OPTIMUM LOUDSPEAKER SPACING FOR ROBUST CROSSTALK CANCELLATION

Darren B. Ward and Gary W. Elko

Acoustics and Speech Research Dept.,
Bell Labs, Lucent Technologies,
600 Mountain Ave, Murray Hill, NJ, 07974.
Email: dbward@research.bell-labs.com

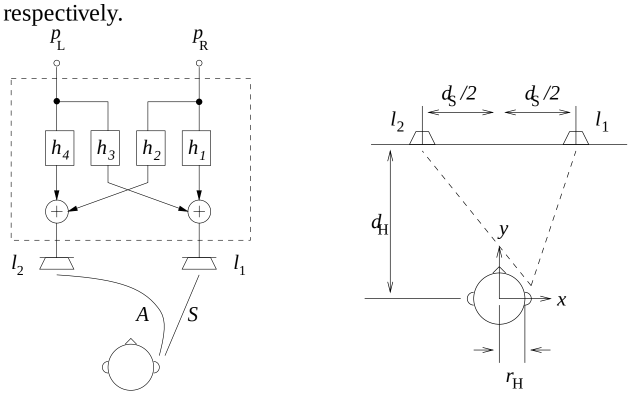
ABSTRACT

Crosstalk cancellation is a signal processing technique whereby two (or more) loudspeakers are used to deliver desired signals exactly at the listener's ears. Such a system is useful for 3D audio applications, and removes the requirement for the listener to wear headphones. However, crosstalk cancelers are notoriously non-robust to slight movements in head position, and there currently exists no clear method for determining the best loudspeaker placement in a given situation. In this paper we propose a robustness measure to evaluate the performance of crosstalk cancelers as a function of loudspeaker spacing. Based on this analysis we conclude that certain loudspeaker spacings give far better robustness performance, and provide a simple empirically-derived equation for determining the optimum loudspeaker spacing in a given situation.

1. INTRODUCTION

Crosstalk cancellation is an old idea, having been invented in the 1960's by Atal and Schroeder [1]. However, with enormous recent interest in 3D audio applications, it is once again the topic of a great amount of current research.

The classic Atal-Schroeder crosstalk canceler [1] is shown in Fig. 1(a), in which p_L and p_R are the left and right program signals respectively, l_1 and l_2 are the loudspeaker signals, and S and A are the transfer functions (TFs) to the same-side and opposite-side ears respectively.



(a) Block diagram (b) Geometry

Figure 1: Atal-Schroeder crosstalk canceler.

The objective is to find the filters h_1, h_2, h_3, h_4 such that: (i) the signals p_L and p_R are reproduced at the left and right ears respectively; and (ii) the crosstalk signals are canceled, i.e., none of the p_L signal is received at the right ear, and similarly for the p_R signal and the left ear.

For simplicity, it is usual to assume the system is acoustically symmetric, so that the TFs from the left loudspeaker to the ears are the same as those for the right loudspeaker. In this case it can be easily shown that the required filter responses are given by:

$$H_1(f) = \frac{S(f)}{S^2(f) - A^2(f)} \tag{1a}$$

$$H_2(f) = \frac{-A(f)}{S^2(f) - A^2(f)} \tag{1b}$$

$$H_3(f) = H_2(f) \tag{1c}$$

$$H_4(f) = H_1(f). \tag{1d}$$

The main problem with this system is that it is extremely sensitive to variations in head position. To demonstrate this, consider the geometry shown in Fig. 1(b), where d_S represents the loudspeaker spacing, d_H represents the distance from the loudspeaker center-line to the default position of the head center, r_H represents the radius of the head, and x and y represent the actual position of the head center relative to the default position (in the figure, x and y are both zero). The dotted lines represent the TFs from the loudspeakers to the right ear (a similar set of TFs for the left ear are not shown).

Assume we design a crosstalk canceler according to (1) for a loudspeaker spacing of $d_S = 0.3$ m, and a default head position of $d_H = 0.5$ m, and we consider its performance at a frequency of 2 kHz.[1] Figure 2 shows the amount of right program signal received at the left ear (i.e., the amount of crosstalk cancellation achieved) as the head moves in steps of 1 cm within the dotted region. The circles indicate the positions of the loudspeakers. We note that effective crosstalk cancellation is only achieved within a small region surrounding the default head position of $(x, y) = (0, 0)$.

