SELECTIVE SIGNAL CANCELLATION FOR MULTIPLE-LISTENER AUDIO APPLICATIONS: AN INFORMATION THEORY APPROACH

Sunil Bharitkar and Chris Kyriakakis

Immersive Audio Laboratory–Integrated Media Systems Center
University of Southern California
Los Angeles, CA 90089-2564

ABSTRACT

Selectively canceling signals at specific locations within an acoustical environment with multiple listeners is of significant importance for home theater, teleconferencing, office, industrial and other applications. The traditional noise cancellation approach is impractical for such applications because it requires sensors that must be placed on the listeners. In this paper we propose an alternative method to minimize signal power in a given location and maximize signal power in another location of interest. A key advantage of this approach would be the need to eliminate sensors. We investigate the use of an information theoretic criterion known as *mutual information* to design filter coefficients that selectively cancel a signal in one audio channel, and transmit it in another (complementary) channel. Our results show an improvement in power gain at one location in the room relative to the other.

1. INTRODUCTION

Selective signal cancellation is required in applications that require a signal of interest to be enhanced while minimizing the effects of noise or other signals. For example, in home theater or television viewing applications a listener in a specific position in a room may not want to listen to the audio signal being transmitted, while another listener at a different position would prefer to listen to the signal. Consequently, if the objective is to keep one listener in a region with a reduced sound pressure level, then one can view this problem as that of signal cancellation in the direction of that listener. Similar applications arise in the automobile or any other environment with multiple listeners in which only a subset wish to listen to the audio signal.

In this paper we investigate the application of mutual information as an optimization criterion to selectively cancel an audio signal in a specific direction (also known as the channel of interest), while leaving it unaltered in another direction. We shall consider an initial case in which two listeners (initially modeled as point receivers) are present in two arbitrary locations in the room. We then derive a set of optimal filter coefficients to achieve selective signal cancellation. In the next section, we briefly discuss some background on information theoretic models. In Section 3, we derive the equations for determining the set of tap weights which would guarantee selective signal cancellation under simplified gaussian noise assumptions. The choice of the tap weights are strongly determined by the channel impulse responses between the transmitter and the two listeners (receivers). In Section 4 we address the results obtained on using this approach and propose some relevant future directions. Section 5 concludes the paper.

This research has been funded by the Integrated Media Systems Center, a National Science Foundation Engineering Research Center, Cooperative Agreement No. EEC-9529152.

IEEE Intl. Conf. Multimedia & Expo 2000,New York.©ICME2000

2. INFORMATION THEORY MODELS

Mutual information $I(x, y)$ (MI) measures arbitrary dependencies between two random variables x, and y with marginal distributions denoted by $f(x)$, and $f(y)$, and their joint distribution $f(x, y)$ (where x may be considered as an input to a channel, and y is the corresponding output). The general form for this measure is,

$$I(x,y) = \sum_x \sum_y f(x,y) \log \frac{f(x,y)}{f(x)f(y)} \tag{1}$$

$I(x, y) \geq 0$, with equality being achieved on general statistical independence between x, and y, and $I(x, y) = I(y, x)$. The computation of this measure is not an easy task, due to the involvement of complicated density functions. However, there is a method of evaluating this measure based on samples of input-output data (x, y), using Fraser's algorithm of mutual information estimation [1]. For the gaussian case the computation of this measure is a simpler task as can be seen with the aid of the following example. Let

$$y(n) = \sum_{k=1}^{P} h(k)x(n-k) + v(n) \tag{2}$$

where $v(n)$ is AWGN (additive white gaussian noise of zero mean) on a simple linear channel $h(n)$. From the entropy based definition of MI, it is well established that

$$I(y(n), x(n)) = \frac{1}{2}\log(\frac{\sigma_y^2(n)}{\sigma_v^2(n)}) \quad (3)$$

The major advantage on using MI is its capability to measure arbitrary general dependence between two variables. There are some distinct advantages on using this type of a measure over the correlation measure used in steepest descent algorithms (on which LMS is based on), where the correlation measure to be minimized is given by,

$$J(n) = R_e(0) = E\{|e(n)|^2\} \quad (4)$$

A detailed investigation of the advantages of MI over correlation is contained in [2],[3].

3. DETERMINATION OF THE OPTIMAL WEIGHTS

Since we are not concerned with source localization and associated head-related transfer functions (HRTF's), but rather with signal minimization at a single point, we can consider the simple model with reference to Fig. 1, where w_k represents the coefficients of the filter we would like to design based on the MI criteria. For the current problem, we assume that the receivers are stationary (i.e., the room impulse response for a certain (S, R) is time invariant and linear, where S, and R, represent a source and a receiver), and the channel (room) impulse response is deterministic at the location of the two listeners. We further assume that the listeners are modeled as point receivers. The listening model is then simply given by (2), where $h(n)$ is the impulse response for a given source-receiver position.

