Audio Engineering Society
Convention Paper
Presented at the 111th Convention
2001 September 21–24 New York, NY, USA

New Factors in Room Equalization Using a Fuzzy Logic Approach

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
Room acoustical modes, particularly in small rooms, cause a significant variation in the room responses measured at different locations. Responses measured only a few cm apart can vary by up to 15-20 dB at certain frequencies. This makes it difficult to equalize an audio system for multiple simultaneous listeners. Previous methods have utilized multiple microphones and spatial averaging with equal weighting. In this paper we present a different multiple point equalization method. We first determine representative prototypical room responses derived from several room responses that share similar characteristics, using the fuzzy unsupervised learning method. These prototypical responses can then be combined to form a general point response. When we use the inverse of the general point response as an equalizing filter, our results show a significant improvement in equalization performance over the spatial averaging methods. This simultaneous equalization is achieved by suppressing the peaks in the room magnitude spectrums. Applications of this method thus include equalization and multiple point sound control at home and in automobiles.

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 form the average from multiple room responses. The minimum phase component from the averaging is stably inverted to form the equalizing filter. This is clearly an *ad hoc* scheme that uses no information provided by the measured responses. Therefore, the equalization performance may not be consistently good, and may depend on the room involved.

In this paper, we propose a nonlinear signal processing method for designing equalizing filters. The proposed method uses the fuzzy c-means clustering technique. We show that this approach results in flatter magnitude responses at several locations simultaneously when compared to the standard spatial averaging approach over the full listening spectrum. We also demonstrate, graphically, the suppression of the resonant peaks in the room transfer function.

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

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

where $\underline{\hat{h}}_i^*$ 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 $\underline{\hat{h}}^*$ 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.

DESIGNING AN EQUALIZING FILTER BASED ON THE ACOUSTICAL ROOM RESPONSE PROTOTYPES

In this section, we primarily focus on designing minimum phase equalizing filters from the room response prototypes (2) for magnitude response equalization. One approach to do this is by using the following model:

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

The corresponding equalizing filter is obtained by stably 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,

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

where, $F : \Re^{d\times n} \rightarrow \Re^{d\times 1}$ 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), $a_j : \Re^{d\times n} \rightarrow [0,1]$ is a mapping function, and $b_j : \Re^{d\times 1} \rightarrow \Re$ 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,

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

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

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

and correspondingly (4) becomes

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

EXPERIMENTAL RESULTS

In this section we present equalization performance results achieved using our proposed method as well as the traditional averaging based method. We evaluate the performance using plots (the spectral deviation measure, [11], has been used in another paper [9] for the evaluation of equalization performance).

For our experiments, we first measured nine room responses with microphones in a small auditorium type room having a single frontally located loudspeaker. The nine responses were obtained at nine chairs (locations) from a group of 15 locations (the choice of the locations were determined randomly).

We also smoothed the corresponding nine response to observe the presence of resonant peaks around 300 Hz.

A. The Non-Smooth Acoustical Room Response Equalization

The non-smoothed magnitude responses of the room transfer functions (RTF's) for the nine locations are shown in Fig. 1 and Fig. 2. We spliced the depiction of the magnitude responses to clearly show the effects of equalization using the spatial averaging technique and the proposed approach. We also approximated a piecewise linear envelope for the resonant peaks in one of the transfer functions. This is also shown in the figures.

For spatial averaging based equalization, the nine responses were averaged (in time) and the minimum phase component was determined for stable inversion. The resulting minimum phase component was then Fourier transformed and inverted to yield the desired equalizing filter. Applying the equalizing filter to each of the nine responses in Figs. 1 and 2, yields the results depicted in Figs. 3 and 4. Clearly, the results on using the spatial averaging based equalizer are poor. However, we also performed two different experiments prior to the experiments discussed herein. The first experiment were done in a desktop type environment (with nine measured responses), whereas the second experiment was done in a regular reverberant room (with six measured responses). The spatial averaging based equalization yielded fairly good results for the latter experiment. However, it did not yield a substantially good result in the desktop type environment. This lack of consistency in adequate equalization may be due to the inherent insufficient information used in forming the spatial averaging based equalizing filter.

In contrast, the proposed approach uses specific knowledge of the room responses (i.e., similarities to group them) for forming the equalizing filter. Thereby, significantly better results are obtained as shown in Figs. 5 and 6. We set the number of clusters to be $c = \sqrt{9} = 3$ (the preferred limit suggested by

Bezdek [5]). Comparing Fig. 5 with Fig. 1, and Fig. 6 with Fig. 2, we see the general behaviour of the proposed equalizing filter, that of suppression of the resonant peaks in the magnitude responses. This suppression leads to much better equalization.

B. The Smoothed Acoustical Room Response Equalization

In this experiment, we smoothed the nine magnitude responses in Fig. 1 to determine the dominant peaks and dips in the 0-5 kHz range. This is depicted in Fig. 7.

Again, we computed the spatial equalizing filter to see if it could suppress the resonances in the magnitude responses. This was not achieved, largely, by the filter (as shown in Fig. 8).

We then computed the proposed fuzzy based equalizing filter, with $c = \sqrt{9} = 3$, for the smoothed responses. We obtain significantly better results on applying the proposed filter to the smoothed responses of Fig. 7. This is shown, again, by the suppression of the resonant peaks in Fig. 9 (comparison done via arrows in Figs. 7 and 9).

