{H}\|_{\max}=\max(\sigma_i)=\max(\sigma_o)=\frac{1}{1-g}. \tag{13}$$

The dynamic range loss given by Eq. (13) as a function of source span is shown in Fig. 5. Since $g\approx 1-\Delta r\sin\theta/l$, then Γ can be approximated as

$$\Gamma\approx\frac{l}{\Delta r\sin\theta} \tag{14}$$

as a function of θ. Figure 5 and Eq. (14) show that the larger the source span, the less is the dynamic range loss. It varies from more than 70 dB when two transducers are very close together to about 15 dB when they are on opposite sides of the ears. When there is a head between the ears, this is relaxed a little.

B. Robustness to error in the plant

A problem is said to be well-conditioned if small changes in parameters produce small changes in the solution, and ill-conditioned if relatively large changes are produced. The extent to which a process can be regarded as well-conditioned can be evaluated by calculating the condition number.[10] Equation (1) implies that the system inversion (which determines $\mathbf{v}$ and leads to the design of the filter matrix $\mathbf{H}$) is very sensitive to small errors in the assumed plant $\mathbf{C}$ (which is often measured and thus small errors are inevitable) where the condition number of $\mathbf{C}$, $\kappa(\mathbf{C})$, is large.[10] Such errors include individual differences of HRTFs,[11–13] and misalignment of the head and loudspeakers.[14] The condition number of $\mathbf{C}$ is given by

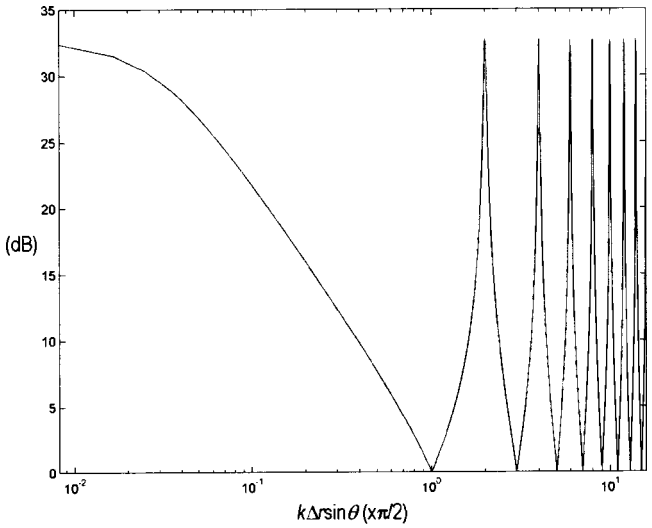

FIG. 6. Condition number $\kappa(\mathbf{C})$ as a function of n.

$$\kappa(\mathbf{C}) = \|\mathbf{C}\|\|\mathbf{C}^{-1}\| = \|\mathbf{C}\|\|\mathbf{H}\| = \|\mathbf{H}^{-1}\|\|\mathbf{H}\| = \max\left(\sqrt{\frac{(1+g\mathrm{e}^{-jk\Delta r\sin\theta})(1+g\mathrm{e}^{jk\Delta r\sin\theta})}{(1-g\mathrm{e}^{-jk\Delta r\sin\theta})(1-g\mathrm{e}^{jk\Delta r\sin\theta})}}, \sqrt{\frac{(1-g\mathrm{e}^{-jk\Delta r\sin\theta})(1-g\mathrm{e}^{jk\Delta r\sin\theta})}{(1+g\mathrm{e}^{-jk\Delta r\sin\theta})(1+g\mathrm{e}^{jk\Delta r\sin\theta})}} \right) \tag{15}$$

and is shown in Fig. 6. As seen in Eq. (15) and Fig. 6, $\kappa(\mathbf{C})$ has peaks where Eq. (12) is satisfied with an even value of the integer number n. The frequencies which give peaks of $\kappa(\mathbf{C})$ are consistent with those which give the peaks of $\|\mathbf{H}\|$.

Around the frequencies where $\kappa(\mathbf{C})$ is large, the system is very sensitive to small errors in $\mathbf{C}$.[15] The calculated inverse filter matrix $\mathbf{H}$ is likely to contain large errors due to small errors in $\mathbf{C}$ and results in large errors in the reproduced signal $\mathbf{w}$ at the receiver. This is because such errors are magnified by the inverse filters but remain uncancelled in the plant. On the contrary, $\kappa(\mathbf{C})$ is small around the frequencies where n is an odd integer number in Eq. (12). For the same value of n, the robust frequency range becomes lower as the source span becomes larger. With a logarithmic frequency scale, which is related to the perceptual attributes of the human auditory system, the frequency range of robust inversion is more or less constant for different source spans for the same value of n, even though it looks wider for smaller source spans on a linear frequency scale.

C. Robustness to error in the inverse filters

In addition, since

$$\mathbf{v} = \mathbf{C}^{-1}\mathbf{w} \tag{16}$$

and $\kappa(\mathbf{C}^{-1}) = \kappa(\mathbf{C})$, a practical and close to ideal inverse filter matrix $\mathbf{H}$ is easily obtained where $\kappa(\mathbf{C})$ is small. However, the reproduced signals $\mathbf{w}$ are less robust to small changes in the inverse of the plant matrix $\mathbf{C}^{-1}$, hence $\mathbf{H}$, where $\kappa(\mathbf{C})$ is large. Even if $\mathbf{C}$ does not contain any errors, the reproduction of the signals at the receiver is too sensitive to the small errors within the inverse filter matrix $\mathbf{H}$ to be useful.

