On the Clustering of Head-Related Transfer Functions Used for 3-D Sound Localization

CHIN-SHYURNG FAHN AND YUEH-CHUAN LO
Department of Electrical Engineering
National Taiwan University of Science and Technology
Taipei, 106 Taiwan

Head-related transfer functions (HRTFs) serve the increasingly dominant role of implementation 3-D audio systems, which have been realized in some commercial applications. However, the cost of a 3-D audio system cannot be brought down because the efficiency of computation, the size of memory, and the synthesis of unmeasured HRTFs remain to be made better. This paper presents a way to moderate the memory requirement and computational complexity in order to reduce the cost of a 3-D audio system. We employ the library of HRTF measurements called Knowles Electronics Mannequin for Acoustic Research (KEMAR) as the original data [8]. First of all, each HRTF measurement has to be approximated in the minimum phase, and the length of the HRTF is limited by use of a modified Hamming window function, if necessary. Second, we propose an improved LBG-based clustering algorithm to lower the huge number of HRTFs. During the clustering, each HRTF is represented by its power cepstrum. Only portions of the HRTFs are reserved, and the others are neglected on condition that the minimal average mismatch distance between measured and synthesized HRTFs is achieved. Before applying localization of 3-D sounds, both unmeasured and removed HRTFs can be synthesized by linear interpolation and interaural time difference insertion. Experimental results reveal that the average and the maximum mismatch distances deriving from our improved LBG-based clustering method are less than those from the uniform clustering and LBG-based clustering methods [13].

Keywords: LBG-based clustering, head-related transfer function, KEMAR, power cepstrum, 3-D sound localization

1. INTRODUCTION

The technology of virtual reality (VR) creates a “cyberspace” by simulating the perception of a human being. Although visual cues play a crucial role in a virtual environment, auditory cues are quite useful to compensate for the insufficiency of visual cues [1]. To respond to auditory fidelity, a 3-D audio system is an important part for constructing a virtual environment. The ability to “directionalize” sounds is the key feature of a 3-D audio system [2]. Both azimuths and elevations, as shown in Fig. 1, are conventions for describing the orientation of a virtual sound source. There are no specially designated reference directions of azimuths and elevations. In this paper, the azimuth is measured from a reference direction corresponding to 0° in right front of a listener, which increases clockwise to 360° along the azimuth circle. And the elevation angle increases upwards from 0° (in front of the listener) to 90° above the listener or decreases downwards to –90° below the listener.

Received July 28, 1999; revised December 1, 2000; accepted November 13, 2001.
Communicated by Wen-Hsiang Tsai.

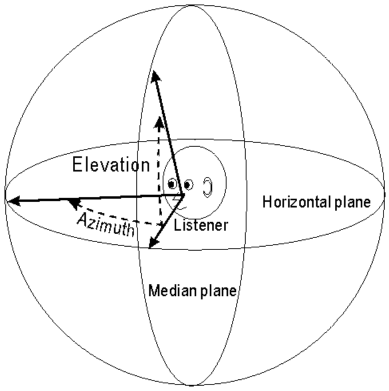


Fig. 1. Illustration of azimuths and elevations.

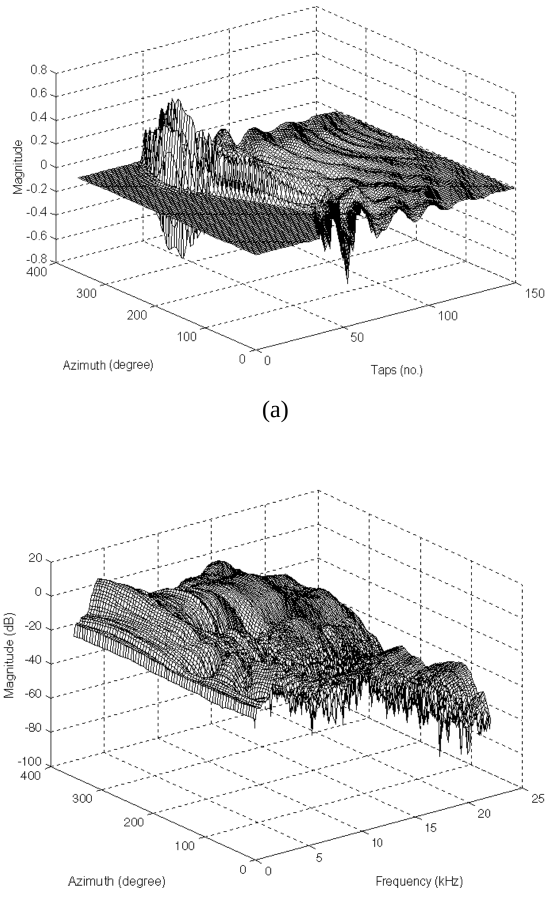
1.1 Head-Related Transfer Functions

The direction-dependent spectral filtering of a sound source before it reaches the eardrum can be viewed based on the idea that the sound source is processed by a set of transfer functions. The interaural time difference (ITD) and interaural intensity difference (IID) are two meaningful transfer functions of responding a sound source to ears. Actually, the outer ears, head, and torso lead to diffraction and reflection on the sound wave entering an ear canal [3, 4]. If the operation of propagating a sound from a pinna to the eardrum is measured as a set of head-related transfer functions (HRTFs), then the perceived location of the sound can be controlled by these transfer functions over headphones [3, 5-7]. Now the study of HRTFs is one of the most significant techniques in the application of 3-D sound effect.

