APRIL 01 2003

Improved sound separation using three loudspeakers

Jun Yang; Woon-Seng Gan; See-Ee Tan

ARLO 4, 47–52 (2003)
https://doi.org/10.1121/1.1566419

Improved sound separation using three loudspeakers

Jun Yang, Woon-Seng Gan, See-Ee Tan

School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Republic of Singapore
ejyang@ntu.edu.sg, ewsgan@ntu.edu.sg, tanse4@pmail.ntu.edu.sg

Abstract: In a virtual sound imaging system, crosstalk cancellation filters are used to create an effective sweet spot for 3D sound reproduction via multiple loudspeakers. A new 3-channel system is proposed to improve system performance on sound separation. Based on the robustness analysis of a crosstalk canceller, a modified-inverse filter technique is explored and demonstrated using two different examples of symmetric speaker positions. The simulation results indicate that the present system is robust over a wider bandwidth compared to a conventional 2-channel system.

PACS numbers: 43.38Hz, 43.38Md, 43.60Pt, 43.66Qp
Date received: 9 July 2002 **Date accepted**: 30 January 2003

1. Introduction

Virtual sound imaging allows the listener to perceive sound at a location other than the actual source position. It has been applied to many areas, including home entertainment, flight simulation and virtual reality. To reproduce the desired signals at the listener's ears via speakers, a preprocessing technique is adopted by compensating for the effect of the physical system, such as the acoustic crosstalk path [1]. This crosstalk arises because each speaker sends sound to the ipsilateral ear, as well as undesired sound to the contralateral ear.

Several methods have been proposed to address crosstalk cancellation so that appropriate binaural signals are delivered to the listener's ears [1]. It was recently shown that the robust solution to this virtual sound system could be obtained by placing the loudspeakers so as to assure the acoustic transmission path or transfer function (TF) matrix is a well-conditioned [2,3]. However, the optimal loudspeaker position is frequency dependent. For a given speaker angle, there are limited frequency bands within which the system is stable. It is well known that this undesirable property causes an excessive amplitude boost at certain frequencies, which may damage the amplifier and the loudspeakers. Moreover, crosstalk cancellation is inherently nonrobust at low frequencies for the 2-channel system.

Recently, Takeuchi et al. [4] have employed a 2-channel multiway sound control system to improve the robustness performance. This system requires multiple speakers with different spans at different frequency bands. In this paper, we propose an alternative solution by using 3 speakers. As an extension to our previous work [3], we focus on the design of a crosstalk cancellation filter to increase the bandwidth over which the proposed system is robust. The proposed method is based on a free-field model that does not account for the presence of a listener in the sound field, and a symmetric geometry with monopole loudspeakers is considered to simplify the filter design.

2. Robustness analysis

Generally, consider multichannel sound reproduction: the K recorded signals are transmitted via N loudspeakers to M target points (ears). Let V be the vector of N source input signals, D is the vector of M program signals, and A is the transfer function (TF) of the $M \times N$ matrix. In the free-field case, the sound field radiated from a monopole source can be written as $p = (S/r)\cdot\exp[j(\omega t - kr)]$, where the time dependence $\exp(j\omega t)$ is henceforth omitted for brevity, ω is the angular frequency, $k = 2\pi/\lambda$ is the wave number, λ being the wave length, and r is the distance from the source to the field point. $S = j\omega\rho_0 q/4\pi$ is defined as strength term, ρ_0 is the density of the medium, and q is the source strength. Then the TF a_{mn} between the n *th* monopole source (loudspeaker) and the m *th* target point (ear) is given by $a_{mn} = e^{-jkr_{mn}}/r_{mn}$. For the 3-channel system illustrated in Fig. 1 ($m = 2$ and $n = 3$), the symmetrical configuration is assumed to simplify the filter design. Two speakers are located on each side, and the third one is placed directly in front of the listener. The symmetrical sound path is defined as: $r_{11} = r_{23}$, $r_{21} = r_{13}$, and $r_{12} = r_{22}$. Therefore, the transfer function matrix A can be written as:

$$A = \begin{bmatrix} e^{-jkr_{11}}/r_{11} & e^{-jkr_{12}}/r_{12} & e^{-jkr_{13}}/r_{13} \\ e^{-jkr_{21}}/r_{21} & e^{-jkr_{22}}/r_{22} & e^{-jkr_{23}}/r_{23} \end{bmatrix} \triangleq \begin{bmatrix} a_1 & a_2 & a_3 \\ a_3 & a_2 & a_1 \end{bmatrix} \quad (1a)$$

