Multimed Tools Appl (2014) 70:2167–2188
DOI 10.1007/s11042-012-1234-6

Optimally truncating head-related impulse response by dynamic programming with its applications

Shingchern D. You · Woei-Kae Chen

Published online: 13 September 2012

Abstract We propose a method to optimally truncate the head-related impulse responses (HRIRs) in this paper. The truncated HRIR consists of a portion of the original HRIR and a flat line. An algorithm based on dynamic programming is used to optimally select the portions of the original HRIRs and the constants of the flat lines to minimize the modeling errors. The truncated HRIRs can be used to reproduce multi-channel sound for headphones with a significantly lower computational cost. The proposed method is compared with another approximation method, the CAPZ (Common-Acoustical-Pole and Zero) approach. The experimental results show that the proposed method yields lower composition as well as modeling errors for the same amount of computation. Compared with the direct implementation, the proposed approach requires about 35 % of the computational cost while maintaining acceptable composition errors.

Keywords HRTF · CAPZ · Multi-channel · Dynamic Programming · Headphones

1 Introduction

With the popularity of DVD (Digital Versatile Disc) and BD (Blu-ray Disc), a video program frequently contains multiple channels of audio signals, encoded with AC-3 [1], DTS [7], MPEG-2/4 AAC [14], or other technologies. A typical audio format in such a program is 5.1-channel, where ‘5’ represents five full-bandwidth channels, namely, left (L), center (C), right (R), left surround (LS), and right surround (RS), and the ‘0.1’ represents the low-frequency effect (LFE) channel. The acoustic sounds of the full-bandwidth channels in a typical setting are individually reproduced by loudspeakers placed at the locations specified

S. D. You (✉) · W.-K. Chen
Department of Computer Science and Information Engineering, National Taipei University of Technology, 1, Sec. 3, Chung-Hsiao East Rd., Taipei, Taiwan
e-mail: you@csie.ntut.edu.tw

W.-K. Chen
e-mail: wkc@csie.ntut.edu.tw

by the channel names. For example, the loudspeaker for LS channel is placed on the back left of the listener's position.

Sometimes, a listener may have to use headphones when watching DVD/BD programs to avoid, among other considerations, disturbing others. In this case, it is necessary to feed multi-channel audio signals into headphones, which typically have only two transducers. The simplest method to reproduce multi-channel audio using headphones is to perform a down-mix operation, under which audio signals of C, L, and LS are mixed into left channel, and C, R, and RS signals are mixed into right channel. Unfortunately, this method is not satisfactory because the front-back spatial information is totally lost during the down-mix process. To preserve spatial sensation, it is required to use head-related impulse responses (HRIRs) [3] during the reproduction process.

One concern of using HRIRs in reproducing spatial sound using headphones is its high computational demand [17], which is expensive to implement for commercial applications. In this paper, we present a method to approximate the HRIR with fewer coefficients so that the computational cost can be reduced. In the proposed method, an HRIR is approximated by a portion of the original response and a flat line. To optimally truncate the original response, an algorithm based on dynamic programming is adopted. With the proposed approach, the required computation for reproducing spatial sound using headphones can be significantly reduced.

The rest of the paper is organized as follows. Section 2 gives a brief review of spatial hearing and HRIR, and also explains how to reproduce multiple-channel audio using headphones. Section 3 gives a brief survey of existing approaches. Section 4 describes the proposed HRIR approximation method and its complexity. An efficient search algorithm to optimally truncate the HRIRs is presented in section 5. In addition to the proposed approach, we also implement the common-acoustical-pole and zero (CAPZ) method [11] as the comparison target of our approach. Section 6 presents the experimental results for both approaches, and section 7 gives the conclusions. For completeness, the complexity analysis of the exhaustive search is given in appendix A.

2 Spatial hearing and reproducing multi-channel audio using headphones

2.1 Spatial hearing

The psychoacoustic studies [3] reveal that the main factors of the spatial hearing are the inter-aural time difference (ITD) and inter-aural intensity difference (IID), among other factors. Whereas it is possible to model the behavior of spatial hearing using mathematics [5], it is much more popular to conduct experiments to obtain empirical numerical models [9, 13]. Specifically, if the path between the sound source and the eardrum is modeled as a linear system, the system's (finite duration) impulse response can be experimentally measured. The concept has been realized by various research groups, including the MIT (Massachusetts Institute of Technology) Media Lab [9] and the group in IRCAM and AKG [13].

In Gardner and Martin's studies [9, 10] in the MIT Media Lab, they placed a dummy, equipped with different sizes of pinnae, and a loudspeaker in an anechoic room. Then, they used the loudspeaker to produce pseudorandom noise. The microphones inside the dummy's ears received the noise, which was used to compute the impulse responses between the loudspeaker and the microphones. The results were many sets of transfer functions with zeros only, called head-related transfer functions (HRTFs). Since there are only zeros in the measured HRTFs, they can also be regarded as head-related impulse responses (HRIRs). The influences of IID, ITD, and the shapes of pinnae are implicitly represented by the coefficients of the HRIRs.

In the following discussion, we will use elevation and azimuth angles to describe the relative direction of the sound source to the head. A positive elevation angle indicates that the sound source is higher than the horizontal plane intercepting the ears. The elevation angle, throughout the discussion of the paper, is always set to zero. The azimuth angle is measured in clockwise direction with the dummy facing zero degree. Figure 1(a) shows a sound source at an azimuth of 45°. From Fig. 1(a), the right ear is closer to the sound source; therefore the sound reaches the right ear faster and stronger. On the other hand, the sound source is farther to the left ear and is partially impaired by the head; thus, the left ear receives the sound with a longer delay and weaker intensity as shown in Fig. 1(b) [9]. Thus, IID and ITD are implicitly represented by the differences of magnitude and time delay in responses between left ear and right ear.

During the MIT's measurements, the dummy was equipped with different sizes of pinnae in left and right ears in order to collect two sets of HRTF datasets corresponding to different sized ears. These two datasets are available in the full-length version (512 coefficients for

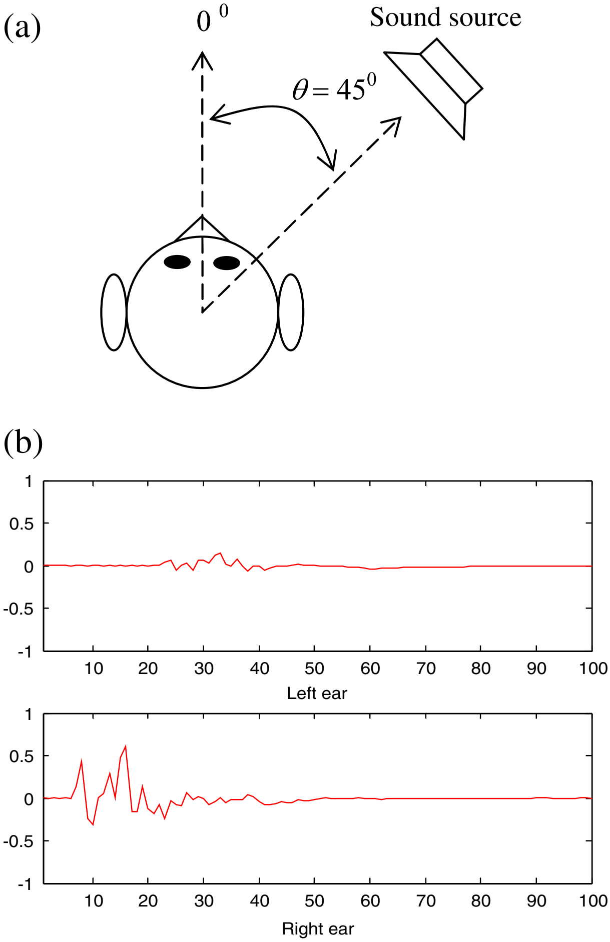


Fig. 1 (**a**) The sound source placed at 45° azimuth angle. (**b**) The corresponding head-related impulse responses

each HRIR). For the compact version (128 coefficients for each HRIR), the symmetry assumption (Eq. 1) is used to produce HRIRs for both ears from one ear, i.e.

$$h_{\alpha,R} = h_{(360-\alpha),L} \tag{1}$$

where $h_{\alpha,R}$ is the impulse response relating the right ear and the sound source at an $\alpha°$ azimuth, and $h_{(360-\alpha),L}$ is the impulse response relating the left ear and the same sound source at a $(360-\alpha)°$ azimuth. Therefore, the compact version of HRIR dataset actually contains measurement results from one ear. In our experiments, we follow the MIT's convention and use the symmetry assumption to obtain the responses of the right ear from left ear.

