المجلة الإلكترونية الشاملة متعددة التخصصات
العدد الرابع و الخمسون شهر (١٢) ٢٠٢٢

Personalization of HRIR measurements based on backpropagation artificial neural networks

By

Eng. Mohammad Abdu Almarhabi

Technical and vocational Training Corporation TVTC, National entrepreneurship institute

Alqunfudhah -Kingdom of Saudi Arabia

malmarhabi@tvtc.gov.sa

Abstract

Head-related impulse response (HRIR) is a key factor in 3-D sound generation over headphones. The main goal of 3D sound applications is the reproduction of the spatial properties of a sound by filtering the sound source with the appropriate HRIR of the listener. Direct measurements are the most important method for HRIR generation because they guarantee that the filtering of 3-D sound systems will have the appropriate impulse response of the listener's ear corresponding to his anthropometric specifications. One of the disadvantages of direct measurement is that it is time-consuming, effort-demanding, and requires special tools and equipment at a high cost. Therefore, accurate reproduction of the spatial properties of sound remains a challenge. This study presents a personalization model based on establishing a non-linear relationship between anthropometric parameters and HRIRs. The HRIR personalization model is based on back-propagation artificial neural networks (BPANN) as a machine learning technique, principal components analysis tools (PCA), and the current HRIR measurements database. The findings of this study outperformed those of recently developed related studies.

Keywords: HRIR personalization, Artificial neural network, Head related transfer function, Head related impulse response.

Keywords: human resource allocation, email, email addresses, the following header: moving to header beats.

المستخلص

تعد الاستجابة النبضية المتعلقة بالرأس (HRIR) عاملاً رئيسيًا في توليد الصوت ثلاثي الأبعاد عبر سماعات الرأس. الهدف الرئيسي لتطبيقات الصوت ثلاثي الأبعاد هو إعادة إنتاج الخصائص المكانية للصوت عن طريق تصفية مصدر الصوت باستخدام HRIR المناسب للمستمع. الطريقة الأكثر أهمية لتوليد HRIR هي القياسات المباشرة لأنها تضمن أن تكون تصفية أنظمة الصوت ثلاثية الأبعاد مع الاستجابة النبضية المناسبة لأذن المستمع المطابقة لمواصفات القياسات البشرية الخاصة به. من عيوب القياس المباشر أنه يستغرق وقتًا طويلاً ويتطلب جهدًا ويتطلب أدوات ومعدات خاصة بتكلفة عالية. لذلك ، لا يزال الاستنساخ الدقيق للخصائص المكانية للصوت يمثل تحديًا. تقدم هذه الدراسة نموذجًا للتخصيص يعتمد على إنشاء علاقة غير خطية بين المعلمات الأنثروبومترية و HRIRs. يعتمد نموذج تخصيص HRIR على الشبكات العصبية الاصطناعية ذات الانتشار العكسي (BPANN) كأسلوب للتعلم الآلي ، وأدوات تحليل المكونات الرئيسية (PCA) ، وقاعدة بيانات قياسات HRIRs الحالية. أسفرت نتائج هذه الدراسة عن نتائج ملحوظة مقارنة بالأنظمة ذات الصلة المطورة حديثًا.

الكلمات المفتاحية: تخصيص HRIR ، الشبكة العصبية الاصطناعية ، وظيفة النقل المتعلقة بالرأس HRTF ، استجابة النبضات المتعلقة بالرأسHRIR.

1-Introduction

The perception of the environment is greatly aided by sound, which provides humans with a constant awareness of their surroundings. Everyday life is full of different sounds that carry information about humans' surroundings. Humans can hear sounds and determine their location with different degrees of accuracy (1). 3-D sound recreation has many applications, such as virtual reality, robotics, teleconferencing, videoconferencing, entertainment, training simulations, helping the visually impaired, and so forth. This 3-D sound detection is a major factor in orienting in space, driving a car, communicating with several people at once, etc.

