Spatial Sound Reproduction

Based on HRTF Auto-Selection Algorithm

Cheng Zhang, Ke'an Chen & Chao Xing

College of Marine, Northwestern Polytechnical University

Xi'an 710072, China

E-mail: zhangcheng001@mail.nwpu.edu.cn

Abstract

The desired beam pattern of the array can be achieved by applying specific array weighting in the eigenbeam space derived from the plane wave decomposition. Head related transfer function (HRTF) data are selected automatically from the KEMAR HRTF measurements database of MIT media lab without the message of the direction of the incoming sound wave. The HRTF for the sound wave direction are derived by combining the directivity of the eigenbeam beamforming and the algebraic expression of HRTF. Computer simulations demonstrate that the HRTF approximation derived from this paper are similar to those of the measurements by MIT media lab in certain frequency band. The higher the order of the eigenbeam used in eigenbeam beamforming, the more similar of the derived HRTF data with those of the measurements. Listening experiments are conducted to evaluate the efficiency of the proposed HRTF auto-selection algorithm subjectively. The results indicate that the audio localization precision for virtual sounds using approximated HRTF is consistent with those of the measured HRTF, which verifies the validity of the proposed algorithm.

Keywords: Planewave decomposition, Eigenbeam beamforming, HRTF, Virtual audition

1. Introduction

Virtual audition has been wildly used in virtual reality, robotics, communication, tele-conferencing, entertainment, and training simulations, etc.. The presentation of binaural signals over headphones is a convenient way for recreating the original auditory scene for a listener. There are two kinds of methods to get binaural signals. One is the traditional binaural technology. It records sound signals using miniature microphones placed at the entrance to the ear canals of a listener or a dummy head. Then replaying these signals via headphones will reproduce the original auditory experience, evoking the auditory perception of a source located at a point in the space surrounding the listener. The other method is based on signal processing technology which gets binaural signals through computer combination. It reproduces the processed sound through head-related transfer function (HRTF) filter (Moller, 1995; Xie, 2008), which has been wildly implemented in practice as it can be realized easily. The direction of the actual or virtual sound source must be known in advance. Then binaural signals can be derived by filtering the sound signal using HRTF data for the sound direction. Reproducing the binaural signals through a pair of headphones can create the three-dimension sound in human auditory perception (Bai, 2004; Savioja, 1999). Array signal processing can improve the signal-to-noise ratio (SNR) and eliminate the influence of the noise and the interferer (validly, 2001). As a result, with the same sound intensity, the distance of the processed sound source is much further using array signal processing technology than listening to the sound signal directly. Furthermore, the sound image of the former in listener's auditory perception is much clear than the latter with the same distance between the sound source place and the listening place. It has great significance to combine the array signal processing technology with human audition system in sound source localization and identification applications as human audition has great ability of sound localization and identification. In practice, the incoming direction of the sound signal is often unknown, and it may be appeared in any direction in three dimensional space blended with a lot of interferer and noise especially in battlefield environment. Thus, in order to reproduce the space sound through HRTF filter, the direction of the sound signal must be known in advance. In present literature, the direction of the incoming sound must be estimated using array signal processing technology, which may increase the computing load and can not be implemented in practice conveniently.

This work deals with the HRTF auto-selection algorithm without the message of the direction of the incoming sound wave. Firstly, the soundfield is decomposed to the eigen-beam space using the planewave decomposition theory based on the array model of a uniform circular array (UCA) mounted around a rigid sphere. Special array beam patterns are derived through eigenbeam beamforming processing using distinct array weighting function. Through filtering the

HRTF measurements data for all the directions by the beam directivity of eigenbeam beamforming, the HRTF approximation are derived without the message of the direction of the sound source based on the HRTF algebraic expression. Then, virtue sound is achieved using the HRTF approximation data. Simulations and subjective experiments are carried out to investigate the validity of the proposed method lastly.

2. Eigenbeam beamforming

Applying the soundfield decomposition theory, the soundfield around the array can be decomposed to the eigenbeam space. Then, the sound signal can be further processed using array signal processing technology in the eigenbeam space (Rafaely, 2004; Teutsch, 2006).

