Perceptual Multiple Location Equalization with Clustering

Sunil Bharitkar* *Student Member, IEEE* and Chris Kyriakakis *Member, IEEE*
E-mail:bharitka@sipi.usc.edu

Section: 4. Speech and Audio Processing

Abstract: Typically, room equalization techniques do not focus on designing filters that equalize the room transfer functions on perceptually relevant spectral features. In this paper we address the problem of room equalization for multiple listeners, *simultaneously*, using a perceptually designed equalization filter based on pattern recognition techniques. Some features of the proposed filter are, its ability to perform simultaneous equalization at multiple locations, a reduced order, and a psychoacoustically motivated design. In summary, the simultaneous multiple location equalization, using a pattern recognition method, is performed over perceptually relevant spectral components derived from the auditory filtering mechanism.

I. 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, $H(e^{j\omega})$, which is clearly a function of frequency (i.e., 20 Hz-20 kHz). Generally, $H(e^{j\omega})$ is also referred to as the room transfer function (RTF). 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). The signal at the receiver consists of direct path components, discrete reflections that arrive a few milliseconds after the direct sound, as well as a reverberant field component. In addition, it is well established that room responses change with source and receiver locations in a room [1], [2]. In other words, 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.

Clearly, the variations in the RTF at a location need to be compensated (equalized). In addition, the variations in the RTF's, at different locations, in a room also need to be compensated. Accordingly, in our previous papers [3], [4], [5], we proposed a novel approach for designing a multiple location equalization filter. The filter was designed on a linear frequency axis (non-perceptual) of the room response function. While the general approach of equalization on a linear frequency scale of the RTF's is justifiable from a strict signal processing perspective, it is not necessarily optimal from a psychoacoustical perspective.

In this paper we propose a perceptual multiple location equalization filter based on the clustering techniques presented in our previous papers (as cited above). Advantages of the proposed method over our prior approach such as, simultaneous equalization at multiple locations, a reduced filter order, tunability of the warping coefficient for desired equalization in specific frequency ranges, and perceptually improved listening experience. In the next section, we briefly discusses the concept of warping the frequency axis. In Section 3 we propose the perceptual multiple location equalization filter. Results are presented in Section 4. Section 5 concludes the paper and proposes future directions.

II. Warped Filters for Psychoacoustical Room Equalization

The concept of warped filters was introduced by Oppenheim and Johnson in their seminal work [7]. Many papers have appeared that discuss warped filter design, such as [8], [9], [10]. In this section we shall briefly discuss psychoacoustically motivated warped filter design for multiple location room equalization (the interested reader can find more details in the citations given above).

The basic idea for warping is done using an FIR chain having all-pass blocks, instead of conventional delay elements. When an all-pass filter, $D_1(z)$, is used, the frequency axis is warped and the resulting frequency response is obtained at nonuniformly sampled points along the unit circle. Thus, for warping

$$D_1(z) = \frac{z^{-1} - \lambda}{1 - \lambda z^{-1}} \tag{1}$$

The group delay of $D_1(z)$ is frequency dependent, so that positive values of the warping coefficient λ yield higher frequency resolutions in the original response for low frequencies,

The authors are with the Immersive Audio Laboratory, Integrated Media Systems Center, University of Southern California, 3740 McClintock Avenue, Los Angeles, CA 90089-2564.

IEEE 36^{th} Asilomar Conference on Signals, Systems & Computers. Pacific Grove. CA. Nov. 3-6, 2002. http://web.nps.navy.mil/∼asilomar

whereas negative values of λ yield higher resolutions for high frequencies.

Clearly, the cascade chain of all-pass filters result in an infinite duration sequence. Typically a windowing is employed that truncates this infinite duration sequence to a finite duration to yield an approximation.