The brain detects the position of sound sources by processing electrical impulses generated by sound waveforms impinging on the eardrums. As a sound waveform travels from the source to the eardrum, it is influenced by many factors, e.g., obstacles in space, shape of the body and head, size of the earlobes, shape of the ear canal, etc. A measure of transfer characteristics can be expressed as head-related impulse response (HRIR) or head-related transfer function (HRTF), i.e., the Fourier transform of the HRIR [2]. These representations enable the generation of synthetic spatial sounds, which can then be played back through headphones. The HRIR depends on the spatial position of the sound source and can be measured at different spatial positions. The impulse response filter describes the sound and the direction-dependent diffusing effects of sound due to the head, torso, and pinna [2]. So, HRIR contains reflections and refractions of sound waves at the ears and head. This leads to the HRIR not only depending on the direction of the sound but also containing some indirect information about the listener. Humans have different-sized heads and ears. Figure 1 shows the different sizes and shapes of the human ear. This entails measuring each listener's HRIRs in order to create an accurate 3D-sound system that gives listeners the sensation of being in a real environment.

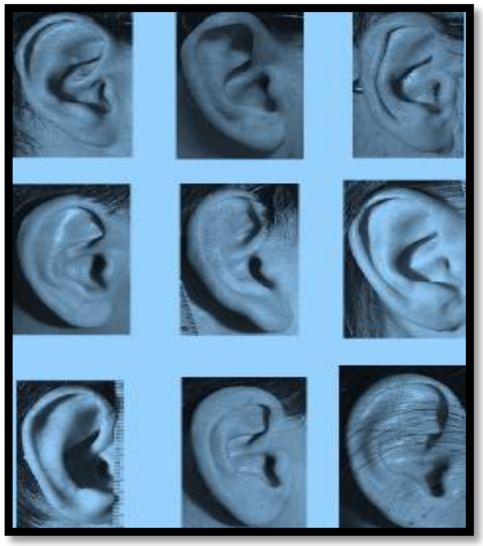


Figure 1. Different sizes and shapes of ear

Among several existing methods of HRIR personalization, the direct measurement procedure is the most accurate and well developed [3-5]. The most significant disadvantage of direct measurement is its complexity. In addition, it is an expensive approach; special equipment and efforts are required; and it is difficult and time-consuming. 3D-sound systems should be adaptable to a wide range of users. Therefore, there is an urgent need to find a method that can estimate the individual HRIR based on the current HRIR measurements. In recent years, several studies have concerned themselves with finding better, more convenient ways for obtaining personalized HRIRs. The structural approach is one of the solutions based on the structural composition of the measured responses of parts of the human body like the head, torso, and pinna that provide the important features of the HRTFs [6]. In [7,8], authors proposed a strategy for HRIR personalization based on matching certain anthropometric ear parameters of the individual with the closet HRIR in the CIPIC database. The traditional statistical methods were used to create a linear or non-linear relationship between HRIR components and the anthropometric parameters of individuals [9-11]. Many researchers have developed geometrical models for the torso, head, and ears [12, 13]. In this approach, it has been proposed to use an accurate geometrical representation of a head, the pinna, obtained by a three-dimensional laser scanner, and use this as the basis for computational acoustic simulation based on the exact solution of the wave equation using boundary element modeling. Machine learning (ML) approaches played a significant role in this direction. The previous HRIR personalization based on ML focused on finding a generation method for the log-magnitude and phases of the Fourier transform of HRIR and then estimating the impulse response as the minimum phase [14-16]. Dealing with HRTF, or impulse response, as log-magnitude and phases requires the computation of the HRTF at a large number of linearly spaced frequencies to invert the Fourier transform and extract the HRIR. For both continuous- and discrete-time signals, the magnitude and phase of the Fourier transform are independent functions. Therefore, the original signal cannot be recovered from knowledge of either one. The minimum phase reconstruction method is used to recover signals in the time domain. Minimum phase reconstruction has been used to obtain an estimated HRIR in the personalization problem by knowing the log-magnitude of the HRIR signal. This work suggests a personalization model that deals with the raw HRIR and bases its estimation directly on two important methods of machine learning, namely the principal component analysis (PCA) and back-propagation neural networks (BPANN). The main objective of this work is to improve the performance of the existing methods of HRIR personalization. With the learning process, the knowledge base will be able to adapt the measurements of HRIR to any listener and will be able to improve its predictions and experiences. Another reason to personalize HRIR is to save time, effort, and resources. In addition, the 3-D sound systems will be adapted for a wide range of listeners. The individual objectives of this work are described below:

