ACOUSTIC CROSSTALK CANCELLATION IN A REVERBERANT ENVIRONMENT

Darren B. Ward
Department of Electrical and Electronic Engineering
Imperial College of Science, Technology and Medicine
Exhibition Road, London SW7 2BT

1. INTRODUCTION

A crosstalk cancellation system (CCS) creates a virtual audio environment by using loudspeakers to deliver appropriate binaural signals to a listener. Typically, the CCS is designed to equalize the direct-path transfer functions between the loudspeakers and the ears, thereby providing the listener with an accurate binaural image. Since most virtual audio systems will be used in a reverberant room, it is important to consider what effect the room reverberation has on the performance of the CCS.

In this paper, statistical room acoustics is used to derive a closed-form expression that predicts the performance of a CCS when used in a reverberant room. In particular, we provide an expression for the expected mean-square response obtained at each of the listener's ears. This expression is parameterised by the loudspeaker/listener geometry, the room dimensions and reverberation time. Image-model simulations are used to verify the results obtained.

The results presented in this paper allow a designer to undertake a preliminary analysis of how well a given CCS will perform in a particular reverberant environment, as well as enabling the comparison of several alternative systems (with different loudspeaker geometries, say) without requiring time-consuming measurements or image-model simulations.

2. ACOUSTIC CROSSTALK CANCELLATION

To recreate a virtual audio environment, it is necessary to recreate at the listener's ears the precise acoustic pressures that would result if the listener were physically present in the virtual environment being simulated. In effect, this serves to produce the same stimuli to the human auditory system as would be produced by a real sound source. A key ingredient of such a system is a crosstalk canceller [1]. This serves to equalize the transmission paths between the loudspeakers and the listener's ears, thereby ensuring that the acoustic pressure at the listener's ears can be accurately controlled.

Consider the standard two-channel CCS shown in Figure 1. (See [2] for results that are applicable to the more general *N*-loudspeaker system.) The aim of the CCS is to accurately deliver the binaural signals b_L and b_R to the listener's ears. Let $l_n, n = 1,2$ denote the locations of the loudspeakers, and e_L and e_R the location of the listener's left and right ears, respectively. At any frequency, the actual signals received at the ears are given by

$$\begin{bmatrix} \hat{b}_L \\ \hat{b}_R \end{bmatrix} = \begin{bmatrix} g(l_1, e_L) & g(l_2, e_L) \\ g(l_1, e_R) & g(l_2, e_R) \end{bmatrix} \begin{bmatrix} h_1 & h_3 \\ h_2 & h_4 \end{bmatrix} \begin{bmatrix} b_L \\ b_R \end{bmatrix}$$

$$\hat{b} = GHb \qquad (1)$$

where $g(l_1, e_L)$ is the acoustic transfer function (TF) between loudspeaker 1 and the left ear (and similarly for the other loudspeaker/ear pairs). We will refer to the matrix G as the acoustic TF matrix. The crosstalk cancellation filters are denoted by $h_n, n = 1 \ldots 4$, and are designed to ensure that the actual ear signals are as close as possible to the desired ear signals.

Without loss of generality, consider the left binaural channel only (a similar analysis holds for the right channel). Let the ideal TFs between b_L (which we will hereafter refer to as the input) and each ear be v_L and v_R. To reproduce the left binaural signal correctly, the left ear should receive the left binaural signal and the right ear should receive no signal. Thus, the TF to the left ear should be a delay (to account for propagation), and the TF to the right ear should be zero, i.e.,

$$\begin{bmatrix} v_L \\ v_R \end{bmatrix} = \begin{bmatrix} e^{-j2\pi f\tau_0} \\ 0 \end{bmatrix} \qquad (2)$$

where τ_0 is the delay to the left ear. Similarly, let the actual TFs between the input and the ears be $\hat{v}_L$ and $\hat{v}_R$ respectively, i.e.,

$$\begin{bmatrix} \hat{v}_L \\ \hat{v}_R \end{bmatrix} = \begin{bmatrix} g(l_1, e_L) & g(l_2, e_L) \\ g(l_1, e_R) & g(l_2, e_R) \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}$$
$$\hat{v} = Gh. \qquad (3)$$

The acoustic TFs between loudspeakers and ears, $g(l,e)$, can be separated into the sum of direct- and reverberant-path parts, i.e., $g(l,e) = g_d(l,e) + g_r(l,e)$, and similarly separate the acoustic TF matrix to give:

$$\hat{v} = [G_d + G_r]h. \qquad (4)$$

In general, the reverberant-path components of the acoustic TFs will be unknown, and so the CCS filters will be designed to equalize only the direct-path TFs, i.e.,

$$v = G_d h. \qquad (5)$$

Assume that the CCS filters exactly equalize the direct paths, so that the actual TFs between the input and the ears become

$$\hat{v} = v + G_r h. \qquad (6)$$

Thus, we note that the actual TFs depend on the reverberant paths between the loudspeakers and the ears.

