HRIR CUSTOMIZATION USING COMMON FACTOR DECOMPOSITION AND JOINT SUPPORT VECTOR REGRESSION

Zhixin Wang and Cheung Fat Chan

City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong
e-mail: zhixiwang2@student.cityu.edu.hk, itcfchan@cityu.edu.hk

ABSTRACT

A two-stage approach for the customization of head-related impulse response (HRIR) for individual subject is proposed. In the first stage, a two-dimension common factor decomposition (2D-CFD) algorithm is applied to extract a subject-dependent impulse response (SDIR) from full HRIR dataset of a subject. The SDIR is then represented as the weighted sum of some principal components using independent component analysis to further reduce the dimensionality of HRIR dataset. In the second stage, joint support vector regression is applied to construct a nonlinear model for mapping the weightings of a target subject from its anthropometric parameters where correlations between different weightings are also exploited. The proposed approach achieves a more accurate and consistent result as compared to the original support vector regression algorithm.

Index Terms— HRIR, CFD, SVR, ICA, Customization

1. INTRODUCTION

Head-Related Impulse Response (HRIR), which captures the filtering effect of human torso, head and pinna to a sound propagating from a specific spatial position to the eardrum of a listener, is the core part in virtual 3D sound synthesis[1]. Although experimental measurements can provide accurate individual HRIR for a perfect virtual 3D sound reproduction, the expensive equipment together with the tedious procedures required in the measurement, has made it impractical for adoptation in commercial applications. It has long been desirable to generate individual HRIR in a more efficient way.

Inspired by the fact that HRIR has closed correlation with the subject's anthropometric parameters, machine learning approach can be applied to train a model between one's HRIR and its anthropometric parameters from a large set of measurements. A target subject's HRIR can be calculated from its anthropometric parameters using the trained model. The CIPIC HRIR database [2] measured by University of California, which contains both HRIR dataset and anthropometric parameters of 37 subjects, has made such idea feasible.

As HRIR is a large dataset, principal component analysis (PCA) or independent component analysis (ICA) is usually applied to reduce its dimensionality before the model training. In [3],[4] PCA is applied to represent each HRIR as the linear combination of some subject-independent principal components and a linear model is trained linking the weightings and the anthropometric parameters. In [5] algorithms for 2-Dimension PCA, Tensor-SVD and generalized low rank approximations of matrices are used for data dimension reduction and a similar model training is conducted after that. As the correlation between HRIR and anthropometric is definitely highly nonlinear, a linear model can only provide inferior performance. To accommodate a more complicated model, support vector regression (SVR) is introduced in [6] to train a nonlinear model and has achieved better result. However, SVR can only train a multiple-to-one mapping model and it needs to train a separate model for the weighting of each components. Different weightings are treated independently in such case though they are clearly correlated. To exploit the correlation between different weightings, a joint SVR (JSVR) algorithm is proposed in this paper. Before JSVR analysis, a 2-dimension common factor decomposition (2D-CFD) algorithm is applied to reduce the HRIR dimensionality by represent it as a subject-dependent impulse response (SDIR) convolved with a common set of direction-dependent impulse responses (DDIRs).

The remaining part of this paper is organized as follows: Section 2 introduces the 2D-CFD algorithm to extract the SDIR and to train DDIRs. Section 3 derived the JSVR algorithm. The proposed approach is evaluated in Section 4.

2. TWO-DIMENSION COMMON FACTOR DECOMPOSITION

In [7], an algorithm based on deconvolution is proposed to extract a common factor for a set of FIR filters, and in [8], a 2D-CFD algorithm is derived to reduce storage and computation requirement in real time 3D sound synthesis. In this section the 2D-CFD algorithm is introduced to train a set of DDIRs common to all subjects from an HRIR dataset containing multiple directions and multiple subjects. An SDIR is extracted simultaneously for each subject.