- Proposing a HRIR personalization method based on backpropagation artificial neural networks and principal components analysis
- Implement the personalization model to illustrate and prove the approach using MathWorks- MATLAB environment.
- Compare the results with related works

2.The Proposed Approach

In general, a backpropagation neural network was trained using the anthropometric parameters for several persons from the CIPIC database as inputs and the principal components of HRIR data (PCs) as outputs. The procedures of the personalization model, which have been followed in this work, are described below. This process applies PCA directly to the raw HRIR data to simplify HRIR data by reducing a multidimensional dataset to lower dimensions with a wide range of variances. Hence, extracting the important HRIR characteristic will be possible, and the prediction accuracy of new impulse responses will be increased. Figure 2 shows the whole model of HRIR personalization. The steps of HRIR customization using PCA are listed below:

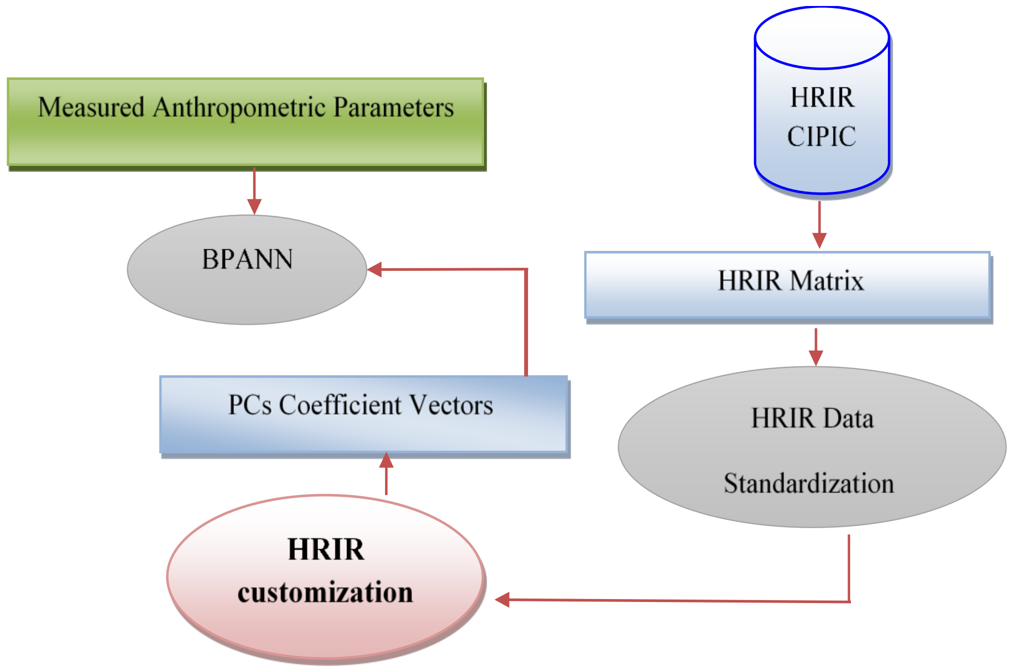


Figure 2. HRIR personalization model

- A data matrix for HRIR is created. Each column in the HRIR matrix contains different samples of HRIR measurements for different individuals.
- Because of the personal differences and likeness in anthropometric properties, the mean is calculated for each column to average all persons' HRIR. The mean vectors (HRIR_Mean) are obtained by

$$\mathbf{HRIR\_Mean} = \mathbf{1/M} \sum_{\mathbf{m=1}}^{\mathbf{M}} \mathbf{HRIR\_matrix[N, M]} \qquad \textbf{Equation 1}$$

Where N is the number of rows and M is the number of columns

- The mean is subtracted from each column to minimize the correlation and increase the variances among the HRIR data. This step produces a new adjusted dataset whose mean is zero.

$$\mathbf{adjusted\_data} = [\mathbf{HRIR\_Matrix}] - [\mathbf{HRIR\_Mean}]$$

- The covariance matrix of the adjusted–data matrix is calculated, and then eigenvectors and eigenvalues are obtained. Eigenvectors are called basis vectors, which represent the principal components of HRIR data. For N rows of the adjusted-data matrix, the covariance is computed by

$$\mathbf{covariance} = \frac{\mathbf{1}}{\mathbf{N-1}} * \sum_{\mathbf{n=1}}^{\mathbf{N}} [\mathbf{adiusted\_data}] * [\mathbf{adjusted\_data}]$$