In this paper we will describe the performance of the CCS in terms of the expected mean-square ear responses, given by:

$$\left\langle |\hat{v}_L|^2 \right\rangle = \left\langle |v_L + r_L^T h|^2 \right\rangle$$
$$\left\langle |\hat{v}_R|^2 \right\rangle = \left\langle |v_R + r_R^T h|^2 \right\rangle \qquad (7)$$

where $\langle \cdot \rangle$ denotes the expectation operator,

$$r_L = [g_r(l_1, e_L) \; g_r(l_2, e_L)]^T \qquad (8)$$

is the vector of reverberant TFs to the left ear, and

$$r_R = [g_r(l_1, e_R) \; g_r(l_2, e_R)]^T \qquad (9)$$

is the vector of reverberant TFs to the right ear. The mean-square ear responses then become:

$$\left\langle |\hat{v}_L|^2 \right\rangle = |v_L|^2 + \left\langle h^H r_L^* \right\rangle + \left\langle r_L^T h \right\rangle + \left\langle h^H r_L^* r_L^T h \right\rangle$$
$$\left\langle |\hat{v}_R|^2 \right\rangle = |v_R|^2 + \left\langle h^H r_R^* \right\rangle + \left\langle r_R^T h \right\rangle + \left\langle h^H r_R^* r_R^T h \right\rangle \quad (10)$$

To determine these ear responses, it is necessary to find an expression for terms of the form:

$$\left\langle g_r^*(l,e) g_r(l,e) \right\rangle$$

for each of the loudspeaker/ear pairs. This can be achieved using the following results from statistical room acoustics.

3. STATISTICAL ROOM ACOUSTICS

Consider a room with volume V, total wall surface area S, and an average absorption coefficient α (defined as the ratio of the intensity of sound pressure absorbed by a wall surface to the intensity incident on the surface). According to the statistical room acoustics method, the sound field at any point in the room can be considered as the superposition of an infinite number of plane waves arriving from all directions [3]. This diffuse model of the room reverberation becomes valid above the Schroeder frequency [3], given by $f_{Sch} = 2000\sqrt{T_{60}/V}$ where $T_{60} = 0.161\,V/(S\alpha)$ is the reverberation time (defined as the time taken for the sound-pressure level to decay by 60 dB once the sound source has stopped).

At a single point x in the room, consider the reverberant-path TFs due to two acoustic sources located at y_1 and y_2 respectively. It was shown in [2] that the correlation between these reverberant-path TFs is given by:

$$\left\langle g_r^*(y_1,x) g_r(y_2,x) \right\rangle = \left(\frac{1-\alpha}{\pi S\alpha}\right) \frac{\sin(k\|y_1 - y_2\|)}{k\|y_1 - y_2\|} \quad (11)$$

where $\langle\cdot\rangle$ denotes the expectation operator, $k = 2\pi f/c$ is the wave number (with c the speed of wave propagation in air), and $\|\cdot\|$ denotes the vector 2-norm.

Relating this result to the CCS, we find that if $\Delta = \|l_1 - l_2\|$ is the distance between the two loudspeakers, then the cross-correlation for the TFs between the loudspeakers and the left ear is given by:

$$\left\langle g_r^*(l_1,e_L) g_r(l_2,e_L) \right\rangle = \left(\frac{1-\alpha}{\pi S\alpha}\right) \frac{\sin(k\Delta)}{k\Delta} \quad (12)$$

and similarly for the right ear. The autocorrelation terms are:

$$\left\langle g_r^*(l_1,e_L) g_r(l_1,e_L) \right\rangle = \left(\frac{1-\alpha}{\pi S\alpha}\right) \quad (13)$$

and similarly for the right ear.

Another well-known result from statistical room acoustics is that the direct and reverberant sound pressures at any point in a room are uncorrelated. Noting from (5) that h is a linear combination of the direct-path TFs, it follows that the cross terms $\langle h^H r^* \rangle$ and $\langle h^H r^* \rangle$ in

(10) are zero. Substituting these expressions into the mean-square ear responses gives the following result.

