

ROBUSTNESS OF THE EIGENFILTER FOR VARIATIONS IN LISTENER RESPONSES FOR SELECTIVE SIGNAL CANCELLATION

Sunil Bharitkar and Chris Kyriakakis

Immersive Audio Laboratory,
NSF-Integrated Media Systems Center,
Univ. of Southern California,
Los Angeles, CA 90089-2564
bharitka@sipi.usc.edu, ckyriak@imsc.usc.edu

Abstract: Selectively cancelling signals at specific locations within an acoustical environment with multiple listeners is of significant importance for home theater, automobile, teleconferencing, office, industrial and other applications. We have proposed the eigenfilter for selectively cancelling signals in one direction, while attempting to retain them at unintentional directions. In this paper we investigate the behaviour of the performance measure (i.e., the gain) for a vowel and an unvoiced fricative, when the listener moves his head, in an automobile type environment. We show that in such a situation, a large energy in the difference between the impulse responses at a listener's location may affect the gain substantially.

1. INTRODUCTION

Integrated media systems are envisioned to have a significant impact on the way groups of people in remote locations communicate with each other. One of the critical elements that help enhance the suspension of disbelief required to convince people that they are truly in the same environment is sound. While a great deal of ongoing research has focused on the problem of delivering high quality sound to a single listener, the problem of delivering the appropriate audio signals to multiple listeners in the same environment has not yet been adequately addressed.

In previous work [1, 2, 3] we focused on presenting an audio signal at a selected direction in a room, while simultaneously minimizing the signal power at another direction. For example, in home theater or television viewing applications a listener in a specific location in the room may not want to listen to the audio signal being transmitted, while another listener at a different location 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 (e.g., when only the driver would prefer to listen to an audio signal), or any other environment with multiple

listeners in which only a subset wish to listen to the audio signal.

An eigenfilter for selective signal cancellation is designed by optimizing an objective function as shown in Section 2. Section 3 summarizes some properties of eigenfilters for stationary signals. In Section 4 we show that the performance function is affected for certain changes in the responses (such as head movements) in a simulated automobile type environment. We confirm these results for simple speech signals, (i) an unvoiced fricative /S/ as in *gat*, (ii) a vowel /AE/ as in *bat*. We conclude this paper in Section 5, and suggest some future directions.

2. THE EIGENFILTER FOR SELECTIVE SIGNAL CANCELLATION

An objective criterion is designed for maximizing the difference in signal power between two different listener locations that have different source-receiver response characteristics. For simplicity we assume that the listeners can be modeled as point receivers. The method can also be extended to take into account ear spacing and head-related transfer function effects. The filter, known as the eigenfilter that is derived by optimizing the objective function, operates on the raw signal before the resulting signal is linearly transformed by the room responses in the direction of the listeners. Such filters aim at increasing the relative gain in signal power between the two listeners with some associated tradeoffs such as: (i) spectral distortion that may arise from the presence of the eigenfilter, and (ii) the sensitivity of the filter to the length of the room impulse response (reverberation), (iii) perceptual coloration, and (iv) sensitivity to spatial variations in the room responses (due to listener head movements). In this paper we focus on the sensitivity issue in a space that has the approximate dimensions of an automobile interior as an example of where this approach could be implemented.

2.1. Determination of the Eigenfilter

Under our assumption of modeling the listeners as point receivers we can set up the problem as shown in Fig. 1, where $w_k; k = 0, 1, \dots, M - 1$ represents the coefficients of the finite impulse response filter to be designed. During the design phase we assume that the listeners are stationary. The listening model is then simply

$$y_i(n) = h_i(n) \otimes \sum_{k=0}^{M-1} w_k x(n-k) + v_i(n) \quad i = 1, 2 \quad (1)$$

where \otimes represents the convolution operation. With this background, we view the signal cancellation problem as a gain maximization problem (between two arbitrary listeners), we can state the performance criterion as,

$$J(n) = \max_{\underline{w}} \frac{1}{2} \left(\frac{\sigma_{y_2(n)}^2}{\sigma_{v_2(n)}^2} \right) - \frac{\lambda}{2} \left(\frac{\sigma_{y_1(n)}^2}{\sigma_{v_1(n)}^2} - \psi \right) \quad (2)$$

in which we would like to maximize the signal to noise ratio (or signal power) in the direction of listener 2, while keeping the power towards listener 1 constrained at $10^{\psi_{dB}}/10$ (where $\psi_{dB} = 10 \log_{10} \psi$). In (2), $\sigma_{y_i(n)}^2 / \sigma_{v_i(n)}^2$ denotes the transmitted signal to ambient noise power at listener R_i with $y_i(n)$ as defined in (1). The quantity λ is the well known Lagrange multiplier.