In contrast to the MIT's experiment where a dummy was used, researchers in IRCAM and AKG [13] used 'real' persons in the experiments. Since the left and right pinnae of a real person are not necessarily identical in size and shape, HRIRs for both ears had to be measured. Therefore, the obtained HRIRs are asymmetric, and therefore Eq. 1 does not hold. In this case, the impulses of $h_{\alpha,R}$ and $h_{\alpha,L}$ are both available in the dataset. With the impulse responses for both ears, we may reproduce spatial sound at the angle of α degrees, as to be discussed next.

2.2 Reproducing multi-channel audio using headphones

Given the HRIRs, it is not difficult to reproduce spatial sound using headphones. Suppose that $h_{\alpha,L}[n]$ and $h_{\alpha,R}[n]$ represent the impulse responses relating left and right ears and a sound source at $\alpha°$ (azimuth). If the signal $x[n]$ is to be reproduced with the sensation of the same azimuth angle, we may use the following equations

$$\begin{aligned} y_L[n] &= x[n]*h_{\alpha,L} \\ y_R[n] &= x[n]*h_{\alpha,R} \end{aligned} \tag{2}$$

where '*' is the convolution operator and $y_L[n]$ and $y_R[n]$ are signals to drive the left and right transducers of the headphones. Note that Eq. 2 does not take the impulse responses of the headphones into account, because, for commercial applications, it is generally not possible to know the impulse responses of the headphones the listener may use in advance. In the following presentation, if appropriate, we'll drop the time index n for brevity.

It is straightforward to extend Eq. 2 to multiple sound sources at different azimuth angles - simply sum up the convolution results for each of the sound sources with its corresponding HRIRs. Let x_L, x_R, x_C, x_{LS} and x_{RS} denote the signals in the L, R, C, LS, and RS channels (with time index n being dropped). If the sound sources are to be reproduced (using headphones) with spatial angles at α,β,0 (zero), γ, and δ degrees respectively, the composite signal y_R can be calculated as

$$\begin{aligned} y_R &= (x_L*h_{\alpha,R} + x_R*h_{\beta,R} + x_C*h_{0,R} + x_{LS}*h_{\gamma,R} + x_{RS}*h_{\delta,R})/5 \\ &= \tfrac{1}{5} \sum_{i=\{L,R,C,LS,RS\}} x_i*h_{\theta(i),R} \end{aligned} \tag{3}$$

where $\theta(i)$ represents one of the azimuth angles. The composite signal y_L can be obtained by an equation similar to Eq. 3. In the above calculation, the LFE channel is not considered because it is not associated with any particular spatial direction. If necessary, it can be easily mixed into y_L and y_R. In the rest of the paper, if not causing confusion, we shall also drop the subscript R and L for brevity. In Eq. 3, the values of α, β, γ and δ may be arbitrarily chosen as long as the angles are reasonable. However, it is a common practice to place the loudspeakers at symmetrical locations. That is,

$$\alpha = 360 - \beta \,\text{and}\, \gamma = 360 - \delta. \tag{4}$$

With these constraints, one set of possible angle values is as follows: α=315, β=45, γ= 125, and δ=235 (degrees).

Though straightforward, direct implementation of Eq. 3 is computationally expensive. The HRIR measurements provided by the MIT's Media Lab contain at least 128 coefficients (compact version) for each impulse response $h[n]$. Therefore, computing a single sample of y_L and y_R requires 128×10=1280 multiply-and-accumulate (MAC) operations. To process an audio program in a DVD with a sample rate of 48 ks/s, a total of 61,440,000 MAC operations per second is required. Since most digital signal processors execute one MAC operation in one instruction, implementation of Eq. 3 requires at least 61.4 MIPS (million instructions per second). As a figure for comparison, a low-cost MP-3 encoder consumes about 35 MIPS [20]. Since an MP-3 encoder needs to perform a series of complicated operations [19, 20] including filter bank, Modified Discrete Cosine Transform, and nonlinear quantization, spending 61.4 MIPS solely for the task of reproducing spatial sound is apparently too expensive. Therefore, it is desirable to reduce the amount of computations.

3 Related work

One of the challenges of using HRIRs is the high computational cost. To reduce the cost, Sakamoto et al. [17] proposed to divide the input signals into three frequency bands, and each band uses a different filter to approximate a portion of the HRTF. By doing so, they achieved a performance of about 50 MIPS for monaural (one channel) inputs. For low-cost digital signal processors, this figure is still too high.

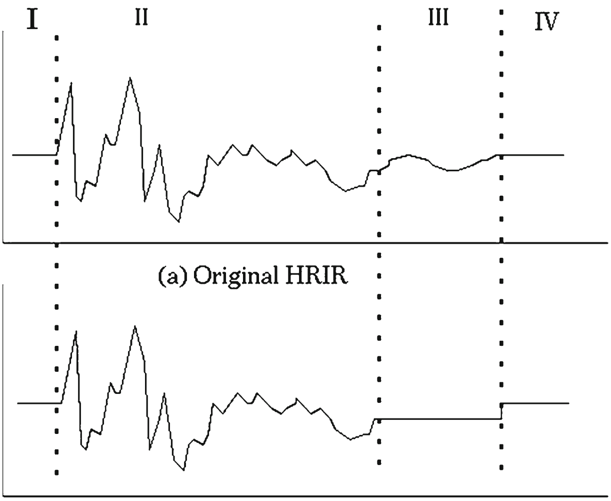
A large number of HRIR measurements has been available and has drawn many attentions to compress or approximate them. By using twelve principal components [12], it is possible to approximate the individual HRIR within 5 % of modeling error. Though this approach offers space savings, it cannot be used to reduce computational cost at run-time, because the convolution operations required in Eq. 3 cannot be performed directly with principal components. Therefore, it is necessary to convert the principal components into the resulting approximated HRIR before the convolution. Since the approximated HRIR has the same number of coefficients as the original one, this method does not provide any computational savings if Eq. 3 is in use.

Since the measured HRIRs are of finite duration, various researchers have proposed using pole-only [15, 16] or pole-zero transfer functions to approximate HRIRs [4, 8]. One concern of the pole-only approximation is that it is more sensitive to numerical errors if fixed-point arithmetic is to be performed. Unfortunately, fixed-point arithmetic is usually the only arithmetic available in a low-cost digital signal processor. For the pole-zero approaches, the locations of poles and zeros can be obtained through various search methods. In this approach, different sets of poles are used for different HRTFs (corresponding to different source locations) [4, 8]. However, in reality there is a set of common poles in all HRTFs due to the resonance of the ear canal. It is the basic idea of the CAPZ (Common-Acoustical-Pole and Zero) model [11]. By using common poles, it is possible to further reduce the computational cost of Eq. 3 (to be described in section 6.1). Since the CAPZ model is a better approach for HRIR approximation, we will use it as a reference to evaluate the performance of the proposed approach in section 6.

4 The proposed approximation approach

When closely examining a typical HRIR, we observe that it can be roughly divided into four sections, as shown in Fig. 2(a). The first and the fourth sections have very small coefficient values, the second section has large coefficient values, and the third section is a small curve. Since the first and fourth sections contain very small coefficient values, these values can be simply set as zero without losing too much spatial information. On the other hand, coefficients of the second section are large values, implying higher importance, and therefore as many of them should be retained as possible. The third section, a small curve, is less important than the second section and can be approximated by a flat line. The approximated impulse response is shown in Fig. 2(b). With the proposed approximation, the computational cost of performing Eq. 3 is reduced. Although the idea looks simple, finding the optimal locations and lengths of the second sections for all of the HRIRs requires special techniques, which will be given in section 5.