1.1.1 The KEMAR

A lot of research projects of institutions and universities have collected some libraries of HRTF measurements in their anechoic chambers. One of the famous libraries is called Knowles Electronics Mannequin for Acoustic Research (KEMAR), which is available on the Internet [8]. The HRTF measurements were made in the anechoic chamber of Massachusetts Institute of Technology. There are 14 azimuth circles sampled at elevations from -40° to +90° in 10° an increment. At each elevation, the entire azimuth circle (from 0° to 360°) is sampled by an equal sized increment. The amount of HRTF measurements in KEMAR for each ear is 710 totally. And the impulse response of each HRTF measurement is 512 taps long. It is notice that the HRTFs of the left and right ears are measured in different styles of pinnae. Hence, the measured HRTFs of a single ear are enough for us to design a 3-D audio system, since it is hypothesized that the left and right ears of normal human being are symmetric.

The impulse and frequency responses of the measured HRTFs of the left ear on the horizontal plane (i.e., at elevation 0°) are shown in Figs. 2(a) and 2(b), respectively. In Fig. 2(a), the value of the x-axis means the serial number of taps. Only the impulse responses of the first 150 taps are plotted because those of the remaining taps approach zero. The value of the y-axis is the azimuth on the horizontal plane, and that of the z-axis is the magnitude of the response. From Fig. 2(a), we observe that the location of the sound source with the longest time delay and the smallest intensity is close to azimuth 90°. Conversely, the sound source with the shortest time delay and the largest intensity is located at about azimuth 270°. It coincides with the properties of the ITD and IID cues. As opposed to Fig. 2(a), the value of the x-axis in Fig. 2(b) is the frequency in kHz. It is easy to see from Fig. 2(b) that the spectra are smoothly varied as the azimuths change. This phenomenon can be thought as the spectral cues which help human beings disambiguate a sound from above to below or from front to back or vice versa. And characteristic of smoothness manifests that we can synthesize the unmeasured HRTFs by interpolation techniques.



(a)

(b)

Fig. 2. The HRTFs of the left ear on the horizontal plane: (a) the impulse responses; (b) the frequency responses.

1.1.2 Pole-zero modeling of HRTFs

In order to alleviate the complexity of HRTFs, some researchers have investigated the pole-zero approximations for HRTFs [9-11]. These methods can be used to synthesize HRTFs with fewer parameters and lower computational complexity. However, the difficulty of pole-zero modeling is to determine where each pole and zero should be positioned along its track to minimize an error associated with approximated HRTFs. An active sensory tuning technique is designed to solve this problem [11]. For mathematical simplicity, we only exert linear interpolation to synthesize the HRTFs that are not really measured in an anechoic chamber.

1.2 Our Proposed Methods

3-D sound effect has become a substantial role in some areas, such as virtual reality, audio, and other entertainment applications. At present, nevertheless, 3-D sound applications are not so popular. One of the reasons is that a 3-D audio system is costly, which needs high level DSP equipment to produce 3-D sound effect in real time because of numerous measurements and complicated computations. If we want to reduce the cost of a 3-D audio system, what we face is the problems of moderating the size of the memory and increasing the efficiency of computation. Although the cost can be economized by using low-priced equipment to implement 3-D sound effect, it will lower the quality of the 3-D audio system. It is a trade-off between cost and quality. Previous studies have discussed on this issue [12, 13], but there are still some improvements that need to be made.

In this paper, we employ the library of the KEMAR as the original data. Just the HRTF measurements of the left ear were adopted. First, all the HRTF measurements have to be approximated in the form of the minimum phase [14]. Such manipulation is convenient for the interpolation of HRTFs in the time domain. Second, the huge number of HRTFs is moderated. Only portions of the HRTFs are reserved and the others are neglected. For this purpose, we propose an improved LBG-based clustering algorithm performed on the power cepstra of HRTFs. The reconstruction error between measured and synthesized HRTFs is redefined to fit our study. The outcomes reveal that the HRTFs synthesized by our algorithm is more precise than those by the Huang's method [12]. After clustering, we try to shorten the length of each reserved HRTF by a truncating approach. To diminish the Gibbs phenomenon, we modify the Hamming window functions through rigorous evaluation [14]. The windowed HRTFs are stored in memory and ready for 3-D sound localization. As Fig. 3 illustrates, the primary sound waves are inputted to a 3-D audio system and convolved with the measured or synthesized HRTFs depending on the sound positions. Then, the 3-D audio can be produced by a couple of headphones.

2. THE CLUSTERING OF HRTFS

The cardinality of a complete set of HRTFs is very tremendous. The larger the number of HRTFs, the finer the resolution of localizing 3-D sounds will be. For some applications, nevertheless, too many HRTFs are inadequate. The easiest way to moderate

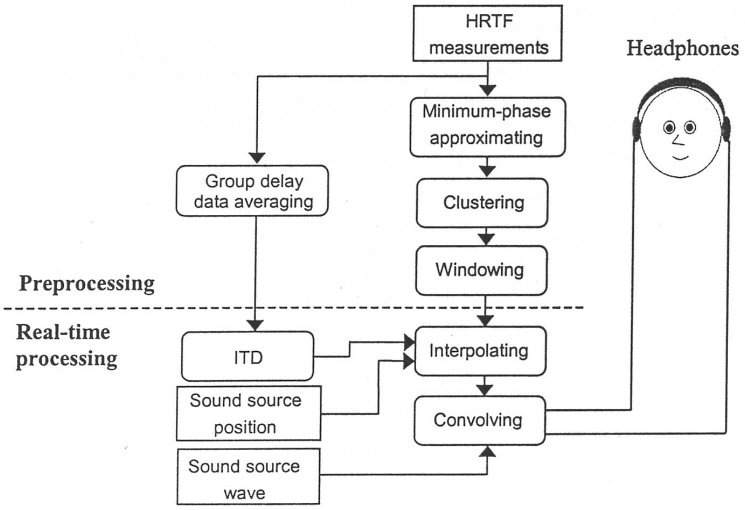


Fig. 3. A 3-D audio producing procedure.

the number of HRTFs is by sampling the original measurements uniformly. Illustrating the KEMAR as an example, there are 72 HRTFs on the horizontal plane where one HRTF is measured for every 5° azimuth. If we keep the first measurement for every 6 neighboring HRTFs, then the cardinality of this set is lowered from 72 to 12. And the other HRTFs that are removed can be restored by the aid of interpolation techniques. Essentially, the capability of human spatial hearing is nonuniform. Therefore, uniform sampling is not the appropriate way to simplify the set of HRTFs. Our goal is to develop a better manner of sampling HRTFs.