2.1. Common Factor Decomposition

An impulse response $h[n]$ of length L can be decomposed into two responses, $c[n]$ of length L_1 and $d[n]$ of length L_2 which satisfies $L = L_1 + L_2 - 1$. Given $h[n]$ and $c[n]$, the best $d[n]$ is calculated by

$$\boldsymbol{d} = C^{\dagger}\boldsymbol{h} \tag{1}$$

where $(\bullet)^{\dagger}$ denotes pseudo inverse and

$$\boldsymbol{h} = \begin{bmatrix} h[1] \\ \vdots \\ h[L] \end{bmatrix} \qquad \boldsymbol{d} = \begin{bmatrix} d[1] \\ \vdots \\ d[N] \end{bmatrix} \tag{2}$$

$$C = \begin{bmatrix} c[1] & & & \mathbf{0} \\ \ddots & c[1] & & \\ c[M] & \ddots & \ddots & \\ & c[M] & \ddots & c[1] \\ & & \ddots & \ddots \\ \mathbf{0} & & & c[M] \end{bmatrix}_{L \times N} \tag{3}$$

Given a set of K impulse responses h_k and one individual factor response (IF) d_k for each of them, a common factor response (CF) $c[n]$for all $h_i[n]$ can be calculated by

$$\boldsymbol{c} = \begin{bmatrix} D_1 \\ \vdots \\ D_K \end{bmatrix}^{\dagger} \times \begin{bmatrix} \boldsymbol{h}_1 \\ \vdots \\ \boldsymbol{h}_K \end{bmatrix} \tag{4}$$

where $\boldsymbol{c}$, $\boldsymbol{h}_i$ is constructed as (2) and D_i as in (3).

2.2. Extension to Two Dimension

Given HRIR dataset $\mathbf{H}_k(n, \theta) \in \mathbf{R}^{L \times I}$, where L is HRIR length, I is the number of directions and $k = 1, \cdots, K$ is the subject index, the 2D-CFD algorithm is described as follows:

For HRIR set of each subject $\mathbf{H}_k(n, \theta)$ $k = 1, \cdots, K$
  Apply CFD to this set of HRIRs and get the CF
End for
All K CFs make up a matrix $S \in \mathbf{R}^{L_1 \times K}$
Repeat
  For each angle θ_i $i = 1, \cdots, I$
    Get HRIR in this direction for all subject $\mathbf{H}(n, \theta_i)$
    Use CFs in S as IFs of this HRIR set
    Extract the CF for this angle by (4)
  End for
  All I CFs make up a matrix $D \in \mathbf{R}^{L_2 \times I}$

  For HRIR set of each subject $\mathbf{H}_k(n, \theta)$ $k = 1, \cdots, K$
    Use CFs in D as IFs of this HRIR set
    Extract the CF for this subject by (4)
  End for
  All K CFs make up a matrix $S \in \mathbf{R}^{L_1 \times K}$
Until Convergence

By using this algorithm, the HRIR dataset $\mathbf{H}_k(n, \theta)$ is decomposed into two sets of common factor impulse responses, $S \in \mathbf{R}^{L_1 \times K}$ and $D \in \mathbf{R}^{L_2 \times I}$, where $L_1 + L_2 = L + 1$. Each column of D is a common factor for HRIRs of all subjects at the same direction and is the desired DDIR. Different DDIRs carry the subject-independent directional information contained in original HRIR. Similarly, each collum of S is a common factor for all HRIRs of a specific subject and is named SDIR. The SDIR fully represents a subject. Given a subject's SDIR, its original HRIR is reconstructed by

$$\widehat{\mathbf{H}}_k(n, \theta_i) = D(n, \theta_i) \bigotimes S_k(n) \tag{5}$$

3. JOINT SUPPORT VECTOR REGRESSION

Support vector regression has been widely used to construct nonlinear prediction models and is adopted in this paper to construct a nonlinear model to predict the SDIR from a subject's anthropometric parameters. Before SVR, independent component analysis is used to further reduce the dimensionality of SDIR. After ICA analysis, each SDIR is represented as the weighted sum of some principal components $pc_j(n)$ which are common to all subjects and statistically independent to each other.

$$S_k(n) = \sum_{j=1}^{N} pc_j(n) w_{k,j} \tag{6}$$

The weighting $\boldsymbol{w}_k = [w_{k,1}, \cdots, w_{k,N}]$ fully represents the original SDIR after ICA analysis.

3.1. Support Vector Regression

A detailed introduction of SVR can be found in [9] and only a brief introduction is given here.