- To obtain an HRIR reduced order model, the basis vectors are selected depending on the descending order of the eigenvalues, and this gives the components of highest significance and ignores the others without much loss of information.
- The number of samples was reduced to represent each 200-sample diffuse-field equalized HRIR in the CIPIC database.
- Once the components have been chosen, the PC coefficient vectors which represent the contribution of each basis vector to the standard HRIR data are obtained by:

$$\mathbf{PC\_Coff} = [\mathbf{basis\_vectors}]' * [\mathbf{adjusted\_data}]$$

2.1 BPANN Training

BPANN training is carried out using the PCs' coefficient vectors and anthropometric parameters. A three-layer BP network is employed in this work, with a tan-sigmoid transfer function in the hidden layer and a linear transfer function in the output layer. The choice of the hidden layer number is very important, but there is no general formula for it. To determine the optimum number of hidden nodes, a series of different topologies were used, in which the nodes varied from 16 to 20. During the training phase, the hidden layer was chosen to achieve the best performance. The implementation has been done using the Mathworks-Matlab environment, and the number of iterations is 5000, which is the number that has given excellent results.

2.2 HRIR Estimation

The estimated HRIR for any person, as shown in figure 3, is given by using his anthropometric measurements and the principal components coefficient vectors, which are predicted from BPANN. Finally, the estimated HRIR is composed of the basis vectors and the predicted PC coefficient vectors. To remove the standardization and generate the estimated personal impulse response, the mean of each column is added to HRIR. This is expressed as

$$\mathbf{Estimated\ HRIR} = ([\mathbf{basis\_vectors}] * [\mathbf{PC\_coff}]) + [\mathbf{HRIR\_Mean}]$$

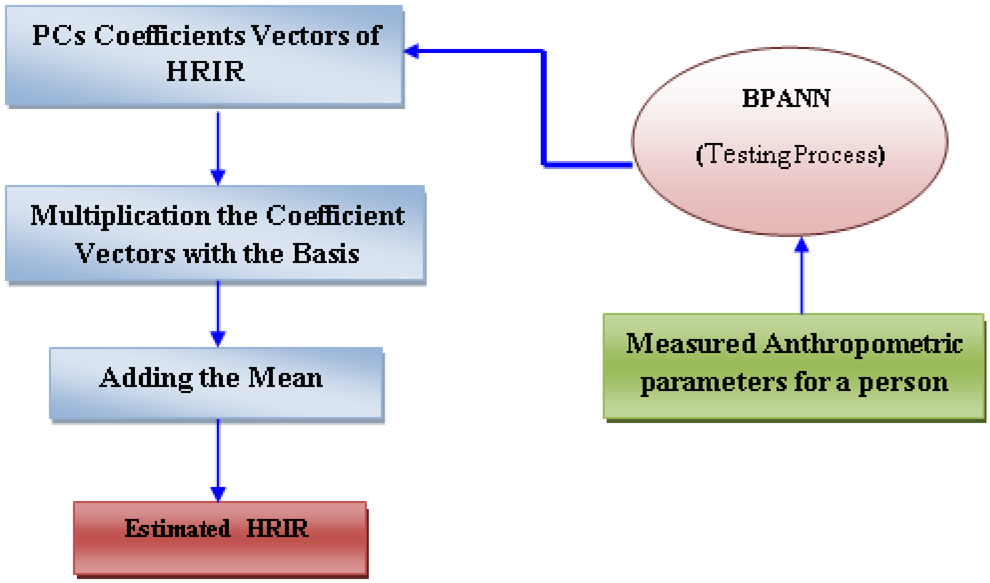


Figure 3. HRIR estimation module.

3- Result and discussion

The final stage of this study is the generation of the estimated HRIR. A comparison between the original and the estimated HRIR is provided at several elevation and azimuth angles. In the objective test method, spectral distortion (SD) was used to assess the performance of the HRIR personalization method. This is done by computing the percentage of the errors in the estimation in decibels. In order to verify that there is no significant difference between the estimated and the original HRIR, HRIR was estimated for different individuals at different positions (azimuth, elevation) depending on their measured anthropometric parameters extracted from CIPIC. Because the objective test has a small SD score, the findings of this study strengthen its effectiveness as an effective approach. Table 1 illustrates the average spectral distortion of the impulse responses in decibels.