Assume a UCA with radius R_0 mounted around a rigid sphere with radius R, and a unit magnitude impinging plane wave comes from the direction of (θ_0, φ_0). The equation describing the resulting sound pressure at (r, ϑ, φ) around the array in spherical coordinates is (Meyer, 2001)

$$P(r,\vartheta,\varphi) = \sum_{n=0}^{\infty}(2n+1)(-j)^n[j_n(kr) - \frac{j_n'(kR_0)}{h_n'(kR_0)}h_n(kr)]\sum_{q=-n}^{n}\frac{(n-|q|)!}{(n+|q|)!}P_n^{|q|}(\cos\vartheta)P_n^{|q|}(\cos\vartheta_0)e^{-jq(\varphi-\varphi_0)} \quad (1)$$

where k is the wave number, $k = 2\pi/\lambda$, λ is the wave length of the plane wave, $j_n(\cdot)$ is the nth-order spherical Bessel function of the first kind, $h_n(\cdot)$ is the nth-order spherical Hankel function, $j_n'(\cdot)$ and $h_n'(\cdot)$ denotes the derivative of the nth-order spherical Bessel function and the nth-order spherical Hankel function with respect to the argument respectively, $P_n^q(\cdot)$ is the Legendre function of order n and degree q. The time dependency is omitted in the equation.

For simplicity, the theoretical construct of a continuous circular aperture and a unit magnitude plane wave impinging from $\vartheta_0 = \pi/2$ are concerned firstly. Assuming the continuous circular aperture is located on the equator of the rigid sphere, i.e. $\vartheta = \pi/2$, the sound pressure at the position of the continuous circular aperture is described by $r = R$. Then, Eq. (1) becomes

$$P(R,\vartheta,\varphi) = \sum_{n=0}^{\infty}(2n+1)(-j)^n[j_n(kR) - \frac{j_n'(kR_0)}{h_n'(kR_0)}h_n(kR)]\sum_{q=-n}^{n}\frac{(n-|q|)!}{(n+|q|)!}\times P_n^{|q|}(\cos\vartheta)P_n^{|q|}(\cos\vartheta_0)e^{-jq(\varphi-\varphi_0)} \quad (2)$$

$P(R,\vartheta,\varphi)$ can be seen as the acoustical transfer function from the source point (ϑ_0, φ_0) to the sensor location (R,ϑ,φ). Assume the aperture weighting function is $\omega(\varphi, f)$, and since $\omega(\varphi, f)$ has to be 2π periodic, it can be expanded in a Fourier series

$$\omega(\varphi, f) = \sum_{p=-\infty}^{\infty} a_p(f)e^{jp\varphi} \quad (3)$$

For simplicity, assume further $a_p(f) = 1$, then the pth eigenbeam caused by the pth term of the aperture weighting function can be represented as

$$P_p = \frac{1}{2\pi}\int_0^{2\pi}P(R,\vartheta,\varphi)\omega(\varphi, f)\mathrm{d}\varphi \quad (4)$$

Using Eq. (2) and Eq. (3), we can get

$$P_p = \frac{1}{2\pi}\int_0^{2\pi}\sum_{n=0}^{\infty}(2n+1)(-j)^n[j_n(kR) - \frac{j_n'(kR_0)}{h_n'(kR_0)}h_n(kR)]\sum_{q=-n}^{n}\frac{(n-|q|)!}{(n+|q|)!}$$
$$\times P_n^{|q|}(\cos\vartheta)P_n^{|q|}(\cos\vartheta_0)e^{-jq(\varphi-\varphi_0)}e^{jp\varphi}\mathrm{d}\varphi \quad (5)$$

From Eq. (5), we can finally get (Meyer, 2001)

$$P_p = b_p e^{jp\varphi_0} \quad (6)$$

where

$$b_p = \sum_{n=p}^{\infty}(2n+1)(-j)^n[j_n(kR) - \frac{j_n'(kR_0)}{h_n'(kR_0)}h_n(kR)]\frac{(n-|p|)!}{(n+|p|)!}P_n^{|p|}(\cos\vartheta)P_n^{|p|}(\cos\vartheta_0) \quad (7)$$

P_p can be seen as the p th order eigen-beam derived from the decomposed soundfield.

