Crosstalk Cancellation

Tim Nufire
7700 Geary Boulevard, #210
San Francisco, CA 94121
415-752-9313

The purpose of this paper is to give the reader a high level understanding of how crosstalk cancellation applies to audio signal processing. After a brief introduction in which I discuss the historical and personal motivations for doing audio crosstalk cancellation, I will present an overview of the problem and then focus on its DSP aspects.

INTRODUCTION

The topic of crosstalk cancellation was first explored by B. S. Atal and M. R. Schroeder[1 2] and has been extensively researched by D. H. Cooper and J. L. Bauck in numerous works.[3] Put simply, the goal of crosstalk cancellation is, given a listener and a couple of loudspeakers, to deliver an audio signal from the right speaker to the right ear and from the left speaker to the left ear while eliminating the *crosstalk* (i.e. the signal from the left speaker to the right ear and vice versa). This is desirable because it allows binaural recordings[4] to be played back via loudspeakers rather than

---

1 B. S. Atal and M. R. Schroeder, "Apparent Sound Source Translator," U.S. patent 3,236,949 (1966 Feb. 22).

2 M. R. Schroeder and B. S. Atal, "Computer Simulation of Sound Transmission in Roms," *IEEE Conv. Rec.*, pt. 7, pp. 150-155 (1963)

3 Duane H. Cooper and Jerald L. Bauck, "Prospects for Transaural Recording," J. Audio Eng. Soc., Vol. 37, No. 1/2, 1989 Jan./Feb.

4 Binaural recordings are those which contain signals representing the sound waves actually present at each ear during a performance and can be created either through the use of a "dummy head" with microphones for ears or they can be synthesized on a computer. In either case, the advantage of binaural recordings is that they reproduce the spatial characteristics of the material much better than traditional stereo; however, this effect is significantly degraded by crosstalk.

being restricted to headphones.

Crosstalk cancellation techniques can also be used to *enhance* traditional stereo recordings. This can be done in two ways: 1) by producing the extreme stereo separation experienced with headphones, and 2) with the addition of binaural synthesis techniques, *virtual* speakers that are farther apart than the physical speakers can be simulated. The latter technique is particularly useful for speakers like those mounted on a computer display or a TV set because these speakers are typically very close together.

Although this paper focuses on a single listener, Cooper and Bauck have worked out a generalized crosstalk cancellation scheme for the case where there are many listeners. For example, by eliminating the crosstalk *between* listeners in a conference room setting, separate audio tracks can be delivered to each participant. Using this technique, Cooper and Bauck suggest that it is possible to simultaneously broadcast several languages at once (without the need for headphones) so that each participant can listen to the proceedings in their native tongue.

OVERVIEW OF CROSSTALK CANCELLATION

Fig. 1 shows the general from of the crosstalk cancellation array. In this array, C(z) can be thought of as the crosstalk canceler and E(z) as an equalizer. Since we know (or can measure) the impulse responses s[n] and a[n], it is desirable to define the filters C(z) and

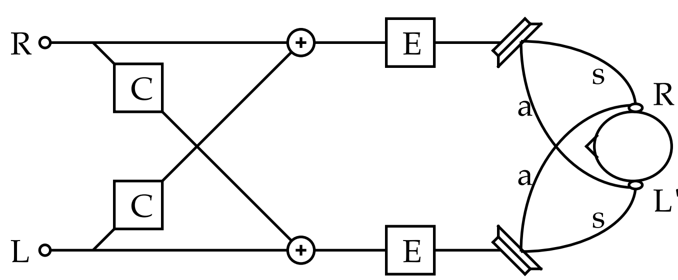

Figure 1: Crosstalk cancellation array

E(z) in terms of S(z) and A(z), the z-transforms of our measured impulse responses. From Fig. 1 we can see that

$$R'(z) = R(z)[E(z)S(z) + C(z)E(z)A(z)] + L(z)[E(z)A(z) + C(z)E(z)S(z)] \tag{1a}$$

$$R'(z) = [R(z)[S(z) + C(z)A(z)] + L(z)[A(z) + C(z)S(z)]]E(z) \tag{1b}$$

Effective crosstalk cancellation requires that R'(z) depend only on R(z). From Eq. (1b) and this requirement we get

$$A(z) + C(z)S(z) = 0 \tag{2a}$$

$$C(z) = -\frac{A(z)}{S(z)} \tag{2b}$$

For good equalization we require that R'(z) be equal to R(z). From Eq. (1b) using this requirement we get

$$[S(z) + C(z)A(z)]E(z) = \left[S(z) + \left(-\frac{A(z)}{S(z)}\right)A(z)\right]E(z) = 1 \tag{3a}$$

$$E(z) = \frac{S(z)}{S^2(z) - A^2(z)} = \frac{1}{S(z)} \cdot \frac{1}{1 - C^2(z)} = E_1(z) \cdot E_2(z) \tag{3b}$$

$$\text{where } E_1(z) = \frac{1}{S(z)} \text{ and } E_2(z) = \frac{1}{1 - C^2(z)} \tag{3c \& 3d}$$

Eq. (2b) and Eq. (3b), taken together, lay the theoretical foundation for crosstalk cancellation. The equalization filter in Eq. (3b) is typically broken up into two filters, one corresponding to Eq. (3c) which cancels the effect of placing the two speakers in front of the listener rather than up against their ears, and the other corresponding to Eq. (3d) which corrects for the distortion introduced by the crosstalk cancellation filter C(z).