2.1 The Clustering Based on Power Cepstra

Given a set of patterns, different representative features chosen for classification usually result in dissimilar consequences. For faithfully reproduce auditory cues, the detailed characteristics of the HRTFs should be enhanced. To achieve this, a power cepstrum is more useful for clustering HRTFs whose dynamic range is less than that of a power spectrum. By taking the power cepstrum as a feature, some clustering methods for grouping the HRTFs into several clusters have been proposed [13, 15]. Such clustering methods first convert the HRTFs into power cepstra, and then choose "pilots" to find clusters using the LBG algorithm [16]. As a result, each pilot represents a cluster, and the other HRTF measurements that can be generated by means of interpolation are ignored.

The power cepstrum is defined as [17]:

$$C(\tau) = \mathrm{F}^{-1}\{\log S(f)\},$$

where $S(f)$ is the power spectrum of a frequency response and $\mathrm{F}^{-1}\{\cdot\}$ is the inverse Fourier transform.

According to the above definition, the power cepstrum of an HRTF is stated as:

$$C[n] = \mathrm{F}^{-1}\{\log S(H_{\min}(j\omega))\}, \tag{1}$$

where $H_{\min}(j\omega)$ is the minimum-phase HRTF in the frequency domain [14]. After calculating the power cepstra of the HRTFs, the LBG algorithm is employed to cluster the set of power cepstra. In this manner, the power cepstrum nearest the centroid of a cluster is selected as a pilot. An average mismatch distance between the pilots and the removed HRTFs is defined as

$$\bar{d} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{N_i} d_{ij}}{\sum_{i=1}^{m} N_i} \quad \text{with} \quad d_{ij} = \left|c_{ij} - p_i\right|, \tag{2}$$

where c_{ij} is the j-th removed HRTF of the i-th cluster, p_i is the pilot representing the i-th cluster, and N_i is the number of HRTFs in the i-th cluster. An example of LBG-based clustering is shown in Fig. 4, where the HRTFs at elevation 0° are projected onto a 2-D plane for the convenience of illustration.

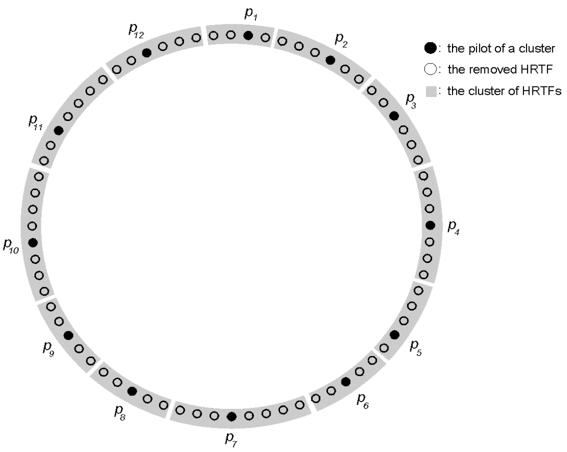


Fig. 4. An example of the LBG-Based clustering of HRTFs.

2.2 The Improved LBG-Based Clustering Algorithm

The clustering problem that we try to surmount is depicted as: 2*m* HRTF measurements are taken from *N* measurements as pilots at a certain elevation in the KEMAR, and the other measurements are replaced by linear interpolation between the 2*m* measurements, where each measurement is located in the hyperspace with 512 dimensions.

2.2.1 The modified average mismatch distance

Using Eq. (2) to evaluate the performance of clustering schemes, the results from the LBG-based clustering algorithm will be better than those from the uniform clustering method. However, Eq. (2) cannot really respond to the mismatch distance, because the distance between c_{ij} and p_i is irrelated to the consequence of linear interpolation. Hence, the average mismatch distance should be modified as:

$$\bar{d} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{N_i} d_{ij}}{\sum_{i=1}^{m} N_i} \quad \text{with} \quad d_{ij} = \left| c_{ij} - \hat{c}_{ij} \right|, \tag{3}$$

where $\hat{c}_{ij}$ is the corresponding interpolated HRTF lying between the pilot representing the i-th cluster and its two neighboring pilots located at the ends of this cluster. By the way, the performance of the improved LBG-based clustering algorithm can be ameliorated.

2.2.2 The necessary conditions for minimization

Since the LBG algorithm cannot be efficiently applied to our clustering problem, it must be modified a little to work out the solution well. In the light of the discussions in [16, 18], the necessary conditions for the minimization of the average distortion are embodied in the LBG algorithm:

Condition 1. Given an input vector x, choose a representative code vector q from an encoder $Q(x)$ to minimize the squared error distortion $\|x - x'\|^2$, where x' is a reproduction vector generated from a decoder $X(q)$.

Condition 2. Given the code vector q, compute the reproduction vector x' as the centroid of those input vectors that satisfy Condition 1.

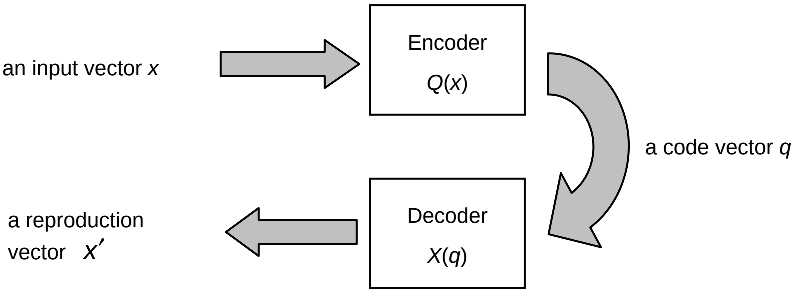


Fig. 5. The encoder-decoder model of formulating the LBG algorithm.