Array received signals can be processed in eigenbeam space by applying eigenbeam beamforming technology based on the planewave decomposition theory. In order to eliminate the frequency influence in the eigenbeam, assume

the *p*th order weighting coefficient as

$$w_p(\varphi) = \frac{1}{b_p} e^{jp\varphi} \tag{8}$$

Then, the *p*th order array output is

$$y_p = \frac{1}{2\pi} \int_0^{2\pi} P(R, \vartheta, \varphi) w_p(\varphi) \mathrm{d}\varphi = e^{jp\varphi_0} \tag{9}$$

Considering a square integrable function on the unite circular, this function can be expressed as the summation of Fourier series

$$f(\varphi) = \sum_{p=-\infty}^{\infty} F_p e^{jp\varphi} \tag{10}$$

The coefficients of Fourier series is defined as

$$F_p = \frac{1}{2\pi} \int_0^{2\pi} f(\varphi) e^{-jp\varphi} d\varphi \tag{11}$$

Assuming an ideal beam pattern $F(\varphi, \varphi_k)$ looking at the direction φ_k, it can be expressed as a delta function

$$F(\varphi, \varphi_k) = \delta(\varphi - \varphi_k) \tag{12}$$

It can also be expanded into a summation of the Fourier series

$$F(\varphi, \varphi_k) = \frac{1}{2\pi} \sum_{p=-\infty}^{\infty} e^{-jp(\varphi - \varphi_k)} \tag{13}$$

From the above analyses, the aperture weighting function to achieve this beam pattern is

$$w(\varphi, \varphi_k) = \frac{1}{2\pi} \sum_{p=-\infty}^{\infty} \frac{1}{b_p} e^{jp(\varphi - \varphi_k)} \tag{14}$$

Sampling the continuous circular array using a UCA result in the same expressions as Eq. (6) when the number of the elements of UCA satisfies $M \geq 2N$. The highest order of the eigen-beams N is satisfied by $N \leq [kR]$, $[\cdot]$ stands for the biggest integer (Mathews, 1994). Then the array weighting function is

$$w(\varphi_s, \varphi_k) = \frac{1}{2\pi} \sum_{p=-N}^{N} \frac{1}{b_p} e^{jp(\varphi_s - \varphi_k)} \tag{15}$$

where φ_s denotes the elements place. The beam pattern can be expressed as

$$F_N(\varphi, \varphi_k) = \frac{1}{2\pi} \sum_{p=-N}^{N} e^{-jp(\varphi - \varphi_k)} \tag{16}$$

The array output is

$$y_{\mathrm{EB}} = \frac{1}{2\pi} \sum_{p=-N}^{N} e^{-jp(\varphi - \varphi_k)} \tag{17}$$

So far, the array beam output pointing to the signal direction φ_k is achieved using the array weighting function defined in Eq. (15). This beamforming technology using eigenbeams derived from the soundfield decomposition process is called as eigenbeam beamforming (EBF). The array received sound signals can be further processed using the directivity of the EBF.

The delay-and-sum beamforming (DSBF) technology can also be applied to the array with baffles in eigenbeam space. All processed microphone outputs can have the same phase by compensating for the delays due to a single plane wave.

A widely used performance measure of array is the directivity factor (DI), which evaluates the improved directivity of the array compared to an omnidirectional microphone (Brandstein, 2001). It can be defined as the logarithm expression of the ratio of the array output in look direction and the array output integrated over all directions. So, the DI of the EBF array can be expressed as

$$\mathrm{DI_{EB}} = 10\lg\left(\left|y_{\mathrm{EB}}(\varphi_k)\right|^2 \middle/ \frac{1}{2\pi}\int_0^{2\pi}\left|y_{\mathrm{EB}}(\phi)\right|^2 d\phi\right)$$

$$= 10\lg\left(\left|\frac{1}{2\pi}\sum_{p=-N}^{N}e^{-jp(\varphi-\varphi_k)}\right|^2 \middle/ \frac{1}{2\pi}\int_0^{2\pi}\left|\frac{1}{2\pi}\sum_{p=-N}^{N}e^{-jp(\varphi-\phi)}\right|^2 d\phi\right) \tag{18}$$

$$= 10\lg\left(\left|\sum_{p=-N}^{N}1\right|^2 \middle/ \sum_{p=-N}^{N}|1|^2\right) = 10\lg(2N+1)$$

It's clearly from Eq. (18) that the DI of the EBF array is independent on the frequency.