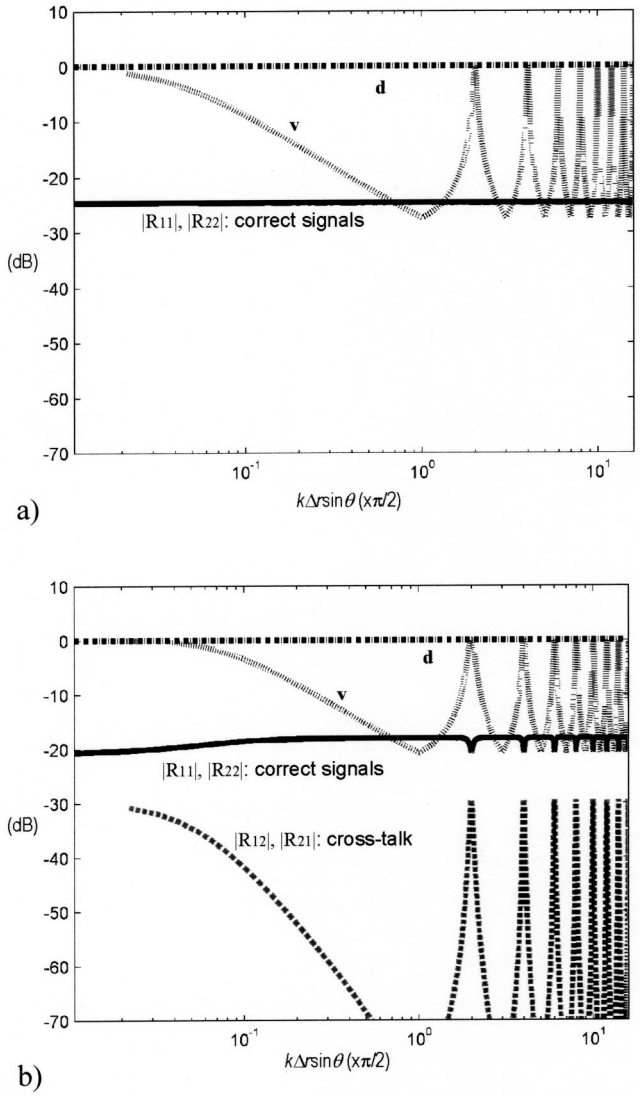

FIG. 7. Dynamic range improvement and loss of control performance with regularization. (a) Without regularization. (b) With regularization.

One common example of such an error is that due to regularisation, where a small error is deliberately introduced to improve the condition of matrix to design practical filters.[16] It is also possible to reduce the excess amplification and hence the dynamic range loss by means of regularization, where the pseudoinverse filter matrix $\mathbf{H}$ is given by

$$\mathbf{H} = [\mathbf{C}^{\mathrm{H}}\mathbf{C} + \beta\mathbf{I}]^{-1}\mathbf{C}^{\mathrm{H}}, \tag{17}$$

where β is a regularization parameter. The regularization parameter penalizes large values of $\mathbf{H}$ and hence limits the dynamic range loss of the system. Since $\|\mathbf{H}\|$ is normalized by the case without system inversion by Eq. (7), the regularization parameter limits the dynamic range loss to less than about

$$\Gamma \approx -10\log_{10}\beta - 6 \ \ (\mathrm{dB}). \tag{18}$$

However, the regularization parameter intentionally, hence inevitably, introduces a small error in the inversion process. This gives rise to a problem for filter design at frequencies where $\kappa(\mathbf{C})$ is large. An example of this is illustrated in Fig. 7. The dynamic range loss is reduced by regularization from

about 27 dB (without regularization) as in Fig. 7(a) to 14 dB as shown in Fig. 7(b) ($\beta=10^{-2}$). However, it can be clearly seen that the control performance of the system deteriorates around the frequencies where n is an even integer number in Eq. (12). The contribution of the correct desired signals (R_{11} and R_{22}) is reduced only slightly but the contribution of the wrong desired signals (R_{12} and R_{21}, the cross-talk component) is increased significantly. In other words, the system has little control (cross-talk cancellation) around these frequencies. This problem is significant at lower frequencies [$n<1$ in Eq. (12)] in the sense that the region without cross-talk suppression is large, and at higher frequencies [$n>1$ in Eq. (12)], in the sense that there are many frequencies at which the plant is ill-conditioned. With an equivalent dynamic range loss, making the source span larger leads to a better control performance at lower frequencies but a poorer performance at higher frequencies. On the contrary, making the source span smaller leads to better control performance at higher frequencies but poorer performance at lower frequencies.

D. Robustness to reflections

Reflections from surrounding objects (e.g., walls, floors, and ceilings) affect the control performance. The effect of reflections on this kind of system has been studied with a simple image source model and subjective experiments.[17] Although the perceptual aspects of sound localization such as the precedence effect suggest that the performance of this kind of system will be retained to some extent,[18] reflected sound with a much larger level than the control sound arriving directly at the listener's ears destroys the correct perception. Therefore, the relative level of sound radiation in directions other than towards listener's ears is a very good measure of the robustness of the system to reflections. Figure 8 shows an example ($n\approx 2$) of far field sound radiation by the control transducers with reference to the receiver directions. The horizontal axis is the intersource axis and the receivers (ears) are at the directions of the vertical axis. At frequencies where Eq. (12) is not satisfied with an odd value of the integer number n, as in this example, the sound radiation in directions other than receiver directions can be significantly larger (typically $+30$ dB$\sim+40$ dB) than those at the receiver directions (0 dB and $-\infty$ dB). The maximum amount of this excessive radiation is the same as the amount of dynamic range loss as in Eq. (13) and Fig. 5. When the environment is not anechoic, as is normally the case, this obviously results in severe reflections and the control performance of the system deteriorates. In addition, the sound radiated in directions other than that of receiver has a peaky frequency response due to the response of inverse filter matrix $\mathbf{H}$ and normally result in severe coloration.

IV. A SYSTEM TO OVERCOME THE PROBLEMS

As discussed above, there is a trade-off between dynamic range, robustness, and control performance. However, a system which aims to overcome these fundamental problems is proposed in what follows.