With this background, we can state the performance criteria as,

$$J(n) = max_w \quad \frac{1}{2}\log(\frac{\sigma^2_{y_2(n)}}{\sigma^2_{v_2(n)}}) - \lambda(\frac{\sigma^2_{y_1(n)}}{\sigma^2_{v_1(n)}} - \psi) \quad (5)$$

where, we would like to maximize the signal in the direction of listener 2, while retaining the power towards listener 1 at least $10^{\psi_{dB}/10}$. We can simplify the computation for the optimal filter coefficients $\underline{w}$, by recognizing the monotonicity of the $log(x)$ function over the domain $x \in [1, \infty)$. In other words, minimizing (maximizing) $log(x)$ implies minimizing (maximizing) x. Hence, the objective function (5) can be recast as,

$$J(n) = max_w \quad \frac{1}{2}(\frac{\sigma^2_{y_2(n)}}{\sigma^2_{v_2(n)}}) - \lambda(\frac{\sigma^2_{y_1(n)}}{\sigma^2_{v_1(n)}} - \psi) \quad (6)$$

Now observe that,

$$y_1(n) = h_1(n) \star \sum_{k=0}^{M-1} w_k x(n-k) + v_1(n) \quad (7)$$

where, $h_1(n)$ is the room response in the direction for listener labeled 1, and $\star$ denotes the linear convolution operator. Let $\underline{w} = (w_0, w_1, ..., w_{M-1})^T$, and $\underline{x}(n) = (x(n), x(n-1), .., x(n-M+1))^T$, then (7) can be expressed as,

$$\begin{aligned} y_1(n) &= h_1(n) \star \underline{w}^T \underline{x}(n) + v_1(n) \\ &= h_1(n) \star z(n) + v_1(n) \\ &= \sum_{p=0}^{L-1} h_1(p) z(n-p) + v_1(n) \end{aligned} \quad (8)$$

where, $z(n) = \underline{w}^T \underline{x}(n)$. We assume that the zero mean noise and signal are statistically independent (and uncorrelated in the gaussian case). In this case signal power in the direction of listener 1 is,

$$\begin{aligned} \sigma^2_{y_1(n)} &= E\{\sum_{p=0}^{L-1}\sum_{q=0}^{L-1} h_1(p)h_1(q) z(n-p) z^T(n-q)\} \\ &+ \sigma^2_{v_1(n)} \\ &= \sum_{p=0}^{L-1}\sum_{q=0}^{L-1} h_1(p)h_1(q)(\underline{w}^T \mathbf{R}_{\underline{x}}(p,q)\underline{w}) \\ &+ \sigma^2_{v_1(n)} \end{aligned} \quad (9)$$

where, $\underline{w} \in \Re^M$, $\mathbf{R}_{\underline{x}}(p,q) \in \Re^{M \times M}$. Similarly,

$$\sigma^2_{y_2(n)} = \sum_{p=0}^{S-1}\sum_{q=0}^{S-1} h_2(p)h_2(q)(\underline{w}^T \mathbf{R}_{\underline{x}}(p,q)\underline{w}) + \sigma^2_{v_2(n)} \quad (10)$$

Solving $\nabla_{\underline{w}} J(n) = 0$ will provide the set of optimal tap coefficients. Hence from (6), (11), and (10), we obtain,

$$\begin{aligned} \frac{\partial J(n)}{\partial \underline{w}} &= \frac{1}{\sigma^2_{v_2(n)}} \sum_{p=0}^{S-1}\sum_{q=0}^{S-1} h_2(p)h_2(q)\mathbf{R}_{\underline{x}}(p,q)\underline{w}^* \\ &- \frac{\lambda}{\sigma^2_{v_1(n)}} \sum_{p=0}^{L-1}\sum_{q=0}^{L-1} h_1(p)h_1(q)\mathbf{R}_{\underline{x}}(p,q)\underline{w}^* = 0; \end{aligned} \quad (11)$$

where $\underline{w}^*$ denotes the optimal coefficients. Let,

$$\begin{aligned} A &= \sum_{p=0}^{S-1}\sum_{q=0}^{S-1} h_2(p)h_2(q)\mathbf{R}_{\underline{x}}(p,q) \\ B &= \sum_{p=0}^{L-1}\sum_{q=0}^{L-1} h_1(p)h_1(q)\mathbf{R}_{\underline{x}}(p,q) \end{aligned} \quad (12)$$