What is required is a system that will provide reasonable crosstalk cancellation, and yet allow for some movement of the head about its default position. Although techniques have been proposed to increase the robustness of the crosstalk canceler (e.g. [2]), these methods use more than two loudspeakers to provide additional degrees-of-freedom to enlarge the cancellation region. In this paper we only consider the more practical case of two loudspeakers, and use the loudspeaker spacing as the means of increasing robustness, i.e., we evaluate the robustness of the crosstalk

[1] For simplicity, we model the head as two points in space at the ear positions, with $r_H = 0.0875$ m, i.e., no HRTFs are included in the model.

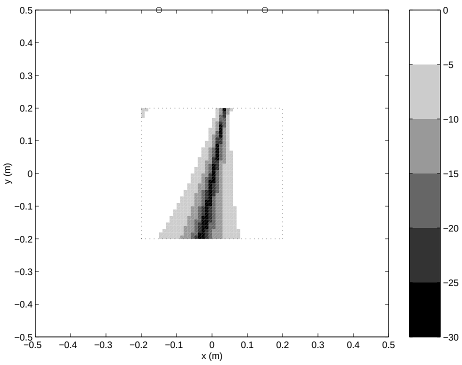


Figure 2: Demonstration of non-robustness of crosstalk cancellation. See text for details of plot.

canceler for varying loudspeaker spacings, and determine the optimum loudspeaker placement for any given situation.

An important point to note from Fig. 2 is that cancellation is more effective as the head moves forwards/backwards, than if it moves sideways. For example, if the head moves 5 cm either side of $(0,0)$, there is less than 5 dB cancellation of the crosstalk signal, whereas if the head moves 5 cm forwards or backwards, crosstalk cancellation of 20 dB is still achieved.[2] Hence, we will only consider robustness to sideways head movement in this paper.

2. PROBLEM FORMULATION

We formulate crosstalk cancellation as a beamforming problem as follows. Consider the response to the p_R signal only, i.e., $p_L = 0$.[3] From Fig. 1(a) we note that the system reduces to a simple two-channel beamformer with filters h_1 and h_2 on the respective channels. To simplify development we consider only a single frequency, and hence, we assume that the filters h_1 and h_2 consist of only a single tap each. Referring to Fig. 1(b), let $a_n^R(x), n = 1, 2$ denote the TF from the nth loudspeaker to the right ear with the head at a position $(x, 0)$, and similarly for $a_n^L(x)$ and the left ear. Without a loss of generality, we assume the default head position is $(x, y) = (0, 0)$.

For the head at a position $(x, 0)$, the spatial response to the right ear is:

$$b_R(x) = \sum_{n=1,2} h_n \, a_n^R(x),$$

and similarly for the left ear

$$b_L(x) = \sum_{n=1,2} h_n \, a_n^L(x).$$

Let $\mathbf{h} = [h_1, h_2]^H$, $\mathbf{a}_R(x) = [a_1^R(x), a_2^R(x)]^T$, and $\mathbf{a}_L(x) = [a_1^L(x), a_2^L(x)]^T$. Hence, we can write the spatial responses as

$$b_R(x) = \mathbf{h}^H \mathbf{a}_R(x) \tag{2a}$$

$$b_L(x) = \mathbf{h}^H \mathbf{a}_L(x). \tag{2b}$$

[2] Note that the region of cancellation is slightly skewed, since we are looking at the cancellation at the left ear. If we were interested in the amount of cancellation achieved at the right ear, the cancellation region would be reflected about $x = 0$.

[3] Because of the symmetry of the problem, a similar solution exists for the p_L signal. Without a loss of generality we will only consider the right program signal here.

Considering the response to the p_R signal only, the crosstalk cancellation problem reduces to finding the weights $\mathbf{h}$ such that

$$b_R(0) = 1 \tag{3a}$$

$$b_L(0) = 0, \tag{3b}$$

where $x = 0$ is the default head position for crosstalk cancellation. Let $\mathbf{A} = [\mathbf{a}_R(0), \mathbf{a}_L(0)]$ and $\mathbf{b} = [1, 0]$. The problem (3) now reduces to

$$\mathbf{h}^H \mathbf{A} = \mathbf{b}, \tag{4}$$

which has a solution

$$\mathbf{h} = \left[\mathbf{A}^H\right]^{-1} \mathbf{b}^H. \tag{5}$$

For a given loudspeaker spacing, (5) gives the appropriate weights to achieve crosstalk cancellation for the default head position $x = 0$. The following question naturally arises: *How robust is the crosstalk canceler to slight movements in head position?* From beamforming theory it is well known that the spatial response, and thus the robustness to head movement, is a physical property of the loudspeaker spacing. Hence, a related question arises: *What is the optimum loudspeaker spacing such that the crosstalk canceler is robust to head movements?* This is the specific question we address in this paper.