and

$$AA^H = \begin{bmatrix} |a_1|^2 + |a_2|^2 + |a_3|^2 & |a_2|^2 + 2|a_1 a_3|\cos(k\Delta) \\ |a_2|^2 + 2|a_1 a_3|\cos(k\Delta) & |a_1|^2 + |a_2|^2 + |a_3|^2 \end{bmatrix} \triangleq \begin{bmatrix} b_1 & b_2 \\ b_2 & b_1 \end{bmatrix} \quad (1b)$$

where $\Delta = |r_{21} - r_{11}|$ denotes the interaural path difference, and the corresponding quantity is the phase difference $\phi = k\Delta$.

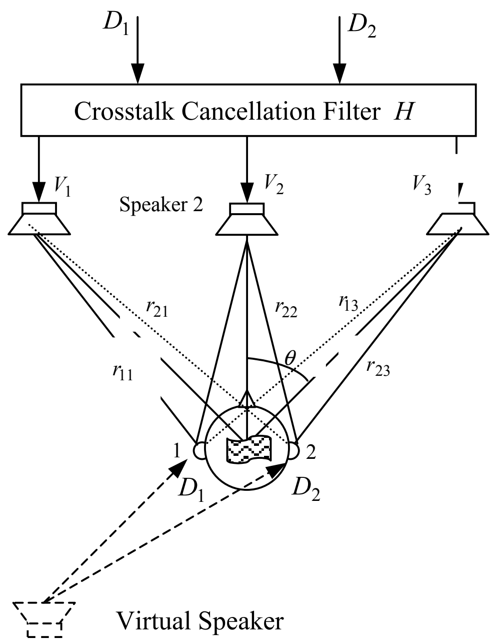

Fig. 1. 3-speaker system with H to modify transmission path of system.

The crosstalk cancellation filter H can be calculated to minimize the cost function $J = e^H \cdot e$, where, $e = D - A \cdot V$ is the error between the desired and reproduced signals and $V = H \cdot D$. It is well known that the pseudoinverse solution is

$$H = A^H \cdot (AA^H)^{-1} \quad (2)$$

where, the superscript H is the Hermitian transpose of a matrix. The condition number $\kappa(H)$ can be used as a robustness measure for the crosstalk canceller, defined as [5]

$$\kappa = \sigma_{\max} / \sigma_{\min} \tag{3}$$

where $\sigma_{\min}$ and $\sigma_{\max}$ represent the smallest and largest singular values of H, respectively, which are the square roots of the respective nonzero eigenvalues of $H^H H$. The value of κ is always greater or equal to 1. A large condition number characterizes the *ill-conditioned* problem, which means the reproduced signals are highly sensitive to small errors or changes in the system. From Eq.(2), we have a 2×2 Hermitian matrix $H^H H = (AA^H)^{-1}$. The decomposition of AA^H is given as

$$AA^H = Q \cdot \Lambda \cdot Q^{-1} \tag{4}$$

where Λ and Q are a diagonal and an orthogonal matrix respectively,

$$\Lambda = \begin{bmatrix} b_1 + b_2 & 0 \\ 0 & b_1 - b_2 \end{bmatrix}, \qquad Q = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \tag{5a, b}$$

Note that $Q = Q^T = Q^{-1}$, we have

$$H^H H = Q \cdot \Lambda^{-1} \cdot Q^{-1} = Q \cdot \begin{bmatrix} \frac{1}{b_1 + b_2} & 0 \\ 0 & \frac{1}{b_1 - b_2} \end{bmatrix} \cdot Q . \tag{6}$$

The ill-conditioned problem can be reflected clearly in Eq.(6). The entries $1/(b_1 \pm b_2)$ are "eigenvalues", denoted as $\varepsilon_{1,2}$, and the singular values become $\sigma_{1,2} = \sqrt{\varepsilon_{1,2}}$. When the magnitude of $b_1 + b_2$ or $b_1 - b_2$ is small at certain frequencies, it results in a large condition number, and thus, large gains in the inverse filter, which is not feasible to implement in practice. According to Eqs.(1), (3) and (6), the condition number is minimum when $b_2 = 0$ or $|a_2|^2 + 2|a_1 a_3|\cos\phi = 0$, and becomes infinity when $b_1 - b_2 = 0$ or $|a_1|^2 + |a_3|^2 - 2|a_1 a_3|\cos\phi$ $= 0$. Note that $b_1 + b_2 = |a_1 + a_3|^2 + 2|a_2|^2$ is a positive value. For simplicity, the variables r_{11} , r_{12} and r_{13} can be regarded as almost equal. The crosstalk canceller will be most robust, $\kappa(H) = 1$ when $\cos\phi = -1/2$, i.e., $\Delta = (p - 1/3)\lambda$ or $(p - 2/3)\lambda$, $p = 1,2,\cdots,n$. On the other hand, if $\cos\phi = 1$, i.e., $\Delta = q\lambda$, $q = 1,2,\cdots,n$, $\kappa(H) \to \infty$, it makes the crosstalk canceller nonrobust. The physical explanation for this singularity is that at certain frequencies in relation to the geometry, the path length difference between a loudspeaker and both ears becomes close or equal to a multiple integer of a wavelength of the input signal.