CONCLUSIONS

In this paper we proposed and demonstrated a fuzzy c-means clustering technique for creating an equalizing filter operating simultaneously at multiple locations. We showed that the proposed filter achieves better equalization than the traditional spatial averaging based equalization filter. This improvement in equalization is obtained due to the suppression of resonant peaks in the RTF's. Furthermore, the proposed method uses information (such as geometric similarity) from the measured room responses, rather than just ad hoc averaging.

There are several directions of research that will be considered in the future, including, (i) simultaneous, multiple location selective equalization in certain frequency regions, (ii) determining appropriate number of clusters using cluster validity measures, and (iii) formulating other methods for combining the prototypes.

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

[4] 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.

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

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

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

[8] S. Bharitkar and C. Kyriakakis, "A Cluster Centroid Method for Room Response Equalization at Multiple Locations," *submitted, 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA),* August 2001.

[9] 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.

[10] 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.

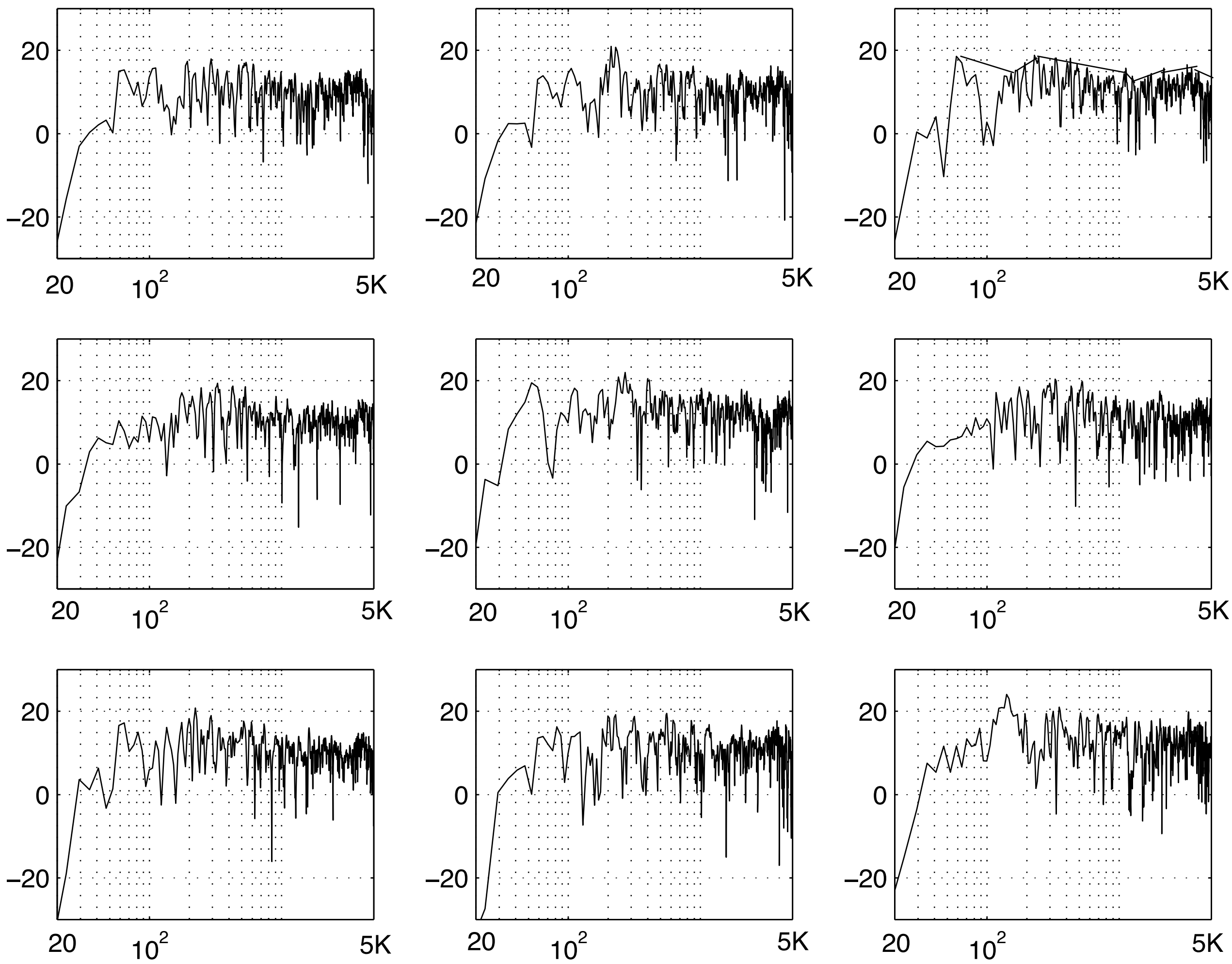

Fig. 1: Magnitude Responses for the nine locations in the 20 Hz-5 kHz range. An approximated piecewise linear envelope of the resonant peaks is shown for the response in the first row and third column.