For DSBF array, assume the pressure on the array is $e^{i\mathbf{k}_0\mathbf{r}}$ due to a plane wave comes from the direction φ_0, where $\mathbf{k}_0$ is the wave vector represents the plane wave arrival direction, and $\mathbf{r} = (z=0, \theta=\pi/2, \varphi)$ denotes the microphone positions. Then the output of the DSBF array can be written as

$$y_{\mathrm{DS}} = \frac{1}{2\pi}\int_0^{2\pi}e^{i\mathbf{k}_0\mathbf{r}}e^{-i\mathbf{k}_l\mathbf{r}}d\varphi \tag{19}$$

where $\mathbf{k}_l$ denotes the wave vector of the array look direction φ_l. The pressure signal can be expressed as the summation of the Fourier series, and the coefficients of the Fourier series can be written as (Rafaely 2004; Teutsch, 2006)

$$P_p = b_p e^{jp\varphi_0} \tag{20}$$

With Eq. (10) and Eq. (20), the DSBF array output becomes

$$y_{\mathrm{DS}} = \frac{1}{2\pi}\int_0^{2\pi}\sum_{p=-N}^{N}P_p e^{jp\varphi}\sum_{p=-N}^{N}P_p^*e^{-jp'\varphi}d\varphi = \frac{1}{2\pi}\int_0^{2\pi}\sum_{p=-N}^{N}b_p e^{jp\varphi_0}e^{jp\varphi}\sum_{p=-N}^{N}b_p^*e^{-jp'\varphi_l}e^{-jp'\varphi}d\varphi$$

$$= \sum_{p=-N}^{N}\left|b_p\right|^2 e^{jp\varphi_0}e^{-jp\varphi_l} = \sum_{p=-N}^{N}\left|b_p\right|^2 e^{-jp(\varphi_l-\varphi_0)} \tag{21}$$

Using the same analysis with the EBF array, the DI of the DSBF array can be written as

$$\mathrm{DI_{DS}} = 10\lg\left(\left|\sum_{p=-N}^{N}\left|b_p\right|^2\right|^2 \middle/ \sum_{p=-N}^{N}\left|b_p\right|^4\right) \tag{22}$$

The DI of the EBF array is compared with that of the DSBF array for different frequencies in Fig.1. It's shown that the DI of the EBF array remains constant throughout the frequency band, which is a benefit for processing low frequency and broad band signals. The DSBF array has approximately the same DI with the EBF array only at higher frequencies. As frequency lowered, the DI of the DSBF array is also reduced until it becomes omidirectional , i.e., DI=0. Figure 2 shows the beam patterns of the EBF array and the DSBF array with kR =2 and 12 respectively. It's obvious from this figure that the beam patterns of the EBF array have frequency-independent characteristic, and the directivity of the DSBF array is very weak at low frequencies ($kR=2$) as the values of b_p in this case are much smaller. It increases gradually as the frequency increase. Approximately the same directivity are derived for the DSBF array and the EBF array at high frequencies ($kR=12$) as the values of b_p are almost equal for different eigen-beams.

3. HRTF auto-selection

As sound propagates from a distant source to the eardrums of a human listener, it is acoustically transformed by interactions with the head, torso, and pinna. This transformation can be modeled as a linear filter. The transfer function is called the head related transfer function (HRTF). HRTF describe the spectral filtering and contain all the information on the sound transmission from a given source position to the ear canals of a human subject including the interaural time difference (ITD), the interaural level difference (ILD), and spectral characteristics. The inverse Fourier transform of a HRTF is termed the head-related impulse response (HRIR) (Xie, 2008). HRTF are used in binaural virtual auditory displays for the synthesis of a sound source with spacial characteristics. The direction of the sound or the virtual sound must be known in advance in order to select the HRTF data for the sound direction. However, the direction of the sound is often unknown and there are a lot of interferers and noise companied with the target sound as described in introduction. So array signal processing technology is needed to estimate the DOA of the sound wave and cancel the influence of the interferers and noise. In order to reduce the DSP computing load, a new method must be investigated to derive the HRTF data for the desired direction without the message of the sound wave direction. We combine the directivity of the beamforming and the algebraic of the HRTF to achieve this goal. Only the eigenbeam beamforming is concerned in the following as it has better performance at low frequencies and for broad band signals.