Furthermore, the 3-channel system reduces to a 2-channel system when $a_2 = 0$. From Eq.(1), $b_1 = 2$ and $b_2 = 2\cos\phi$. The 2-channel system is the most robust when $\cos\phi = 0$ ($\kappa(H) = 1$) and nonroubst when $\cos\phi = \pm 1$ ($\kappa(H) \to \infty$), it corresponds to $\Delta = (2p - 1) \cdot \lambda / 4$ and $\Delta = p \cdot \lambda / 2$, $p = 1,2,\cdots,n$. The simulation results of the condition number will be given in the section 4.

3. New crosstalk canceller design

The crosstalk canceller H described in Eq. (2) is obtained in a least-square manner. To enlarge the robust bandwidth, a modified, inverse filter technique is explored by relaxing the requirement of minimum power input into the speakers, and two inverse solutions are derived under different constraints. Since the reproduced signal depends on the physical system A and the input $V = H \cdot D$ into the speakers, we can always ensure minimum error at any adjustment of A^H in Eq. (2) without changing the physical system A. For example, if a new matrix C^H is used to replace A^H in the inverse filter form, i.e., $H' = C^H \cdot (AC^H)^{-1}$, then the error $e = [I - AC^H \cdot (AC^H)^{-1}] \cdot D = 0$, provided (AC^H) is nonsingular. It implies that there are many solutions to ensure accurate reproduction at target points. Assuming the unknown matrix C has a form similar to A, $C = \begin{bmatrix} c_1 & c_2 & c_3 \\ c_3 & c_2 & c_1 \end{bmatrix}$, we have

$$AC^H = \begin{bmatrix} a_1c_1^* + a_2c_2^* + a_3c_3^* & a_1c_3^* + a_2c_2^* + a_3c_1^* \\ a_1c_3^* + a_2c_2^* + a_3c_1^* & a_1c_1^* + a_2c_2^* + a_3c_3^* \end{bmatrix} \triangleq \begin{bmatrix} d_1 & d_2 \\ d_2 & d_1 \end{bmatrix} \tag{7a}$$

and

$$(AC^H)^{-1} = \frac{1}{d_1^2 - d_2^2} \cdot \begin{bmatrix} d_1 & -d_2 \\ -d_2 & d_1 \end{bmatrix}. \tag{7b}$$

According to the previous robustness analysis, the system should be robust when the condition number $\kappa(AC^H) = 1$. It means $d_2 = 0$, i.e.,

$$a_1c_3^* + a_2c_2^* + a_3c_1^* = 0. \tag{8}$$

For simplicity, set $c_1 = -c_3$, Eq. (8) can be written as

$$(a_1 - a_3) \cdot c_3^* = -a_2 \cdot c_2^*. \tag{9}$$

Using Eqs. (7) and (9), we get the modified inverse filter form:

$$H_1 = C^H \cdot (AC^H)^{-1} = \frac{1}{2} \begin{bmatrix} 1/(a_1 - a_3) & -1/(a_1 - a_3) \\ 1/a_2 & 1/a_2 \\ -1/(a_1 - a_3) & 1/(a_1 - a_3) \end{bmatrix} \tag{10}$$

and

$$H_1^H H_1 = \frac{1}{4} \cdot \begin{bmatrix} 1+g & 1-g \\ 1-g & 1+g \end{bmatrix} \tag{11}$$

where $g = 2/|a_1 - a_3|^2 = 1/(1 - \cos\phi)$.