Given training data $\{(w_1, \boldsymbol{a}_1)\ \cdots,\ (w_K, \boldsymbol{a}_K)\}$, SVR is aimed to find a function $f(\boldsymbol{a}) = \langle \boldsymbol{m}, \Phi(\boldsymbol{a}) \rangle + b$ with the following objective criteria

$$\text{minimize} \quad \frac{1}{2}\|\boldsymbol{m}\|^2 + C\sum_{k=1}^{K}(\xi_k + \xi_k^*) \tag{7}$$

$$\text{subject to} \begin{cases} w_k - f(\boldsymbol{a}_k) & \leq & \varepsilon + \xi_k \\ f(\boldsymbol{a}_k) - w_k & \leq & \varepsilon + \xi_k^* \\ \xi_k, \xi_k^* & \geq & 0 \end{cases}$$

where, $\Phi(\boldsymbol{a})$ denotes a nonlinear mapping of $\boldsymbol{a}$ to a higher dimension and $\langle \boldsymbol{m}, \Phi(\boldsymbol{a}) \rangle$ means the inner product. $C > 0$

is a constant parameter that determines the trade-off between the flatness of f and cost of the prediction error, and ε is the error tolerance introduced to avoid over fitting in the trained model. This optimization problem can be solved in its dual formulation

$$\text{maximize} - \frac{1}{2}\sum_{i,j=1}^{K}\lambda_i\lambda_j\kappa(\boldsymbol{a}_i,\boldsymbol{a}_j) + \sum_{i=1}^{K}(\lambda_i w_i - \varepsilon\eta_i) \quad (8)$$

$$\text{subject to} \sum_{i=1}^{K}\lambda_i = 0 \text{ and } \lambda_i + \eta_i, \lambda_i - \eta_i \in [0, C] \quad (9)$$

where, $\kappa(\boldsymbol{a}_i,\boldsymbol{a}_j) \equiv \langle\Phi(\boldsymbol{a}_i),\Phi(\boldsymbol{a}_j)\rangle$ is called the kernel function and λ_i, η_i are dual variables. This optimization problem is a convex problem and can be solved using convex optimization. The result of (7) is

$$\boldsymbol{m} = \sum_{i=1}^{K}\lambda_i\Phi(\boldsymbol{a}_i) \text{ and } f(\boldsymbol{a}) = \sum_{i=1}^{K}\lambda_i\kappa(\boldsymbol{a},\boldsymbol{a}_j) + b \quad (10)$$

3.2. Joint Optimization for Vector Regression

As SVR can only construct multiple-to-one mapping functions, a function has to be trained for the weighting of each basic component if it is adopted for SDIR customization. Each training only considers the weighting being dealt with and disregard all other weightings. To exploit the correlation between weightings of different basic components, a joint SVR algorithm is proposed so that all weightings are considered in when finding a mapping function.

Given training data $\{(\boldsymbol{w}_1,\boldsymbol{a}_1)\ \cdots,\ (\boldsymbol{w}_K,\boldsymbol{a}_K)\}$, for the p^{th} component of $\boldsymbol{w}$, joint SVR is aimed to find a function $f_p(\boldsymbol{a}_k) = \langle\boldsymbol{m}_p,\Phi(\boldsymbol{a}_k)\rangle + \langle\boldsymbol{n}_p,\bar{\boldsymbol{w}}_{k,p}\rangle + b$ with the following objective function

$$\text{minimize } \frac{1}{2}(\|\boldsymbol{m}_p\|^2 + \|\boldsymbol{n}_p\|^2) + C\sum_{k=1}^{K}(\xi_k + \xi_k^*) \quad (11)$$

$$\text{subject to } \begin{cases} w_{k,p} - f_p(\boldsymbol{a}_k) & \leq & \varepsilon + \xi_k \\ f_p(\boldsymbol{a}_k) - w_{k,p} & \leq & \varepsilon + \xi_k^* \\ \xi_k, \xi_k^* & \geq & 0 \end{cases}$$

where, $\bar{\boldsymbol{w}}_{k,p}$ denotes $[w_{k,1},\cdots,w_{k,p-1},w_{k,p+1},\cdots,w_{k,N}]$

The dual problem of the Eq. (11) is

$$-\frac{1}{2}\sum_{i,j=1}^{K}\lambda_i\lambda_j[\kappa(\boldsymbol{a}_i,\boldsymbol{a}_j) + \langle\bar{\boldsymbol{w}}_{i,p},\bar{\boldsymbol{w}}_{j,p}\rangle] + \sum_{i=1}^{K}(\lambda_i w_{i,p} - \varepsilon\eta_i) \quad (12)$$

and the result becomes

$$w_{t,p} = f_p(\boldsymbol{a}_t) = \sum_{i=1}^{K}\lambda_i\kappa(\boldsymbol{a}_t,\boldsymbol{a}_j) + \langle\bar{\boldsymbol{w}}_{t,p},\boldsymbol{n}_p\rangle + b \quad (13)$$