Assume the sound wave comes from the point (r, ϑ, φ), then the HRTF can be expressed as (Duraiswami, 2004)

$$h(\vartheta,\varphi) = \sum_{n=0}^{\infty} \sum_{p=-n}^{n} \alpha_{nm} h_n(kr) Y_n^m(\vartheta,\varphi) \qquad (23)$$

where α_{nm} are the fitting coefficients which be determined using the discrete HRTF measurements. $Y_n^m(\vartheta,\varphi)$ is the spherical harmonics of order n and degree m. Here the incoming planewave paralleling to the array is concerned only for simplicity, i.e., only the HRTF in the array plane is discussed. Then, $\vartheta = \pi/2$, and the following expression is derived by expanded the spherical harmonics

$$h(\varphi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \alpha_{nm} h_n(kr) \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(0) e^{jm\varphi} \qquad (24)$$

Assume the sound wave comes from the direction φ_k, we need to select the HRTF for the direction φ_k to combine the virtual sound. In order to achieve this goal, we use the delta function (ideal beam pattern) to select the HRTF data. Using Eq. (12), we can get

$$\int_0^{2\pi} h(\varphi) F(\varphi,\varphi_k) d\varphi = \int_0^{2\pi} h(\varphi) \delta(\varphi-\varphi_k) d\varphi = h(\varphi_k) \qquad (25)$$

From Eq. (13), we can also derive

$$\begin{aligned} \int_0^{2\pi} h(\varphi) F(\varphi,\varphi_k) d\varphi &= \int_0^{2\pi} \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \alpha_{nm} h_n(kr) \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(0) e^{jm\varphi} \times \frac{1}{2\pi} \sum_{p=-\infty}^{\infty} e^{-jp(\varphi-\varphi_k)} d\varphi \\ &= \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \alpha_{nm} h_n(kr) \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(0) e^{jm\varphi_k} \\ &= h(\varphi_k) \end{aligned} \qquad (26)$$

In practice, HRTF are measured only at finite points at the array plane. Assume the highest order of the eigenbeam derived from the soundfield decomposition is N, and the number of measurements points is M, which are uniformly distributed around the circle. Then

$$h(\varphi_l) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \alpha_{nm} h_n(kr) \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(0) e^{jm\varphi_l} \ , \quad l=1,2,\cdots,M \qquad (27)$$

Equation (27) can be expand into two groups as

$$\begin{aligned} h(\varphi_l) &= \sum_{n=0}^{N} \sum_{m=-n}^{n} \alpha_{nm} h_n(kr) \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(0) e^{jm\varphi_l} \\ &+ \sum_{n=N+1}^{\infty} \sum_{m=-n}^{n} \alpha_{nm} h_n(kr) \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(0) e^{jm\varphi_l} \\ &= h_N(\varphi_l) + h_{N+1}(\varphi_l) \end{aligned} \qquad (28)$$

Rewrite Eq. (26) as a summation form

$$\begin{aligned} \sum_{l=1}^{M} h(\varphi_l) F_N(\varphi_l,\varphi_k) &= \sum_{l=1}^{M} [h_N(\varphi_l) + h_{N+1}(\varphi_l)] \times \frac{1}{2\pi} \sum_{p=-N}^{N} e^{-jp(\varphi_l-\varphi_k)} \\ &= \sum_{l=1}^{M} [\sum_{n=0}^{N} \sum_{m=-n}^{n} \alpha_{nm} h_n(kr) \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(0) e^{jm\varphi_l} \\ &+ \sum_{n=N+1}^{\infty} \sum_{m=-n}^{n} \alpha_{nm} h_n(kr) \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(0) e^{jm\varphi_l}] \times \frac{1}{2\pi} \sum_{p=-N}^{N} e^{-jp(\varphi_l-\varphi_k)} \\ &= \frac{M}{2\pi} \sum_{n=0}^{N} \sum_{m=-n}^{n} \alpha_{nm} h_n(kr) \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(0) e^{jm\varphi_k} + \varepsilon \\ &= \frac{M}{2\pi} h_N(\varphi_k) + \varepsilon \end{aligned} \qquad (29)$$