The eigenvalues of $H_1^H H_1$ are $1/2$ and $1/2(1-\cos\phi)$, and the crosstalk canceller will be most robust as in the 2-channel system, $\kappa(H_1) = 1$ when $\cos\phi = 0$, i.e., $\Delta = (2p-1) \cdot \lambda/4$, $p = 1,2,\cdots,n$. On the other hand, it will be the least robust as in the 3-channel system, $\kappa(H_1) \to \infty$ when $\cos\phi = 1$, i.e., $\Delta = q\lambda$, $q = 1,2,\cdots,n$. Therefore, a wider robust region is achieved by using the presented modified inverse filter.

An alternative solution to Eq. (8) can be obtained by setting $C = A + \delta A$, or $c_i = a_i + \delta a_i$, $i = 1,2,3$. Using this configuration has the advantage of requiring little effort in deriving a robust system, and also the solution obtained is likely to be near the least norm. The Eq. (8) can be rewritten as

$$a_1\delta a_3^* + a_3\delta a_1^* + a_2\delta a_2^* = -|a_2|^2 - (a_1a_3^* + a_1^*a_3). \tag{12}$$

Similarly, let $a_2\delta a_2^* = -(a_1a_3^* + a_1^*a_3)$, one gets $\delta a_2 = -(a_1a_3^* + a_1^*a_3)/a_2^*$. Substituting it into Eq. (8), we have

$$a_1\delta a_3^* + a_3\delta a_1^* = -|a_2|^2. \tag{13}$$

If $a_2\delta a_2^* = -|a_2|^2$, $\delta a_2 = -a_2$ and $c_2 = 0$. It reduces to the 2-channel system. Assuming

$\delta a_1 = -\delta a_3$, it is derived that $\delta a_1 = -\delta a_3 = |a_2|^2 / (a_1^* - a_3^*)$, and the modified inverse becomes

$$H_2 = \frac{1}{2 \cdot (2 - \cos\phi)} \begin{bmatrix} (2 - a_1^* a_3)/(a_1 - a_3) & (a_1 a_3^* - 2)/(a_1 - a_3) \\ (1 - 2\cos\phi)/a_2 & (1 - 2\cos\phi)/a_2 \\ (a_1 a_3^* - 2)/(a_1 - a_3) & (2 - a_1^* a_3)/(a_1 - a_3) \end{bmatrix}. \quad (14)$$

The eigenvalues of $H_2^H H_2$ are respectively $\varepsilon_1 = (4\cos^2\phi - 3\cos\phi + 2)/2(2 - \cos\phi)^2$ and $\varepsilon_2 = 1/2(1 - \cos\phi)$. If $\cos\phi = 0$, $\kappa(H_2) \to \infty$, it makes the crosstalk canceller nonrobust as the modified inverse H_1 . However, this crosstalk canceller has a better robust performance than filter H_1 at low frequencies as shown in the next section. It is a desirable design, because crosstalk cancellation at low frequencies is fundamentally difficult [2].

It should be noted that, unlike in the pseudoinverse case $b_1 > 0$, d_1 in Eq. (7a) might be zero at certain situations and raise an ill-conditioned problem. From Eqs.(7) and (8), one gets $d_1 = (a_1 - a_3)(c_1^* - c_3^*)$ when $d_2 = 0$. Obviously, $d_1 = 0$, if $a_1 - a_3 = 0$ or $\cos\phi = 1$. It shows the intrinsic limitation of the physical system, regardless of the designed crosstalk canceller. Thus, a constraint such as a regularization technique can be introduced to overcome this problem, and a proper modeling delay is added to the system design to guarantee a causal filter for practical application [6].

4. Simulation and discussion

For the model shown in Fig.1, all the loudspeakers are placed at a distance of $1m$ from the center of the head, and a standard head radius of $0.0875m$ is used. The condition numbers can be calculated from Eqs.(2), (3) (10) and (14) corresponding to 2-channel and 3-channel systems using pseudoinverse and modified inverse filters. The maximum condition number is limited to five. In addition, for frequencies above 6 kHz, signals from the loudspeakers are generally assumed to add incoherence [1], and a crosstalk canceller that is based on the coherent addition of a signal will not be employed. The condition number plots for different azimuth angles and frequencies below 6 kHz are presented in Fig. 2(a)-(d).