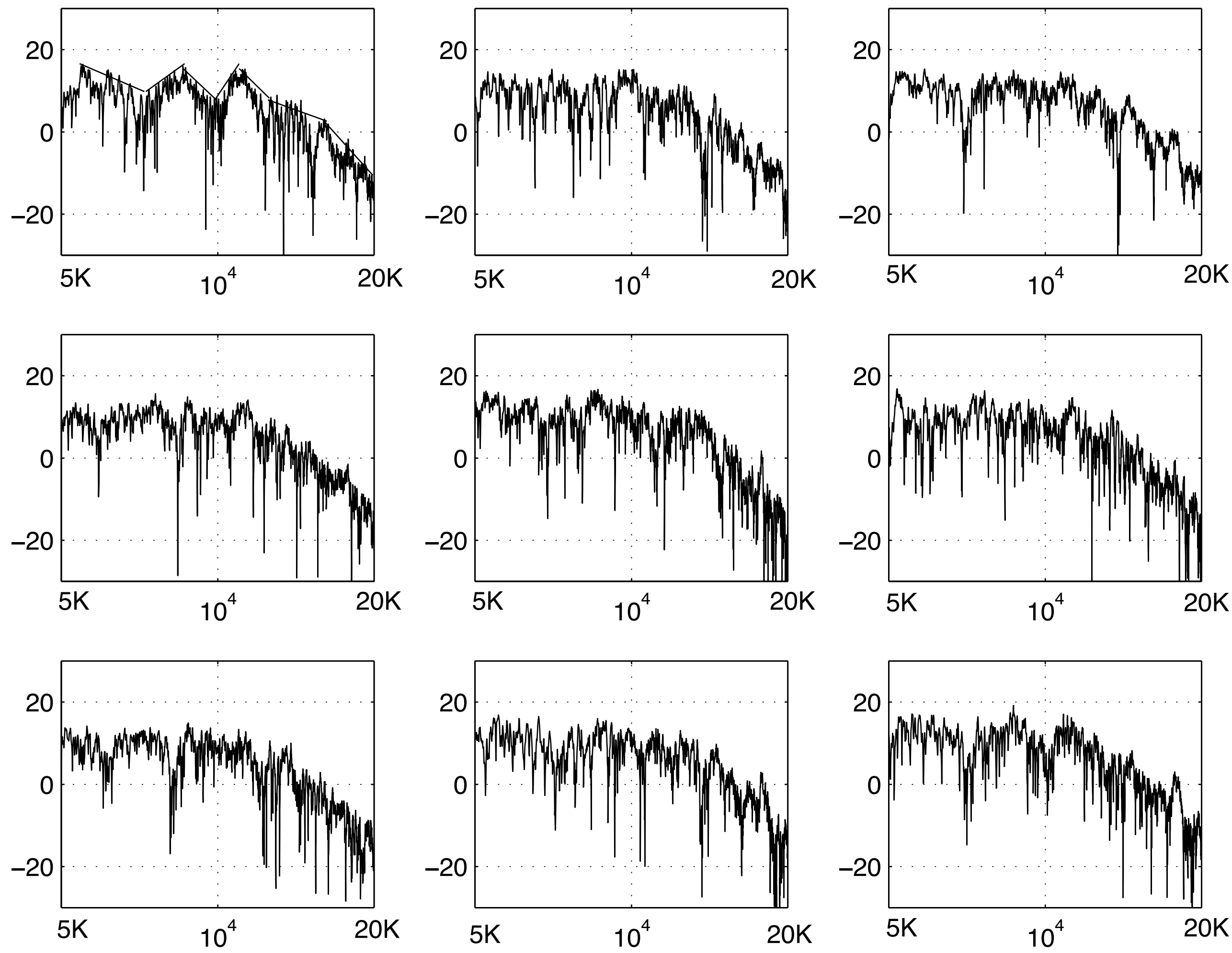

Fig. 2: Magnitude Responses for the nine locations in the 5-20 kHz range. An approximated piecewise linear envelope of the resonant peaks is shown for the response in the first row and first column.