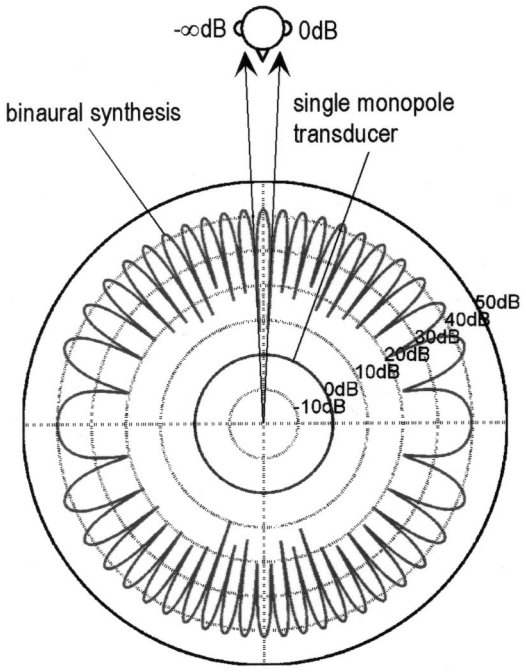

FIG. 8. Sound radiation by the control transducer pairs with reference to the receiver directions (0 dB and $-\infty$ dB).

A. Principle of the optimal source distribution

Equation (12) can be rewritten in terms of the source span Θ as

$$\Theta=2\,\theta=2\,\arcsin\left(\frac{n\pi}{2k\Delta r}\right). \tag{19}$$

As seen from the analysis above, systems with the source span where n is an odd integer number in Eq. (19) gives the best control performance as well as robustness. This implies that the optimal source span must vary as a function of frequency.

We now consider a pair of conceptual monopole transducers whose span varies continuously as a function of frequency (Fig. 9) in order to satisfy the requirement for n to be an odd integer number in Eq. (19) (Fig. 10). This relationship is where σ_i and σ_o are balanced and the source span becomes smaller as frequency becomes higher. With this concept, the expression for the inverse filter matrix $\mathbf{H}$ [Eq. (9)] becomes very simple as

$$\mathbf{H}=\frac{1}{1+g^2}\begin{bmatrix}1 & -jg\\ -jg & 1\end{bmatrix}. \tag{20}$$

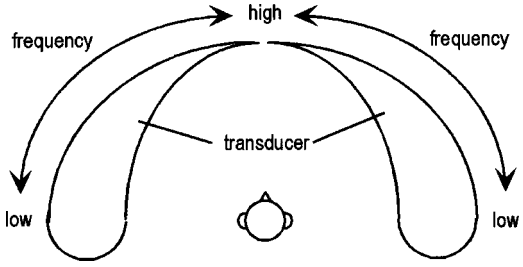

FIG. 9. Principle of the "OSD" system.

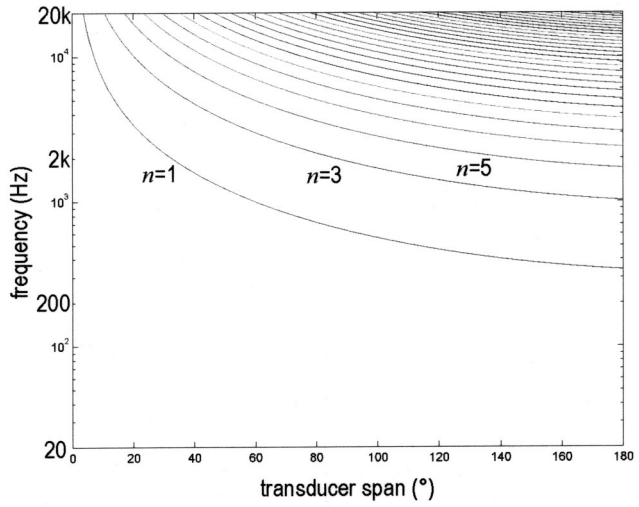

FIG. 10. Relationship between source span and frequency for different odd integer number n.

Note that $\|\mathbf{H}\|=1/\sqrt{2}$ for all frequencies. Therefore, there is no dynamic range loss compared to the case without system inversion. In fact, there is a dynamic range gain of 3 dB since the two orthogonal components of the desired signals are $\pi/2$ out of phase. This means the system has good signal to noise ratio and advantage over distortion or fatigue of transducers. The inverse filters have flat frequency response so there is no coloration, in excess of that produced by room response, at any location in the listening room even outside the sweet area. When the listener is far away from the sweet spot, the spatial information perceived may not be ideal. However, the spectrum of the sound signals are not changed by the inverse filters. Therefore, the listener can still enjoy the natural production of sound together with some remaining spatial aspects. The sound radiation by the transducer pair in all directions is always smaller than those at the receiver directions, which is also smaller than the sound radiation by a single monopole transducer producing the same sound level at the ears as shown in Fig. 11. In contrast to Fig. 8, the system

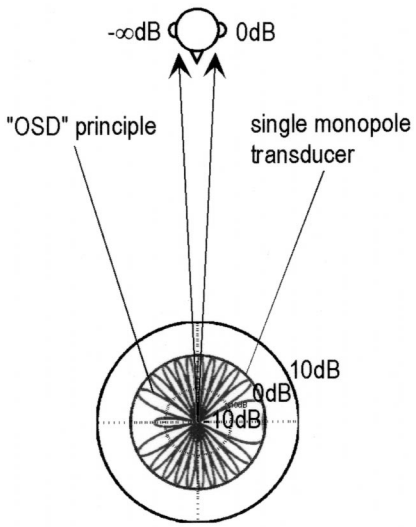

FIG. 11. Sound radiation by the ''OSD'' loudspeakers with reference to the receiver directions (0 dB and $-\infty$ dB).

does not radiate excessive sound all around so it is also robust to reflections in a reverberant environment, and these small reflections do not have any coloration other than that caused by the reflecting materials. Note also that $\kappa(\mathbf{C})=1$ which is constant over all frequencies and which is the smallest possible value. The error in calculating the inverse filter is small and the system has very good control over the reproduced signals. The system is also very robust to the changes in plant matrix.

Also note that when $l \gg \Delta r$, $g \approx 1$ therefore,

$$\mathbf{H} \approx \frac{1}{2}\begin{bmatrix} 1 & -j \\ -j & 1 \end{bmatrix}. \tag{21}$$

This implies that independent control of the two signals is nearly achieved just by addition of the desired signals with a $\pi/2$ relative phase shift between them.