where ε are residual errors due to the sampling operation. The orthonormality property of the discrete Fourier transform is used in Eq. (29). Assuming an impinging wavefield is decomposed to eigenbeams up to a finite order N, the residual errors can be omitted in Eq. (29) when $M \geq 2N$ is satisfied (Mathews, 1994). Then

$$\sum_{l=1}^{M} h(\varphi_l) F_N(\varphi_l,\varphi_k) = \frac{M}{2\pi} h_N(\varphi_k) \qquad (30)$$

The HRTF approximation for the direction φ_k can be achieved by multiplying a constant $2\pi/M$ to the summation in Eq. (30). It is obviously that the quality of the HRTF selection is determined by the highest order of the eigenbeams derived

from the soundfield decomposition.

The beam pattern looking at the sound incoming direction is arrived using EBF technology described above, and the HRTF auto-selection is achieved by filtering the HRTF measurements using the output of this beam pattern without the message of the incoming sound direction.

4. Virtual sound combination

The virtual audition space can be created by filtering the sound wave signal with appropriate HRTF data and replaying it through a pair of stereo headphone. The principal of the virtual sound combination using the HRTF data is calculating the convolution of the appropriate pair of HRIR with the sound signal (Xie, 2008). Here $h_l(n)$ and $h_r(n)$ denote the left and right ear HRIR respectively, and $x(n)$ denotes the input signal series. Then, the outputs after HRIR filtering are

$$y_r = h_r(n) * x(n) \tag{31}$$

$$y_l = h_l(n) * x(n) \tag{32}$$

Thus, the direction of the sound wave must be known in advance in order to select the appropriate HRIR. Using the method proposed in this paper, the virtual sound can be combined without the knowledge of the direction of the sound wave through filtering the HRTF measurements using the beam output of the EBF to yield the binaural signals for a pair of stereo headphone. Figure 3 shows the diagram of the proposed procedure of virtual sound combination.

5. Simulation analysis

In simulations, a UCA of 24 omnidirectional microphones with radius R =0.093m is mounted around a rigid sphere with radius R_0 =0.085m. Therefore, the maximum order of the eigen-beams derived from the soundfield decomposition is N=12. The HRTF auto-selection algorithm proposed in this paper is used to get the HRTF for the sound direction. Figure 4 and 5 show the HRTF approximation and the KEMAR HRTF measurements of MIT (MIT Media Lab, 1994) of the left and right ear for the direction of 35°, where the dotted line denotes the approximation data, and the solid line denotes the KEMAR measurements. The order of the eigenbeams used in the eigenbeam beamforming is 6. It can be seen that there is a good approximation for frequencies until about 3.5kHz. The order of the eigenbeams used in Fig. 6 and 7 is 12. There is a good approximation for frequencies until about 7.5kHz. It can be concluded that the HRTF auto-selection algorithm proposed in this paper is valid and can be used to synthesize virtual sound. Furthermore, the higher the order of the eigenbeams used in the EBF, the more similar of the approximation with the measurements at a given broad frequency band.

6. Experiments

There are traditionally two kinds of evaluation methods to evaluate the validity of the HRTF approximation. The first one is the objective evaluation, which compares the consistency of the amplitude and phase responses in the whole frequency band. The other one is the subjective evaluation, which compares the consistency of listeners' auditory perception caused by the virtual sound with those of the real sound. In order to verify the validity of the proposed method, a subjective listening experiment was conducted. The listening test involved 12 human subjects whose age range from 22 to 34 years, and all of them have normal hearing.