By assuming equal ambient noise powers at the two receivers (i.e., $\sigma^2_{v_2(n)} = \sigma^2_{v_1(n)}$), (11) can be written as

$$\frac{\partial J(n)}{\partial \underline{w}} \mid_{\underline{w}=\underline{w}^*} = (B^{-1}A - \lambda I)\underline{w}^* = 0 \qquad (13)$$

The reason for arranging the optimality condition in this fashion is to demonstrate that the maximization is in the form of an eigenvalue problem, (i.e., the eigenvalues corresponding to the matrix $B^{-1}A$), with the eigenvectors being $\underline{w}^*$. There are in general M distinct eigenvalues for the $M \times M$ matrix - $B^{-1}A$, with the largest eigenvalue corresponding to the maximization of the ratio of the signal powers between receiver 2 and receiver 1. The optimal filter that yields this maximization is given by,

$$\underline{w}^* = \underline{e}_{\lambda_{max}[B^{-1}A]} \qquad (14)$$

where, $\underline{e}_{\lambda_{max}[B^{-1}A]}$ denotes the eigenvector corresponding to the maximum eigenvalue λ_{max} of $B^{-1}A$. An FIR filter whose impulse response corresponds to the elements of an eigenvector is called an *eigenfilter* [4],[5].

Clearly it can be seen from (14) that, the optimal filter coefficients are determined by the channel responses between the source and the two listeners. We are currently devising adaptive versions of this filter that will be presented in forthcoming papers. In the following section we test the design of this filter for a stochastic model and show the performance improvement.

4. RESULTS

In this section we generate a stochastic autoregressive model of order one, denoted as $AR(1)$. This process is generated by filtering white noise with a linear time-invariant filter with a rational system function [6].

$$x(n) = 0.3x(n-1) + 0.7t(n) \qquad (15)$$

where, the $t(n)$ describes a white noise process of zero mean and unit variance. It is well established from the Yule-Walker equations (with $x(n)$ and $t(n)$ being wide-sense stationary) that the correlation function for (15) satisfies the following relation,

$$r_x(k) = \frac{0.49}{0.91}(-0.3)^{|k|} \quad k = .., -1, 0, 1, .. \qquad (16)$$

The optimal filter coefficients $\underline{w}^*$ for the system described by (15) can be found by using (14), with A, and B as given in (12), with $\mathbf{R}_{\underline{\mathbf{x}}}(p,q)$ being a symmetric(asymmetric) Toeplitz matrix for $p = q$ ($p \neq q$)-containing the correlation function given in (16).

The performance measure (expressed as the ratio of average signal powers between receiver 2 and receiver 1) for the statistical model is given as,

$$G_{dB} = 10 \log_{10} \frac{\underline{w}^{*T} A \underline{w}^*}{\underline{w}^{*T} B \underline{w}^*} \qquad (17)$$

We also define a performance measure for the individual realizations obtained from (15). Let

$$P_i^{\kappa}(n) = (h_i(n) \star w_n^* \star x(n))^2$$
$$i = 1, 2; \kappa = 0, 1, 2, ..., N-1 \qquad (18)$$

where, κ denotes a realization for (15). Then the sample based average ratio for the signal powers is given by,

$$\widehat{G}_{dB} = 10 \log_{10} \frac{\overline{P}_2(n)}{\overline{P}_1(n)} \qquad (19)$$

$$\overline{P}_i(n) = \frac{1}{N} \sum_{\kappa=0}^{N-1} P_i^{\kappa}(n) \quad i = 1, 2.$$

The impulse responses $h_1(n)$, and $h_2(n)$ (comprising of 8192 point) were obtained in a room from microphones placed at a radial distance of 3m, with azimuth angle of ± 30 degrees and elevation of 0 degrees relative to a loudspeaker, and are shown in Fig 2 (a),(b). We obtained the optimal filter for two cases, (a)$S = L = M = 100$, and (b)$S = L = M = 128$, (i.e., the duration of the impulse responses corresponded to the first $100(128)$ samples measured from the arrival along the direct path). Clearly, we need not have $S = L = M$. For the aforementioned cases we obtained the following gains,

$$G_{dB}^{(a)} \approx 4$$
$$G_{dB}^{(b)} \approx 5 \qquad (20)$$
$$G_{dB}^{\text{no filter}} = 0.03$$

In Fig. 3(a), we have shown the $\widehat{G}_{dB}^{(a)}$ for a single realization (i.e., $N = 1$ in (21)); and, $G_{dB}^{(a)}$. Fig. 3(b) displays $\widehat{G}_{dB}^{(a)}$ for $N = 100$, along with $G_{dB}^{(a)}$. In Fig. 4 (a), we display the results for $\widehat{G}_{dB}^{(b)}$ for $N = 1$, along with $G_{dB}^{(b)}$; and, in Fig. 4 (b), we have $\widehat{G}_{dB}^{(b)}$ for $N = 100$, along with $G_{dB}^{(b)}$.