Smith *et al.* [12] proposed a bilinear conformal map based on the all-pass transformation (1) that achieves a frequency warping nearly identical to the Bark frequency scale (also called the critical band rate) [13], [6]. They found a closed form expression that related the warping coefficient of the all pass transformation to the sampling frequencies $f_s \in$ (1kHz, 54kHz] that achieved this psychoacoustically motivated warping transformation. Specifically it is established that

$$\lambda = 0.8517[arctan(0.06583 f_s)]^{1/2} - 0.1916 \tag{2}$$

We use the closed form expression, as given above, for designing the warped simultaneous multiple listener equalization filter. In the subsequent details, it is to be understood that we used $\lambda = 0.7707$ (for an f_s of 50 kHz). The warping induced between two frequency axis by (1) is depicted in Fig. 1 for different values of the warping coefficient-λ.

III. Warped Filters for Psychoacoustical Room Equalization

The design of the simultaneous multiple listener equalizer is done via the fuzzy c-means clustering technique as demonstrated in our previous papers [3], [4], [5]. This is summarized below,

A. The Fuzzy c-means clustering

We use the following system of equations for determining the cluster centroids, $\underline{\hat{h}}_i^*$,

$$\begin{aligned}
\underline{\hat{h}}_i^* &= \frac{\sum_{k=1}^N (\mu_i(\underline{\tilde{h}}_k))^2 \underline{\tilde{h}}_k}{\sum_{k=1}^N (\mu_i(\underline{\tilde{h}}_k))^2} \\
\mu_i(\underline{\tilde{h}}_k) &= [\sum_{j=1}^c (\frac{d_{ik}^2}{d_{jk}^2})]^{-1} = \frac{\frac{1}{d_{ik}^2}}{\sum_{j=1}^c \frac{1}{d_{jk}^2}}; \\
d_{ik}^2 &= \|\underline{\tilde{h}}_k - \underline{\hat{h}}_i^*\|^2 \\
i = 1, 2, ..., c; &\qquad k = 1, 2, ..., N
\end{aligned} \tag{3}$$

where, $\tilde{h}_k$ denotes the k-th warped minimum phase room response sequence. An iterative scheme as proposed by Bezdek [11] is used for clustering the $N = 6$ responses.

A final centroid (prototype) is then constructed based on the Fuzzy SAM of Kosko [14], [15].

$$\underline{h}_{final} = \frac{\sum_{j=1}^c (\sum_{k=1}^N \mu_j(\underline{\tilde{h}}_k)) \underline{\hat{h}}_j^*}{\sum_{j=1}^c (\sum_{k=1}^N \mu_j(\tilde{h}_k))} \tag{4}$$

The final response $\underline{h}_{final}$ is comprised of a linear combination of the cluster centroids, where the coefficients in the linear expansion are the fuzzy weights of the corresponding cluster. The result being normalized by the sum of the weights of all clusters. The corresponding multiple listener equalizing filter is obtained by inverting the minimum phase component, $h_{min,final}$, of the final prototype $h_{final} = h_{min,final} \otimes h_{ap,final}$ ($h_{ap,final}$ is the all pass component). The minimum phase sequence $h_{min,final}$ is obtained from the real periodic approximation to the cepstrum, $c_p(n)$ as follows:

$$\begin{aligned}
H_{final}(k) = DFT[h_{final}(n)] &= \sum_{n=0}^{d-1} h_{final}(n) e^{-jk(\frac{2\pi}{d})n} \\
C(k) &= log|H(k)| \\
c_p(n) = DFT^{-1}[C(k)] &= \frac{1}{d} \sum_{k=0}^{d-1} C(k) e^{jk(\frac{2\pi}{d})n} \\
\hat{m}(n) &= \begin{cases} c_p(n)) & n = 0, d/2 \\ 2c_p(n) & 1 \le n < d/2 \\ 0 & d/2 < n \le d-1 \end{cases} \\
\hat{M}(k) = DFT[\hat{m}(n)] &= \sum_{n=0}^{d-1} \hat{m}(n) e^{-jk(\frac{2\pi}{d})n} \\
H_{min,final}(k) &= e^{\hat{M}(k)} \\
h_{min,final}(n) &= DFT^{-1}[H_{min,final}(k)] \\
&= \frac{1}{d} \sum_{k=0}^{d-1} H_{min,final}(k) e^{jk(\frac{2\pi}{d})n}
\end{aligned} \tag{5}$$

B. LPC modeling of the Equalization Filter

Linear Predictive Coding (LPC) is a powerful technique to model a spectrum, such as in coding applications. It is reasonably motivated from the human hearing perspective, since it provides an all-pole spectral representation which focuses on modeling spectral peaks. Furthermore, the representation of spectral information by a small set of parameters which can be quantized efficiently is a beneficial feature, especially in coding applications, of LPC. Details of this approach can be found in, for example, Rabiner *et al.* [16].