B. Aspects of the proposed system

From Eq. (19), the range of variable source span Θ is given by the frequency range of interest as can be seen from Fig. 10. A smaller value of n gives a smaller source span for the same frequency. Therefore, the smallest source span Θ_h for the same high frequency limit is given by $n=1$ and this is about 4° to give control of the sound field at two positions separated by the distance between two ears (about 0.13 m for KEMAR dummy head[19]) up to a frequency of 20 kHz.

Equation (12) can also be rewritten in terms of frequency as

$$f = \frac{nc_0}{4\Delta r \sin\theta}, \tag{22}$$

The smallest value of n gives the lowest frequency limit for a given source span. Since $\sin\theta \leqslant 1$,

$$f \geqslant \frac{nc_0}{4\Delta r}, \tag{23}$$

i.e., the physically maximum source span of $\Theta=2\theta=180°$ gives the lowest frequency limit, f_1, associated with this principle. A smaller value of n gives a lower low frequency limit so the system given by $n=1$ is normally the most useful among those with an odd integer number n. The low frequency limit given by $n=1$ of a system designed to control the sound field at two positions separated by the distance between two ears is about $f_1=300\sim400$ Hz.

C. Consideration of the head related transfer function model

The condition number $\kappa(\mathbf{C})$ of the plant matrix plotted as a function of frequency and source span is shown in Fig. 12 for the audible frequency range (20 Hz~20 kHz). Figure 13 shows the condition number of the more realistic plant matrix with HRTFs. The HRTFs were measured with the KEMAR dummy head at MIT Media Lab[20] and the loudspeaker response was deconvolved later. Those between sampled directions are obtained by bilinear interpolation on the virtual spherical surface of magnitude and phase spectra in the frequency domain.[14] A similar trend can clearly be seen as in the free field case. However, additional ''ill-

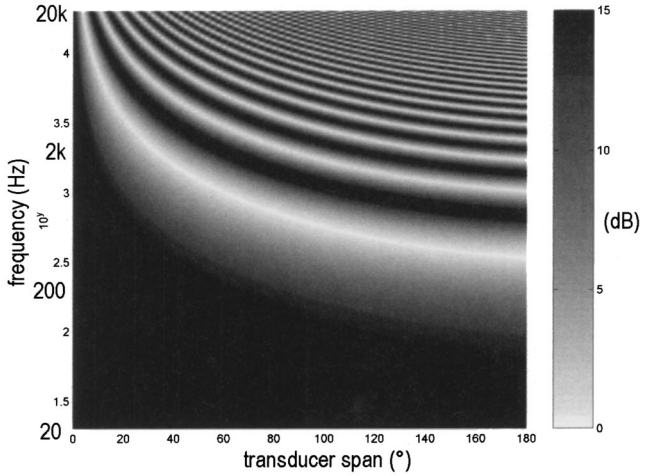

FIG. 12. Condition number $\kappa(\mathbf{C})$ of a free field plant matrix $\mathbf{C}$ as a function of source span and frequency.

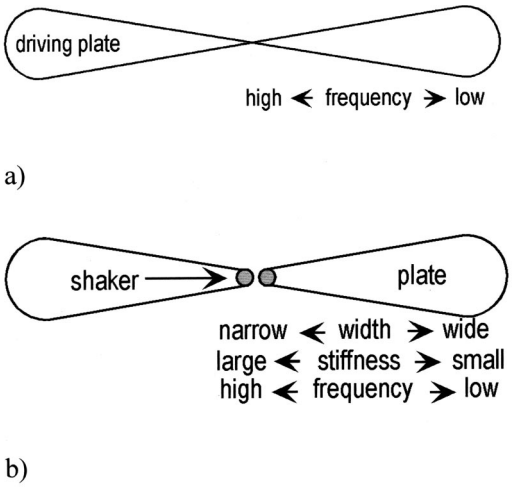

FIG. 14. Flat panel transducers. (a) Individual excitation. (b) Point excitation.

conditioned frequencies'' can be observed around 9 kHz and 13 kHz where the HRTFs have minima. It is possible that the signal-to-noise ratio of the measured data around these frequencies is poor.

It should also be noted that where the incidence angle θ is small, the peak frequencies obtained with the HRTF plant matrix are similar to that of the free field plant with the receiver distance $\Delta r \approx 0.13$. This corresponds to the shortest distance between the entrances of the ear canals of the KEMAR dummy head. However, where the incidence angle θ is large, the peak frequencies obtained with the HRTF plant matrix are similar to that of the free field plant with the receiver distance $\Delta r \approx 0.25$. This is a much larger distance than the shortest distance between the entrances of the ear canals of the KEMAR dummy head and is a result of diffraction around the head. A correction to the receiver distance Δr can be made in order to match the frequency-span characteristics of the free field model. Following is an example of a linear approximation,

$$\Delta r = \Delta r_0 (1 + \Theta/\pi), \tag{24}$$

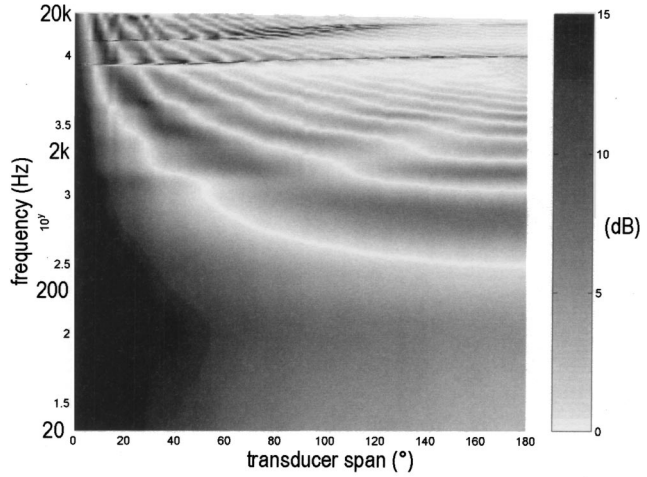

FIG. 13. Condition number $\kappa(\mathbf{C})$ of a plant matrix $\mathbf{C}$ with HRTFs as a function of source span and frequency.

where Δr_0 is the geometrical distance between the ears. Figure 12 incorporates Eq. (24) and shows good agreement with Fig. 13 which is obtained from more realistic KEMAR data.

