A CLUSTER CENTROID METHOD FOR ROOM RESPONSE EQUALIZATION AT MULTIPLE LOCATIONS

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

In this paper we address the problem of simultaneous room response equalization for multiple listeners. Traditional approaches to this problem have used a single microphone at the listening position to measure impulse responses from a loudspeaker and then use an inverse filter to correct the frequency response. The problem with that approach is that it only works well for that one point and in most cases is not practical even for one listener with a typical ear spacing of 18 cm. It does not work at all for other listeners in the room, or if the listener changes positions even slightly. We propose a new approach that is based on the Fuzzy c-means clustering technique. We use this method to design equalization filters and demonstrate that we can achieve better equalization performance for several locations in the room simultaneously as compared to single point or simple averaging methods.

1. INTRODUCTION

Room equalization has traditionally been approached as a classic inverse filter problem. Although this may work well in simulations or highly-controlled experimental conditions, once the complexities of real-world listening environments are factored in, the problem becomes significantly more difficult. This is particularly true for small rooms in which standing waves at low frequencies cause significant variations in the frequency response at the listening position. A typical room is an acoustic enclosure that can be modeled as a linear system whose behavior at a particular listening position is characterized by an impulse response, $h(n); n \in \{0, 1, 2, ...\}$. This is generally called the room impulse response and has an associated frequency response, $H(e^{j\omega})$. The impulse response yields a complete description of the changes a sound signal undergoes when it travels from a source to a receiver (microphone/listener).

It is well established that room responses change with source and receiver locations in a room [1, 2]. A room response can be uniquely defined by a set of spatial co-ordinates $l_i \triangleq (x_i, y_i, z_i)$. This assumes that the source is at origin and the receiver i is at the spatial co-ordinates, x_i, y_i and z_i, relative to a source in the room.

Now, when sound is transmitted in a room from a source to a specific receiver, the frequency response of the audio signal is distorted at the receiving position mainly due to interactions with room boundaries and the buildup of standing waves at low frequencies. One scheme to minimize these distortions is to introduce an *equalizing* filter that is an inverse of the room impulse response. This equalizing filter is applied to the source signal before it is transmitted. If $h_{eq}(n)$ is the equalizing filter for $h(n)$, then, for perfect equalization $h_{eq}(n) \otimes h(n) = \delta(n)$; where $\otimes$ is the convolution operator and $\delta(n) = 1, n = 0; 0, n \neq 0$ is the Kronecker delta function. However, two problems arise when using this approach, (i) the room response is not necessarily invertible (i.e., it is not minimum phase), and (ii) designing an equalizing filter for a specific receiver will produce poor equalization performance at other locations in the room. In other words, multiple-point equalization cannot be achieved by a single equalizing filter that is designed for equalizing the response at only one location.

To address this problem, standard multiple point equalization techniques typically use the average from multiple room responses and invert the resulting minimum phase part to form the equalizing filter.

We have previously proposed a fuzzy c-means based clustering approach for identifying the representative (prototype) response in each cluster [3]. In this paper we examine methods for combining such prototypes so that one can design an equalization filter based on this single representative. We show that this approach results in flatter responses at each of the cluster members (room responses) as compared to the standard single point and spatial averaging approach.

2. THE PROPOSED FUZZY C-MEANS TECHNIQUE FOR GENERATING ACOUSTICAL ROOM RESPONSE PROTOTYPES

A. Review of Cluster Analysis in Relation to Acoustical Room Responses

Broadly speaking, clustering procedures yield a data description in terms of clusters having centroids or *prototypes* . The clusters are formed from data points (room responses in the present case) having strong *similarities*. Clustering procedures use a criterion function, such as a sum of squared distances from the prototypes, and seek a grouping (cluster formation) that extremizes the criterion function.

Specifically, clustering refers to identifying the number of subclasses of c clusters in a data universe X^d comprised of N room responses $\{h_i(n); i = 1, 2, ..., N; n = 0, 1, ..., d-1\}$, and partitioning X^d into c clusters $(2 \leq c \leq P < N)$. The trivial case of $c = 1$ denotes a rejection of the hypothesis that there are clusters in the data comprising the room responses, whereas $c = N$ constitutes the case where each room response vector

$\underline{h}_i \triangleq (h_i(0), h_i(1), ..., h_i(d-1))^T$ is in a cluster by itself. Upon clustering, the room responses bearing strong *similarity* to each other should be grouped in the same cluster. The similarity between the room responses is decided indirectly through the cluster prototype. One of the simplest similarity measures in clustering is the distance between pairs of room responses, in which case the euclidean distance metric is commonly used. If the clustering algorithm yields clusters that are well formed then, the euclidean distance between samples in the same cluster is significantly less than the distance between samples in different clusters.