In the proposed warping approach, the frequency responses of each of the measured room responses, at different locations, are warped via the Bark scale. The warped responses are then clustered in the time domain, and a final centroid is formed via the Fuzzy SAM model. Once the final centroid is obtained in the warped domain, the stable component is inverted and approximated by the LPC model of order $p = 512$. To test the equalization performance, the response based on the Fuzzy SAM/LPC is unwarped and applied to each of the measured responses.

IV. Results

We have compared the proposed warping and LPC based equalization performance to the non-warped LPC based equalization. Both methods use the proposed fuzzy c-means clustering. Based on this, we are able to demonstrate the effectiveness of using warping for simultaneous multiple listener equalization, in terms of (i) reduced LPC coefficients, (ii) better overall, (specifically low frequency) equalization. Since the warping is based on the psychoacoustically motivated Bark scale, an improvement in the hearing experience is expected.

Fig. 2 shows the magnitude responses of six responses measured at six different locations in a typically large room. The general measurement setup is shown in Fig. 3. The distance of the microphone at the sweet spot (denoted by the number 1) from the speaker is about 4 meters. The microphone spacings is around 0.5 meters. The plot panels are organized sequentially in a row manner in Fig. 2. Location 1 magnitude response is shown in the first (top) row, first (left) column in the plot, whereas location 2 response is shown at first row and second (right) column and so on.

In Fig. 4, we have shown the effect of the Bark warping on the magnitude responses. As expected, there is a better resolution in the low frequency region (non-uniform sampling on the unit circle and dense in the low frequency region) as is the case in human hearing.

The equalization filter based on the fuzzy c-means clustering, but with no warping is plotted in Fig. 5. This filter was obtained by approximating the uniformly sampled 8192 point equalization filter spectrum with an LPC model of order $p = 512$. Comparing Fig. 5 with Fig. 6 (which is the equalization filter based on warping and $p = 512$), it can be seen that the proposed warping filter is able to resolve more "equalizing" peaks and dips in the low frequency (which are traditionally harder to equalize).

Fig. 7 and Fig. 8 depict the equalized responses at the six locations using the LPC model fitted to the inverse of the FCM clustering based centroid, and the LPC model fitted to the inverse of the warped FCM clustering based centroid. Clearly significant improvement in performance is exhibited by the warp based LPC model over the non-warp based LPC model for the same model order at all the listeners (locations).

The results are confirmed by the frequency domain error measure (we computed this over the full frequency range, 20Hz-20kHz),

$$
\begin{aligned}
\sigma_E &= \sqrt{[\frac{1}{P}\sum_{i=0}^{P-1}(10\log_{10}|E(e^{j\omega_i})| - \Phi)^2]} \\
\Phi &= \frac{1}{P}\sum_{i=0}^{P-1} 10\log_{10}|E(e^{j\omega_i})| \\
|E(e^{j\omega})| &= |H(e^{j\omega})||\hat{H}_{inv}(e^{j\omega})|;
\end{aligned}
$$

where, $\hat{h}_{inv}(n)$ is the designed filter. The lower the σ_E, the better the equalization performance. The tabulated results comparing the equalized responses of Fig. 7 and Fig. 8 are given below. We also show that the performance of Warped Fuzzy SAM/LPC of order 170 is about the same as the performance of conventional Fuzzy SAM/LPC of order 512 (i.e., $\sigma_E^{p_{warpedLPC}=170} \approx \sigma_E^{p_{LPC}=512}, p_{warpedLPC} \approx p_{LPC}/3$)

Loc.	Warp LPC (p=512)	LPC	Warp LPC (p=170)
1	0.9062	0.9527	0.9445
2	0.6187	0.6668	0.6656
3	0.5772	0.6264	0.6223
4	0.5578	0.5902	0.5998
5	0.6273	0.6430	0.6635
6	0.5461	0.5856	0.5875

V. Conclusions

In this paper, we developed a psychoacoustically motivated multiple listener equalization by warping the frequency axis to obtain improved equalization performance especially at low frequencies. Several advantages exist based on the proposed approach, viz., (i) a low order (smaller number of parameters) LPC model, (ii) simultaneous equalization at multiple listeners, and (iii) psychoacoustically motivated warping based on the Bark scale. Future work will be directed towards conducting formal listening tests for a more comprehensive performance evaluation.