The experiment was conducted in an anechoic chamber to minimize unwanted reflections. The sound used in the experiment was 5s music sound, bandpass filtered between 50Hz and 6kHz. In the experiment, 9 positions, ranging from -60° to 60° with 15° intervals, were pre-selected to position the source, which were at the same height with the head of the listeners and the arrays. The right side of the listener's head was defined as positive angle, and the left side of the listener's head was defined as negative angle. The distance between the microphone array and the source was 2 m. Binaural signals were combined using the measured HRTF data and the approximation data respectively. The number of eigenbeams used in HRTF approximation was 12. The binaural signals were send to a pair of stereo headphones to reproduce the space sound. In order to have the proper listening perception, the listeners received extensive training with the real sound prior to the actual data collection. During the experiment, the listeners judged the directions of the virtual sound played in the headphones and recorded the directions. The sound was played stochastically at the 9 positions. Stochastic analysis was conducted to the experimental data. Figure 8 and 9 show the result of the subjective localization experiment. The judgment angle is plotted as a function of the target angle. The size of each symbol represents the percentage of judgments from each target-angle that fall within each judgment-angle. The positive-slope diagonal is also plotted in the panel. If the judgment angle was identical to the target angle, then all points would fall along this line. It shows that the perceived angles of the source are in very good agreement with the presented angles of the source since most data points fall on the diagonal of the plot. Another view of these data is shown in Fig. 10, which indicate the statistic of the localization errors in a bar chart. It can be seen in this result that the mean localization errors using the HRTF approximation data are very similar to the measured data. The largest localization error is 13.75°, less than one test interval. It can be concluded that the validity of the proposed method is verified by the subjective listening experiments.

7. Conclusion

In the application of virtual audition technology to the sound source localization and identification in the battlefield, the sound wave direction must be estimated in advance in order to select the proper HRTF to combine the binaural signals as this message is often unknown. Replaying the binaural signals through a pair of stereo headphones can reproduce the space sound in the listener's sensation. Traditional virtual sound combination technology need estimate the DOA of the sound wave which will increase the computing load of the system. An HRTF auto-selection algorithm is proposed in this paper using the directivity of EBF and the algebraic expression of the HRTF. Moreover, the EBF technology used in this paper has higher directivity than DSBF technology at low frequencies. Computer simulations demonstrate that there are good consistency between the HRTF approximation and the measurements under certain frequency band. Subjective listening experiment results indicate that the subjects are capable of localizing the presented source direction within an average error of 13.75°, which are consistent with the results using the measured HRTF data. Thus, the HRTF auto-selection algorithm proposed in this paper provides a resolution for the virtual sound applications with unknown sound wave directions.

References

Bai, M. R., & Lin, C. P. (2004). Microphone array signal processing with application in three- dimensional spatial hearing. *J. Acoust. Soc. Am*., 117(4): 2112-2121.

Brandstein, M., & Ward, D. (2001). Microphone Arrays: Signal Processing Techniques and Applications. New York: Springer-Verlag.

Duraiswami, R., Zotkin, D., & Gumerov, N. (2004). Interpolation and range extrapolation of HRTFs. *IEEE ICASSP*, IV: 45-48.

Mathews, C. P. & Zoltowski, M. D. (1994). Eigen-structure techniques for 2D angle estimation with uniform circular arrays. *IEEE Trans. Signal Process*., 42(9): 2395-2407.

Meyer, J. (2001). Beamforming for a circular microphone array mounted on spherically shaped objects. *J. Acoust. Soc. Am.*, 109(1): 185-193.

MIT Media Lab, (1994) HRTF measurements of a KEMAR dummy-head microphone. [Online] Available: http://sound.media. mit.edu./ resources/KEMAR.html (May 18, 1994).

Moller, H., Soerensen, M. F., Hammershoei, D., & Jensen C. B. (1995). Head-related transfer functions of human subjects. *J. Audio Engineer. Soc.*, 43(5): 300–321.

Rafaely, B. (2004). Plane-wave decomposition of the sound field on a sphere by spherical convolution. *J. Acoust. Soc. Am*., 116(4): 2149-2157.

Savioja, L., Huopaniemi, J., Lokki, T., & Vaananen, R. (1999). Creating interactive virtual acoustic environments. *J. Audio Engineer. Soc.*, 47(9): 675-705.

Teutsch, H., Kellermann, W. (2006). Acoustic source detection and localization based on wavefield decomposition using circular microphone arrays. *J. Acoust. Soc. Am*. 120(5): 2724-2736.