A cluster room response *prototype* is a compact representation of the room responses that are grouped in the cluster, and play a fundamental role in the proposed multiple-point equalization technique.

B. The Proposed Fuzzy c-means Algorithm for Determining The Cluster Prototypes

In the Hard c- means clustering algorithm, a given room response, $\underline{h}_j$, can strictly belong to one and only one cluster. This is accomplished by the binary membership function $\mu_i(\underline{h}_j) \in \{0, 1\}$ which indicates the presence or absence of the response $\underline{h}_j$ within a cluster i.

However, in fuzzy clustering, a room response $\underline{h}_j$ may belong to more than one cluster by different "degrees". This is accomplished by a continuous membership function- $\mu_i(\underline{h}_j) \in [0, 1]$. There are some interesting viewpoints on the advantages of fuzzy clustering over hard clustering (see the example of clustering a peach, a plum, and a nectarine in [5] pp. 13). Importing this viewpoint to the clustering of room responses, it can be argued that it is possible to find a room response $\underline{h}_i$ that is similar to two differing responses $\underline{h}_j$ and $\underline{h}_k$ (for example, it may so happen that response $\underline{h}_i$ exhibits a similar response as $\underline{h}_j$ in its direct and early reflection components, whereas $\underline{h}_i$ may show a similar response to $\underline{h}_k$ in its reverberant components). Then, surely the hard clustering algorithm, during clustering, will mis-cluster $\underline{h}_i$ as strictly belonging to the same cluster as $\underline{h}_j$, or to the same cluster as $\underline{h}_k$. However, fuzzy clustering overcomes this limitation by assigning degrees of membership of the room responses to the clusters via continuous membership functions.

It can be shown that the centroids (prototypes) and membership functions are given by

$$\begin{aligned} \hat{\underline{h}}_i^* &= \frac{\sum_{k=1}^N (\mu_i(\underline{h}_k))^2 \underline{h}_k}{\sum_{k=1}^N (\mu_i(\underline{h}_k))^2} \\ \mu_i(\underline{h}_k) &= [\sum_{j=1}^c (\frac{d_{ik}^2}{d_{jk}^2})]^{-1} = \frac{\frac{1}{d_{ik}^2}}{\sum_{j=1}^c \frac{1}{d_{jk}^2}}; \\ d_{ik}^2 &= \|\underline{h}_k - \hat{\underline{h}}_i^*\|^2 \qquad (1) \\ i = 1, 2, ..., c; & \qquad k = 1, 2, ..., N \qquad (2) \end{aligned}$$

where $\hat{\underline{h}}_i^*$ denotes the i-th cluster room response prototype.

An iterative optimization procedure proposed by Bezdek [4] was used for determining the quantites in (2).

In the trivial case when all the room responses belong to a single cluster, the single cluster room response prototype $\hat{\underline{h}}^*$ in (2) is the average (spatial) of the room responses since, $\mu(\underline{h}_k) = 1, \forall k$. In a traditional approach for room response equalization, the resulting room response formed from spatially averaging the individual room responses at multiple locations is stably inverted to form a multiple-point equalizing filter.

3. DESIGNING EQUALIZING FILTERS BASED ON THE ACOUSTICAL ROOM RESPONSE PROTOTYPES

In this section, we investigate several approaches for designing and implementing multiple point equalizing filters. We primarily focus on designing minimum phase equalizing filters from the room response prototypes (2) for magnitude response equalization.

A. Combining the Acoustical Room Responses using Fuzzy Membership Functions

The objective here is to design a single equalizing filter, using the prototypes (2), for multiple point equalization. One approach to do this is by using the following model:

$$\underline{h}_{final} = \frac{\sum_{j=1}^c (\sum_{k=1}^N (\mu_j(\underline{h}_k))^2) \hat{\underline{h}}_j^*}{\sum_{j=1}^c (\sum_{k=1}^N (\mu_j(h_k))^2)} \qquad (3)$$

The corresponding equalizing filter is obtained by inverting the minimum phase component, $\underline{h}_{min,final}$, of the final prototype $\underline{h}_{final} = \underline{h}_{min,final} \otimes \underline{h}_{ap,final}$ ($\underline{h}_{ap,final}$ is the all pass component). The minimum phase sequence $\underline{h}_{min,final}$ is obtained from the cepstrum.

