COMMON FACTOR DECOMPOSITION OF HRTF WITH EFFICIENT IIR MODELING OF DIRECTIONAL FACTORS

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

In this paper, an algorithm to decompose Head-Related impulse response(HRIR) into the convolution of two factor responses is presented. By this Common Factor Decomposition (CFD) algorithm a group of HRIRs can share a common factor(CF) response while differ only in the other factor response which carries all directional information. The directional factor(DF) response is modeled as infinite impulse response (IIR) filter in order to reduce storage. The CF is further represented as IIR filter using Balanced Model Truncation(BMT) algorithm to reduce the computation complexity. The proposed method is shown to be computation and storage efficient for virtual 3D sound synthesis.

Index Terms— HRIR, CFD, IIR, BMT, 3D sound

1. INTRODUCTION

Virtual 3D sound synthesis has been a hot research topic in the past two decades due to the expanding application in surround system, games, home theaters and human aid system. 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]. However, measured HRIR is usually a large dataset. For example, the HRIR database measured by CIPIC has 1250 pairs of HRIR and each HRIR has 200 samples[2]. Nevertheless, this dataset still covers only a finite set of positions. Synthesis of spatial sound not in this set of positions requires interpolation to get unmeasured HRIR first and then convolution to synthesis the spatial sound. This is both storage-consuming and computation-demanding.

As most of the filtering effect caused by reflection and diffraction of torso and head is similar for sound originated from different spatial positions, there is a lot of redundancy in HRIR dataset. Figure 1 shows three HRIRs in left,front and up directions(the differences in time delay are removed), which are very similar. In such case, by representing HRIR as the convolution of a common factor (CF) and a shorter directional factor (DF) the storage can be reduced remarkably. In [3] an algorithm based on deconvolution have been proposed for this purpose with both CF and DF being finite impulse response (FIR) filter. To further reduce the storage, this algorithm is modified such that DF is derived from the response of an efficient IIR filter.

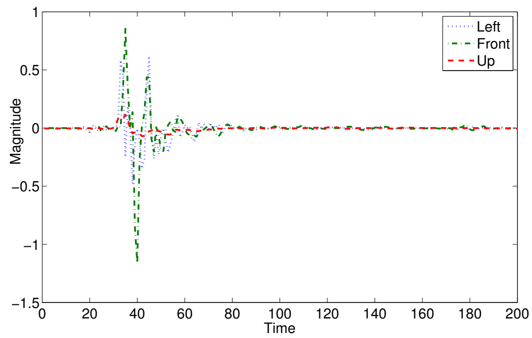

Fig. 1. HRIR of 3 directions

BMT has also been used to reduce the storage and computation in 3D sound synthesis by representing the high order FIR HRIR with more efficient IIR filter[4]. However, problem arises in interpolation if BMT is applied to each HRIR individually. Besides, a comparably high order is required to have an acceptable model distortion and thus the storage requirement is still comparably high. In this paper BMT is applied to the CF after CFD analysis and thus avoid the problem of IIR filter interpolation.

The remaining part of this paper is organized as follows: section 2 outlines the CFD algorithm and in section 3 an algorithm of CFD is derived to get IIR DF, BMT is combined with CFD in section 4, the combined algorithm is detailed in section 5 and its performance evaluated in section 6.

2. COMMON FACTOR DECOMPOSITION

2.1. Factor Decomposition

An impulse response $h[n]$ of length L can be decomposed into two responses, $c[n]$ of length M and another response $d[n]$ of length N which satisfies $L = M + N - 1$. The factorization can be done in the following way such that the serial

 - ISSN 2076-1465

connection of the two factor responses $c[n] \bigotimes d[n]$ has the least disparity compared with $h[n]$.

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

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

$$\boldsymbol{h} = \begin{bmatrix} h[1] \\ h[2] \\ h[3] \\ \vdots \\ h[L] \end{bmatrix} \qquad \boldsymbol{d} = \begin{bmatrix} d[1] \\ d[2] \\ d[3] \\ \vdots \\ d[N] \end{bmatrix} \tag{2}$$

$$C = \begin{bmatrix} c[1] & & & \mathbf{0} \\ c[2] & c[1] & & \\ \ddots & c[2] & \ddots & \\ c[M] & \ddots & \ddots & c[1] \\ & c[M] & \ddots & c[2] \\ & & \ddots & \ddots \\ \mathbf{0} & & & c[M] \end{bmatrix}_{L\times N} \tag{3}$$

After getting $d[n]$, $c[n]$ can be re-optimized in the same way. By such an iteration two factor responses that can model $h[n]$ with very low distortion can be found.