It can be easily shown, under equal ambient noise, that the optimal filter, \underline{w}^* , is an eigenfilter given by

$$\begin{aligned} \underline{w}^* &= \underline{e}_{\lambda_{\max}[B^{-1}A]} \quad (3) \\ A &= \sum_{p=0}^{S-1} \sum_{q=0}^{S-1} h_2(p)h_2(q)\mathbf{R}_{\underline{x}}(p,q) \\ B &= \sum_{p=0}^{L-1} \sum_{q=0}^{L-1} h_1(p)h_1(q)\mathbf{R}_{\underline{x}}(p,q) \end{aligned}$$

where, $\underline{e}_{\lambda_{\max}[B^{-1}A]}$ denotes the eigenvector corresponding to the maximum eigenvalue λ_{\max} of $B^{-1}A$.

The performance is the gain G_{dB} expressed as,

$$\begin{aligned} G_{dB} &= 10 \log_{10} \frac{\sigma_{y_2(n)}^2}{\sigma_{y_1(n)}^2} \\ &= 10 \log_{10} \frac{\underline{w}^{*T} A \underline{w}^*}{\underline{w}^{*T} B \underline{w}^*} \quad (4) \end{aligned}$$

Fundamentally, by casting the signal cancellation problem as a gain maximization problem, we aim at introducing a large gain of Q dB between two listeners, R_1 and R_2 . This Q dB gain is equivalent to virtually positioning listener R_1 at a distance which is $\sqrt{10^{Q/10}}$ times the distance of listener R_2 from a fixed sound source*.

*Strictly speaking, in the free field, the gain based on the inverse square law, is expressed as, $Q = 10 \log_{10} r_1^2 / r_2^2$ (dB), where r_1, r_2 are the radial distances of listeners R_1 and R_2 from the source.

3. SOME PROPERTIES OF EIGENFILTERS

A couple of interesting properties of the proposed eigenfilter under wide-sense stationary (WSS) assumptions are restated below .

Property 1 : For a WSS processes $x(n)$, and $y(n)$ with finite variances, the matrix $\mathbf{R}_{\underline{x}}(p, q)$ is toeplitz, and the gain (4) can be expressed as,

$$G_{dB} = 10 \log_{10} \frac{\int_{2\pi} |W^*(e^{j\omega})|^2 |H_2(e^{j\omega})|^2 S_x(e^{j\omega}) d\omega}{\int_{2\pi} |W^*(e^{j\omega})|^2 |H_1(e^{j\omega})|^2 S_x(e^{j\omega}) d\omega} \quad (5)$$

where, $r_x(k) \in \mathbf{R}_{\underline{x}}(k)$ and $S_x(e^{j\omega})$ form a fourier transform pair, and $h_1(n)$ and $h_2(n)$ are stable responses. Moreover, since we are focusing on real processes in this chapter, the matrix $\mathbf{R}_{\underline{x}}(k)$ is a symmetric matrix, with

$$r_x(k) = r_x(-k) \quad (6)$$

Property 2 (Linear phase) : The optimal eigenfilter (4) is a linear phase FIR filter having a constant phase and group delay, or a constant group delay.

4. SENSITIVITY TO SPATIAL VARIATIONS OF LISTENERS

The goal in this experiment is to observe the robustness of the designed optimal eigenfilter to variations in room responses. For the present situation, we generated synthetic room responses, at the direction of the two listeners, using the image method [4] for an automobile enclosure (dimensions of 2m \times 2m \times 2m). The relative locations of the source and the two listeners is shown in Fig. 2, where a single source is assumed to be operating below and to the left of the driver (e.g., a speaker located on the driver side door). Listener 2 is assumed to be the driver, whereas listener 1 is assumed to be the passenger for designing the eigenfilter. The normal (nominal) positions of the driver and passenger are denoted by an asterisk. An eigenfilter was designed for these two locations (having different responses). A set of four responses were also synthesized around each of the listeners head, depicting head movements of the listeners (indicated by circles). Two eigenfilters were designed. The first design involved an unvoiced fricative /S/ as an input to the automobile enclosure (shown in Fig. 3), whereas the second design involved a vowel /AE/ as an input to the eigenfilter (shown in Fig. 4). Once the eigenfilter was determined for the nominal head locations, the gain (4) was obtained for the nominal positions, as well as for positions corresponding to the head variations (while keeping the eigenfilter fixed). Ideally, it is preferred that the gain changes are negligible with listener variations. The order M of the eigenfilter was set at 100. We are currently investigating the perceptual effects of filter length on sound quality and will report those results in the near future.