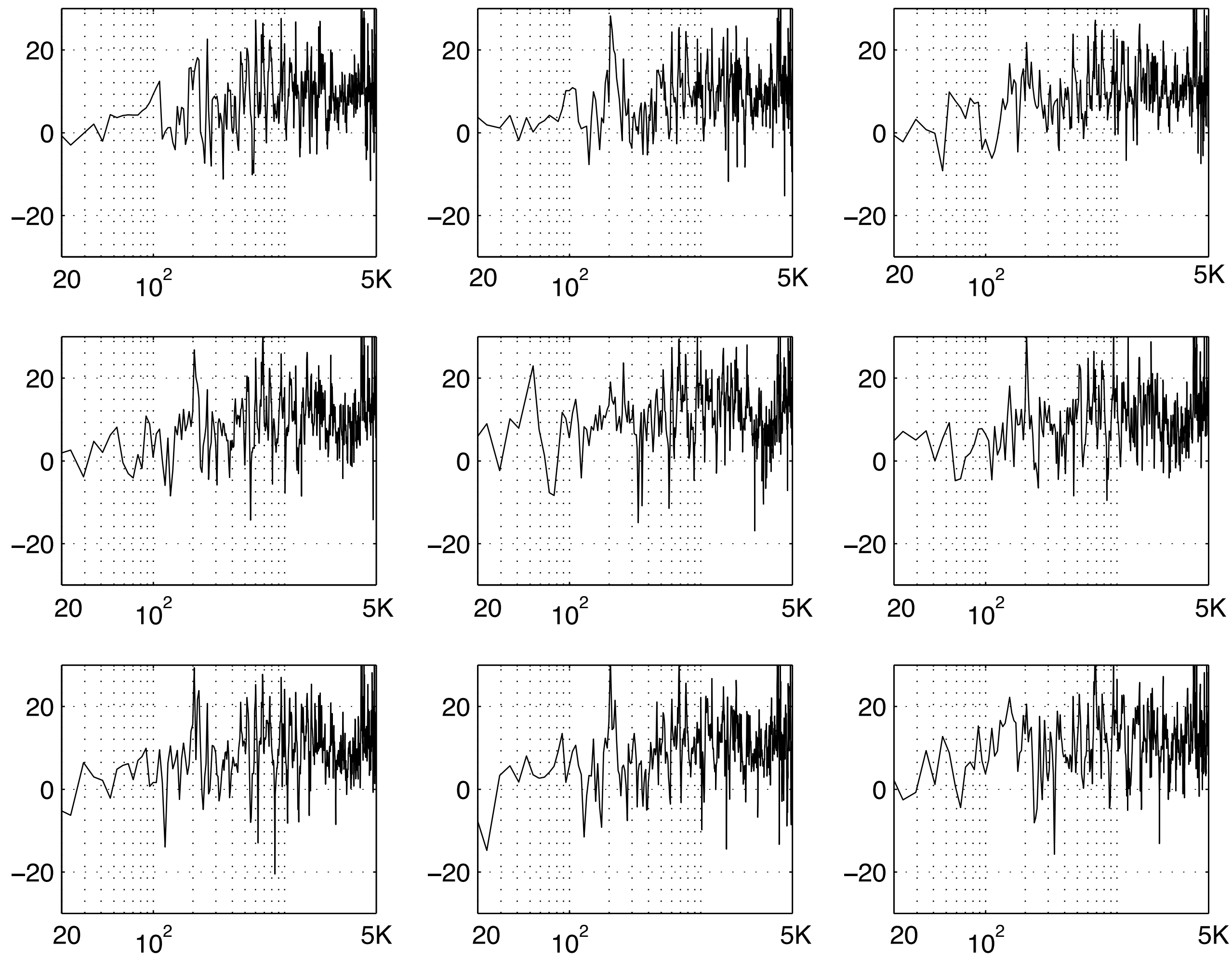

Fig. 3: Deviation from flatness of the magnitude responses for the nine locations, of Fig. 1, in the 20 Hz-5 kHz range using spatial averaging based equalization.