The model of (3) employs a weighting indicating "the level of activation" of a prototype depending upon the degrees of assignment of the room responses to the cluster containing the prototype. One interpretation of this model can be understood in relation to the Standard Additive Model (SAM) of Kosko [6, 7]. The SAM allows combining fuzzy systems by combining the throughputs of fuzzy systems *before* defuzzification. The advantage of SAM (as any additive fuzzy model) lies in its ability to approximate any continuous function on a compact (closed and bounded) domain.

The functional form for the SAM is given as,

$$F(x) = \frac{\sum_{j=1}^m a_j(x) V_j c_j}{\sum_{j=1}^m a_j(x) V_j}; \quad x = (\underline{h}_1, \underline{h}_2, ..., \underline{h}_N) \qquad (4)$$

where, $F : \Re^{d \times n} \to \Re^{d \times 1}$ is the convex sum of centroids c_j of the m *then* (consequent) part fuzzy sets. Specifically, any additive fuzzy system [8] (including the SAM) stores m if-then rules of a word form. In (4), $a_j : \Re^{d \times n} \to [0, 1]$ is a mapping function, and $b_j : \Re^{d \times 1} \to \Re$ is a set function of multivalued consequent fuzzy sets. The volumes V_j and the centroids c_j of each of the m rules as expressed by Kosko are,

$$\begin{aligned} V_j &= \int_{-\infty}^{\infty} b_j(y) dy \\ c_j &= \frac{\int_{-\infty}^{\infty} y b_j(y) dy}{\int_{-\infty}^{\infty} b_j(y) dy}; \quad j = 1, 2, ..., m \qquad (5) \end{aligned}$$

Comparing (4) and (3) we see an equivalent relationship between the SAM and the proposed model (3). The equivalence is obtained by (i) setting m to be the number of clusters, (ii) setting $a_j(x) = 1, \forall j$ (we shall experiment other forms of the joint set functions a_j, for equalization, in future research), and (iii) setting $b_j(y) = (\mu_j(\underline{h}_s))^2$. Then the discrete version of (5) is

$$\begin{aligned} V_j &= (\sum_{k=1}^N (\mu_j(\underline{h}_k))^2) \\ c_j &= \frac{\sum_{k=1}^N (\mu_j(\underline{h}_k))^2 \underline{h}_k}{\sum_{k=1}^N (\mu_j(\underline{h}_k))^2} = \hat{\underline{h}}_j^* \qquad (6) \end{aligned}$$

and correspondingly (4) becomes

$$F(x) = \underline{h}_{final} \tag{7}$$

C. Combining the Acoustical Prototype Room Responses using Least Mean Squares

The proposed prototype combining method is shown in Fig. 4. The filter $w_k, k = 0, 1, ..., M-1$ or $\underline{w} = (w_0, w_1, ..., w_{M-1})^T$ is adapted to minimize the sum square errors between the outputs from each of the prototypes and the original signal (for designing minimum phase inverse filters). The fundamental equations guiding the adaptive filter are

$$\begin{aligned}
\underline{e}(n) &= \underline{d}(n) - \mathbf{R}(n)\underline{w} \\
\underline{e}(n) &= (e_1(n)e_2(n), ..., e_c(n))^T \\
\underline{d}(n) &= (x(n), x(n), ..., x(n))^T \in \Re^{c\times 1} \\
\mathbf{R}(n) &= (\underline{r}_1(n), \underline{r}_2(n), ..., \underline{r}_c(n))^T \\
\underline{r}_i^T(n) &= [r_i(n) \quad r_i(n-1) \quad ... \quad r_i(n-M+1)] \\
r_i(n) &= \sum_{l=0}^{d-1} \hat{\underline{h}}_i^*(l)x(n-l)
\end{aligned} \tag{8}$$

where, c is the number of prototypes or clusters, d is the duration of the prototype room response, and $x(n)$ is a preferably a white training sequence. The adaptive filter update equation is then

$$\underline{w}(n+1) = \underline{w}(n) + \alpha\mathbf{R}^T(n)\underline{e}(n) \tag{9}$$

with α being the learning rate.

4. EXPERIMENTAL RESULTS

In this section we present the equalization results achieved using our proposed methods, and evaluate their performance using a spectral deviation measure (e.g., [11]).

A. Spectral Deviation Measure

Assuming a response $h(n) \stackrel{Fourier}{\longleftrightarrow} H(e^{j\omega})$ that is equalized by its approximate inverse $\hat{h}_{inv}(n) \stackrel{Fourier}{\longleftrightarrow} \hat{H}_{inv}(e^{j\omega})$, wherein the equalized response $E(e^{j\omega})$ is

$$e(n) \stackrel{Fourier}{\longleftrightarrow} |E(e^{j\omega})| = |H(e^{j\omega})||\hat{H}_{inv}(e^{j\omega})|; \tag{10}$$

then the spectral deviation measure is

$$\sigma_E = \sqrt{[\frac{1}{P}\sum_{i=0}^{P-1}(10\log_{10}|E(e^{j\omega_i})| - \frac{1}{P}\sum_{i=0}^{P-1}10\log_{10}|E(e^{j\omega_i})|)^2]} \tag{11}$$

This measure, (11), provides a measure of residual spectral distortion from a constant level.