VI. Acknowledgements

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

This research was also funded in part by the United States Department of the Army under contract number DAAD 19-99-D-0046. Any opinions, findings and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the Department of the Army.

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, "A Cluster Centroid Method for Room Response Equalization at Multiple Locations," *Proc. 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA '01),* New York. Oct. 2001.

[4] S. Bharitkar and C. Kyriakakis, "New Factors in Room Equalization Using a Fuzzy Logic Approach," *Proc. Audio Eng. Soc. 111th Conv.,* New York. Nov. 2001.

[5] S. Bharitkar and C. Kyriakakis, "Multiple-point room response equalization using clustering," submitted *IEEE Trans. on Speech and Audio Proc.*, Aug. 2001

[6] B. C. J. Moore, *An Introduction to the Psychology of Hearing,* Academic Press New York, 1997.

[7] A. Oppenheim, D. Johnson, and K. Steiglitz, "Computation of spectra with unequal resolution using the fast fourier transform," *Proc. IEEE,* 59, pp. 299-301, 1971.

[8] K. Steiglitz, "A note on variable recursive digital filters," *Proc. IEEE ASSP* 28(1), pp. 111-112, Feb. 1980.

[9] B. C. J. Moore, *An Introduction to the Psychology of Hearing,* Academic Press New York, 1997.

[10] M. Karjalainen, A. Härmä, U. K. Laine, and J. Huopaniemi, "Warped filters and their audio applications," *Proc. 1997 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA '97),* New York, 1997.

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

[12] J. O. Smith, and J. S. Abel, "Bark and ERB Bilinear Transforms," *IEEE Trans. on Speech and Audio Proc.* 7(6), pp. 697-708, Nov. 1999.

[13] E. Zwicker, and H. Fastl, *Psychoacoustics, Facts and Models,* Springer-Verlag Berlin, 1990.

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

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

[16] L. Rabiner and B-H. Huang, *Fundamentals of Speech Recognition,* Prentice Hall, New Jersey. 1993.

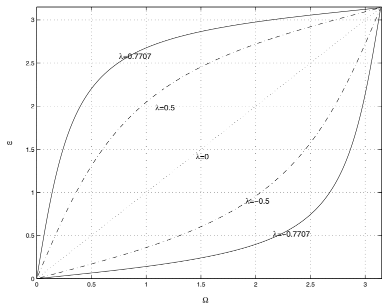

Fig. 1. Warping of frequency for various values of λ.