where $\boldsymbol{n}_p = \sum_{i=1}^{K}\lambda_i\bar{\boldsymbol{w}}_{i,p}$, $\boldsymbol{a}_t$ is the anthropometric parameters of a target subject and $\boldsymbol{w}_t$ is the weighting of the subject. Define $h_p(\boldsymbol{a}_t) = \sum_{i=1}^{K}\lambda_i\kappa(\boldsymbol{a}_t,\boldsymbol{a}_j) + b$, Eq. (13) can be reformulated as

$$w_{t,p} - \langle\bar{\boldsymbol{w}}_{t,p},\boldsymbol{n}_p\rangle = h_p(\boldsymbol{a}_t) \quad (14)$$

This set of N linear equations has N variables and $\boldsymbol{w}_t$ can be readily solved.

As shown in Eq.14, the proposed JSVR algorithm can only consider a linear correlation between different weightings. However, a nonlinear expansion (NJSVR) can be readily obtained by adding some nonlinear terms, such as $w_{i,1}w_{i,2}$, in $\bar{\boldsymbol{w}}_{i,p}$.

4. EXPERIMENT AND RESULT

4.1. Experiment Setup

The publicly available database which contains HRIRs along with anthropometric parameters on pinna, head and torso sizes measured on 37 subjects by CIPIC is used as source data in this work. There are 1250 pairs of HRIRs measured at a regular position grid containing 50 elevation and 25 azimuth angles for each subject[2]. A minimum-phase version is constructed so that 2D-CFD algorithm performs better. Such version of HRIR is believed to be perceptually indistinguishable with original version[10]. As the spacial resolution at the two sides is much larger, the HRIRs at these positions show less similarity. Besides, HRIRs at low elevation also show less similarity than that at high elevation. In this research, HRIRs at elevation larger than 60 degree are considered and HRIRs at the two sides are excluded in 2D-CFD analysis to reduce distortion.

A 100-sample long SDIR is extracted for each training subject and these SDIRs are modeled as the linear combination of two independent components using ICA algorithm. A nonlinear model between these component weightings and the subjects' anthropometric parameters is then trained using JSVR. Although it is suggested in [9] that radical kernel function $\kappa(\boldsymbol{a}_i,\boldsymbol{a}_j) = e^{-\frac{\|\boldsymbol{a}_i-\boldsymbol{a}_j\|^2}{2\sigma^2}}$ is applicable to most modeling problem, we find that order-2 polynominal kernel $\kappa(\boldsymbol{a}_i,\boldsymbol{a}_j) = \langle\boldsymbol{a}_i,\boldsymbol{a}_j\rangle^2$ has the best performance in this work. The cost parameter C and the error tolerance parameter ε are set to 5 and 0.3 by experiment, respectively.

As there are only 30 subjects for training and 25 anthropometric parameters are measured, to avoid over fitting, the anthropometric parameters are reduced to 10 using PCA analysis.

4.2. Result and Analysis

The performance of the proposed algorithm is evaluated by Spectrum Distortion score(SD) and Waveform Fit score in time domain. Their definitions are:

$$SD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(20\log_{10}\frac{|H_i|}{|\hat{H}_i|})^2} \qquad Fit = 1 - \frac{\sum_{i=1}^{N} e[i]^2}{\sum_{i=1}^{N} x[i]^2}$$

where $e[n] = x[n] - \hat{x}[n]$

The first 30 subjects in the database are used as training subjects to train the common DDIR set and a nonlinear prediction model. The SDIRs of the other 7 subjects, which are used as test subjects, are calculated from their anthropometric parameters by using the trained model. These SDIRs are convolved with the common DDIR to construct customized HRIRs of the test subjects. The results are compared with the measured HRIRs of these subjects.

The proposed JSVR, NJSVR algoritms together with the original SVR implemented in [6] are evaluated here. The average distortions of all 7 subjects over all customized directions are listed in Table 1 and the average distortions of both ears for each subject are shown in Fig. 1 and Fig. 2.