2.2. Common Factor Decomposition

Given a set of K impulse responses $h_1[n], \cdots, h_K[n]$ and one factor response $d_1[n], \cdots, d_K[n]$ for each of them (namely DF), another factor response $c[n]$ which is common for all $h_i[n]$ (namely CF) can be found in the following way

$$\boldsymbol{c} = \begin{bmatrix} D_1 \\ D_2 \\ \vdots \\ D_K \end{bmatrix}^{\dagger} \times \begin{bmatrix} \boldsymbol{h}_1 \\ \boldsymbol{h}_2 \\ \vdots \\ \boldsymbol{h}_K \end{bmatrix} \tag{4}$$

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

After getting the CF $c[n]$, the DF $d_i[n]$ can further be optimized one by one in the way mentioned in section 2.1. By repeating this optimization procedure the set of impulse responses can finally be decomposed into the convolution of a CF and a set of DFs with both of the factors being FIR filters.

3. CFD WITH IIR DIRECTIONAL FACTOR

Although CFD can reduce storage of HRIR dramatically, it still requires a little more storage than principal component analysis(PCA) algorithm. To further reduce this storage, CFD is modified in this section such that DFs are derived as more efficient IIR filters.

Given a length-L impulse response $h[n]$ and a length-M FIR factor response $c[n]$, an order-P all-pole IIR filter $\frac{1}{1+\sum_{i=1}^{P} a_i z^{-i}}$ can be found so that the serial connection of the two factors responses can model $h[n]$ well. The response of this IIR filter $d[n]$ is a function of a_i, which is denoted as

$$d[n] = f_n(a_i), i = 1, \cdots, P, \ n = 1, 2, \cdots, \infty \tag{5}$$

Then $d[n]$ is truncated to length N such that $M+N = L+1$ and the impulse response of the serially connected system can be written as $\hat{h}[n] = c[n] \bigotimes d[n]$ approximately. The square error between this impulse response and $h[n]$ is

$$\xi = \sum_{i=1}^{L} \xi_i = \sum_{i=1}^{L} (h[i] - \hat{h}[i])^2 \tag{6}$$

To use Newton iteration to minimize ξ, the derivative of ξ over a_k should be calculated

$$\frac{\partial \xi}{\partial a_k} = \sum_{i=1}^{L} \frac{\partial \xi_i}{\partial a_k} = \sum_{i=1}^{L} [-2(h[i] - \hat{h}[i]) \frac{\partial \hat{h}[i]}{\partial a_k}] \tag{7}$$

as

$$\frac{\partial \hat{\mathbf{h}}}{\partial a_k} \equiv \begin{bmatrix} \frac{\partial \hat{h}[1]}{\partial a_k} \\ \frac{\partial \hat{h}[2]}{\partial a_k} \\ \vdots \\ \frac{\partial \hat{h}[L]}{\partial a_k} \end{bmatrix} = \frac{\partial C\boldsymbol{d}}{\partial a_k} = C \frac{\partial \boldsymbol{d}}{\partial a_k} \equiv C \begin{bmatrix} \frac{\partial f_1}{\partial a_k} \\ \frac{\partial f_2}{\partial a_k} \\ \vdots \\ \frac{\partial f_N}{\partial a_k} \end{bmatrix} \tag{8}$$

here C and $\boldsymbol{d}$ is constructed as (3) and (2), respectively. So

$$\frac{\partial \hat{h}[n]}{\partial a_k} = c[n] \bigotimes \frac{\partial d[n]}{\partial a_k} \tag{9}$$

where $\frac{\partial d[n]}{\partial a_k}$ for all n and k is calculated using (10)

$$\frac{\partial d[n]}{\partial a_k} = \sum_{i=1}^{P} a_i \frac{\partial d[n-i]}{\partial a_k} + d[n-k] \tag{10}$$

$$\frac{\partial d[n]}{\partial a_k} = \frac{\partial d[n-I]}{\partial a_{k-I}} \quad I = 1, \cdots, k-1 \tag{11}$$

formula (11) helps to reduce this computation by simplifying the calculation of $\frac{\partial d[n]}{\partial a_k}$ for $k \geq 2$. It is proved in Appendix.

After all the optimum IIR DFs are found, their impulse responses are calculated and truncated to FIR filters. Then the CF is optimized again using the method in section 2.2.

4. COMBINE CFD WITH BMT ANALYSIS

Although CFD can reduce the storage of HRIR dramatically, the computation complexity in the synthesis stage remains unchanged. In this paper, BMT analysis which can model a high order FIR filter with a low order pole-zero IIR filter, is applied to the CF to reduce the computation complexity.