3. OPTIMUM LOUDSPEAKER SPACING

Define the following cost functions for the right and left ears respectively:

$$J_R = \int_{\chi} |1 - b_R(x)|^2 \; dx \tag{6a}$$

$$J_L = \int_{\chi} |b_L(x)|^2 \; dx, \tag{6b}$$

where χ is a region surrounding $x = 0$. Hence, J_R measures the variation in the right program signal reaching the right ear as the head moves about the default position, i.e., how much distortion is introduced. Similarly, J_L measures how much of the right program signal reaches the left ear as the head moves, i.e., how effective is the cancellation. Define the combined cost function:

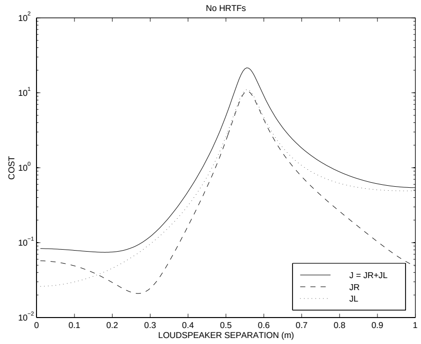
$$J = \alpha_R \, J_R + \alpha_L \, J_L, \tag{7}$$

where α_R and α_L are constants that trade off signal distortion for cancellation performance. Using this cost function we can evaluate the robustness of a crosstalk canceler designed using (5) for various loudspeaker spacings.

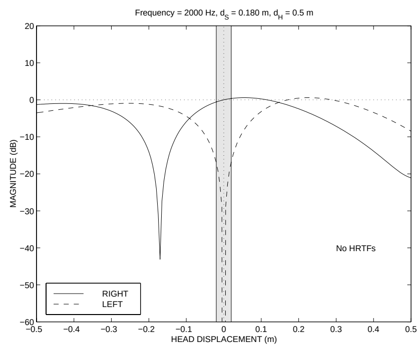
Unfortunately, the spatial responses $b_R(x)$ and $b_L(x)$ (and hence the cost function J) are highly-nonlinear functions of loudspeaker spacing, and no analytical solution is possible. We therefore propose the following numerical procedure:

Proposed Method:

1. *Given a set $D_S = \{d_k\}, k = 1, \ldots, K$, of K loudspeaker spacings which we want to test.*
2. *For each value in D_S, use (5) to calculate the appropriate weights to give crosstalk cancellation.*
3. *For each crosstalk canceler, calculate the robustness cost (7).*
4. *The optimum loudspeaker spacing corresponds to the value of D_S for which the robustness cost J is minimized.*



(a)



(b)

Figure 3: Robustness analysis at a frequency of 2 kHz and a head distance of $d_H = 0.5$ m. (a) Variation of robustness cost with loudspeaker separation. (b) Spatial responses at each ear using optimum loudspeaker separation.

4. SIMULATIONS

4.1. Optimum Loudspeaker Spacing at a Single Frequency

Consider the optimum loudspeaker spacing at a frequency of 2 kHz, with $d_H = 0.5$ m, and $r_H = 0.0875$ m (which is the commonly cited radius of the average adult human head). The default head position is $(x, y) = (0, 0)$, and we will evaluate the robustness of the crosstalk canceler within a region $\chi = \pm 2$ cm. Initially, we will assume very simple TFs consisting only of attenuation and delay, i.e., no HRTFs are included.

The variation in robustness cost (7) is shown in Fig. 3(a) as a function of loudspeaker spacing, and the spatial responses at the optimum spacing of $d_S = 0.18$ m are shown in Fig. 3(b), in which the shaded region corresponds to $x \in \chi$. Note that, as required, the right ear signal is unity (0 dB) for the default head position $x = 0$, and the left ear signal is zero at this position.