4.1. Unvoiced Fricative /S/

The gain (dB) matrix as a function of spatial variations is given below. In the matrix, a gain at the i -th row and the j -th column provides a gain at an i -th location of the driver head against the j -th location of the passenger head around the nominal position ($i = j = 1$ indicates the gain at nominal locations of the head for which the eigenfilter was designed). The nominal positions of the driver and passenger head are marked by an asterisk. The numbers in the parenthesis depict the energy in the difference between the room responses for nominal positions and room responses for head variations.

	1* (0%)	2 (0%)	3 (82%)	4 (8%)	5 (59%)
1*(0%)	10.8	10.8	-.57	6.7	0.5
2 (10.9%)	11.7	11.7	.37	7.6	1.4
3 (30.3%)	12	12	.6	.8	1.6
4 (30.3%)	12	12	.6	7.8	1.6
5 (10.9%)	11.7	11.7	.37	7.6	1.4

4.2. Vowel /AE/

The gain matrix for this case is given below:

	1* (0%)	2 (0%)	3 (82%)	4 (8%)	5 (59%)
1*(0%)	11.3	11.3	-.8	6.8	.44
2 (10.9%)	12.4	12.4	.17	7.8	1.4
3 (30.3%)	12.7	12.7	.5	8.1	1.8
4 (30.3%)	12.7	12.7	.5	8.1	1.8
5 (10.9%)	12.4	12.4	.17	7.8	1.4

The largest changes in the gain occur when the passenger head location varies. This is mapped in Fig. 5, which depicts the energy in the difference between the room responses for nominal positions and room responses for head variations[†]. In summary largest changes in the gain occur for large energy differences between room responses at the passengers (listener 1) head. This seems intuitive, since the driver's response has a dominant direct field component which is not substantially affected due to the closeness of the driver to the source. The passenger's response has dominant reflective components which vary significantly with variations in the head locations.

5. CONCLUSIONS

In this paper we investigated the robustness of the eigenfilter to changes in head locations of two listeners (i.e., changes in

[†]The energy difference between room responses \underline{h}_1 and \underline{h}_2 is given by $\|\underline{h}_1 - \underline{h}_2\|^2 = \frac{1}{2\pi} \int_{2\pi} |H_1(e^{j\omega}) - H_2(e^{j\omega})|^2 d\omega$.

the responses at the two listeners) in an automobile type enclosure for simple speech signals, in terms of the gain. We observed that the performance is affected largely due to the passenger (listener 1) head movements than the driver head movements. We believe that this is because of larger energy in the difference between room responses corresponding to head movements and nominal responses. We plan to use perturbation theory to further investigate and quantify this behavior. Future research will also be directed to more complex signals, and perceptual aspects of designing eigenfilters (i.e., gain-perceptual coloration tradeoffs).

6. REFERENCES

- [1] S. Bharitkar and C. Kyriakakis, "Selective Signal Cancellation for Multiple-Listener Audio Applications: An Information Theory Approach," *Proc. IEEE Conf. on Multimedia*, New York, July 2000.
- [2] S. Bharitkar and C. Kyriakakis, "Eigenfilters for Audio Signal Cancellation," *Proc. IEEE International Symposium on Intelligent Signal Processing and Communication Systems*, Hawaii, November 2000.
- [3] S. Bharitkar and C. Kyriakakis, "Selective Signal Cancellation For Multiple-Listener Audio Applications Using Eigenfilters," *IEEE Trans. on Multimedia*, conditionally accepted with revisions.
- [4] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America* vol. 65, pp. 943-950, 1979.

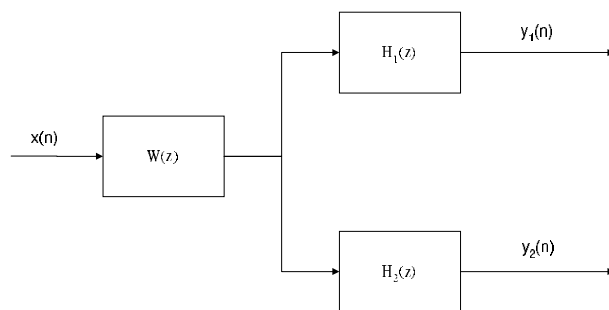


Figure 1: The source-two listener model.