By choosing a larger impulse response duration, L, M, it may be possible to increase the gain for the filtered cases, (i.e., considering the effects due to reverberation). Finally, we also plan to consider adaptive implementations of the filter to compensate the computational difficulties encountered in solving (13) when the length of the filter is large, and when the input data rate is high. We shall consider this aspect in the future.

5. CONCLUSIONS

In this paper, we propose a method for selectively cancelling signals in the presence of multiple listeners. This approach is useful in environments with differing listening requirements. The proposed method resulted in a satisfactory improvement in the objective function. However, we have not addressed the issues related to the spectral variations due to the introduction of such filters. There could be a possible tradeoff in subjective listening tests (psychoacoustical) between sound quality and signal cancellation in certain types of environments (e.g., automobiles). In forthcoming papers, we shall address these issues, and investigate other relevant objective functions, along with adaptive techniques for filter design that will include the CAPZ model [7]. We will also consider the effect of multiple transmitters which would increase the degree of freedom in the choice of the filter.

6. REFERENCES

[1] A. M. Fraser and H. L. Sweeney "Independent coordinates for strange attractors from mutual information," *Physical Review A,* vol. 33(2), pp.1134–1140, 1986.

[2] A. M. Fraser, "Reconstructing attractors from scalar time series:a comparison of singular system and redundancy criteria," *Physica D,* vol. 34, pp. 391–404, 1989.

[3] W. Li, "Mutual Information versus correlation functions," *Journal of Statistical Physics,* vol. 60(5/6), pp. 823–837, 1990.

[4] J. Makhoul, "On the eigenvectors of symmetric Toeplitz matrices," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. ASSP-29, pp. 868–872, 1981.

[5] S. Haykin, *Adaptive Filter Theory,* Prentice-Hall Inc. 1996

[6] G. U. Yule, "On a method of investigating periodicities in disturbed series, with special reference to Wölfer's sunspot numbers," *Philos. Trans. Royal Soc. London,* vol. A226, pp. 267–298, 1927.

[7] Y. Haneda and S. Makino and Y. Kaneda, "Common Acoustical Pole and Zero modeling of room transfer functions," *IEEE Transactions on Speech and Audio Proc.,* vol. 2(2), pp. 320–328, Apr. 1994.

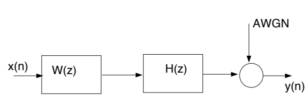

Figure 1: The source-receiver model

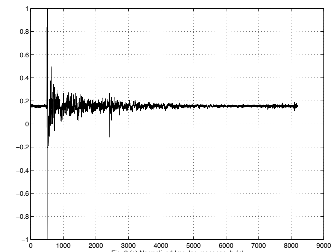

Fig. 2:(a) Normalized Impulse response, $h_1(n)$

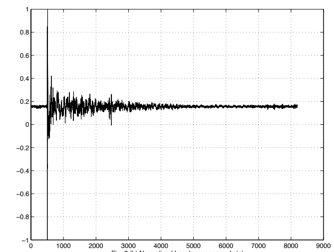

Fig. 2:(b) Normalized Impulse response, $h_2(n)$

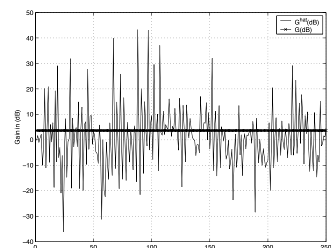

Fig. 3(a): Plot of sample gain for N=1, and average gain; M=100

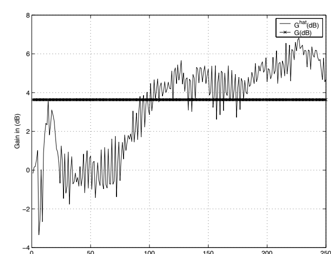

Fig. 3(b): Plot of sample gain for N=100, and average gain; M=100

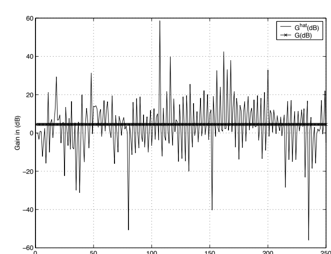

Fig. 4(a): Plot of sample gain for N=1, and average gain; M=128

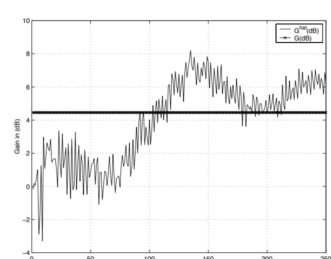

Fig. 4(b): Plot of sample gain for N=100, and average gain; M=128