4. RESULTS AND DISCUSSION

Consider the two-channel CCS shown in Figure 1. Let h denote the CCS filters that equalize the direct-path acoustic transfer functions between the loudspeakers and ears for a specific position of the listener. Also let $\nu_L = e^{-j2\pi f\tau_0}$ and $\nu_R = 0$ be the ideal ear responses for the left binaural channel. If this CCS is located in a room with total wall surface area S and an average absorption coefficient α, the mean-square responses at the listener's ears are:

(14)

where

(15)

is the error, and is the distance between the loudspeakers. We make the following comments regarding (14) and (15):

1. For each ear, the effect of reverberation is to add an error term to the ideal ear response, where the error term is the same for both ears.
2. This error term is scaled by a constant determined by the room parameters, and consists of two separate parts. The first part is the L_2 norm of the filter weights, i.e., , thereby suggesting that the effect of reverberation can be reduced by reducing the norm of the filter weights. The second part (consisting of the terms) falls off rapidly with increasing frequency or with increasing loudspeaker spacing, thereby suggesting that increasing the loudspeaker spacing can reduce the effect of reverberation.

Thus, the expression for the mean-square ear response suggests two ways in which the effect of reverberation can be reduced, either by reducing the norm of the filter weights or by increasing the loudspeaker spacing. Note from (5) that the norm of the filter weights will be related to the conditioning of the direct-path TF matrix : if has a high condition number, then the filter weights will have a high norm, and vice versa. Furthermore, the conditioning of depends on the loudspeaker geometry. As shown in [4], the direct-path TF matrix is generally well conditioned if the loudspeakers are positioned close together. It has also been shown in [5] that a CCS with closely spaced loudspeakers is less sensitive to the exact position of the listener. In other words, closely spaced loudspeakers help enlarge the sweet spot. Thus, although increasing loudspeaker spacing will reduce the second part of the error term in (15), large loudspeaker spacing has two detrimental effects: first, it causes the filter weights to have a large norm and thereby increases the first part of the error term in (15); second, it reduces the size of the sweet spot. Therefore, it seems that close loudspeaker spacing is to be preferred for the two-channel CCS.

5. SIMULATIONS

To verify the theoretical results, we now present simulations. The transfer functions in a room with dimensions of metres, having a reverberation time of seconds, were simulated using the image model [6]. Referring to Figure 1, the loudspeakers were separated by a distance of m, and the listener was located one metre back along the line bisecting the loudspeakers. No head-related transfer function (HRTF) effects were included, and the listener's head was therefore modelled as having two point-receiver ears separated by a distance of 0.175 m. The CCS filters were designed according to (5), with the direct-path TF from the first loudspeaker to the left ear given by

and similarly for the other loudspeaker/ear pairs. For each simulation run, this overall geometry was placed randomly within the room. A total of 100 simulation runs was made.

The average mean-square ear responses (averaged over 100 runs) are shown in Figure 2 (solid lines) together with the predicted results (dashed lines). The average results are in good agreement with the predicted, thereby verifying the validity of the derived theoretical expression.

6. CONCLUSIONS

Virtual audio systems that use acoustic crosstalk cancellation are invariably designed to only equalize the direct-path transfer functions between loudspeakers and ears. In previous work, it has been shown that the size of the sweet spot depends on the loudspeaker geometry used. In this paper we have derived a closed-form expression for the reduction in performance that one could expect to obtain when using a two-channel CCS in a reverberant room. It was shown that the loudspeaker geometry plays an important role in determining system performance. In particular, the distance between the loudspeakers should be chosen to give a well-conditioned matrix of direct-path transfer functions. A generalisation of these results to the *N*-channel case can be found in [2].

7. REFERENCES

[1] B.S. Atal and M.R. Schroeder, "Apparent sound source translator", U.S. Patent 3,236,949, Feb. 1966.

[2] D.B. Ward, "On the performance of acoustic crosstalk cancellation in a reverberant environment", *J. Acoust. Soc. Am.*, vol. 110, no. 2, pp. 1195-1198, Aug. 2001.

[3] P.A. Nelson and S.J. Elliot, *Active Control of Sound*, Academic Press, London, 1992.

[4] D.B. Ward and G.W. Elko, "Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation", *IEEE Signal Processing Lett.*, vol. 6, pp. 106-108, May 1999.

[5] O. Kirkeby, P.A. Nelson and H. Hamada, "The stereo dipole – A virtual source imaging system using two closely spaced loudspeakers", *J. Audio Eng. Soc.*, vol. 46, pp. 387-395, May 1998.

[6] J. Allen and D. Berkley, "Image method for efficiently simulating small-room acoustics", *J. Acoust. Soc. Am.*, vol. 65, pp. 943-950, 1979.

Figure 1: Two-loudspeaker crosstalk cancellation system.

Figure 2: Spatially averaged ear responses as a function of frequency, for the geometry shown in Figure 1 using the image model.