D. Transducers for the optimal source distribution

This principle requires a pair of monopole type transducers whose position from which sound is radiated varies continuously as frequency varies. This might, for example, be realized by exciting a plate at each position individually [Fig. 14(a)]. The requirement of such a transducer is that a certain frequency of vibration is excited most at a particular position such that sound of that frequency is radiated mostly from that position. Relatively large damping would be required to suppress unwanted plate vibration modes. Such characteristics may also be achieved by exciting a triangular shaped plate with large damping at one end whose width and stiffness varies along its length in a controlled manner [Fig. 14(b)]. The narrow and stiff excited end radiates most high frequency sound whereas the wide and floppy end of the plate radiates the lower frequency sound. This principle is very similar to the way the basilar membrane of the human auditory system performs frequency decomposition. Alternatively, a similar effect might be obtained by changing the width of a slot along an acoustic waveguide (Fig. 15). Again a relatively large damping would be necessary in order to suppress peaks at resonance. In both cases, the vibration characteristics of the plate or air particles would differ along the length, and so as the radiation impedance. Then, transducers which effectively distribute each frequency components to desired position may be designed.

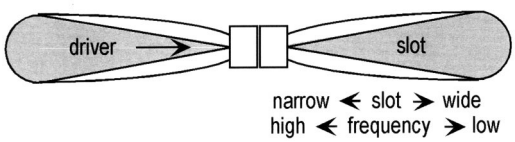

FIG. 15. Acoustic waveguide type transducers.

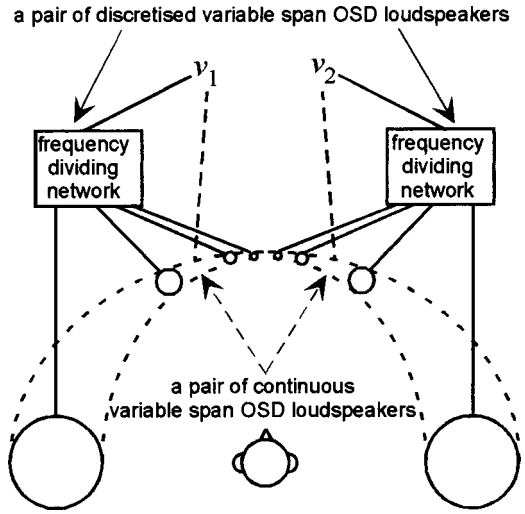

FIG. 16. Discretized variable frequency/span transducers.

V. DISCRETE SYSTEM

A. Discretization of frequency-span relationship

In practice, a monopole transducer whose position varies continuously as a function of frequency is not easily available. However, it is possible to realize a practical system based on this principle by discretizing the transducer span as illustrated in Fig. 16 and Fig. 17. With a given span, the frequency region where the amplification is relatively small and plant matrix $\mathbf{C}$ is well conditioned is relatively wide around the optimal frequency. Therefore, by allowing n to have some width, say $\pm\nu$ $(0<\nu<1)$, which results in a small amount of dynamic range loss and slightly reduced robustness, a certain transducer span can nevertheless be allocated to cover a certain range of frequencies where control performance and robustness of the system is still reasonably good (Fig. 17). Consequently, it is possible to discretize the continuously varying transducer span into a finite number of discrete transducer spans. A system with a smaller n gives a wider region with the same performance on a logarithmic scale as can be seen in Figs. 12 and 13.

It is important to design the system to ensure $\|\mathbf{H}\|$ and $\kappa(\mathbf{C})$ that are as small as possible over a frequency range that is as wide as possible. Therefore, the transducer spans for each pair of transducers in each frequency range can be decided to ensure that the smallest possible values of n are used over all frequency ranges of interest above f_1.

It is also possible to discretize, i.e., decide the transducer spans and frequency ranges to be covered by each pair of driver units (i.e., range of n), in terms of a tolerable dynamic range loss. The dynamic range loss of the entire system is now given by the maximum value among the values given by each discretized transducer span.

A number of examples of the discrete OSD system have been simulated.[21] The effect has been calculated of discretization number and the allocated dynamic range loss on the cross-talk cancellation performance. Studies have been undertaken of the spans for each discretized control transducers arrangement, crossover frequencies, and width of ν. Practical consideration above a number of possibilities for designing crossover filters and inverse filters have also been described in detail. In short, the dynamic range loss is more than 40 dB when the discretization is one way. It reduces to 18 dB when the discretization is two way and to less than 7 dB when discretization is three way.

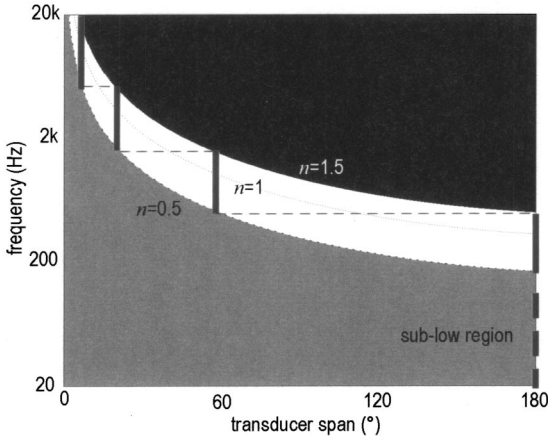

FIG. 17. An example of frequency/span region and discretization.