The idea of dividing one HRIR into four sections is based on the following physical phenomena. First, left and right ears of a listener do not hear the sound at the same time if the azimuth angle of the sound direction is nonzero. The time difference is known as the ITD (given in section 2.1). Therefore, the pair of HRIRs associated with the left and right ears have different 'silent' time before exhibiting large magnitude responses. Consequently, any HRIR has the first section (called section I in the following). Since the diameter of a human head is around 20 cm, the time difference for the impulse sound reaching two ears in room temperature is (at most) around 0.6 ms, or equivalently 26 samples with a sample rate of 44.1 ks/s. Once the impulse sound reaches the ear, the sound wave is partially reflected by the pinna, head, shoulder, and torso. Through the ear canal, finally the sound wave reaches the ear drum. The overall effect of these factors normally produces an impulse response with duration of (less than) around 1.5 ms [12] (or 66 samples with 44.1 ks/s sample rate). As the ear canal acts as a resonant system, its magnitude impulse response fluctuates after receiving the impulse and then gradually decays. The part of response with large magnitude constitutes the second section (section II), and that with smaller magnitude constitutes the third section (section III). After the 1.5 ms duration, the response is very small and may be ignored.



(a) Original HRIR

(b) Truncated HRIR with flat-line

Fig. 2 Original HRIR and approximated HRIR

This part is the forth section (section IV). Based on our discussion, an HRIR may need up to 26+66=92 samples to represent. In the MIT's compact version, each HRIR has 128 samples (with a sample rate of 44.1 ks/s). Therefore, every HRIR in the compact version satisfies the above requirement. In fact, when we examine the HRIR dataset provided by MIT, we find that all of the HRIRs possess four sections.

We now consider the computational complexity of the proposed approach. Let the original impulse response be

$$h = h^{I} + h^{II} + h^{III} + h^{IV} \tag{5}$$

where h^{I} to h^{IV} contain the coefficients of h in sections I to IV. Then, the approximated impulse response is

$$\widehat{h} = h^{II} + \widehat{h}^{III} \tag{6}$$

where $\widehat{h}^{III}$ is the flat-line approximation of section III. Therefore, the convolution becomes

$$\widehat{y} = \widehat{h}*x = h^{II}*x + \widehat{h}^{III}*x \tag{7}$$

For the first term $h^{II}*x$, it is difficult to reduce the complexity. However, for second term $\widehat{h}^{III}*x$, we may efficiently compute it. Since $\widehat{h}^{III}$ is a flat line, it can be represented as $\widehat{h}^{III}[n] = k_0, m^{III} \le n < m^{III} + N^{III}$, where k_0 is a constant, m^{III} is the starting point of section III, and N^{III} is the length of section III. With this notation, we know

$$\begin{aligned}
\widehat{y}^{III}[n] &= \widehat{h}^{III}[n]*x[n] \\
&= \sum_{m=m^{III}}^{m^{III}+N^{III}-1} \widehat{h}^{III}[n]x[n-m] \\
&= k_0\left(\sum_{m=m^{III}}^{m^{III}+N^{III}-1} x[n-m]\right) \\
&= k_0\left(\sum_{m=m^{III}+1}^{m^{III}+N^{III}} x[n-m] + x[n-m^{III}] - x[n-m^{III}-N^{III}]\right)
\end{aligned} \tag{8}$$

Since m^{III} and N^{III} are independent of n, we know $\widehat{y}^{III}[n-1] = k_0\left(\sum_{m=m^{III}}^{m^{III}+N^{III}-1} x[n-m-1]\right) = k_0\left(\sum_{m=m^{III}+1}^{m^{III}+N^{III}} x[n-m]\right)$. Therefore, from Eq. 8, we have $y^{III}[n] = k_0(y^{III}[n-1] + x[n-m^{III}] - x[n-m^{III}-N^{III}])$, which can be efficiently computed with one multiplication, one addition, and one subtraction. Therefore, the cost of computing Eq. 6 is mainly determined by the length of section II. Consider the following case. If $N^{II}_{\theta(i)}$ is the number of coefficients in $h^{II}_{\theta(i)}$ and F is the number of impulse responses involved in Eq. 3, then the required number of MAC operations to generate a pair of output samples for both channels is (with F subtraction operations ignored)

$$C_{prop} = 2 \cdot \sum_{i=1}^{F} \left(N^{II}_{\theta(i)} + 1\right) = 2 \cdot (M + F) \tag{9}$$

where M is the sum of all $N^{II}_{\theta(i)}$. Please note that in an actual application, the exact location of each division is pre-computed. Therefore, the parameter M is the only factor related to the computational cost of reproducing multi-channel sound for headphones.

In the experiments shown in Section 6, we use Eq.1 to obtain all necessary impulse responses from the MIT's dataset. However, in order to cover the general case, the presentation of this paper does not take the advantage of the symmetry property (i.e. Eq. 1) in computing y_L and y_R. Doing so can further reduce the computational cost [18]. It is straightforward to modify the proposed approach to incorporate such a change.

Based on Eq. 9, if M=220 and F=5, the proposed approach reduces the computation cost to about 35.2 % of the direct implementation. Thus, if a direct implementation requires 61.4 MIPS (given in section 2.2), the proposed approach with M=220 can be implemented with 21.6 MIPS, which is acceptable for many embedded devices. The modeling and composition errors in this case will be given in section 6. If the approach of [18] is also used, the computation cost is further reduced to around 10.8 MIPS.

5 Search algorithm for optimal truncation

5.1 Objective function for optimality

The key issue of the proposed approach is that, given M, how to optimally truncate the original HRIRs to minimize the error between h and $\widehat{h}$ (called modeling error). To do so, we need to define the objective (cost) function to be minimized first. For demonstration purposes, we use mean-square error to represent the modeling error, although other objective functions can also be easily applied to the proposed search algorithm. That is, we define the individual modeling error of an HRIR as either (without normalization)

$$e_i = \sum_{n=0}^{N-1} \left(h_{\theta(i)}[n] - \widehat{h}_{\theta(i)}[n] \right)^2 \tag{10}$$

or (with normalization)

$$e_i = \frac{\sum_{n=0}^{N-1} \left(h_{\theta(i)}[n] - \widehat{h}_{\theta(i)}[n] \right)^2}{\sum_{n=0}^{N-1} h^2_{\theta(i)}[n]} \tag{11}$$

where N is the length of the impulse response. We'll then examine whether normalization is necessary based on experiments. The average modeling error is calculated as

$$e_{MD} = \frac{1}{F} \sum_{i=1}^{F} e_i \tag{12}$$

where F is the number of impulse responses to be approximated. Therefore, given a desired value of M, an optimal approximation is achieved if e_{MD} is minimized. Since taking the logarithm operation on e_{MD} does not affect the optimization criterion, the modeling error can also be represented in dB.

5.2 Computational complexity for minimization

In the proposed approach, a good approximation of h heavily depends on how h is partitioned into h^I, h^{II}, h^{III}, and h^{IV}, especially on the length and the starting point of h^{II}.

Once h^{II} is determined, h^{I} is fixed, and since h^{IV} can be easily determined by a predefined threshold, h^{III} can also be computed.

We now consider the complexity of the search problem that determines the optimal lengths and starting points for all $h^{II}_{\theta(i)}$. Following the notation in Eq. 9, let M be the sum of all $N^{II}_{\theta(i)}$. The most straightforward method is to compute the modeling errors for all possible lengths and starting points of $N^{II}_{\theta(i)}$ and then select the best one, a method known as the exhaustive search. By doing so, the complexity for approximating five impulse responses is $\Omega(M^9)$, where Ω denotes the lower bound asymptotic growth rate [6] (see Appendix A for justification). Such a complexity would require a prohibitively long search time and cannot be used in practice, even though the search algorithm is executed off-line with a high performance computer (see also Appendix A).

Instead of using the exhaustive search method, we propose an algorithm (given in next subsection) based on dynamic programming [6] to find the lengths and starting points for all h^{II}_{i} . The proposed search algorithm has a time complexity of $O(FM^3)$, where F is the number of impulse responses to be approximated. Therefore, a set of optimal $\widehat{h}$ can be found within a few seconds using a personal computer.