The LBG algorithm in the form of an encoder-decoder model is shown in Fig. 5, which basically operates in a batch mode by first adjusting the encoder $Q(x)$ in accordance with Condition 1, and then adjusting the decoder $X(q)$ in accordance with Condition 2, until the average distortion converges to a minimum value. To raise the efficiency, our improved LBG-based clustering algorithm is iteratively executed via such an encoder-decoder model.

What follows designates the notations used in the improved LBG-based clustering algorithm. Given a set of HRTFs $H = \{H_1, H_2, \ldots, H_N\}$ and its set of power cepstra $C = \{C_1, C_2, \ldots, C_N\}$, we intend to select a set of pilots $P = \{p_1, p_2, \ldots, p_K\}$ from H, so that the remaining H_i is removed and can be synthesized from P by linear interpolation. To facilitate the clustering, we divide P into two parts: one is the set consisting of the "centroids" of clusters $A = \{a_1, a_2, \ldots, a_{K/2}\} = \{p_1, p_3, \ldots, p_{2m-1}\}$ as the input vector and another is the set of "separators" $B = \{b_1, b_2, \ldots, b_{K/2}\} = \{p_2, p_4, \ldots, p_{2m}\}$ as the code vector for $K = 2m \leq N$. The separators are employed to segment H into m partitions: $U = \{U_i : i = 1, 2, \ldots, m\}$, where U_i is composed of some adjacent HRTFs in H, i.e., a cluster of H. When B is determined, U is subsequently done. The improved LBG-based clustering for the HRTFs on a certain plane is illustrated in Fig. 6. It is note that: $U_i, A, B, P \subset H, P = A \cup B, A \cap B = \varnothing, U_i \cap B = \varnothing$, and $a_i \in U_i$.

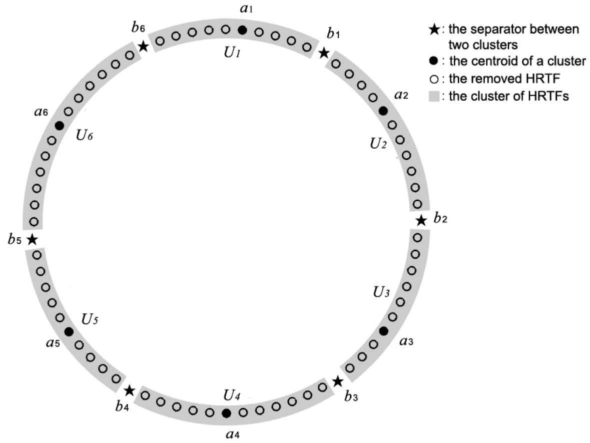


Fig. 6. An example of the improved LBG-based clustering of HRTFs.

In particular, we adopt the same function for both the encoder and decoder, i.e., set $F(\cdot) = Q(\cdot) = X(\cdot)$ to alternately optimize the centroids and separators in the improved LBG-based clustering algorithm. If A is given, find a new set of separators $B = F(A)$ such that the average mismatch distance reaches a minimum value. In this case, A plays the role of "pseudoseparators" to group H into m "pseudolusters" U_i^* consisting of some adjacent HRTFs. Conversely, if B is given, find a new set of centroids $A = F(B)$ such that the average mismatch distance attains a minimum value. It is obvious that $U_i^* \cap A = \varnothing$ and $F(A) \cap F(B) = \varnothing$.

2.2.3 The steps of the improved LBG-based clustering algorithm

The improved LBG-based clustering algorithm is summarized as follows:

Step 1: Given a set of HRTFs $H = \{H_1, H_2, \ldots, H_N\}$, compute its power cepstra $C = \{C_1, C_2, \ldots, C_N\}$, for $N \geq 2$. Given the number of pilots K that must be a multiple of 2, assign the set of initial pilots $P^{(0)} = \{p_1, p_2, \ldots, p_K\}$.Let the set of the initial centroids in all clusters be $A^{(0)} = \{a_1, a_2, \ldots, a_{K/2}\} = \{p_1, p_3, \ldots, p_{K-1}\}$, and the set of the initial separators be $B^{(0)} = \{b_1, b_2, \ldots, b_{K/2}\} = \{p_1, p_4, \ldots, p_K\}$ then the set of the initial clusters is $U^{(0)} = \{U_1, U_2, \ldots, U_{K/2}\}$, where the element U_i comprises adjacent HRTFs between two neighboring separators b_{i-1} and b_i, $i = 1, 2, \ldots, K/2$ with $b_0 = b_{K/2}$. Set $j = 0$.

Step 2: For every power cepstrum C_i , $i = 1, 2, \ldots, N$, find the new set of separators $B^{(j+1)} = F(A^{(j)})$ such that the average mismatch distance $\overline{d}$ defined in Eq. (3) falls into a minimum value, where $K/2$ pseudoclusters U_i^* are partitioned by $A^{(j)}$.

Step 3: For every power cepstrum C_i , $i = 1, 2, \ldots, N$, find the new set of centroids $A^{(j+1)} = F(B^{(j+1)})$ such that the average mismatch distance $\overline{d}$ defined in Eq. (3) falls into a minimum value, where $K/2$ clusters U_i are partitioned by $B^{(j+1)}$.

Step 4: The new set of pilots is derived from $P^{(j+1)} = A^{(j+1)} \cup B^{(j+1)}$. If $P^{(j+1)} = P^{(j)}$, the algorithm is terminated; otherwise, set $j = j + 1$ and go to step 2.