FILTER DESIGN CONSIDERATIONS

Using experimental data for a[n] and s[n] like that show in Fig. 2, we can model the filter C(z) as the rational polynomial P(z)/Q(z). One can think of this

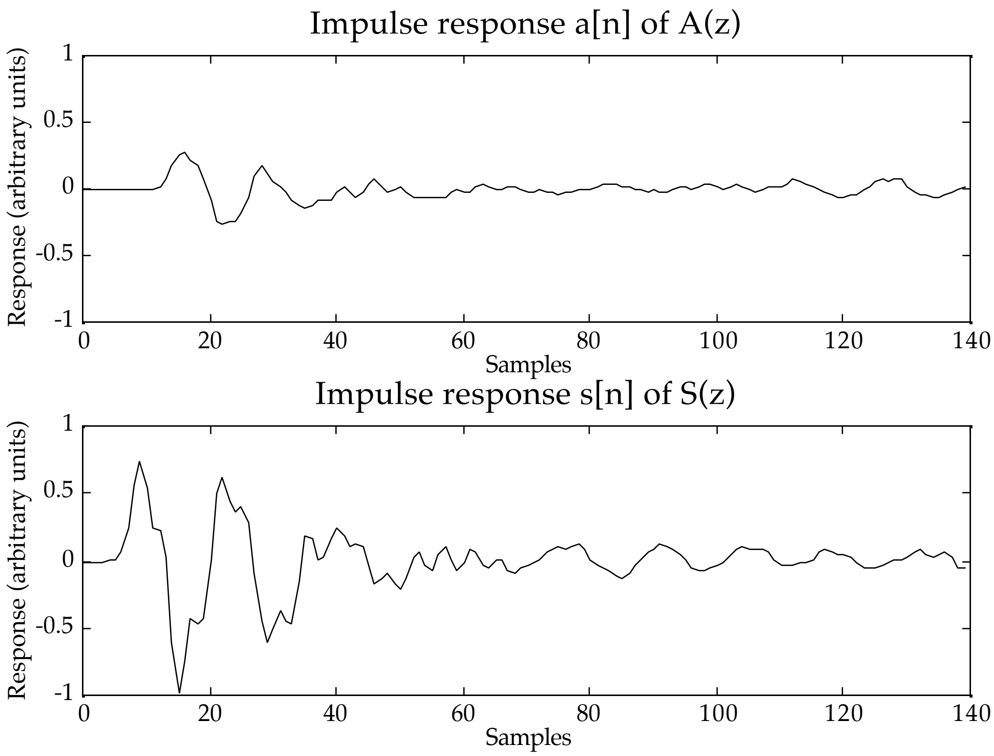

Figure 2: Experimental data for a[n] & s[n]

model as identifying a system which, given the sequence s[n] as its input, will produce a[n] as its output. Once identified, this model can be used to realize Eq. (3d) by placing it, cascaded with itself, in a positive feedback loop as shown in Fig. 3. It should be noted that, since we are dealing with discrete time systems, this realization is only possible if C(z) has at least a one sample delay. Although this may seem like a significant limitation, from the geometry of Fig. 1 it is clear that C(z) will naturally contain some delay because the "a" path is longer than the "s" path.

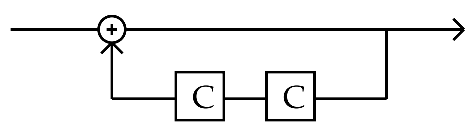

Figure 3: $E_2(z)$ realization

The most important constraint on the design of C(z) is stability. Because C(z) is used to realize Eq. (3d), ensuring the stability of C(z) alone is not enough – our choice of C(z) must also ensure the stability of $E_2(z)$. To understand how this constrains the design of C(z), recall that a discrete time system will be stable if and only if all of its poles are all within the unit circle in the z-plane. This means that, in order for C(z) to be stable, we must put the following constraint on Q(z), the denominator of C(z)

$$|\text{roots of } Q(z)| < 1 \qquad (4)$$

In addition, requiring the poles of $E_2(z)$ to be within the unit circle further constrains our choice of C(z). Specifically, by substituting $P(z)/Q(z)$ for C(z) in Eq. (3d) we have

$$E_2(z) = \frac{1}{1 - C^2(z)} = \frac{1}{1 - \left(\frac{P(z)}{Q(z)}\right)^2} \qquad (5a)$$

$$E_2(z) = \frac{Q^2(z)}{Q^2(z) - P^2(z)} \qquad (5b)$$

Forcing the poles of Eq. (5b) to be inside the unit circle requires

$$|\text{roots of } [Q^2(z) - P^2(z)]| < 1 \qquad (6)$$

This result, combined with Eq. (4), represents the necessary (and sufficient) stability constraints on our model.