5.3 Search algorithm based on dynamic programming

For brevity, the search problem is called *optimal allocation* problem and is expressed simply as 'allocating M coefficients into F impulse responses' without emphasizing that the goal is to find the lengths and starting points of section IIs so that the overall modeling error is minimized (using either Eq. 10 or Eq. 11). Correspondingly, the proposed search algorithm finds the optimal solution of allocating M coefficients into F impulse responses. We use $A=\{L,S\}$ to denote a solution to the problem, where $L=\{l_1,l_2,\ldots,l_F\}$ and $S=\{s_1,s_2,\ldots,s_F\}$ indicate the lengths and starting points of the section IIs, respectively. For example, suppose 20 coefficients are to be allocated into 2 impulse responses. Then, $A=\{\{12,8\},\{15,20\}\}$ represents a solution, which allocates 12 and 8 coefficients to the first and second impulse responses, respectively, and the starting points of the first and second impulse responses are at 15 and 20, respectively.

Note that, the optimal allocation problem is non-trivial, because an optimal search algorithm must minimize the overall error of *several* impulses responses simultaneously, which requires considering too many combinations of lengths and starting points. However, when $F=1$, the problem becomes trivial. In this case, since all M coefficients are to be allocated into the same impulse response, an optimal algorithm can simply search for all possible starting points and find the best one, which can be done in $O(M^2)$time (there are $O(M)$ starting points; each requires $O(M)$ time computing the modeling error, assuming the number of coefficients of the impulse response is proportional to M). The proposed algorithm uses a divide-and-conquer strategy, which breaks the problem into smaller subproblems until F becomes 1, which can be solved easily.

Following the treatment of dynamic programming algorithms given in [6], the key observation is that the problem exhibits optimal substructure, i.e., an optimal solution to the optimal allocation problem contains within it optimal solutions to the subproblems. Given j coefficients for allocation, an optimal solution must split j coefficients into 2 parts, one for the last impulse response and the other for the first (F-1) impulse responses. Let the two parts to have k and (j-k) coefficients. Then, k coefficients are allocated to the F^{th} impulse response and (j-k) coefficients are allocated to the 1st ,…,

(F-1) $^{\text{th}}$ impulse responses. In other words, once k is determined, the problem is reduced to two smaller subproblems, one allocates k coefficients and the other allocates (j-k) coefficients. The following Lemma states and proves the optimal substructure of the problem.

Lemma: *Suppose A={L,S} is an optimal solution of allocating j coefficients into F impulse responses, where* $L = \{l_1, l_2, \ldots, l_F\}$ *and* $S = \{s_1, s_2, \ldots, s_F\}$ *indicate the lengths and starting points of the impulse response, and* $j = \sum_{i=1}^{F} l_i$. Let k=l_F and $j - k = \sum_{i=1}^{F-1} l_i$. *Then,* $A' = \{\{l_F\}, \{s_F\}\}$ *is an optimal solution to the subproblem of allocating k coefficients into the F*$^{\text{th}}$ *impulse response. And,* $A'' = \{\{l_1, \cdots, l_{F-1}\}, \{s_, , \cdots, s_{F-1}\}\}$ *is an optimal solution to the subproblem of allocating (j-k) coefficients into 1st ,..., (F-1)* $^{\text{th}}$ *impulse responses.*

Proof: We will prove that $A' = \{\{l_F\}, \{s_F\}\}$ is an optimal solution to the subproblem of allocating k coefficients into the F^{th} impulse response first. By contradiction, suppose that A' is not an optimal solution to the subproblem. Then, there exists another solution $B' = \{\{l_b\}, \{s_b\}\}$ having a lower modeling error than A'. Let e(X) denote the modeling error of applying a solution X. Then, $e(B') < e(A')$. We can combine B' and A'' to create a new solution $B = \{\{l_1, \cdots, l_{F-1}, l_b\}, \{s_, , \cdots, s_{F-1}, s_b\}\}$ to the original problem. Since B also uses j coefficients, B is a legal solution to the original problem. But, $e(B) = e(\{\{l_1, \cdots, l_{F-1}, l_b\}, \{s_, , \cdots, s_{F-1}, s_b\}\}) = e(A'') + e(B') < e(A'') + e(A') = e(A)$, which contracts the assumption that A is an optimal solution. A similar argument can be used to prove that $A'' = \{\{l_1, \cdots, l_{F-1}\}, \{s_, , \cdots, s_{F-1}\}\}$ is an optimal solution to the subproblem of allocating (j-k) coefficients into 1st ,..., (F-1) $^{\text{th}}$ impulse responses. For brevity, we will omit the proof.

The above optimal substructure shows that we can build an optimal solution by splitting the original problem into two subproblems, finding optimal solutions to the two subproblems and then combining these optimal subproblem solutions. Assuming k is known (how to efficiently find k will be given later), the first subproblem, allocating k coefficients into the F^{th} impulse response, can be solved easily by searching all possible starting points of the F^{th} impulse response. The second subproblem, allocating (j-k) coefficients into 1st,…, (F-1) $^{\text{th}}$ impulse responses, can be solved by recursion. Let $e[i][j]$ denote the minimal error of allocating j coefficients into totally i impulse responses with $1 \leq i \leq F$, and $error[i][j]$ be the minimal error of the i^{th} impulse response when j coefficients are used. From the two subproblems, $e[i][j]$ can be defined recursively as

$$e[i][j] = error[i][k] + e[i-1][j-k] \tag{13}$$

where $error[i][k]$ and $e[i-1][j-k]$ are the errors of the first and second subproblems, respectively. But, using Eq. 13 implies that we know the value of k, which we do not. Since k can be as small as 0 and as large as j, we must examine all possible values of k to ensure that we find the optimal one. Thus, the recurrence relation for computing minimal error $e[i][j]$ becomes

$$e[i][j] = \begin{cases} error[1][j] & \text{if } i = 1, \\ error[i][0] + e[i-1][0] & \text{if } j = 0 \\ Min_{0 \leq k \leq j}(error[i][k] + e[i-1][j-k]) & \text{if } i > 1. \end{cases} \tag{14}$$

Since the value of $error[i][j]$ can be obtained by searching all possible starting points of the i^{th} impulse response, the recurrence relation can be used to solve the optimal allocation problem.

Instead of computing e[i][j] in a top-down fashion by a recursive program, it is more efficient to perform the computation in a bottom-up fashion, i.e., storing $e[i-1][j-k]$ in a table and using it to compute $e[i][j]$. The algorithm to implement (14) is given below.

```
procedure OptimalAllocation() is
// Input:
//     M: The desired total section II length (e.g., 220)
//     N: Number of coefficients of each HRIR (e.g., 512) and N >= M
//     F: Number of HRIRs to be approximated (e.g., 5)
//     h[i][j]: the coefficient $h_i[j]$
// Output:
//     n[i][j]: $h_i$'s optimal $N_i^{II}$, when j coefficients are used
//     s[i][j]: optimal $h_i^{II}$'s starting point when j coefficients are used
begin
    // Compute error[i][j] and s[i][j]
    for i := 1 to F do
        for j := 0 to N do
            min := ∞;
            for k := 0 to N-j+1 do  // each k is a starting point
                h' := An approximation of $h_i$ that uses only
                     h[i][k]..h[i][k+M-1] and the rest of
                     the coefficients are set as either zeros
                     or constant values (flat line);
                h'_error := error between h' and h[i] defined
                     by Eq. (10) or (11);
                if (h'_error < min) then
                    error[i][j] := h'_error;
                    s[i][j] := k;
                endif
            end do
        end do
    end do
    // Initialize e[1][j] and n[1][j]; boundary condition i = 1
    for j := 0 to M do
        e[1][j] := error[1][j];
        n[1][j] := j;
    end do
    // Compute e[i][j] with dynamic programming
    for i := 2 to F do
        e[i][0] := e[i - 1][0] + error[i][0];
        n[i][0] := 0;
        for j := 1 to M do
            min := ∞; // infinity
            for k := 0 to j do
                t := e[i - 1][j - k] + error[i][k];
                if (t < min) then
                    min := t;
                    e[i][j] := min; // store the best e[i][j]
                    n[i][j] := k; // store the best n[i][j]
                endif
            end do
        end do
    end do
end
```

In the algorithm, lines 12–28 compute the values of *error*[*i*][*j*] by searching all possible starting points. In addition, when an *error*[*i*][*j*] is found, the entry *s*[*i*][*j*] records the best starting point associated with the *error*[*i*][*j*]. Lines 30–33 initialize *e*[1][*j*] based on the *i*=1 boundary condition. Line 36 initializes *e*[*i*][0] based on the *j*=0 boundary condition. Lines 38–48 search for the best *k* that gives the minimal error.