The variation of robustness cost with loudspeaker spacing can be explained as follows. In the formulation of the crosstalk cancellation problem (4), all degrees of freedom (i.e., the filters h_1 and h_2) are used in imposing the constraints (3). However, it is a physical property of beamformers that beamwidth is inversely proportional to element spacing. Thus, as d_S varies and the beamwidth in Fig. 3(b) varies, the position of the beam will move to counteract the beamwidth variation and yet still maintain the constraints. This is demonstrated in Fig. 4 (over page), which shows the spatial responses for several different loudspeaker spacings. In each case the weights $\mathbf{h}$ were calculated from (5). At $d_S = 0.1$ m (see Fig. 4(a)), the beam is very spread out, resulting in a wide beam and a corresponding large cancellation region. As d_S increases (see Fig. 4(b)), the beam narrows (as does the cancellation region) and the position of the beam peak moves. This narrowing and movement of the beam continues until $d_S = 0.555$ m (see Fig. 4(c)) when the beam is precisely wide enough such that both ear positions fall between peaks of the spatial response. At this loudspeaker spacing, any slight movement in head position causes major variation in spatial response, thereby producing large peaks in the robustness measures, J_R and J_L. This corresponds to the large peak in Fig. 3(a). As an aside, the large cancellation region shown in Fig. 4(a) is the basis of the "stereo-dipole", which consists of two closely-spaced loudspeakers [3].

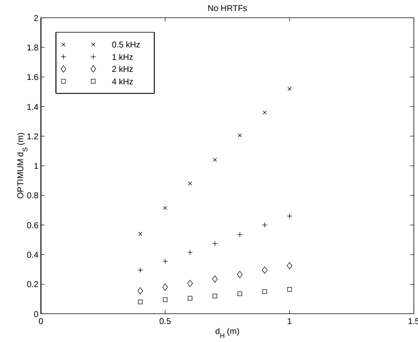


Figure 5: Variation of optimum loudspeaker spacing with head distance, at different frequencies.

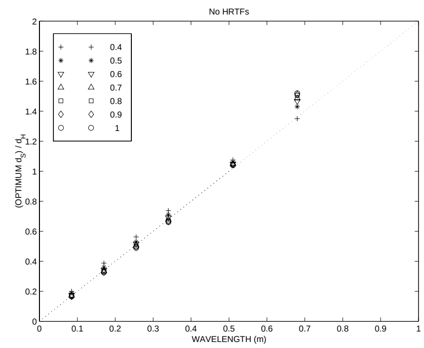
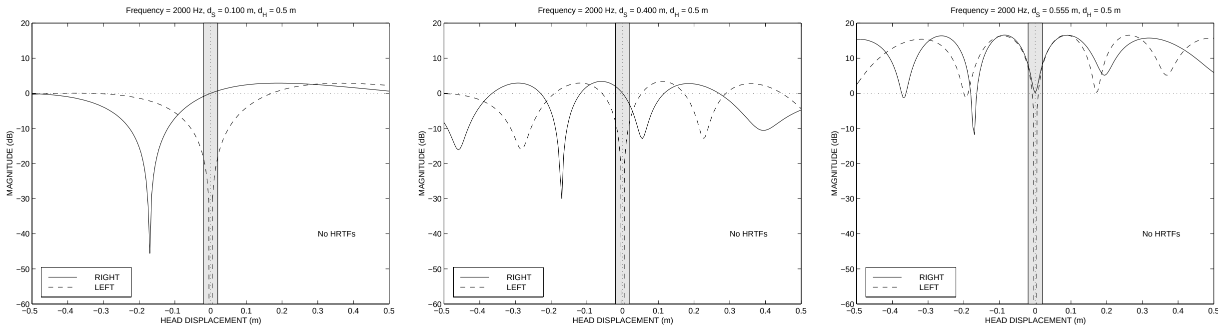


Figure 6: Ratio of optimum loudspeaker spacing to head distance, as a function of wavelength, for various head distances.

In summary, it is clear from Fig. 3(a) that there is an optimum loudspeaker spacing that will give good robustness to head movement, but there are also loudspeaker spacings (corresponding to the large peak in Fig. 3(a)) which will be extremely non-robust and should be avoided. Note that spacings less than the optimum will also provide good robustness, although implementation problems will inevitably arise for extremely close loudspeaker spacings.