Location (Azimuth, Elevation)	The average SD score
Front(0,0)	3.6
Back (0, 180)	5.4
Right (80, 0)	5.6
Left (-80, 0)	5.5
Top (0, 90)	4.5
Right-top (80, 45)	4.24
Left- top (-80, 45)	5.9
Bottom-back (0, 225)	3.8
Bottom-front (0, -45)	3.5
Top-back (0,135)	5.2
decibleLeft-bottom (-80,-45)	4.5
Right-bottom (80,-45)	5.2

Table 1. Average spectral distortion for several persons from CIPIC.

The figures below depict various samples of estimated and original HRIR and their magnitude spectra for several people at various directions from CIPIC.

Spectral distortion is 4.24 dB

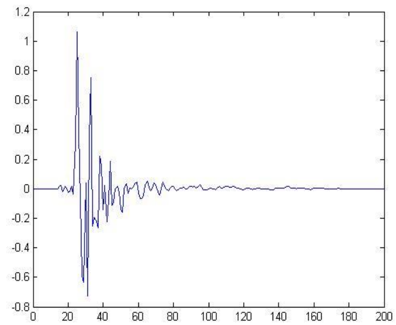
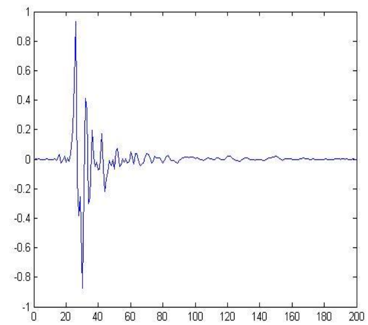
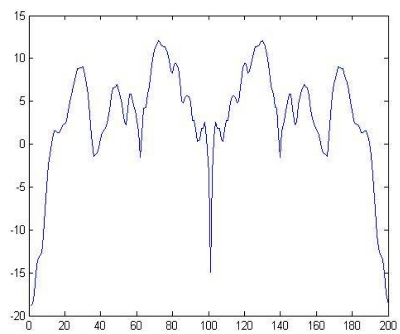
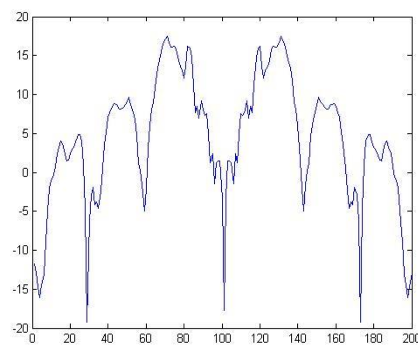
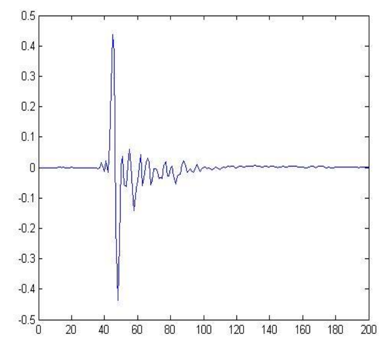
Original HRIR at(80,45)	**Estimated HRIR at(80,45)**
	
Spectrum of the original HRIR at(80,45)	**Spectrum of the estimated HRIR at(80,45)**
	

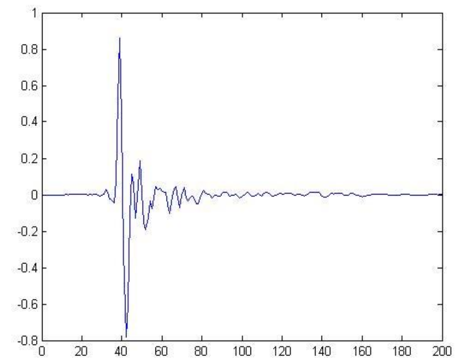
Figure 4. HRIR and its magnitude spectrum at (0,180).