When the algorithm finishes, the table *n* and *s* can be used to retrieve the optimal section II length and starting point of each impulse response. Each *n*[*i*][*j*] stores the optimal value of *k* (line 45), when *j* coefficients are allocated into *i* impulse responses. That is, the optimal solution allocates the i^{th} impulse response with *n*[*i*][*j*] coefficients starting at the point *s*[*i*][*n*[*i*][*j*]]. The rest of lengths and starting points can be determined similarly. For example, suppose *M*=220 and *F*=5, the optimal is stored in *n*[5][220] with its optimal starting point stored in *s*[5][*n*[5][220]]; the optimal N_4^{II} is stored in *n*[4][220−*n*[5][220]] with its optimal starting point stored in *s*[4][*n*[4][220−*n*[5][220]]; and so on.

In the following, we use a simple example to illustrate how the algorithm works. For simplicity, the example allocates only *M*=4 coefficients into *F*=3 impulse responses. The coefficients of the three impulse responses to be approximated are given in Table 1(a) labeled as *h*[*i*][*j*], where each HRIR has *N*=5 coefficients and each coefficient is an integer between −32768 to 32767 corresponding to −1 to 1 in the floating-point representation. In the actual computation, the integer value is converted into a floating-point number by dividing 32768. From *h*[*i*][*j*], the algorithm obtains *error*[*i*][*j*] (Table 1(b)) and *s*[*i*][*j*] (Table 1 (c)) using Eq. 10 (lines 12–28). For example, *error*[3][2] stores the minimal error of using 2 points for the 3rd HRIR, i.e., the coefficients {0, 100, 20, 10, 0} are approximated as {0, 100, 20, 5, 5} (the middle two values 100 and 20 are preserved and the last two values 10 and 0 in the original HRIR are averaged into 5 and 5). Following Eq. 10, we have $error[3][2] = (10/32768 - 5/32768)^2 - (0 - 5/32768)^2 \approx 4.7 \times 10^8$. The special case of *error*[*i*][0] is considered as approximating all coefficients with a flat line (since no coefficients are preserved). Table 1(d) shows the resulting *e*[*i*][*j*]. Note that *e*[1][*j*] is the same as *error*[1][*j*] (line 31). The computation of the rest of *e*[*i*][*j*] follows lines 38–48 (or the recurrence). For example, *e*[3][4] is determined by the minimal value among (*error*[3][0]+*e*[2][4]), (*error*[3][1]+*e*[2][3]), (*error*[3][2]+*e*[2][2]), (*error*[3][3]+*e*[2][1]), and (*error*[3][4]+*e*[2][0]). Therefore, $e[3][4] = Min(6.6 \times 10^{-6} + 0, 1.9 \times 10^{-7} + 5.6 \times 10^{-7}, 4.7 \times 10^{-8} + 7 \times 10^{-6}\ ,\ \ 0 + 2.6\times 10^{-5}, 0 + 5.2 \times 10^{-5}) = 1.9 \times 10^{-7} + 5.6 \times 10^{-7} = 7.5 \times 10^{-7}$. Since (*error*[3][1]+*e*[2][3]) gives the minimal error, the value of *k* is 1 and is stored in *n*[3][4] (line 45). The resulting *n*[*i*][*j*] is shown in Table 1(e). From *n*[3][4] (marked by a circle), we know that 1 coefficient is allocated to the 3rd impulse and the starting point is *s*[3][1]=2. The rest of the 4−1= 3 coefficients are allocated to the first two impulse responses. Again, from *n*[2][3] (marked by a circle), we know that 2 coefficients are allocated to the 2nd impulse response, and the starting point is *s*[2][2]=3. Finally, the 1st impulse response is allocated with *n*[1][1]=1 coefficient and starts at *s*[1][1]=2.

The time complexity of the algorithm can be easily determined from the algorithm. Since lines 16–21 can be computed in *O*(*N*) time, the algorithm has a complexity of $O(FN^3+FM^2)$. In practice, *M* is proportional to *N*. Therefore, the complexity can also be rewritten as $O(FM^3)$. Note that the widths of table *e* and *error* are not the same (*M* and *N*, respectively). To make the algorithm easier to understand, we assume *N*≥*M*, which is true for MIT's full-length HRIR dataset (*N* is 512 and a typical *M* is 220). In case that *M*>*N*, the algorithm needs a slight modification. Line 31 could access *error*[1][*M*], which does not exist; therefore, it should be modified to use *error*[1][*N*] instead of *error*[1][*j*], when *j*>*N*. For the same reason, line 40 should be modified to limit the range of *k* not to exceed *N*, when *j*>*N*.

Table 1 An example of allocating M=4 coefficients into F=3 impulse responses

(a) $h[i][j]$: the coefficient of $h_i[j]$

	j=1	j=2	j=3	j=4	j=5
i=1	0	100	30	0	0
i=2	0	0	200	200	0
i=3	0	100	20	10	0

(b) $error[i][j]$

	j=0	j=1	j=2	j=3	j=4	j=5
i=1	7.0×10^{-6}	5.6×10^{-7}	0	0	0	0
i=2	4.5×10^{-5}	1.9×10^{-5}	0	0	0	0
i=3	6.6×10^{-6}	1.9×10^{-7}	4.7×10^{-8}	0	0	0

(c) $s[i][j]$: the resulting optimal starting point of each HRIR is marked by a circle

	j=0	j=1	j=2	j=3	j=4	j=5
i=1	x	②	2	1	1	1
i=2	x	3	③	2	1	1
i=3	x	②	2	2	1	1

(d) $e[i][j]$

	j=0	j=1	j=2	j=3	j=4
i=1	7.0×10^{-6}	5.6×10^{-7}	0	0	0
i=2	5.2×10^{-5}	2.6×10^{-5}	7.0×10^{-6}	5.6×10^{-7}	6.2×10^{-7}
i=3	5.8×10^{-5}	3.2×10^{-5}	1.4×10^{-5}	7.2×10^{-6}	7.5×10^{-7}

(e) $n[i][j]$: the optimal length of each HRIR is marked by a circle

	j=0	j=1	j=2	j=3	j=4
i=1	0	①	2	3	4
i=2	0	1	2	②	2
i=3	0	0	0	0	①

6 Experiments and results

This section compares the performance of the proposed approach and the CAPZ method. The performance is measured by modeling error and composition error, to be given below.

6.1 Reducing computational cost for the CAPZ approach

Since the CAPZ (Common-Acoustical-Pole and Zero) model [11] is used in the experiments for performance comparison, we briefly explain why the common poles offer computational savings when Eq. 3 is applied. To see the reason, let the CAPZ-approximated HRIRs be written as

$$\widetilde{h}_{\theta(i)} = p_{CM} * z_{\theta(i)} \tag{15}$$

where p_{CM} and $z_{\theta(i)}$ represent the impulse responses due to the poles and zeros in the approximated transfer function with an azimuth of $\theta(i)$. Thus, when using Eq. 3, we obtain

$$\widehat{y}_R = \sum_{i=\{L,R,C,LS,RS\}} x_i * \widetilde{h}_{\theta(i),R} = \left(\sum_{i=\{L,R,C,LS,RS\}} x_i * z_{\theta(i),R} \right) * p_{CM} \tag{16}$$

Using Eq. 16 implies savings in arithmetic operations by taking advantage of associativity. Considering that the common poles can be implemented as an IIR filter, the implementation of Eq. 16 with N_{pole} common poles and N_{zero} zeros requires $N_{pole} + 5 \cdot (N_{zero} + 1)$ MAC operations. More generally, suppose that there are F impulse responses, the computational cost of the CAPZ approach can be expressed as the function $C_{CAPZ} = 2 \cdot (N_{pole} + F \cdot N_{zero})$ for a pair of output samples (for left and right channels). For F=5, N_{pole}=20, and N_{zero}=40, the required computation is 21.6 MIPS, which is the same as M=220 coefficients of the proposed approach (see Section 4). Therefore, when comparing CAPZ with the proposed approach, we will also use M to represent the computation cost of the CAPZ approach with $M = N_{pole} + F \cdot N_{zero}$.