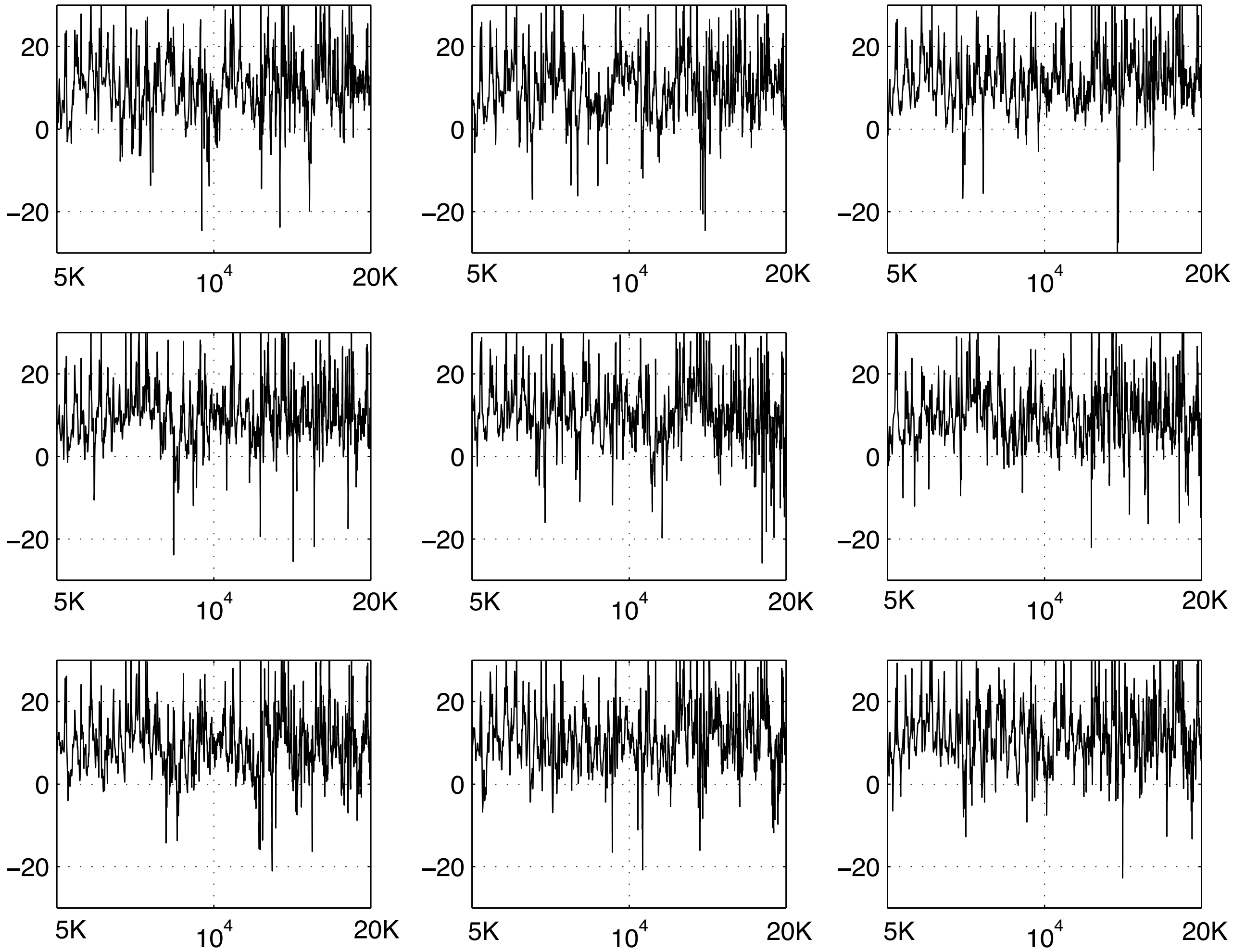

Fig. 4: Deviation from flatness of the magnitude responses for the nine locations, of Fig. 2, in the 5 kHz-20 kHz range using spatial averaging based equalization.

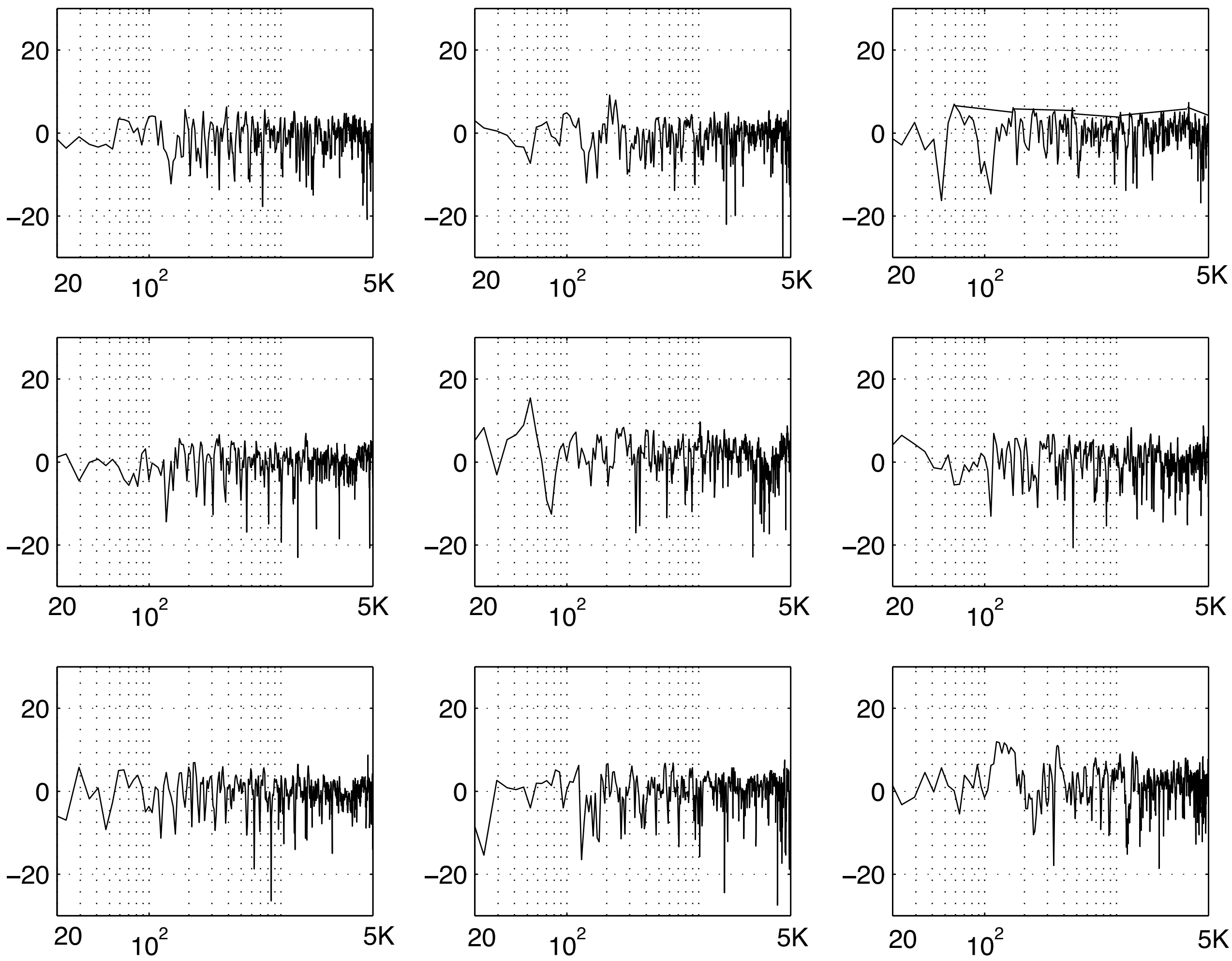

Fig. 5: Deviation from flatness of the magnitude responses for the nine locations, of Fig. 1, in the 20 Hz-5 kHz range using the proposed fuzzy c-means clustering approach. An approximated piecewise linear envelope of the resonant peaks is shown for the response in the first row and third column.