Notice that different initial inputs may lead to distinct clustering outputs. To find preferable results, it is necessary to run the improved LBG-based clustering algorithm several times with different initial inputs.

2.3 The Clustering Test of HRTF Measurements on the Horizontal Plane

In the KEMAR, there are 72 HRTF measurements on the horizontal plane. Suppose that we want to choose 12 HRTFs from the 72 ones. To prevent from converging to local minima, we experimentally ran 6 initial inputs, and compared their mismatch distances. The smallest average mismatch distance corresponds to a better result. For instance, the 6 initial inputs and their clustering outputs together with the average mismatch distances are depicted in Table 1. Remark that the initial inputs are assigned uniformly, and all such combinations are tested for the set of HRTF measurements. In the table, each element of the set of pilots represents the location of an HRTF at an azimuth and the superscript of the symbol means the number of iterations in the clustering process. For each set of pilots, the element in the odd sequence is a centroid and another in the even sequence is a separator between two clusters.

It can be seen that the clustering result from the Input 4 has the minimum average mismatch distance, which is available for synthesizing the removed HRTFs by linear interpolation. Fig. 7(a) shows the interpolated HRTF at azimuth 205°, which is compared with the original HRTF at the same azimuth. Its mismatch distance is 0.4368 that approaches the average one. Fig. 7(b) shows the interpolated HRTF compared to the original HRTF at azimuth 120°. Its mismatch distance is 1.2583. This is the worst case of interpolation obtained from the clustering result of Input 4.

2.4 Comparison of Uniform Clustering and LBG-Based Clustering Results

For the 72 HRTFs of the KEMAR, Table 1 records the mismatch distances derived from the best outcomes of uniform clustering, LBG-based clustering, and improved

LBG-based clustering after running six times for different numbers of pilots. From Table 2, it is evident that the uniform clustering method even has better performance than the LBG-based clustering method using Eq. (2) does [13]. As expected, we also observe that the number of pilots increases, the mismatch distance continuously will be lower and lower.

Table 1. The clustering outputs and average mismatch distances derived from different initial inputs of the HRTF measurements on the horizontal plane.

Input 1: $P^{(0)}$ ={0°, 30°, 60°, 90°, 120°, 150°, 180°, 210°, 240°, 270°, 300°, 330°}
Output 1: $P^{(6)}$ = {15°, 45°, 75°, 90°, 120°, 140°, 180°, 205°, 230°, 265°, 310°, 340°}
and $\bar{d}$ = 0.4347.

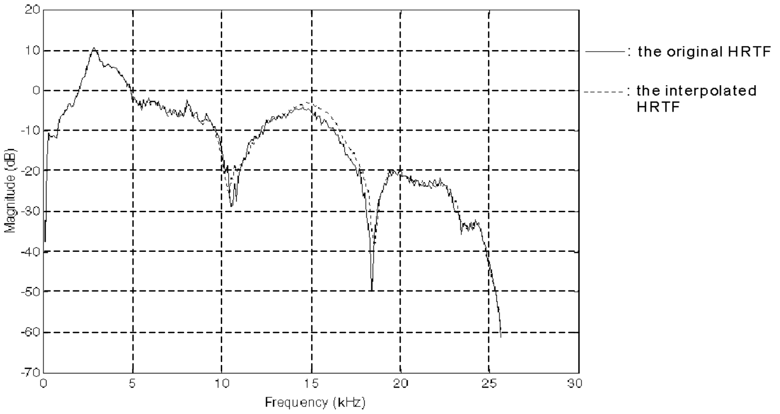
Input 2: $P^{(0)}$ = {5°, 35°, 65°, 95°, 125°, 155°, 185°, 215°, 245°, 275°, 305°, 335°}
Output 2: $P^{(4)}$ = {15°, 45°, 75°, 100°, 120°, 160°, 185°, 210°, 245°, 285°, 315°, 340°}
and $\bar{d}$ = 0.4329.

Input 3: $P^{(0)}$ = {10°, 40°, 70°, 100°, 130°, 160°, 190°, 220°, 250°, 280°, 310°, 340°}
Output 3: $P^{(4)}$ ={15°, 45°, 75°, 100°, 120°, 170°, 195°, 220°, 250°, 285°, 315°, 340°}
and $\bar{d}$ = 0.4242.

Input 4: $P^{(0)}$ = {15°, 45°, 75°, 105°, 135°, 165°, 195°, 225°, 255°, 285°, 315°, 345°}
Output 4: $P^{(4)}$ = {15°, 45°, 75°, 105°, 125°, 180°, 200°, 225°, 255°, 285°, 315°, 340°}
and $\bar{d}$ = 0.4212.

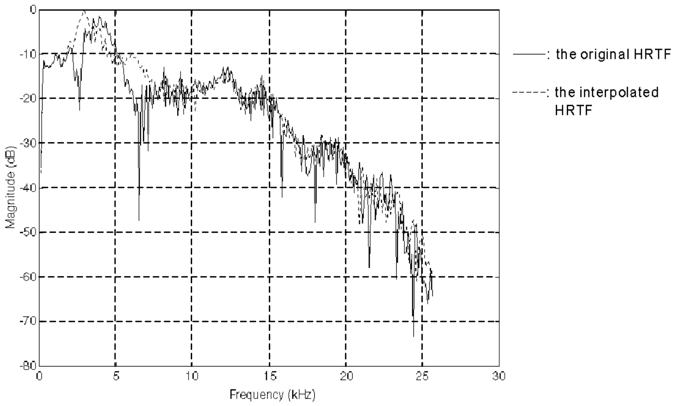
Input 5: $P^{(0)}$ = {20°, 50°, 80°, 110°, 140°, 170°, 200°, 230°, 260°, 290°, 320°, 350°}
Output 5: $P^{(4)}$ = {15°, 45°, 75°, 120°, 140°, 180°, 205°, 230°, 265°, 290°, 315°, 340°}
and $\bar{d}$ = 0.4235.