Spectral distortion is 5.2 dB

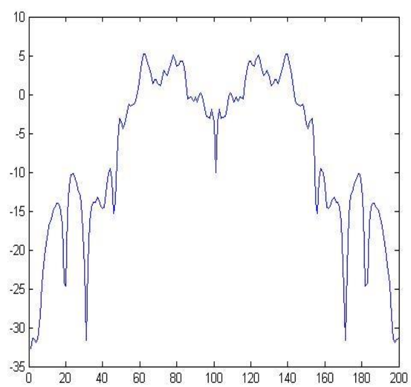
Original HRIR at(0,135)



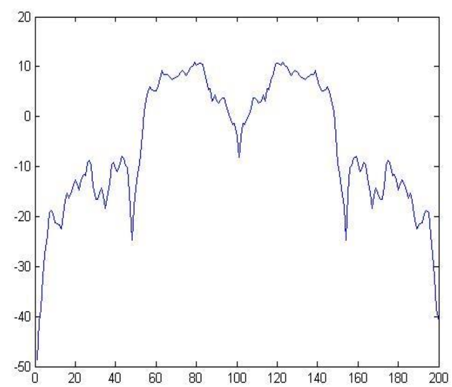
Estimated HRIR at(0,135)



spectrum of the original HRIR at(0,135)



spectrum of the estimated HRIR at(0,135)



9 HRIR and its magnitude spectrum at (0,135).

4. **Conclusion**

According to the importance of the personalization of HRIR for developing virtual reality applications and 3D sound systems, a convenient and efficient method based on three layers (BPANNN) was proposed to estimate an intended listener's HRIR in this study. The results show that there is no significant difference in perception between estimated and retrieved measured HRIRs from the CIPIC database. The personalization method has avoided the following problems:

- Inaccuracy of the HRIR estimation problem.
- Unavailability and difficulty of performing HRIR measurements.
- Avoiding the drawbacks of the previous studies of impulse response personalization.

5. References

1. Begault D. (2000), 3-D Sound for Virtual Reality and Multimedia, Academic Press, California, USA, NASA/TM . 4

2. Cheng CI, Wakefield GH. (2001), Introduction to head-related transfer functions (HRTF): representations of HRTFs in time, frequency, and space (invited tutorial). J Audio Eng Soc 49(4):231–49

3. Gardner W G, Martin K D. HRTF measurements of a KEMAR. J Acoust Soc Am, 1995, 97(6): 3907—3908

4. CIPIC Interface laboratory (2001), CIPIC HRTF Database Files, from: http://Interface.cipic.ucdavis.edu. 9

5. BoSun X. and ZhiQiang L.,(2007), Head-related transfer function database and its analyses , National Natural Science Of China.

6. Algazi V. and Duda R. and Thompson D.(2002),Structural composition and decomposition of HRTF's, IEEE WASPAA01, New Paltz

7. Zotkin D., Duraiswami R., Davis L., Mohan A. and Raykar V.(2002),Virtual_Audio System Customization Using Visual Matching of Ear Parameters , Proceeding IEEE ICPR ,Quebec City, Canada.

8. Zotkin, et al., (2003), HRTF Personalization Using Anthropometric Measurements, Workshop on Applications of Signal Processing to Audio and Acoustic ,IEEE.

9. Hongmei H, et al (2006), Head Related Transfer Function Personalization Based on Multiple Regression Analysis, International Conference on Computational Intelligence and Security.

10. Hongmei Hu1, 2, Liang Chen1, 2, Zhen yang Wu2 (2008), The Estimation of Personalized HRTFs in Individual VAS, IEEE.

11. T. Nishino et al. (2007), Estimation of HRTFs on the horizontal plane using physical features, Applied Acoustics 68.

12. Algazi V.and Duda R.(2002), Approximating the head-related transfer function using simple geometric models of the head and torso", Acoustical Society of America.

13. Kahana Y. (2006), Boundary Element Simulations of the Transfer Function of Human Heads and Baffled Pinnae Using Accurate Geometric Models, Journal of Sound and Vibration.

14. Olfat A. and Zadeh H. (2001), A Neural Network Approach to Magnitude Retrieval, ELSIVER Science, Signal Processing.

15. Hongmei H., Lin Z., Hao M., Zhenyang W. (2008), HRTF Personalization Based on Artificial Neural Network in Individual Virtual Auditory Space, Applied Acoustics.

16. Manlin Zhao , Zhichao Sheng * and Yong Fang, (2022), Magnitude Modeling of Personalized HRTF Based on Ear Images and Anthropometric Measurements , Shanghai Institute for Advanced Communication and Data Science.