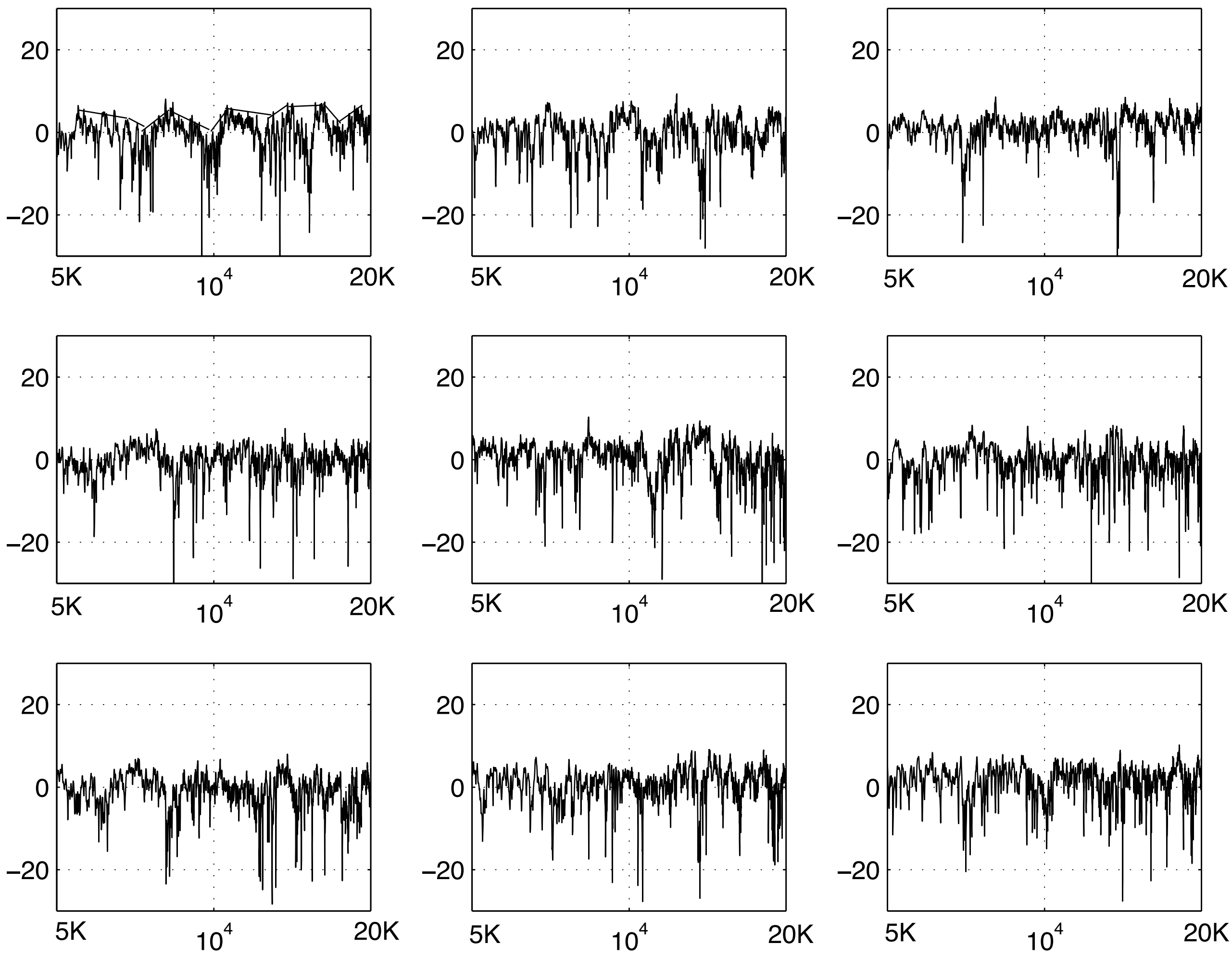

Fig. 6: Deviation from flatness of the magnitude responses for the nine locations, of Fig. 1, in the 5 kHz-20 kHz range using the proposed fuzzy c-means clustering based equalization. An approximated piecewise linear envelope of the resonant peaks is shown for the response in the first row and first column.