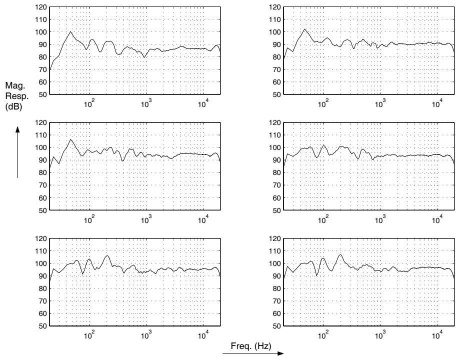

Fig. 2. Mag. Resp. at the six locations in a reverberant room

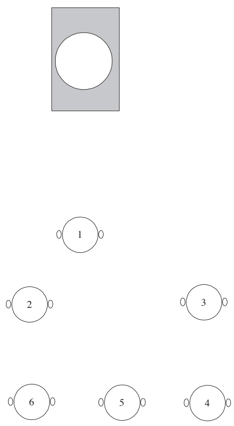

Fig. 3. Measurement setup in a reverberant room

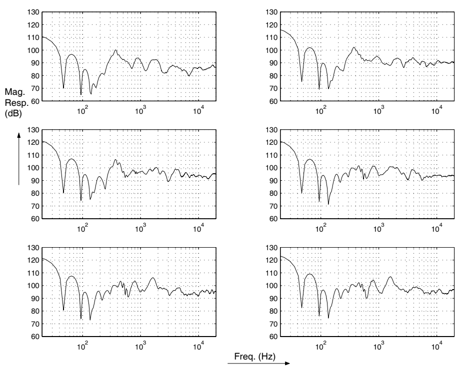

Fig. 4. Mag. Resp. corresponding to the warped RTF's at the six locations.

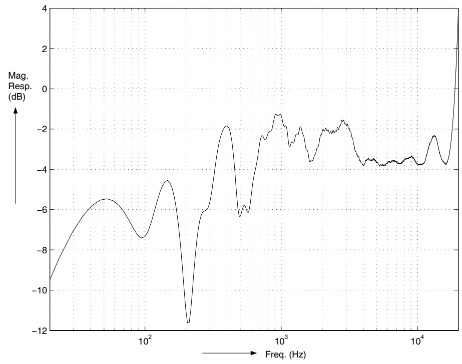

Fig. 5. Non-warped LPC ($p = 512$) Equalization filter.

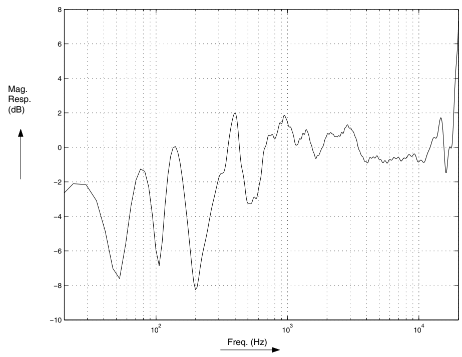

Fig. 6. Warped RTF based LPC ($p = 512$) Equalization filter.

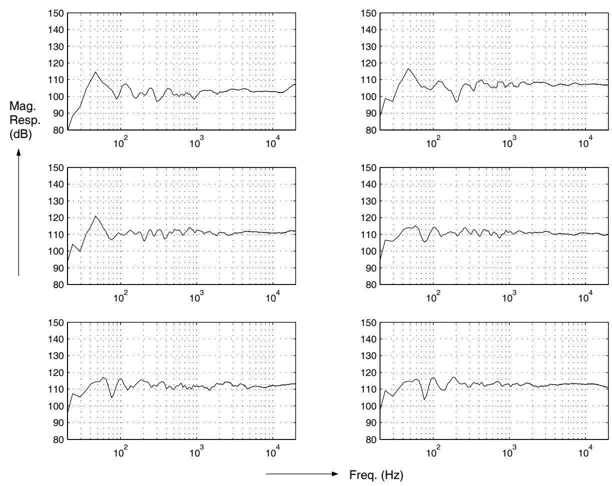

Fig. 7. Equalized responses at the six locations using standard LPC ($p = 512$) filter.

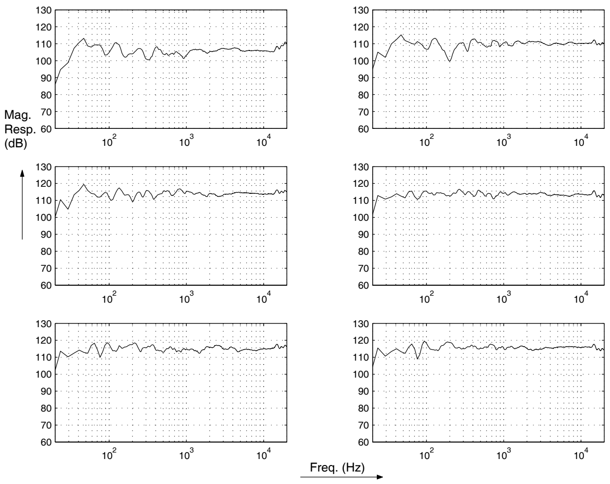

Fig. 8. Equalized responses at the six locations using warping ($p = 512$).