B. Large Microphone Spacing

The microphones were arranged in a rectangular grid at $N = 6$ locations in a reverberant enclosure. The spacings between the microphone in both directions of the grid were roughly the same and about 1 meter. The loudspeaker was placed symmetrically with reference to the grid, directly to the front, and at a distance of about 10 meters at roughly the same height as the microphones.

The number of clusters determined were $c = \sqrt{6}$. We used the nearest integer 3, for the number of clusters. Room responses obtained from each of these microphone locations were clustered (2) and combined by the proposed mechanisms.

The results in the form of the spectral deviation measure (11) are tabulated below for the proposed method using the SAM based combiner. The table also displays the results for single point (an arbitrary location is equalized) and spatial averaging based equalization (i.e, averaging the responses, determining the resulting minimum phase response, and inverting this response). Our Fuzzy SAM method yielded better results (a lower number indicates a better result) than the single point and the spatial averaging method.

Location	Sing. Pt.	Spat. Avg.	Our Fuzzy SAM
1	2.16	1.46	1.39
2	0.0	1.72	1.65
3	2.22	1.71	1.61
4	2.42	1.8	1.7
5	2.3	1.54	1.49
6	2.7	1.75	1.5

The LMS combiner for the proposed method, with $M = 100$ (where M is the filter order) , yielded measures of the order of approximately 2.5, whereas the conventional LMS yielded measures of the order of approximately 7 for the same filter duration. We believe these results may be improved by fine-tuning the LMS method by controlling the adaptation rate, and increasing the duration of the filter.

C. Small Microphone Spacing

The microphones were arranged in a rectangular grid at $N = 9$ locations in a reverberant enclosure. The spacings between the microphone in both directions of the grid were roughly the same and about 8 cm. The loudspeaker was placed about 30 degrees to the left off a horizontal axis at a distance of about 1 m. The axis passed through the center location of the grid. The microphone and the loudspeakers were arranged at roughly the same height. This configuration could correspond to a listener's head located in front of a computer monitor in a desktop environment with changing head orientations.

The number of clusters determined were $c = \sqrt{9} = 3$. Room responses obtained from each of these microphone locations were clustered (2) and combined by the proposed mechanisms.

The results in the form of the spectral deviation measure (11) are tabulated below for the proposed method using the SAM based combiner. The table also displays the results for single point and spatial averaging based equalization. Our Fuzzy SAM method yielded better results (a lower number indicates a better result) than the single point and the spatial averaging method.

Location	Sing. Pt.	Spat. Avg.	Our Fuzzy SAM
1	2.93	3.55	2.08
2	0.0	3.65	2.12
3	3.14	3.92	2.42
4	3.27	3.93	2.30
5	3.14	3.79	2.24
6	3.10	3.70	2.33
7	3.22	3.32	2.12
8	3.35	3.78	2.12
9	3.82	3.73	2.54

Clearly, the proposed combining methods perform better than the standard methods for small microphone spacings.

The magnitude responses at the nine microphone locations with small spacings are shown in Fig. 1. Spatial averaged equalization is shown in Fig. 2, whereas the Fuzzy SAM equalized results are depicted in Fig. 3.

5. CONCLUSIONS

In this paper we proposed a fuzzy c-means clustering technique for creating prototypes of multiple acoustical room responses. By using different methods to combine these prototypes, we designed minimum phase inverse filters that achieve effective multiple point room response equalization. The best results are obtained on using the proposed Fuzzy SAM related model. The LMS based combiner performed lower than the other schemes. But this should improve by fine-tuning the LMS (size of the filter, learning rate).

There are several directions of research that will be considered in the future, including, (i) determining appropriate cluster validity measures, (ii) clustering the zeros of the CAPZ [10] (common acoustical pole and zero) model, (iii) formulating other methods for combining the prototypes, (iv) clustering of room acoustical responses in certain frequency ranges.