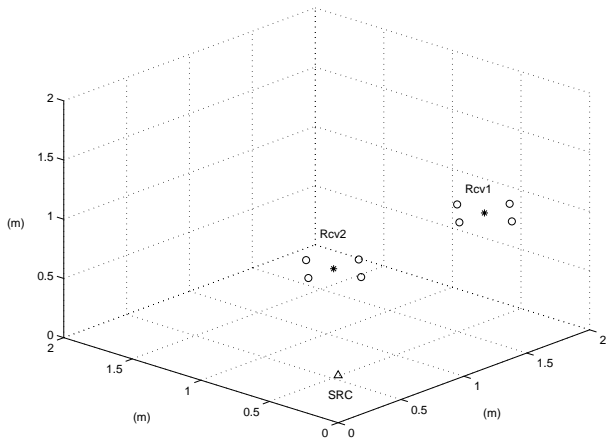


Figure 2: An automobile enclosure with a source and a driver and passenger.

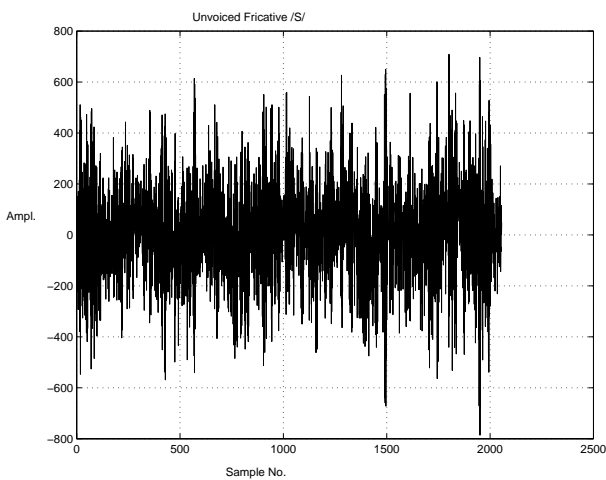


Figure 3: Unvoiced fricative speech signal /S/ as in sat.

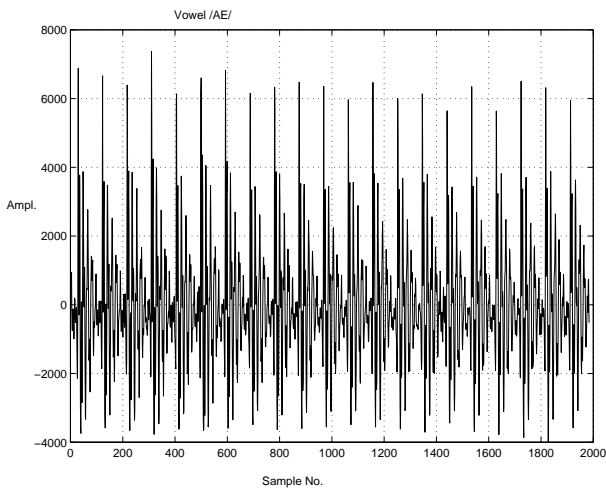


Figure 4: Vowel /AE/ as in bat

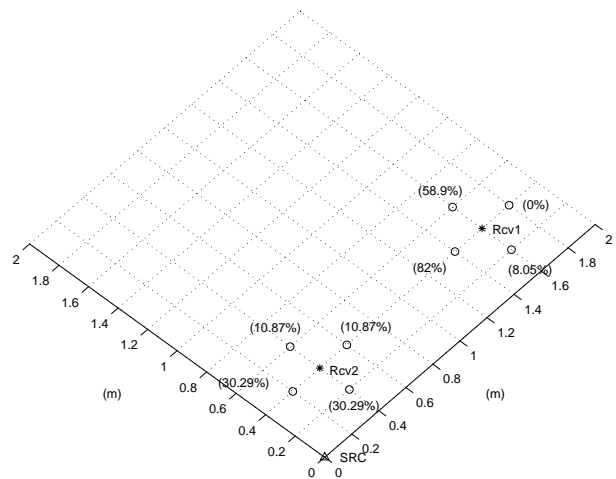


Figure 5: Energy in the difference between room responses corresponding to head movements and nominal responses.