6.2 Experiment overview

Before conducting experiments, we need to establish a criterion to assess the relative performance of the proposed approach and the CAPZ approach. In addition to evaluating modeling errors, we also consider composition error, the error of the composite signals (y_L and y_R) produced by a particular approach. Judging from the application, we define the composition error as

$$e_{CP} = \frac{1}{2} \left(10 \log \frac{\sum_n (y_L[n] - \widehat{y}_L[n])^2}{\sum_n y_L^2[n]} + 10 \log \frac{\sum_n (y_R[n] - \widehat{y}_R[n])^2}{\sum_n y_R^2[n]} \right) \tag{17}$$

where $y[n]$ is the result obtained using Eq. 3 and $\widehat{y}[n]$ is from the approximation method.

In the experiments, unless otherwise stated, we assume that the number of audio channels is five (F=5), the angles are α=45, β=315, γ=235 and δ=125 (degrees), and the HRIRs to be approximated contain the coefficients of the compact version (N=128) provided by MIT's media lab [9]. For the evaluation of composition errors, except for the last experiment, we use the 5.1-channel audio extracted from a Dolby digital trailer (Broadway) as the source signal. The duration of the signal is around 30 s.

The sequence of the experiments is as follows. The first experiment evaluates the composition error produced by the proposed approach. The second and third experiments compare the modeling errors and composition errors, respectively, of the proposed approach with the CAPZ approach. The fourth experiment is similar to the second and third experiments except that the full-length version (N=512) of HRIR dataset and a different set of azimuth angles are used. To study the performance with more varieties of audio signals, the last experiment compares composition errors for more multi-channel audio signals. The experimental results will show that the proposed approach is better than the CAPZ approach.

Before the first experiment, we perform visual inspection of section IIs obtained by the proposed search algorithm. We use M=220 coefficients to obtain the five approximated HRIRs. Figure 3 shows each original HRIR and its section II marked with vertical lines. Note that the starting points and lengths of the section IIs are all different. A thorough examination by the authors confirms that the proposed search algorithm indeed minimizes the overall modeling error and gives the optimal number of coefficients and starting points for each response.

6.3 Experiment one

The first experiment evaluates the composition error of the proposed approach. In this experiment, we investigate the following three problems: (i) Of the Eqs. 10 and 11 used in the search algorithm, which will produce lower composition errors? (ii) Does a nonzero value in section III actually improve the composition accuracies? (iii) How does the increase of M affect composition errors? Are there any suitable values of M that can be used for practical applications? To answer these questions, we will plot a chart showing the composition error under different values of M, with or without section III for both modeling error criteria.

The results are shown in Fig. 4, where “without normalization” and “with normalization” indicate that the results are obtained by using Eqs. 10 and 11 , respectively, and “with sec. III” and “without sec. III” indicate that the results are obtained using a nonzero value or zero in section III, respectively. From Fig. 4, we know that using Eq. 10 yields better composition errors. This is because, with normalization, an HRIR with smaller energy and one with larger energy are considered equally important when the optimization algorithm determines the lengths of section IIs. But during the signal composition process in Eq. 3 or Eq. 7, an HRIR with larger energy actually contributes a larger portion in the composition error and, thus, needs a longer length to reduce composition error. Figure 4 also indicates that using flat lines

Fig. 3 The chosen section IIs in the impulse responses

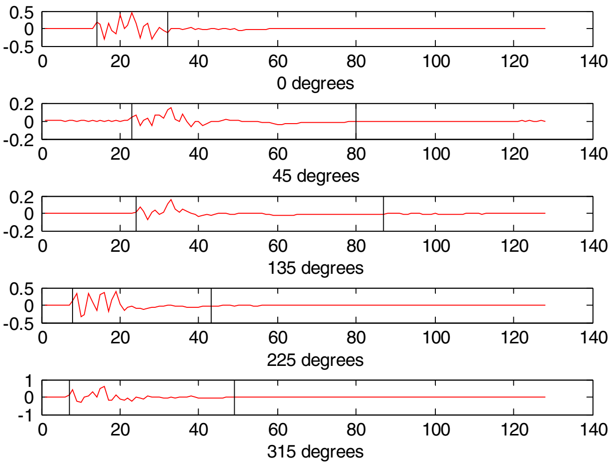
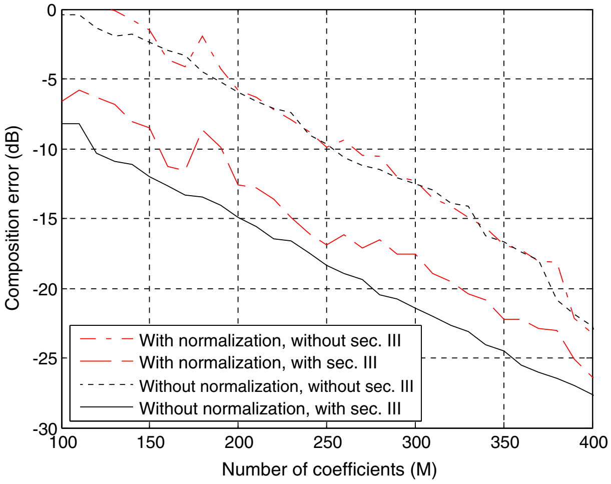


Fig. 4 The composition error of the proposed approach with or without section III



(section IIIs) can reduce the composition error, and a larger M (longer section IIs) gives a smaller composition error. Overall, the proposed approach yields a composition error of less than −15 dB for M=220. Since using Eq. 10 and flat lines give better results, the following experiments will always use Eq. 10 and flat lines (section IIIs).