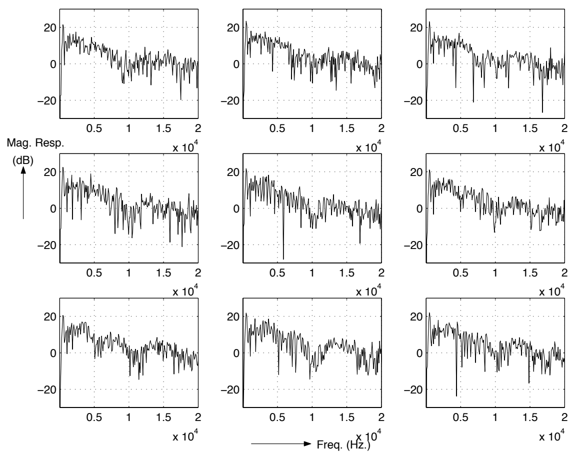

Figure 1: *Deviation of magnitude responses from flatness at the nine locations for small microphone spacings.*

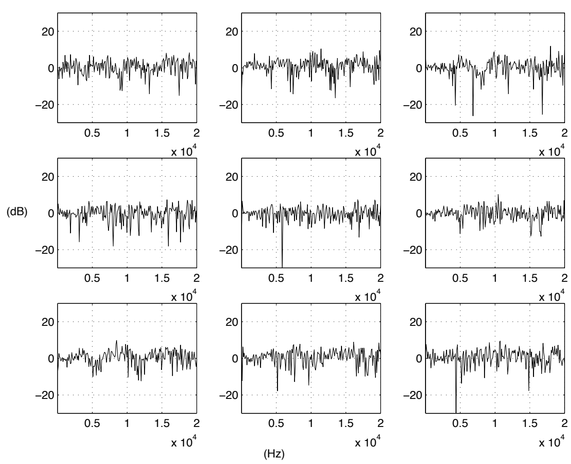

Figure 2: *Deviation of magnitude responses from flatness at the nine locations for small microphone spacing using the proposed Fuzzy SAM equalization filter.*

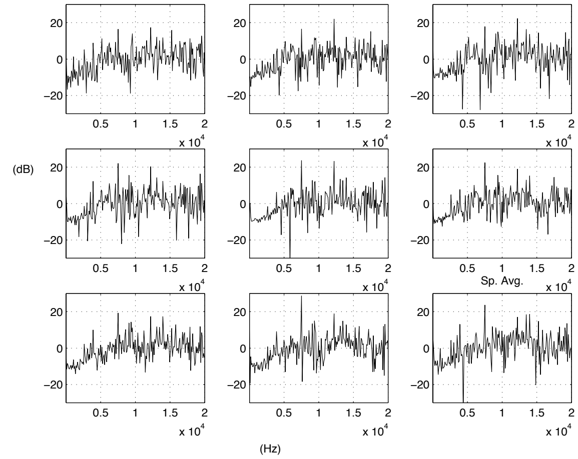

Figure 3: *Deviation of magnitude response from flatness at the nine locations for small microphone spacing using the spatial averaging equalizing filter.*

6. REFERENCES

[1] H. Kuttruff, *Room Acoustics,* Elseiver Applied Science, 3rd ed., New York. 1991.

[2] J. Mourjopoulos, "On the variation and invertibility of room impulse response functions," *Journal of Sound and Vibration,* vol. 102(2), pp. 217–228, 1985.

[3] S. Bharitkar and C. Kyriakakis, "New Factors in Room Equalization using a Fuzzy Logic Approach," *Proc. Audio Eng. Soc.* 111^{th} *Conv.,* New York. Sept. 21-24, 2001.

[4] J. Bezdek, *Pattern Recognition with Fuzzy Objectiver Function Algorithms,* Plenum Press, New York. 1981.

[5] J. Bezdek, J. Keller, R. Krishnapuram, and N. R. Pal, *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing,* Kluwer cademic Publishers. Boston. 1999.

[6] B. Kosko, *Fuzzy Logic and Neural Network Handbook,* (Ed.: C. Chen), McGraw-Hill. New York. 1996.

[7] B. Kosko, "Combining Fuzzy Systems," *Proc. of IEEE FUZZ-95,* vol. IV, pp. 1855-1863, March 1995.

[8] B. Kosko, "Fuzzy System as Universal Approximators," *IEEE Trans. on Neural Networks,* vol. 43(11), pp. 1329-1333, Nov. 1994.

[9] B. Widrow and S. Stearns, *Adaptive Signal Processing,* Prentice Hall, Englewood Cliffs. 1985.

[10] 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.

[11] B. Radlović, and R. Kennedy, "Nonminimum-phase Equalization and its Subjective Importance in Room Acoustics," *IEEE Trans. on Speech and Audio Proc.,* vol. 8(6), pp. 728-737, Nov. 2000.