B. Consequence of the discretization of variable source span

The discretization is extremely useful and practical because a single transducer which can cover the whole audible frequency range is not practically available either. Therefore, this principle also gives the ideal background for multiway systems for binaural reproduction over loudspeakers which maximize the frequency range to be produced and controlled. Conventional driver units and crossover filters can easily be accommodated to be used for this system. It should be noted that this is still a simple ''2 channel'' control system where only two independent control signals are necessary to control any form of virtual auditory space. This in principle can synthesize an infinite number of virtual source locations with different source signals with any type of acoustic response of the space. The difference for this discrete system from the conventional 2-channel system is that the two control signals are divided into multiple frequency bands and fed into the different pairs of driver units with different spans. Ironically, substantial effort has been invested in conventional multiway loudspeakers for stereophony in order to approximate a point source by multiple driver units. The discrete OSD system requires just the opposite; different driver units are required to be at different locations. A ''poor'' performance unit in the sense of stereophony which has relatively narrow operational frequency range may perform very well with this principle.

It should be noted that the low frequency limit f_1 given by odd integer numbers n in Eq. (23) is extended towards a lower frequency by discretization because now the region for frequency and transducer span where n is not an integer number is also used. For example, a practical system discretized from the ideal system with $n=1$ can now make use of the region $1-\nu<n<1+\nu$ so that the low frequency limit is given by $n=1-\nu$.

As can be seen from Figs. 10 and 17, in the higher frequency range where the source span is very small, the frequency range to be covered is very sensitive to small

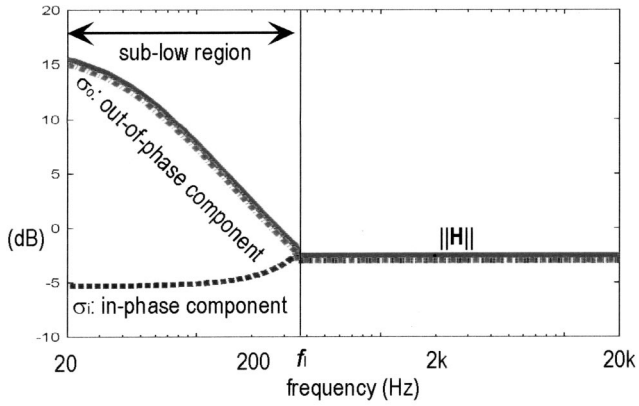

FIG. 18. Norm and two singular values of the inverse filter matrix **H** with "OSD" principle.

differences in transducer span. On the contrary, it is very insensitive to the source span at lower frequencies. Consequently, the range of practical span for the low frequency units is very large, which can practically be anywhere from about 60° to 180° with only a very slight increase of low frequency limit.

VI. CONSIDERATIONS FOR THE SUB-LOW-FREQUENCY REGION

At the frequencies below f_1 ($n<1-\nu$) where $\|\mathbf{H}\|$ and $\kappa(\mathbf{C})$ is larger than other frequencies, the requirement for dynamic range loss and robustness of the system are more severe than at other frequencies. Figure 18 illustrates the 2-norm of **H** and the two singular values (σ_i and σ_o) with the "OSD" principle. As described in Sec. IV A, $\|\mathbf{H}\|$ shows the flat amplitude response of the inverse filters above f_1. However, below f_1, it still increases moderately as frequency becomes lower. In this region, although the system has difficulty in reproducing the out-of-phase component of the desired signal, it still can produce the in-phase component as well as before.

When f_1 is reasonably low, where interaural difference may not be crucial for binaural reproduction, one can avoid system inversion and simply add a single subwoofer unit for this frequency region to avoid the extra dynamic range loss required by this region. As seen in Eq. (10), adding two channels of signals results in complete cancellation of the out-of-phase component of the binaural signals and producing the in-phase component only. Then, there is no independent control of binaural signals in this region.

It is possible to cover this sub-low-frequency region with the lowest frequency pair of units without sacrificing performance for other frequencies. The large σ_o, the out-of-phase component, in this region can be regularized with the method described in Sec. III C. In such case, even though little cross-talk suppression is available, the low-frequency pair can still work as a subwoofer mostly producing the in-phase component, while it is working perfectly within the OSD frequency range. In the sub-low region, the control performance deteriorates severely due to heavy regularization. However, $\|\mathbf{R}\|$ and hence the norm of the reproduced signal, is the same as that without regularization. This may be acceptable in binaural reproduction since the difference between the two desired signals is normally not so large and sometimes negligible in the very low-frequency range.

When slight dynamic range loss is acceptable, the regularization can be used to limit the amplification, and hence avoid too much dynamic range loss, without sacrificing robustness for other frequencies. The cross-talk performance with regularization in the frequency range below f_1 is not as good as at the other frequencies. However, there can still be reasonable cross-talk suppression available. If more dynamic range loss is allowed, a relatively small regularization can be used to suppress the out-of-phase component in the sub-low region. The cross-talk cancellation performance in this region is very sensitive to the allocated dynamic range loss. Therefore, it is possible to design the system by selecting the required low frequency cross-talk cancellation performance. The amount of the dynamic range loss required by the discretization often gives relatively good control performance also in the sub-low-frequency region, especially when the discretization is coarse.

One might choose to allow all the dynamic range loss necessary for the full control of the sub-low-frequency region. The overall dynamic range loss is determined by the lowest frequency pair, which has the largest span. As discussed in Sec. III A, the dynamic range loss by the largest span is the smallest value among all other pairs.

VII. CONCLUSIONS

Analysis with a free field model and more realistic plant matrix with head related transfer functions reveals a number of fundamental problems related to multichannel sound control with system inversion such as binaural synthesis over loudspeakers. A principle of 2-channel (binaural) sound control with loudspeakers is proposed which overcomes the fundamental problems with system inversion by utilizing a variable transducer span. Practical ways to tackle the sub-low-frequency region are also described where outside the OSD principle.

The proposed principle has various advantages. No dynamic range loss due to system inversion directly means good signal-to-noise ratio but also leads to less distortion and longer life of transducers. The robustness to errors has advantages in many respects, e.g., incorrect inverse filters due to restriction of hardware resources, differences between individuals or products, and the misalignments that is inevitable in practical use. The minimal sound radiation in directions other than receiver directions reduces the chance of the 3D effect being destroyed by reflections from surrounding objects. The system inversion does not result in coloration because of the flat response of the inverse filters, and this adds practicality by enabling the listener to enjoy the reproduced sound signals even outside the "sweet region." As a natural consequence of this, the reflections or reverberation of the room are not colored either.