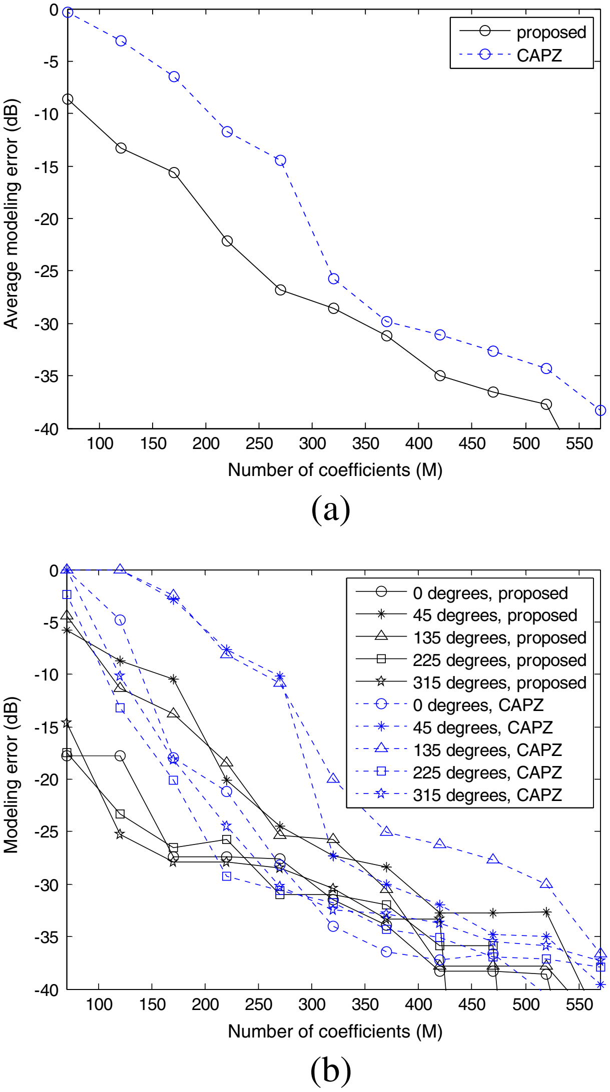
6.4 Experiment two

The second experiment compares the modeling errors of the proposed approach with the CAPZ approach. Due to the use of symmetrical HRIRs, reporting modeling errors for ten HRIRs (two HRIRs per direction) is not necessary, as there are only five sets of modeling errors. Therefore, we report HRIRs associated with left ear only. To simplify the comparison, the number of poles used in the CAPZ approach is set to 20, and only the number of zeros varies. In the CAPZ case, M is $20+5{\cdot}N_{zero}$ (see Section 6.1). For example, when M=120, the number of zeros used in each impulse response is $(120-20)\div 5=20$. The results are plotted in Fig. 5(a), indicating that given the same amount of computation (M), the proposed approach provides smaller average modeling errors than the CAPZ approach. To gain more insights, the individual modeling errors are also given in Fig. 5(b), which shows that the CAPZ approach has a larger variation of errors among different HRIRs than the proposed approach. Because the HRIR having the largest modeling error eventually dominates the composition error, a good approximation approach should have small variations of errors among different HRIRs. In this regard, our approach is better. From the viewpoint of computational cost, our approach produces individual modeling errors below −15 dB with 220 coefficients, but the CAPZ approach requires about 300 coefficients to achieve the same result.

6.5 Experiment three

The third experiment compares the relative composition errors between the proposed approach and the CAPZ approach. To further investigate the influence of the number of common poles on the composition error of the CAPZ approach, we also evaluate the composition errors with 15 and 30 common poles. The results, given in Fig. 6, show that

Fig. 5 The modeling errors of the proposed approach and the CAPZ approach. (**a**) Average modeling error. (**b**) Individual modeling errors



(a)

(b)

the proposed approach has a better performance on composition errors. The reason that the CAPZ approach has larger composition errors is partially due to its larger variation of modeling errors shown in the previous experiment. The results also indicate that using more poles in the CAPZ approach does not significantly improve its composition errors. Therefore, for the CAPZ approach, increasing the number of zeros is the only effective method to reduce composition errors.

6.6 Experiment four

The fourth experiment studies the error performance of approximating the full-length HRIRs (N=512) with a different set of azimuth angles, namely, α=30, β=330, γ=240, and δ=120

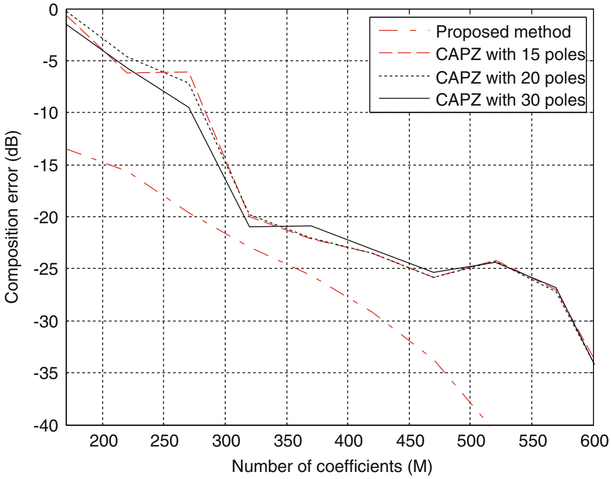


Fig. 6 The relative composition error of the proposed method and the CAPZ method

(degrees). The results are shown in Fig. 7, which reveals that the characteristics of the composition errors of the first and the second sets of angles are very similar. For most of the computational regions ($M<1000$ and $M>1300$), the proposed approach has lower composition errors than those of the CAPZ approach. The composition errors of the full-length version are much higher than those of the compact version, because the values of $h_{\theta(i)}[n]$ in the full-length version are not zeros for $n>128$. These values, though very small, still contribute to making errors. Overall speaking, we can conclude that the proposed approach gives a composition error better than the CAPZ approach when the computing resources are limited ($M<1000$).

6.7 Experiment five

This experiment studies whether the proposed approach is effective for audio signals of different varieties. Several test items from different sources are used. The first and second one are also Dolby-digital trailers designed to demonstrate the capability of the five-channel surround sound. The third and fourth items are musical works from Internet

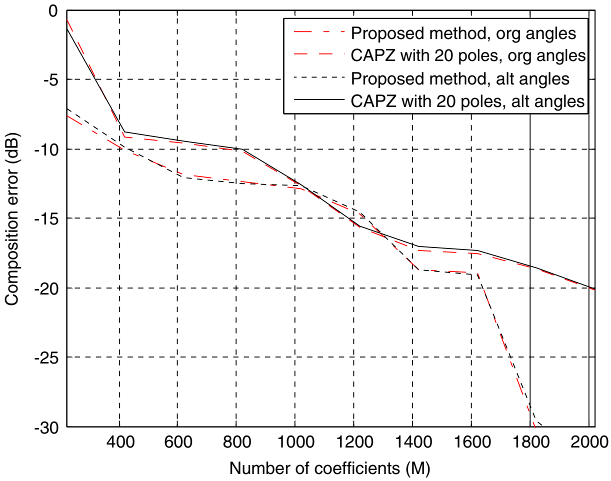


Fig. 7 Composition errors when modeling HRIRs with 512 coefficients

Table 2 The test items used in experiment five

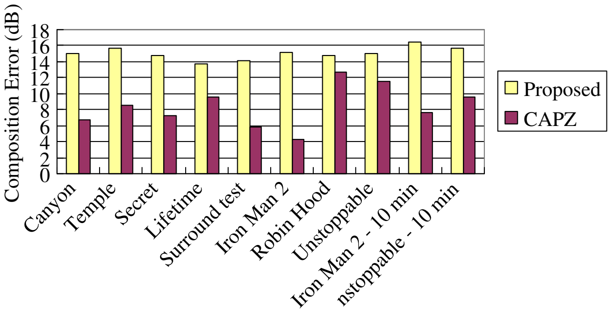
Index	Content
1	Dolby Digital Trailer: Canyon
2	Dolby Digital Trailer: Temple
3	Secret World [2], first 30 s
4	A Lifetime of Moments [2], first 30 s
5	5.1 Surround Test File [2]
6	DVD Release of Iron Man II (with dialogue and background music, 30 s)
7	DVD Release of Robin Hood (with dialogue and background sound of forest, 30 s)
8	DVD Release of Unstoppable (with dialogue and background noise, 30 s)
9	Same as item 6, duration 10 min
10	Same as item 8, duration 10 min.

[2]. These works contain different types of instruments played at different locations (azimuth angles). However, most instruments are located in front of the audience (i.e., in the front channels). The fifth one is a surround test file containing a female voice saying ‘front left’, ‘center’, etc. for each channel. Items sixth to ten are video clips retrieved from DVD discs. The first eight items have duration of 30 s or less, whereas the last two items have 10 min. The details of the test pieces are listed in Table 2. Overall, these test pieces cover many real scenarios a player may encounter. To simplify the comparison, we use M=220 for all test items. The resulting composition errors are given in Fig. 8, which shows that the proposed approach constantly outperform the CAPZ approach for all test items. We also note that the composition error is a function of audio contents, a reasonable situation. However, the proposed approach produces composition errors with much less fluctuation than the CAPZ does. Therefore, the proposed approach not only performs better, but also more consistently.

7 Conclusions

In this paper, we propose a method to optimally truncate HRIRs for reproducing spatial sound for headphones with a lower computational cost. The approximated impulse response contains a portion of the original impulse response plus a flat-line. The numbers of coefficients preserved in the approximated impulse responses are determined by a dynamic programming algorithm. The experimental results show that the proposed approach yields modeling errors to less than −15 dB when the amount of computation is about 35 % of that required in the direction computation. We

Fig. 8 Composition errors of various test items. The values are reported by discarding the negative signs



also implement the CAPZ approach as an alternative approach for comparison. Given the same amount of computation, the proposed approach is better than the CAPZ approach in terms of both modeling and composition errors. Therefore, when computing resources are limited, the proposed approach is a better choice to reproduce spatial sound for headphones.