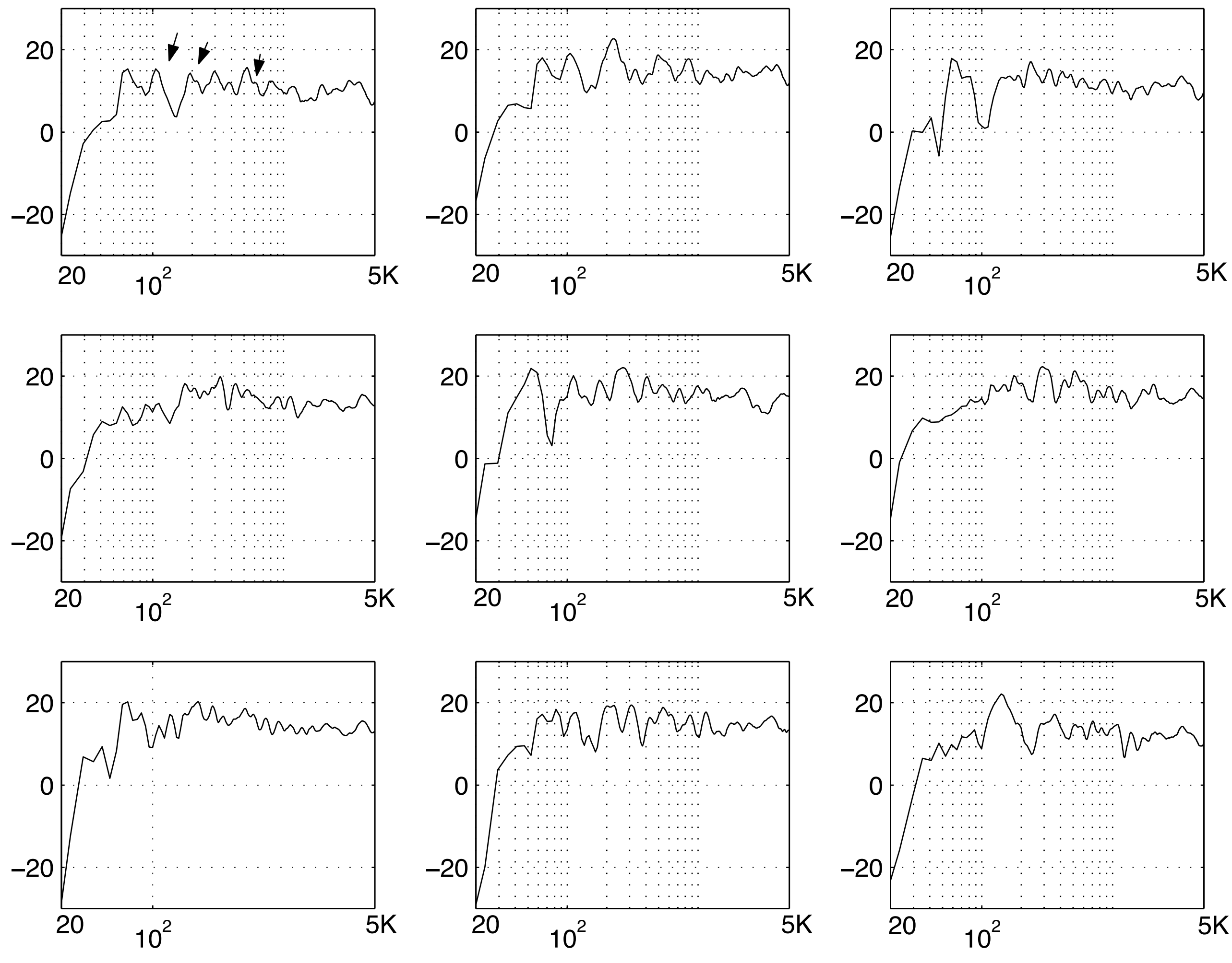

Fig. 7: Smoothed magnitude responses of Fig. 1. An approximated piecewise linear envelope of the resonant peaks is shown for the response in the first row and first column.