It is observed that the 3-speaker configuration gives a better robustness performance compared to the 2-speaker case. The arrows shown in plots represent the start and end points of the robust bandwidth for a fixed speaker angle such as $15°$. There is a significantly wider robust margin (the lower condition number is represented in white) in all the 3-channel systems. It is the extra degree of freedom that improves the system performance. Fig. 2 shows a clear trend that, for the wider speaker placement, the system is robust over a narrow range of frequencies, whereas the robust region is increased with the decrease in speaker angle. However, the system robustness degrades at low frequencies, because the difference between the direct and cross path is very small. An ill-conditioned problem exists and is particularly severe in the case of the two side loudspeakers placed closely. From Figs. 2(b), (c) and (d), it can also be seen that the robust performance of modified inverse filters is superior to that of pseudoinverse filter. The proposed crosstalk cancellers H_1 and H_2 further enlarge the robust region compared to the pseudoinverse filter H . The modified inverse H_1 provides a wider robust region than that of H_2 , but the filter H_2 extends the robust zone to low frequencies. As shown in Fig. 2(d), for the side speakers' placement of $15°$, a good robustness is achieved from 300 Hz up to 6 kHz. This is especially attractive because crosstalk cancellation at low frequencies is fundamentally difficult. In general, considering the robustness at low and high frequencies, a practical solution would be to use a modified inverse filter with side speaker angle around $15°$.

 1529-7853/03/4(2)/47/6/$19.00

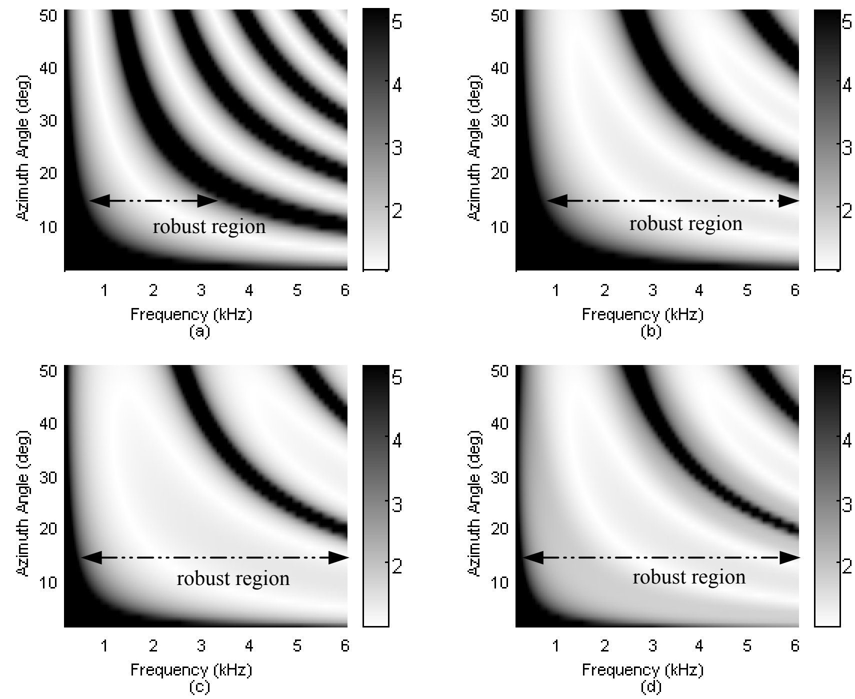

Fig. 2. Condition number plot for (a) 2-speaker configuration; and 3-speaker configuration (b) pseudoinverse H ; (c) modified inverse H_1 ; and (d) modified inverse H_2 .

6. Summary

A 3-channel system is proposed for virtual sound imaging to improve the performance over the 2-channel case. Based on robustness analysis, a new crosstalk cancellation filter is derived to further enlarge the robust bandwidth, which results in a flexible speaker placement. The present design is relatively easy and economical to implement. It should be noted that one may be able to obtain even greater improvements in robustness by proper selection of the C matrix, provided the optimal solution is affordable in practical application.

Acknowledgments

The authors are grateful to the referees for their meticulous reading of the manuscript and many helpful recommendations.

References

[1]W.G. Gardner, *3D audio using loudspeakers*, (Kluwer Academic Pub, Boston, 1998).

[2]D. B. Ward and G.W. Elko, "Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation," IEEE Signal Processing Letter, **6**, 106-108 (1999).

[3]J. Yang and W.S. Gan, "Speaker placement for robust 3D audio system," IEE Electr. Letter. **36**, 683-686 (2000).

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

[5]G.H. Golub and C.F. Van Loan, *Matrix computation*, (3rd edition, Baltimore: The Johns Hopkins Univ. Press, 1996).

[6]O. Kirkeby, P.A. Nelson, H. Hamada and F. Orduna-Bustamante, "Fast deconvolution of multichannel systems using regularization," IEEE Trans. Speech Audio Processing, **6**, 189-194 (1998).