The practical system can be realized in a number of ways including discretizing the theoretical continuously variable transducer span that results in multiway sound control system. The discretization enables the use of conventional

transducer units and crossover filter networks. The relationship between the position of a driver unit and the frequency region to be covered can be determined easily. Further developments to realize ideal continuously distributed transducer will be beneficial to improve the performance of such systems.

ACKNOWLEDGMENTS

This research was supported by Yamaha Corporation, Alpine Electronics, Hitachi Ltd., and Kajima Corporation.

APPENDIX

When the desired signals are defined as Eq. (7), this effectively normalizes the plant transfer function matrix $\mathbf{C}$ with respect to the acoustic pressure signals which would have been produced by the closer of two sound sources to the receiver point. Then this normalized plant transfer function matrix $\mathbf{C}$ can be written as

$$\mathbf{C}=\begin{bmatrix} 1 & g\mathrm{e}^{-jk\Delta l} \\ g\mathrm{e}^{-jk\Delta l} & 1 \end{bmatrix}. \tag{A1}$$

It is possible to express $\mathbf{C}$ with unitary matrices $\mathbf{U}$ and $\mathbf{V}$ such that

$$\mathbf{C}=\mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}}, \tag{A2}$$

where $\mathbf{\Sigma}$ is the diagonal matrix whose elements are singular values of $\mathbf{C}$, σ_1, and σ_2. The singular values of $\mathbf{C}$ can be found from the square roots of eigenvalues of $\mathbf{C}^{\mathrm{H}}\mathbf{C}$,

$$\begin{aligned}\mathbf{C}^{\mathrm{H}}\mathbf{C}&=\begin{bmatrix} 1 & g\mathrm{e}^{jk\Delta l} \\ g\mathrm{e}^{jk\Delta l} & 1 \end{bmatrix}\begin{bmatrix} 1 & g\mathrm{e}^{-jk\Delta l} \\ g\mathrm{e}^{-jk\Delta l} & 1 \end{bmatrix} \\ &=\begin{bmatrix} 1+g^2 & g(\mathrm{e}^{jk\Delta l}+\mathrm{e}^{-jk\Delta l}) \\ g(\mathrm{e}^{jk\Delta l}+\mathrm{e}^{-jk\Delta l}) & 1+g^2 \end{bmatrix}.\end{aligned} \tag{A3}$$

The eigenvalues of $\mathbf{C}^{\mathrm{H}}\mathbf{C}$ are given by

$$\lambda_{1,2}=(1+g\mathrm{e}^{jk\Delta l})(1+g\mathrm{e}^{-jk\Delta l}),(1-g\mathrm{e}^{jk\Delta l})(1-g\mathrm{e}^{-jk\Delta l}). \tag{A4}$$

Therefore, the singular values of $\mathbf{C}$ are given by

$$\begin{aligned}\sigma_{1,2}=\sqrt{\lambda_{1,2}}=&\sqrt{(1+g\mathrm{e}^{jk\Delta l})(1+g\mathrm{e}^{-jk\Delta l})}, \\ &\sqrt{(1-g\mathrm{e}^{jk\Delta l})(1-g\mathrm{e}^{-jk\Delta l})}.\end{aligned} \tag{A5}$$

It is possible to find an orthonormal set of eigenvectors $\mathbf{x}_i$ such that

$$\mathbf{C}^{\mathrm{H}}\mathbf{C}\mathbf{x}_i=\sigma_i^2\mathbf{x}_i. \tag{A6}$$

Therefore,

$$\mathbf{V}=\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}. \tag{A7}$$

The vectors comprising $\mathbf{U}$ are given by

$$\mathbf{y}_i=\frac{1}{\sigma_i}\mathbf{C}\mathbf{x}_i. \tag{A8}$$

Hence

$$\mathbf{U}=\frac{1}{\sqrt{2}}\begin{bmatrix} \sqrt{\frac{1+g\mathrm{e}^{-jk\Delta l}}{1+g\mathrm{e}^{jk\Delta l}}} & \sqrt{\frac{1-g\mathrm{e}^{-jk\Delta l}}{1-g\mathrm{e}^{jk\Delta l}}} \\ \sqrt{\frac{1+g\mathrm{e}^{-jk\Delta l}}{1+g\mathrm{e}^{jk\Delta l}}} & -\sqrt{\frac{1-g\mathrm{e}^{-jk\Delta l}}{1-g\mathrm{e}^{jk\Delta l}}} \end{bmatrix}. \tag{A9}$$

The singular value decomposition of $\mathbf{C}$ may therefore be written as

$$\begin{aligned}\mathbf{C}=\mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}}=&\frac{1}{\sqrt{2}}\begin{bmatrix} \sqrt{\frac{1+g\mathrm{e}^{-jk\Delta l}}{1+g\mathrm{e}^{jk\Delta l}}} & \sqrt{\frac{1-g\mathrm{e}^{-jk\Delta l}}{1-g\mathrm{e}^{jk\Delta l}}} \\ \sqrt{\frac{1+g\mathrm{e}^{-jk\Delta l}}{1+g\mathrm{e}^{jk\Delta l}}} & -\sqrt{\frac{1-g\mathrm{e}^{-jk\Delta l}}{1-g\mathrm{e}^{jk\Delta l}}} \end{bmatrix}\begin{bmatrix} \sqrt{(1+g\mathrm{e}^{jk\Delta l})(1+g\mathrm{e}^{-jk\Delta l})} & 0 \\ 0 & \sqrt{(1-g\mathrm{e}^{jk\Delta l})(1-g\mathrm{e}^{-jk\Delta l})} \end{bmatrix} \\ &\times\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}.\end{aligned} \tag{A10}$$