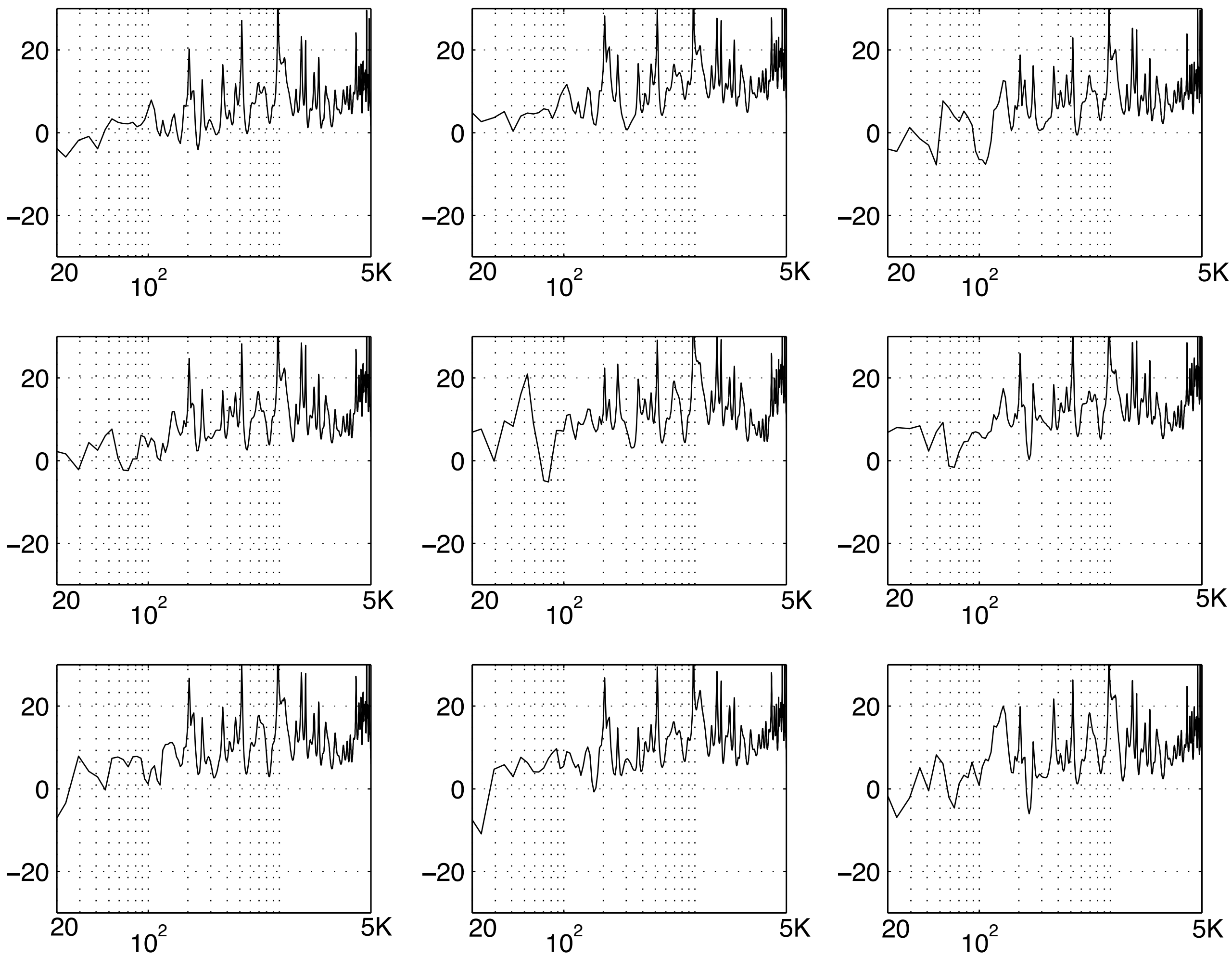

Fig. 8: Deviation from flatness of the magnitude responses for the nine locations, of Fig. 7, in the 20 Hz-5 kHz range using spatial averaging based equalization.

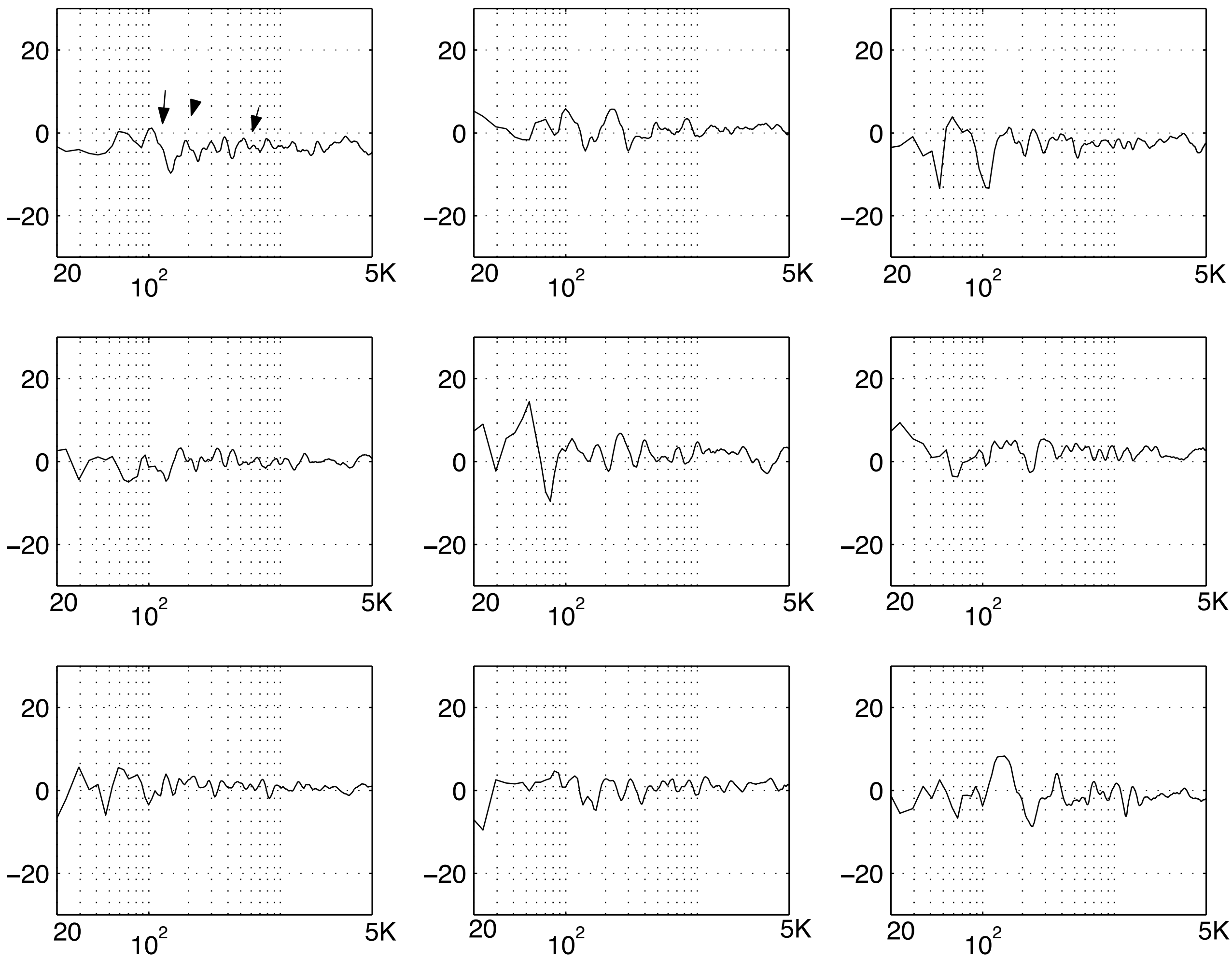

Fig. 9: Deviation from flatness of the magnitude responses for the nine locations, of Fig. 7, in the 20 Hz-5 kHz range using using the proposed fuzzy c-means clustering based equalization. An approximated piecewise linear envelope of the resonant peaks is shown for the response in the first row and first column.