4.1. Balanced Model Truncation

The detail of BMT can be found in [5] and only a principal idea is given here. A system can be represented in state-space model:

$$y[n] = C \cdot s[n] \tag{12}$$

$$s[n+1] = A \cdot s[n] + B \cdot x[n] \tag{13}$$

here, $x[n]$ is the system input, $y[n]$ is the system output. This system is fully determined by matrix triplet (A, B,C) has transfer function

$$f(z) = C(z \cdot I - A)^{-1}B \tag{14}$$

where I is unit matrix.

By a transform T that doesn't change its transfer function

$$\hat{A} = T^{-1}AT^{-1} \quad \hat{B} = T^{-1}B \quad \hat{C} = CT^{-1} \tag{15}$$

we can change the original system into a balanced system in which the state that contributes little to the output of the system also requires a large input to drive it from zero. Such states can be removed without greatly affecting the output of the system and the order of the system is thus reduced.

4.2. Combining CFD with BMT

Figure 2 shows the synthesis structure after CFD analysis, where φ and θ are azimuth and elevation angles respectively. This approach requires as much computation as using measure HRIR directly.

$$\text{input} \rightarrow \boxed{c[n]} \rightarrow \boxed{d_{\varphi,\theta}[n]} \rightarrow \text{output}$$

Fig. 2. Synthesis Structure of CFD model

It is propose that the high order FIR CF is modeled with a low order IIR filter by BMT analysis to reduce the computation complexity. The new synthesis structure is shown in Figure 3.

$$\text{input} \rightarrow \boxed{\frac{A[n]}{1+B[n]}} \rightarrow \boxed{d_{\varphi,\theta}[n]} \rightarrow \text{output}$$

Fig. 3. Synthesis Structure of CFD+BMT Model

BMT modeling will inevitably deviate the CF from the optimum one. To reduce the distortion brought along by BMT modeling to the lowest level, an optimization procedure is performed after BMT modeling: the zero part of the pole-zero IIR filter A[n] is convolved together with the DF first and apply CFD again to recompute $A[n]$ and all DFs.

5. COMBINED ALGORITHM

The algorithm for CFD with IIR DF is a two level iteration: Newton iteration is the inner loop to optimize the IIR coefficients of the DF and the outer loop is to optimize the FIR CF. The initial value of CF is derived from the truncation of the first PCA components of the given HRIR dataset.

Initialize the CF and Calculate FIR DFs using (1)
Model DF with IIR filter using LPC

repeat (outer loop)
  for each HRIR $h[n]$
    repeat (inner loop)
      Obtain $\frac{\partial \xi}{\partial a_k}$ from (7,9)
      Update $a_k \leftarrow a_k + step \times \frac{\partial \xi}{\partial a_k}$
      Check stability and truncation error of $\frac{1}{1-\sum a_k z^{-k}}$
      if unstable or large truncation error
        Restore previous a_k
        Exit inner loop
      endif
    until Convergence
    Truncate the response of $\frac{1}{1-\sum a_k z^{-k}}$ to FIR filter
  endfor
  Optimize CF using (4)
until Convergency

Represent CF with IIR filter $\frac{A[n]}{1+B[n]}$
Convolve $A[n]$ with $d_{\varphi,\theta}[n]$
Recompute $A[n]$ and all DFs again using (4,7,9)

When DF is derived as IIR filter in CFD, the problem of DF interpolation emerges. However, this problem can be avoided by interpolating on the filter output rather than on the filter coefficients. In this way, if a 3D sound is to be generated at an unmeasured position, the output at the four neighbor positions are calculated and then interpolated. As the impulse responses of these positions share a CF and differ only by a low order IIR filter, the increase in computation to calculate all the output of the neighbor positions is small.

6. EXPERIMENT AND RESULT

The Subject 3 HRIR in the publicly available HRIR database measured 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]. Before CFD analysis, all the 2500 HRIRs are aligned using covariance analysis to remove the differences in time delay. Then a minimum-phase version is constructed so that BMT algorithm can perform better[5]. Such version of HRIR is believed to be perceptually indistinguishable with original version[6]. To exploit the most redundancy, HRIR is divided into groups according to azimuth angle as in[7] and a CF is extracted for each group.

In CFD with FIR DF (CFD1) analysis, an order-195 FIR CF is extracted for each group and an order-6 FIR DF for each HRIR. While in CFD with IIR DF(CFD2) analysis, an order-150 FIR CF is extracted for each group and an order-4 IIR DF for each HRIR. In CFD2 iteration the impulse response of IIR DF is truncated to length 51. After CFD analysis, the CF is modeled by an order-10 IIR filter using BMT.