Note that

$$\mathbf{C}^{-1}=[\mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}}]^{-1}=\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^{\mathrm{H}}, \tag{A11}$$

since $\mathbf{V}^{\mathrm{H}}\mathbf{V}=\mathbf{I}$, $[\mathbf{V}^{\mathrm{H}}]^{-1}=\mathbf{V}$, $\mathbf{U}^{\mathrm{H}}\mathbf{U}=\mathbf{I}$, and $\mathbf{U}^{-1}=\mathbf{U}^{\mathrm{H}}$. Therefore,

$$\mathbf{H}=\mathbf{C}^{-1}=\mathbf{V}\boldsymbol{\Sigma}^{-1}\mathbf{U}^{\mathrm{H}}=\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{(1+g\mathrm{e}^{jk\Delta l})(1+g\mathrm{e}^{-jk\Delta l})}} & 0\\ 0 & \frac{1}{\sqrt{(1-g\mathrm{e}^{jk\Delta l})(1-g\mathrm{e}^{-jk\Delta l})}}\end{bmatrix}$$
$$\times\frac{1}{\sqrt{2}}\begin{bmatrix}\sqrt{\frac{1+g\mathrm{e}^{jk\Delta l}}{1+g\mathrm{e}^{-jk\Delta l}}} & \sqrt{\frac{1+g\mathrm{e}^{jk\Delta l}}{1+g\mathrm{e}^{-jk\Delta l}}}\\ \sqrt{\frac{1-g\mathrm{e}^{jk\Delta l}}{1-g\mathrm{e}^{-jk\Delta l}}} & -\sqrt{\frac{1-g\mathrm{e}^{jk\Delta l}}{1-g\mathrm{e}^{-jk\Delta l}}}\end{bmatrix}. \tag{A12}$$

Hence

$$\sigma_i=\frac{1}{\sqrt{(1+g\mathrm{e}^{-jk\Delta r\sin\theta})(1+g\mathrm{e}^{jk\Delta r\sin\theta})}},\quad \sigma_o=\frac{1}{\sqrt{(1-g\mathrm{e}^{-jk\Delta r\sin\theta})(1-g\mathrm{e}^{jk\Delta r\sin\theta})}} \tag{A13}$$

are the singular values of the inverse filter matrix **H**.

[1] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA, 1997), Chap. 2, pp. 50–136.
[2] H. Møller, ''Fundamentals of Binaural Technology,'' Appl. Acoust. **36**, 171–218 (1992).
[3] D. R. Begault, *3-D Sound for Virtual Reality and Multimedia* (AP Professional, Cambridge, MA, 1994).
[4] M. R. Schroeder and B. S. Atal, ''Computer simulation of sound transmission in rooms,'' IEEE Intercon. Rec. Pt7, 150–155, 1963.
[5] P. Damaske, ''Head-related two-channel stereophony with reproduction,'' J. Acoust. Soc. Am. **50**, 1109–1115 (1971).
[6] J. L. Bauck and D. H. Cooper, ''Generalized transaural stereo and applications,'' J. Audio Eng. Soc., **44**, 683–705 (1996).
[7] P. A. Nelson, O. Kirkeby, T. Takeuchi, and H. Hamada, ''Sound fields for the production of virtual acoustic images,'' J. Sound Vib. **204**, 386–396 (1997).
[8] P. A. Nelson, F. Orduna-Bustamante, and H. Hamada, ''Inverse filter design and equalisation zones in multi-channel sound reproduction,'' IEEE Trans. Speech Audio Process. **3**, 185–192 (1995).
[9] O. Kirkeby, P. A. Nelson, F. Orduna-Bustamante, and H. Hamada, ''Local sound field reproduction using digital signal processing,'' J. Acoust. Soc. Am. **100**, 1584–1593 (1996).
[10] S. Barnett, *Matrices—Methods and Applications* (Oxford University Press, Oxford, 1990), Chap. 8, pp. 218–225, 396–401.
[11] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, ''Localisation using nonindividualized head-related transfer functions,'' J. Acoust. Soc. Am. **94**, 111–123 (1993).
[12] H. Møller, M. F. Sørensen, D. Hammershøi, and C. B. Jensen, ''Head-related transfer functions on human subjects,'' J. Audio Eng. Soc., **43**, 300–321 (1995).
[13] T. Takeuchi, P. A. Nelson, O. Kirkeby, and H. Hamada, ''Influence of individual head related transfer function on the performance of virtual acoustic imaging systems,'' 104th AES Convention Preprint 4700 (P4-3), 1998.
[14] T. Takeuchi, P. A. Nelson, and H. Hamada, ''Robustness to Head Misalignment of Virtual Sound Imaging Systems,'' J. Acoust. Soc. Am. **109**, 958–971 (2001).
[15] D. B. Ward and G. W. Elko, ''Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation,'' IEEE Signal Process. Lett. **6**, 106–108 (1999).
[16] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, 2nd ed. (Cambridge University Press, Cambridge, 1992).
[17] T. Takeuchi, P. A. Nelson, O. Kirkeby, and H. Hamada, ''The effects of reflections on the performance of virtual acoustic imaging systems'' in Proceedings of the Active 97, The International Symposium on Active Control of Sound and Vibration, Budapest, Hungary, 1997, pp. 955–966.
[18] B. Rakerd, and W. M. Hartmann, ''Localization of sound in rooms. II. The effects of a single reflecting surface,'' J. Acoust. Soc. Am. **78**, 524–533 (1985).
[19] U. Burandt, C. Poesselt, S. Ambrozus, M. Hosenfeld, and V. Knauff, ''Anthropometric contribution to standardising manikins for artificial head microphones and to measuring headphones and ear protectors,'' Appl. Ergonomics **22**, 373–378 (1991).
[20] B. Gardner and K. Martin, ''HRTF Measurements of a KEMAR Dummy-Head Microphone,'' MIT Media Lab Perceptual Computing—Technical Report No. 280, 1994.
[21] T. Takeuchi and P. A. Nelson, ''Optimal source distribution for virtual acoustic imaging,'' ISVR Technical Report No. 288, University of Southampton, 2000.