Input 6: $P^{(0)}$ = {25°, 55°, 85°, 115°, 145°, 175°, 205°, 235°, 265°, 295°, 325°, 355°}
Output 6: $P^{(5)}$ = {15°, 55°, 90°, 110°, 140°, 180°, 205°, 230°, 265°, 290°, 315°, 340°}
and $\bar{d}$ = 0.4291.



(a)

Fig. 7. Comparison of the frequency responses of the interpolated and original HRTFs: (a) at azimuth 205° (d = 0.4368); (b) at azimuth 120° (d = 1.2583).



(b)

Fig. 7. (Cont'd) Comparison of the frequency responses of the interpolated and original HRTFs: (a) at azimuth 205° (d = 0.4368); (b) at azimuth 120° (d = 1.2583).

Table 2. The average mismatch distances derived from the best outcomes of three different clustering methods.

Method / No. of pilots	Uniform clustering	LBG-based clustering	Improved LBG-based clustering
6	0.7093	0.7321	0.6689
8	0.5707	0.6244	0.5530
12	0.4338	0.4831	0.4212
18	0.3362	0.3395	0.3213
24	0.2710	0.2847	0.2545

Table 3. The maximum mismatch distances derived from the best outcomes of three different clustering methods.

Method / No. of pilots	Uniform clustering	LBG-based clustering	Improved LBG-based clustering
6	1.5270	2.0275	1.3772
8	1.2853	1.8183	1.2583
12	1.3200	1.4804	1.2853
18	1.3941	1.2742	1.0376
24	1.0784	1.2820	1.0366

Table 3 reveals the mismatch distances of the worst cases derived from the best outcomes of the three different clustering methods. Our method is still the best one of them. The number of iterations and the execution time needed to acquire the best outcomes of the three methods are shown in Table 4 and Table 5, respectively. The disadvantage of our method is that the execution time is far longer than those of the other two. It is worthwhile doing so, because the clustering of HRTFs can be achieved at an off-line preprocessing stage, which does not slow the speed of localizing 3-D sounds in real time.

3. EXPERIMENTAL RESULTS

To demonstrate the effectiveness of our proposed methods, many experiments are made by employing the library of the KEMAR as the original data. Just the HRTF measurements of the left ear are adopted. The working platform is equipped with Intel Pentium 133 CPU and 64MB RAM. The development software is MATLAB for Microsoft Windows 3.1 with the Signal Processing Toolbox, which is run in Windows 95. First illustrated are the experiments of comparing the clustering results of the HRTFs on the horizontal plane by the modified Hamming window function using different filter lengths and numbers of pilots. The subsequent experiments apply the improved LBG-based algorithm to cluster the whole HRTFs of the KEMAR.

3.1 The Clustering Results of the HRTF Measurements on the Horizontal Plane

Following experiments are all based on the minimum-phase HRTFs on the horizontal plane. For localizing a 3-D sound, the ITD information that needs to be artificially inserted into the HRTFs is obtained from computing the average group delay.

3.1.1 Number of pilots

In [19], the study had shown that the HRTFs interpolated between the measured ones even far apart by the azimuth of 60 degrees, the subjects cannot distinguish the interpolated HRTFs from actual measurements. Consequently, even the number of pilots is assigned as 6, interpolated results are still available. To compare with the results from different numbers of pilots, we choose 6, 12, and 24 pilots as our illustrative examples.

3.1.2 Windowing and filter lengths

Through the evaluation, the performance of the modified Hamming window function is better than that of the modified Bartlett, Hanning, and Blackman ones. Hence, we choose the modified Hamming window function which filter length is assigned to be 64, 128, and 256 in our experiments.

3.1.3 Performance evaluation

A reconstruction error is applied to evaluate the performance of our improved LBG-based clustering algorithm with various numbers of clusters and different filter lengths of the window function. This reconstruction error is defined as

$$\text{Reconstruction error} = \frac{\sum_{k=1}^{N} R\left(S_{k,measured}, S_{k,synthesized}\right)}{N} \tag{4}$$

with $R\left(S_{k,measured}, S_{k,synthesized}\right) = \dfrac{\sum_{i=1}^{l}\left(S_{k,measured}(i) - S_{k,synthesized}(i)\right)^2}{\sum_{i=1}^{l} S_{k,measured}^{2}(i)}$,

where l is the range of the frequency responses,
N is the total number of HRTFs on the horizontal plane,
$S_{k,\,measured}$ is the frequency response of the k-th measured HRTF,
and $S_{k,\text{synthesized}}$ is the frequency response of the k-th HRTF synthesized by means of linear interpolation.

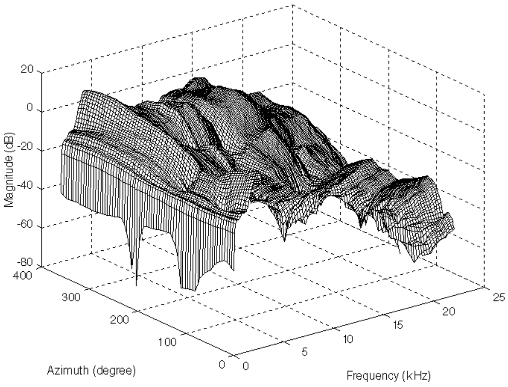
Table 4 shows the reconstruction errors received from the best performance of our improved LBG-based clustering algorithm using 6, 12, and 24 pilots with the filter lengths of 64, 128, and 256. From Table 6, we see that the larger the number of pilots, the smaller the reconstruction error gets. Besides this, the larger the filter length of the window function is, the smaller the reconstruction error obtains as expected.

Table 4. The reconstruction errors obtained from our clustering algorithm using the specified numbers of pilots and different filter lengths.