Although not necessary, a stricter constraint can be placed on C(z) by requiring that

$$|C(e^{j\omega})| < 1 \qquad \text{for all } \omega \qquad (7)$$

Taking a step back and looking at the problem that we are trying to solve, it is clear why this additional requirement might make sense. Since C(z) is a model of the ratio of A(z) to S(z), Eq. (7) is equivalent to assuming that the magnitude frequency response $|S(e^{j\omega})|$ is greater than $|A(e^{j\omega})|$ for all frequencies. From the geometry of Fig. 1 we can see that this assumption makes sense because the left ear is closer to (and has a more direct path to) the left speaker than it does to the right. As a result, it is reasonable to assume that the *volume* from the left speaker will be greater than that from the right across all frequencies.

Returning again to the issue of stability, it can be shown that if Eq. (7) and Eq. (4) are satisfied then Eq. (6) will be as well. In other words, Eq. (7) can be used instead of Eq. (6) as a sufficient (but not necessary) requirement of stability. To see why this is true, note that, because we have required C(z) to have at least a one sample delay, the order of P(z) must be less than that of Q(z). From this it follows that

$$\lim_{|z|\to\infty} |M(z)| = \lim_{|z|\to\infty} \left|\frac{P(z)}{Q(z)}\right| = 0 \qquad (8)$$

To see why this is important, it is useful to rewrite Eq. (6) (the stability requirement of $E_2(z)$) as

$$C(z) \neq \pm 1 \text{ for } |z| > 1 \qquad (9)$$

which follows directly from Eq. (5a).

Now, given that Eq. (7) is satisfied, we know that $|C(z)|$ is less than 1 on the unit circle. From Eq. (9) we see that in order for Eq. (6) not to be true, $|C(z)|$ would have to grow back up to 1 at some point outside the unit circle. Although I will not present a rigorous proof, Eq. (8) gives a sense for why this is not possible. Because Q(z) grows faster than P(z), once $|C(z)|$ drops below 1 for all z on a contour of constant radius R around the origin (and within the region of convergence of C(z)) it will remain less than 1 for all $|z| > R$.

Another consideration in the design of C(z) has to do with how it behaves at high frequencies (those above about 10 kHz). Although it is possible to create a model that performs effective crosstalk cancellation at these frequencies, according to Cooper and Bauck this is not desirable.[5] This is because at high frequencies the wavelengths are small relative to the human head and, as a result, slight head movements will result in severe phase effects and a very narrow "sweet spot."

In modeling C(z) we have taken care of Eq. (2b) and Eq. (3d) – two thirds of the problem is solved! This leaves Eq. (3c), the $E_2(z) = 1/S(z)$ term in E(z). Strictly speaking, this term is not causal because S(z) contains delay and, as a result, a real-time model of Eq. (3c) will need to be padded with enough delay to force causality. Aside from this requirement, however, there are not as many constraints on our model of $E_2(z)$ as there were on our model of C(z). In fact, the only significant constraint is stability and, assuming that $E_2(z)$ is modeled as the rational polynomial $U(z)/V(z)$, this constraint boils down to requiring that

$$|\text{roots of } V(z)| < 1 \tag{10}$$

An alternative to modeling $E_2(z)$ directly is to model S(z) as the rational polynomial $V'(z)/U'(z)$ and then invert to get $E_2(z) = 1/S(z) = U'(z)/V'(z)$. With this approach, we need to require that our model be minimum phase. That is

$$|\text{roots of } V'(z)| < 1 \tag{11}$$

This requirement arises because only minimum phase filters are invertible.

---

5 Cooper et al., "Head diffraction Compensated Stereo System," U.S. patent 4,893,342 (1990 Jan. 9).

SUMMARY

In this paper I have presented an overview of crosstalk cancellation and have discussed some of the constraints on the filters in a crosstalk cancellation array. This has not been an exhaustive look at the subject and there are many additional practical concerns like performance that were not covered. In addition, I glossed over the most difficult task in developing a crosstalk canceler – that of actually creating the models for C(z) and $E_2(z)$. Issues like these were beyond the scope of this paper because they depend on the requirements of a particular application and on the tools available to aid the development process. I instead focused on the theoretical foundations for crosstalk cancellation and detailed the stability constraints on our models. The intent was to provide tools to aid understanding and evaluation in the hope that this would lead to better models regardless of the particular application.