Appendix A. Computational complexity of exhaustive search

This appendix briefly discusses the lower-bound time complexity of using an exhaustive search to find optimal section IIs, i.e., the optimal starting points and lengths (N_i^{II}) of section IIs. Let the number of impulse responses be F, the number of coefficients in a response be N, $M = \sum_{i=1}^{F} N_i^{II}$, and, for simplicity, $N=M$ (M is proportional to N in practice). An allocation of M is an assignment of the values of N_I^{II} … N_F^{II} satisfying the constraint that $M = \sum_{i=1}^{F} N_i^{II}$. For example, if $F=5$ and $M=100$, a possible allocation is $N_1^{II}=1$, $N_2^{II}=1$, $N_2^{II}=1$, $N_3^{II}=1$, $N_4^{II}=96$. In this case, since $N_1^{II}=1$, the section II of the first impulse response has a total of N possible starting points. An exhaustive search must compute the modeling errors of all possible distinct allocations and starting points, and record the one with the smallest error. We will show that there exists at least $\Omega(M^{F-1})$ distinct allocations, and each of the distinct allocation has at least $\Omega(M^{F-1})$ distinct combinations of starting points. Therefore, an exhaustive search must compute $(M^{F-1} \times M^{F-1}) = (M^{2F-2})$ distinct modeling errors. Since computing a particular modeling error requires $\Omega(N)=\Omega(M)$ time (Eq. 10 or Eq. 11). The complexity of an exhaustive search is $\Omega(M^{2F-1})$.

Let's consider the number of distinct allocations first. Since we are concerned with the lower bound, we do not need to calculate all possible distinct allocations. Instead, we consider a subset of all possible distinct allocations, one with the restriction that $1 \le N_1^{II} \le \frac{M}{F}$, …, $1 \le N_{F-1}^{II} \le \frac{M}{F}$. Under this restriction, N_I^{II} have $\frac{M}{F}$ different possible values; so are N_2^{II} …N_{F-1}^{II} . Note that this restriction requires $F-1 \le \sum\limits_{i=1}^{F-1} N_i^{II} \le \frac{(F-1)}{F} M$, and therefore there must exist a value of N_F^{II} that satisfies $M = \sum_{i=1}^{F} N_i^{II}$. In other words, all combinations of possible values of N_I^{II}…N_{F-I}^{II} are legal and are distinct allocations. Therefore, there exist $\Omega\left(\frac{M^{F-1}}{F^{F-1}}\right)$ distinct allocations. Since F is typically a constant (e.g., 5), F^{F-1} is also a constant (e.g., $5^4=625$). The number of distinct allocations can be simplified as $\Omega(M^{F-1})$.

We now calculate the number of possible starting points for each distinct allocation. Normally, for the i^{th} impulse response, depending on the value of N_i^{II}, the number of distinct starting points can be as small as 1 (when $N_i^{II}=N$), and as large as N (when $N_i^{II}=1$). However, for the first (F−1) impulse responses, we have made the restriction that $1 \le N_i^{II} \le \frac{M}{F}$. When $N_i^{II}=1$ and $N_i^{II} = \frac{M}{F}$, the number of distinct starting points are N and $N - \frac{M}{F}$, respectively. Thus, each of the first (F−1) impulse response has at least $\Omega(M)$ distinct starting points. Therefore, the first (F−1) impulse responses alone have $\Omega(M^{F-1})$ possible combinations of starting points. Note that we did not count the starting points of the F^{th} impulse response. This is safe because we are calculating a lower bound.

Since there are $\Omega(M^{F-1})$ distinct allocations and each allocation has $\Omega(M^{F-1})$ distinct combination of starting points, an exhaustive search must compute $(M^{F-1} \times M^{F-1}) = \Omega(M^{2F-2})$ different possible modeling errors. The modeling error of each allocation can be calculated in $\Omega(N)=\Omega(M)$ time according to Eq. 10 or Eq. 11. Therefore, the complexity of an exhaustive search becomes $\Omega(M^{2F-2} \times M) = \Omega(M^{2F-1})$. If $F=5$ (five-channel audio) and $M=220$, the number of computations is at least proportional to $M^9 = 1.2 \times 10^{21}$. For a computer that can compute one square and one addition in Eq. 10 in 10^{-8}s, it would take 3.8×10^5 years to find an optimal solution, which is clearly impractical.

References

1. Advanced Television Systems Committee (1995) Digital Audio Compression Standard (AC-3), Doc. A/52
2. Available at http://www.lynnemusic.com/surround.html
3. Blaurt J (1997) Spatial hearing – the psychophysics of human sound localization, Revisedth edn. MIT press, Cambridge
4. Blommer MA, Wakefield GH (1997) Pole-zero approximations for head-related transfer functions using a logarithmic error criterion, IEEE Trans. Speech and Audio Processing 5(3):278–287
5. Brown CP, Duda RO (1998) A structural model for binaural sound synthesis. IEEE Trans Speech Audio Process 6(5):476–488
6. Cormen TH, Leiserson CE, Rivest RL, Stein C (2001) Introduction to algorithms, 2nd edn. MIT Press, Cambridge, MA, USA
7. DTS standard is not open to public; however, introductory materials are available at http://dts.com
8. Durant EA, Wakefield GH (2002) Efficient model fitting using a genetic algorithm: pole-zero approximations of HRTFs. IEEE Trans Speech and Audio Processing 10(1):18–27
9. Gardner WG, Martin KD (1994) HRTF measurements of a KERMAR dummy-head microphone, MIT Media Lab. Available at (http://sound.media.mit.edu/resources/KEMAR.html)
10. Gardner WG, Martin KD (1995) HRTF measurements of a KERMAR. J Acoust Soc Am 97(6):3907–3908
11. Haneda Y, Makino S, Kaneda Y, Kitawaki N (1999) Common-acoustical-pole and zero modeling of head-related transfer function. IEEE Trans Speech and Audio Processing 7(2):188–196
12. Huang S, Park Y (2008) Interpretations on principal components analysis of head-related impulse responses in the median plane. J Acoust Soc Am 123(4):1–7
13. Listen HRTF DATABASE by IRCAM and AKG (2003) available at http://recherche.ircam.fr/equipes/salles/listen/index.html
14. ISO/IEC (2005) Information Technology – Coding of Audio-visual Objects, Part 3: Audio, IS 14496-3
15. Kulkarni A, Colburn HS (2004) Infinite-impulse-response models of the head-related transfer function. J Acoust Soc Am 115(4):1714–1728
16. Mackenzie J, Huopaniemi J, Välimäki V, Kale I (1997) Low-order modeling of head-related transfer functions using balanced model truncation. IEEE Signal Processing Lett 4(2):39–41
17. Sakamoto N, Kobayashi W, Onoye T, Shirakawa I (2003) Single DSP implementation of real time 3D sound synthesis algorithm. Journal of Circuits, Systems, and Computers 12(1):55–73
18. Shen Y-C, You SD (2003) Rendering spatial sound on headsets for five-channel audio. Proc. of the Fourth Int'l Conf. on Info., Com. and Signal Proc. and Fourth Pacific-Rim Conf. on Multimedia (ICICS-PCM 2003), Singapore, 1–5
19. Yen C-H, Lin Y-S, Wu B-F (2007) An efficient implementation of a low-complexity MP3 algorithm with stream cipher. Multimedia Tools and Applications 35(3):335–355
20. You SD, Chen W-K (2008) Efficient quantization algorithm for real-time MP-3 encoders. Multimedia Tools and Applications 40(3):341–359

Shingchern D. You received the Ph.D. degree in Electrical Engineering from the University of California, Davis, CA, USA in 1993. Dr. You's research interests include audio signal processing and recognition, applications of digital signal processing to communication systems, and intelligent systems.

Woei-Kae Chen received the diploma in Electronic Engineering from National Taipei Institute of Technology, Republic of China, in 1984, the M.S. and Ph.D. degrees in Computer Engineering from North Carolina State University, NC, USA in 1988 and 1991. Dr. Chen's research interests include software engineering, distributed computing, and graph algorithms.