No. of pilots / Filter length	6	12	24
64	0.0396	0.0226	0.0202
128	0.0361	0.0182	0.0155
256	0.0350	0.0171	0.0154

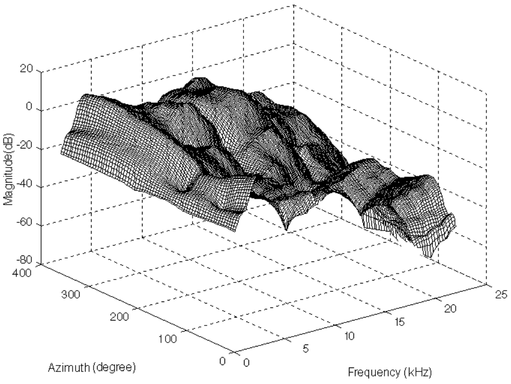
3.1.4 Simulation results

Two representative simulation results of the above experiments are illustrated in Fig. 8. We find that they are very similar to Fig. 2(b). From these figures, we observe that the details of the frequency responses become more and more in the axis of azimuths as the filter length increases. And the more subtle changes are obviously acquired in the axis of frequencies as the number of pilots increases.



(a)

Fig. 8. The frequency responses of the interpolated HRTFs on the horizontal plane resulting from our clustering algorithm: (a) 12 pilots and M = 256; (b) M = 64.



(b)

Fig. 8. (Cont'd) The frequency responses of the interpolated HRTRs on the horizontal plane resulting from our clustering algorithm: (a) 12 polots and M=256; (b) 24 pilots and M=64.

3.2 The Overall Clustering Results of the KEMAR

Prior experiments are all concentrated on the clustering of the HRTF measurements on the horizontal plane. Now we want to find the whole pilots of the 710 HRTF measurements of the KEMAR. Because the numbers of HRTF measurements at different elevations are not the same, the number of the pilots reserved at each elevation varies. Another problem is the arrangement of the initial inputs for the improved LBG-based clustering algorithm. What this algorithm requires is the even number of pilots, but the number of HRTF measurements at elevation 50° is odd. Therefore, the initial inputs at such an elevation cannot be uniformly assigned, just near uniformly. Generally speaking, we select one fourth of the measurements as the pilots. Note that all the truncation errors are zero since the minimum-phase HRTFs are not filtered by the window function in these experiments. Fig. 9 illustrates the positions of pilots resulting from the clustering of the HRTF measurements of the KEMAR. The denser the positions of pilots are, the more variations of the frequency responses of the corresponding HRTFs exist. In this experiment, the total number of pilots is 181; that is, the data compression ratio is about 25%. At an elevation, the HRTF between two adjacent pilots can be synthesized by linear interpolation. Between two adjacent elevations, the HRTF can be furthermore synthesized by bilinear interpolation using four neighboring pilots.

4. CONCLUSIONS AND FUTURE WORK

This paper has presented a way to moderate the memory requirement and computational complexity in order to reduce the cost of a 3-D audio system. Through many experiments, the results show that our method can remedy the imperfections and improve the performance of both Huang's and Chuang's methods on the clustering of HRTFs [12, 13].

For the purpose of moderating the memory size, we propose an improved LBG-based clustering method. This method is applied to select a certain number of impulse responses from a huge number of HRTF measurements. Our method is similar to the Chuang's clustering method, which groups the HRTF measurements according to their corresponding power cepstra. As opposed to the Chuang's method, we redefine a different mismatch distance used for evaluating the approximation error. Comparing with the synthesized HRTFs, the experimental results reveal our method is better than the Chuang's method.

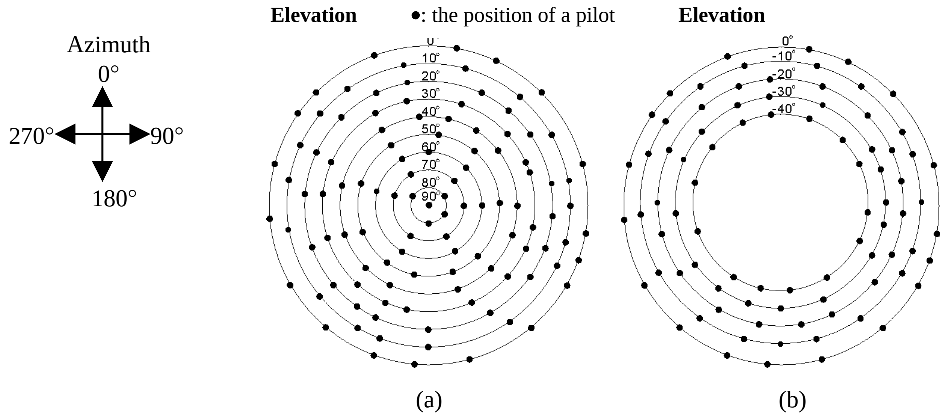


Fig. 9. The overall clustering result of the HRTFs of the KEMAR at each elevation represented by concentric circles: (a) the Northern Hemisphere (above the listener); (b) the Southern Hemisphere (below the listener).

Most of our studies focus on the preprocessing stage as shown in Fig. 3. It is enough for us to simulate one or more sound sources moving in a virtual space. However, the head movement of a listener is also a significant cue for sound localization. If head movement is considered, a tracker system is necessarily combined with the 3-D audio system. We shall investigate the tracker technology in 3-D audio.

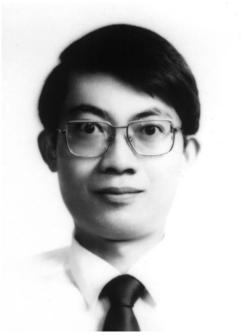
In this paper, the environmental context is neglected, too. Without regarding the environment of sound sources, a 3-D audio system can only imitate the sounds heard in an anechoic chamber. The 3-D audio system that incorporates the techniques of room modeling and auralization to simulate the environmental context [20] will be much more complex and realistic. In the future, we shall develop such techniques to realize 3-D audio system.