The performance of this model is evaluated by Spectrum Distortion score(SD) in spectrum domain and waveform Fit score in time domain. The definitions for them are:

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

$$Fit = (1 - \frac{\sum_{i=1}^{N} e[n]^2}{\sum_{i=1}^{N} x[n]^2}) \times 100\%$$
$$where\ e[n] = x[n] - \hat{x}[n]$$

Two other implementation schemes are also introduced here: in [4] BMT is applied to each HRIR directly and in [7] BMT is combined with PCA to solve the problem in IIR filter interpolation. In this paper, an order-4 IIR filter is extracted for each HRIR in BMT implementation scheme. The interpolation problem in this scheme is also avoided by interpolating on the filter output. In PCA+BMT scheme 4 principal components are extracted for each group and each components is modeled with an order-12 IIR filter. Their distortion together with storage requirement and computation complexity are shown in Table 1, in which the computation refer to the synthesis complexity in term of the number of multiplication and addition needed to output a sound sample.

	Fit(%)	SD(dB)	Storage	Computation
BMT	89.6	6.38	8	32
PCA+BMT	90.9	4.86	4	96
CFD1	91.6	4.28	6	200
CFD1+BMT	90.8	4.61	6	26
CFD2+BMT	90.0	4.63	4	40

Table 1. Efficiency Comparison of Five Implementations

PCA+BMT is very efficient in storage. However, it still suffers from comparably high computation complexity. Furthermore, if less distortion is required this complexity will further increase remarkably(increased by 24 with every increase of PCA component).

CFD with FIR DF can reduce the computation complexity dramatically. Besides, little increase in computation complexity is required if we want to improve the distortion performance. For example, when the computation complexity is increased from 26 to 31 multiplication and addition, the Fit score can be improved to 94.9% while 6 principal components are needed for PCA+BMT scheme to get the same model accuracy performance, which requires 144 multiplication and addition. However, this scheme suffers from slightly increased storage requirement.

When DF in CFD analysis is extracted as IIR filter, the storage is reduced due to the highly efficient IIR representation. Although the computation complexity is increased because we have to calculate all the output of four neighbor positions, it's still much small than PCA+BMT.

Besides, we can see that combining BMT to CFD only causes slight degradation in model accuracy while the computation complexity is reduced dramatically. This is due to the re-optimization step after BMT analysis which is aimed to bring down the distortion caused by BMT modeling.

7. CONCLUSION

In this paper, an algorithm based on CFD and BMT is proposed to synthesis virtual 3D sound which is more computation efficient as compared with PCA+BMT algorithm. By using IIR DF in CFD analysis, this algorithm performs as good as PCA+BMT in reducing storage requirement while requires much less computation complexity.

8. REFERENCES

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

[4] J. Mackenzie, J. Huopaniemi, V. Valimaki, and I. Kale, "Low-order modeling of head-related transfer functions using balanced model truncation," *Signal Processing Letters, IEEE*, vol. 4, no. 2, pp. 39 –41, feb 1997.

[5] B. Beliczynski, I. Kale, and G.D. Cain, "Approximation of fir by iir digital filters: an algorithm based on balanced model reduction," *Signal Processing, IEEE Transactions on*, vol. 40, no. 3, mar 1992.

[6] N A Schellart and J C De Munck, "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, 1987.

[7] Zhixin Wang and Cheung-Fat Chan, "Efficient implementation of virtual 3d sound synthesis based on combining grouped pca and bmt," in *Acoustics, Speech and Sig-*

nal Processing (ICASSP), 2011 IEEE International Conference on, may 2011, pp. 449 –452.

9. APPENDIX

$$\frac{\partial d[n]}{\partial a_k} = \frac{\partial d[n-I]}{\partial a_{k-I}} \quad k = 2, \cdots, K \quad I = 1, \cdots, k-1$$

Proof

This equation can be proved by mathematical induction. It is obvious that

$$d[n] = 0 \quad n < 0$$

so

$$\frac{\partial d[n]}{\partial a_k} = 0 = \frac{\partial d[n-1]}{\partial a_{k-1}} \quad n < 0, k = 2, \cdots, K$$

Suppose the equation holds for $n < N$,then

$$\begin{aligned} \frac{\partial d[N]}{\partial a_k} &= \frac{\partial(\delta[n] + \sum_{i=1}^{K} a_i d[N-i])}{\partial a_k} \\ &= \sum_{i=1}^{K} a_i \frac{\partial d[N-i]}{\partial a_k} + d[N-k] \\ &= \sum_{i=1}^{K} a_i \frac{\partial d[N-i-1]}{\partial a_{k-1}} + d[N-k] \\ &= \frac{\partial d[N-1]}{\partial a_{k-1}} \end{aligned}$$

By simple induction the above equation can be obtained.