4.2. Variation with Frequency

Just as beamwidth is dependent on element separation, it is also dependent on frequency. Hence, it is instructive to repeat the above experiment for different frequencies. We evaluated the robustness of the crosstalk canceler at several frequencies for several different values of d_H. The results of this investigation are shown in



(a) $d_S = 0.1$ m (b) $d_S = 0.4$ m (c) $d_S = 0.555$ m

Figure 4: Spatial responses at each ear for various loudspeaker spacings, at a frequency of 2 kHz and with a head distance of $d_H = 0.5$ m.

Fig. 5, plotted as the optimum value of d_S versus d_H for various frequencies.

Observe that, as one might expect, there is an almost linear variation of d_S with d_H, and that the slope of each curve is inversely proportional to frequency. This effect is more apparent if we re-plot the data, as shown in Fig. 6. The results are displayed as the ratio of the optimum d_S to d_H, versus wavelength (assuming a wave propagation speed of $c = 340$ m/s). A separate set of points is drawn for each value of d_H. We note that, except at large wavelengths (or frequencies below about 600 Hz), there is a linear relationship between wavelength and d_S/d_H, which is given by:

$$\frac{d_S}{d_H} = 2\lambda, \tag{8}$$

and is shown dotted in Fig. 6.

4.3. Effect of HRTFs

For all the results shown so far, we have used a simple delay and attenuation model for the TFs between the loudspeakers and the ear positions. Naturally, in a real environment we must consider the effect of the head, and so we now include HRTFs. We use a simple HRTF model proposed by Brown and Duda [4]. This model uses a pole-zero TF to approximate the effect of head shadow, based on calculations made for an ideal sphere. Although the model is very simple, it does have the basic features necessary to approximate a real HRTF. Results from this model are shown in Fig. 7, which should be compared with Fig. 6 for the delay-attenuation model. Again we note that (except at large wavelengths) the ratio of optimum d_S to d_H is largely independent of d_H. The relationship $d_S = 2\lambda d_H$ is shown dotted in Fig. 7, and again provides a good approximation for the optimum loudspeaker spacing for frequencies above about 600 Hz.

We have found that the results are reasonably independent of the width of cancellation region χ used to calculate (7). Hence, equation (8) will give a good approximation of the optimum loudspeaker spacing for most practical situations.

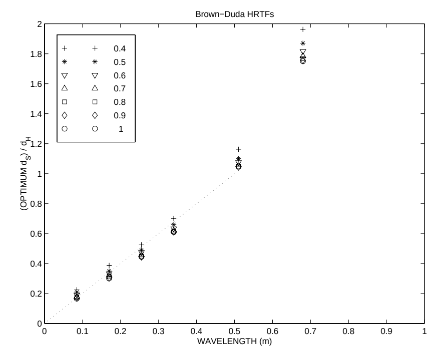


Figure 7: As in Fig. 6, but including HRTFs.

5. CONCLUSIONS

We have presented an evaluation of the robustness of crosstalk cancelers for various loudspeaker spacings. From this investigation, we noted that a good rule-of-thumb for the optimum loudspeaker spacing is given by $d_S = 2\ \lambda\ d_H$, where d_H is the distance of the head from the loudspeaker center-line (see Fig. 1), and λ is the wavelength of operation.

For broadband signals, we would ideally like the loudspeaker spacing to vary with frequency. This suggests use of a loudspeaker array, with frequency selective filters on each loudspeaker which select an appropriate loudspeaker spacing for each frequency, analogous to the design of frequency independent sensor arrays [5]. Alternately, a more practical solution is to simply use the optimum spacing corresponding to the upper frequency of interest.

6. REFERENCES

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

[2] F. Asano, Y. Suzuki, and T. Sone, "Sound equalization using derivative constraints", *Acustica*, vol. 82, pp. 311–320, 1996.

[3] Y. Watanabe, H. Tokuno, H. Hamada, O. Kirkeby, and P.A. Nelson, "Subjective investigation of new sound reproduction system (stereo dipole)", in *Proc. 3rd Joint Meeting, Acoust. Soc. Amer., Acoust. Soc. Japan*, Hawaii, 1996, pp. 307–310.

[4] C.P. Brown and R.O. Duda, "A structural model for binaural sound synthesis", *IEEE Trans. Speech Audio Processing*, (in press).

[5] D.B. Ward, R.A. Kennedy, and R.C. Williamson, "Theory and design of broadband sensor arrays with frequency invariant far-field beam patterns", *J. Acoust. Soc. Amer.*, vol. 97, no. 2, pp. 1023–1034, Feb. 1995.