REFERENCES

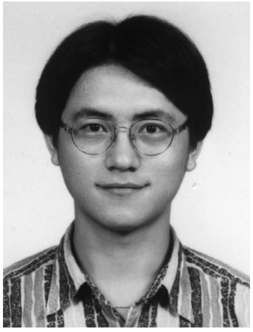
1. J. Vince, *Virtual Reality Systems*, Addison-Wesley, Reading, Massachusetts, 1995.
2. G. S. Kendall, "A 3-D sound primer: directional hearing and stereo reproduction," *Computer Music Journal*, Vol. 19, 1995, pp. 23-46.
3. E. A. G. Shaw, "The external ear," in W. D. Keidel and W. D. Neff, eds., *Handbook of Sensory Physiology*, Springer-Verleg, Heidelberg, Germany, Vol. V/I, 1974, pp. 455-490.
4. G. Plenge, "On the differences between localization and lateralization," *Journal of*

the Acoustical Society of America, Vol. 56, 1974, pp. 944-951.

5. F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. I: stimulus synthesis," *Journal of the Acoustical Society of America*, Vol. 85, 1989, pp. 858-867.
6. F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. II: psychophysical validation," *Journal of the Acoustical Society of America*, Vol. 85, 1989, pp. 868-878.
7. F. Asano, Y. Suzuki, and T. Sone, "Role of spectral cues in median plane localization," *Journal of the Acoustical Society of America*, Vol. 88, 1990, pp. 159-168.
8. W. G. Gardner and K. D. Martin, "HRTF Measurements of a KEMAR," *Journal of the Acoustical Society of America*, Vol. 97, 1995, pp. 3907-3908.
9. R. L. Jenison, "A spherical basis function neural network for pole-zero modeling of head-related transfer functions," in *Proceedings of IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, 1995, pp. 92-95.
10. M. A. Blommer and G. H. Wakefield, "Pole-zero approximations for head-related transfer functions using a logarithmic error criterion," *IEEE Transactions on Speech and Audio Processing*, Vol. 5, 1997, pp. 278-287.
11. P. R. Runkle, M. A. Blommer, and G. H. Wakefield, "A comparison of head related transfer function interpolation methods," in *Proceedings of IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, 1995, pp. 88-91.
12. J. J. Huang, "Realization of lower-ordered 3-D sound system," Master Thesis, Department of Communications Engineering, National Chiao Tung University, Hsinchu, Taiwan, 1996.
13. C. C. Chuang, "Study on HRTF clustering and synthesis with 3-D sound applications," Master Thesis, Department of Communications Engineering, National Chiao Tung University, Hsinchu, Taiwan, 1995.
14. Y. C. Lo, "A clustering and synthesis method for the head-related transfer functions in the minimum-phase approximation," Master Thesis, Department of Electrical Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan, 1998.
15. S. Shimada, N. Hayashi, and S. Hayashi, "A clustering method for sound localization transfer functions," *Journal of the Audio Engineering Society*, Vol. 42, 1994, pp. 577-583.
16. Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Transactions on Communications*, Vol. 28, 1980, pp. 84-95.
17. A. M. Noll, "Short time spectrum and cepstrum techniques for voice pitch detection," *Journal of the Acoustical Society of America*, Vol. 36, 1964, pp. 296-302.
18. S. Haykin, *Neural Networks: A Comprehensive Foundation*, Macmillan College, New York, 1994.
19. E. M. Wenzel and S. H. Foster, "Perceptual consequences of interpolating head-related transfer functions during spatial synthesis," in *Proceedings of IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, 1993, pp. 102-105.
20. Y. Haneda, S. Makino, and Y. Kaneda, "Common acoustical pole and zero modeling of room transfer functions," *IEEE Transactions on Speech and Audio Processing*, Vol. 2, 1994, pp. 320-328.



Chin-Shyurng Fahn (范欽雄) was born in Tainan, Taiwan, Republic of China, on October 15, 1958. He received the B.S. degree in electronic engineering from National Taiwan Ocean University, Keelung, Taiwan, in 1981, and the M.S. and Ph.D. degrees both in electrical engineering from National Cheng Kung University, Tainan, Taiwan, in 1983 and 1989, respectively. From August 1983 to July 1984, he served as a Teaching Assistant in the Department of Electrical Engineering at National Cheng Kung University. From August 1983 to July 1989, he worked as an Adjunct Research Assistant in the Institute of Electrical Engineering of this university. From November 1989 to May 1991, he served in the Chinese Army as a Signal Officer. In the meantime, he served as a Lecturer in the Department of Mechanical Engineering at the Chinese Army Engineering School, Yenchao, Kaohsiung Hsien, Taiwan. Since August 1991, he has been an Associate Professor in the Department of Electrical Engineering at National Taiwan University of Science and Technology, Taipei, Taiwan. His current research fields of interest include fuzzy systems, neural networks, and evolutionary computing applied in the areas of image processing, pattern recognition, computer vision, computer graphics, and virtual reality.



Yueh-Chuan Lo (羅悅全) was born in Tainan, Taiwan, Republic of China, on January 5, 1969. He received the B.S. degree in electronic engineering from Fu Jen Catholic University, Hsinchuang, Taipei Hsien, Taiwan, in 1992, and the M.S. degree in electrical engineering from National Taiwan University of Science and Technology, Taipei, Taiwan, in 1998. From June 1995 to August 1996, he worked as an engineer on music I.C. design for Holtek Microelectronics Inc., Hsinchu, Taiwan. From July 1999 to January 2000, he was a technical editor for Pots Weekly, Taipei, Taiwan. Since then, he has set up a studio for Internet and multimedia use in Vancouver, Canada. His current research interests are in the areas of computer music, 3-D sound, and multimedia.