Table 1. Distortion of Customized HRIR

	Left Ear			Right Ear		
	SVR	JSVR	NJSVR	SVR	JSVR	NJSVR
Fit	84.7	90.7	91.3	80.9	87.5	87.4
SD	4.72	4.57	4.48	5.12	4.75	4.75

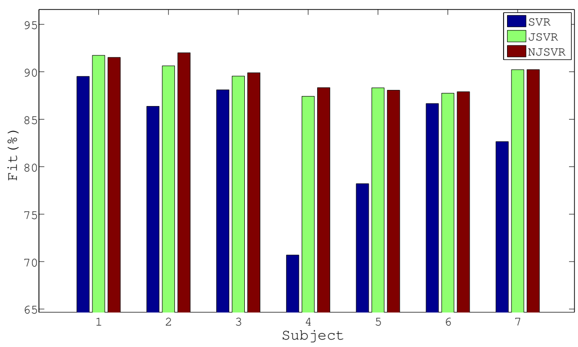

Fig. 1. Fit Score of different subject

From the Table 1, we can see that the proposed JSVR algorithm has better performance than the original SVR algorithm. As shown in Fig. 1, the JSVR algorithm also achieves a more consistent results for different subjects. Beside, we can see that the NJSVR algorithm, which explores nonlinear correlation between different weightings, only has slight improvement compared with JSVR. This may indicate that different weightings are correlated linearly. This is a result of the additive relationship between different weightings as shown in Eq. (6). It's worth while noticing that there is a consistent difference between the distortions of the left ear and the right ear. The same result is also reported in [5] and the cause of this has yet to be determined.

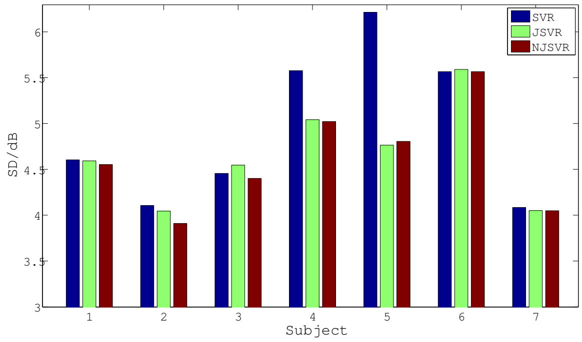

Fig. 2. SD Score of different subject

5. CONCLUSION

A HRIR customization approach based on a two-dimension common factor decomposition (2D-CFD) algorithm and joint support vector regression (JSVR) algorithm is proposed. 2D-CFD is applied first to reduce the HRIR dataset dimensionality for a more efficient customization. A JSVR algorithm is then applied to construct a nonlinear model to predict the weightings that represent a subject's HRIR from its anthropometric parameters. By exploiting the correlation between different weightings, JSVR can achieve a better and more consistent performance.

6. REFERENCES

[1] J. Blauert, *Spatial Hearing*, 55 Hayward Street Cambridge, MA02142-1315 USA, MIT Press, 1983.

[2] V.R. Algazi, R.O. Duda, D.M. Thompson, and C. Avendano, "The cipic hrtf database," in *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, 2001, pp. 99 –102.

[3] H.M. Hu, L. Zhou, J. Zhang, H. Ma, and Z.Y. Wu, "Head related transfer function personalization based on multiple regression analysis," in *Computational Intelligence and Security, 2006 International Conference on*, Nov. 2006, vol. 2, pp. 1829–1832.

[4] S.M. Hwang and Park Y.J., "Hrir customization in the median plane via principal components analysis," in *Audio Engineering Society Conference: 31st International Conference: New Directions in High Resolution Audio*, 6 2007.

[5] M. Rothbucher, M. Dukovic, H. Shen, and K. Diepold, "Hrtf customization using multiway array analysis," in

Signal Processing Conference (EUSIPCO), 2010 Proceedings of the 18th European, aug. 2010, pp. 229 –233.

[6] Q.H. Huang and Q.L. Zhuang, "Hrir personalisation using support vector regression in independent feature space," *Electronics Letters*, vol. 45, no. 19, pp. 1002 –1003, 10 2009.

[7] Claire Masterson, Stephen Adams, Gavin Kearney, and Frank Bol, "A method for head related impulse response simplification.," *17th European Signal Processing Conference (EUSIPCO-2009)*, pp. 2569–2573, Aug. 2009.

[8] Wang Z.X. and C.F. Chan, "Two-dimension common factor decomposition of head-related impulse response," *Signal Processing Letters*, 2013.

[9] Alex J. Smola and Bernhard Scholkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, pp. 199–222, Sep. 2003.

[10] Kistler D. J. and Wightman F. L., "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction.," *Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1637–1647, May 1992.