Xie, B. S. (2008). *Head-related transfer function and virtual audition* . Beijing: National defence industry press.
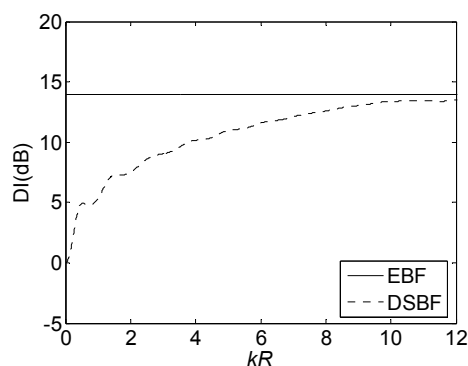
Figure 1. DI of the EBF and DSBF array.
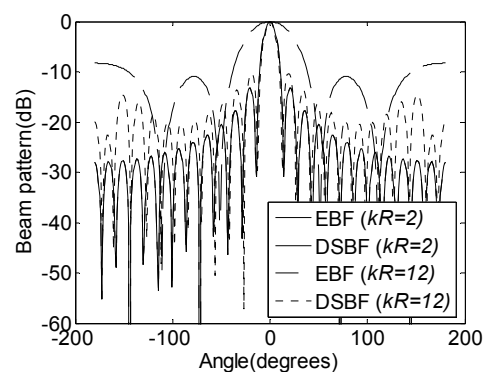
Figure 2. Beam pattern of the EBF and DSBF array.
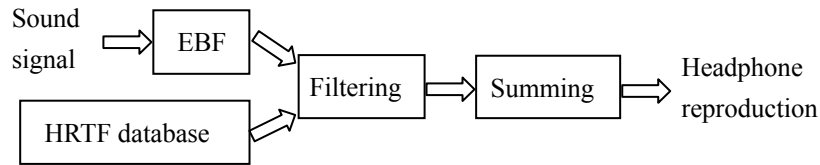
Figure 3. the diagram of virtual sound combination based on HRTF auto-selection
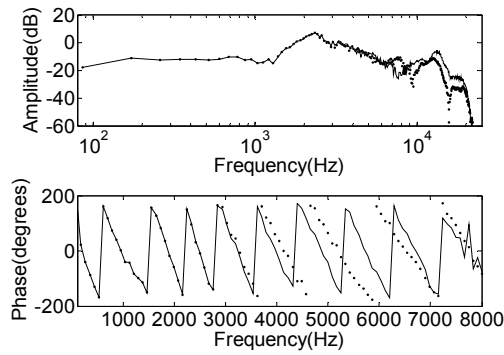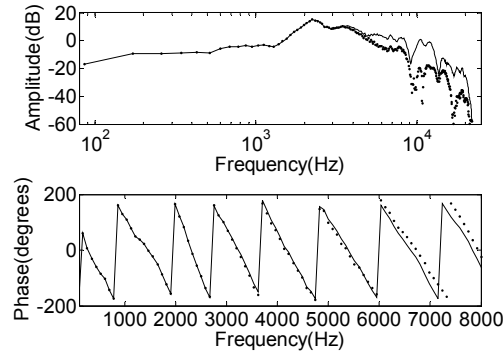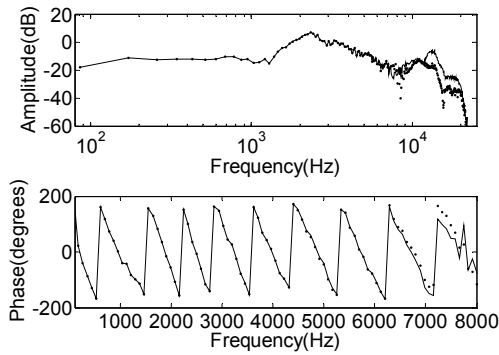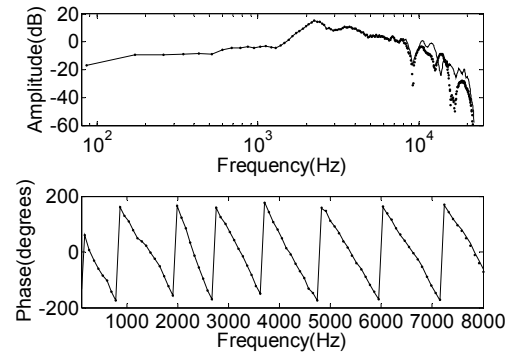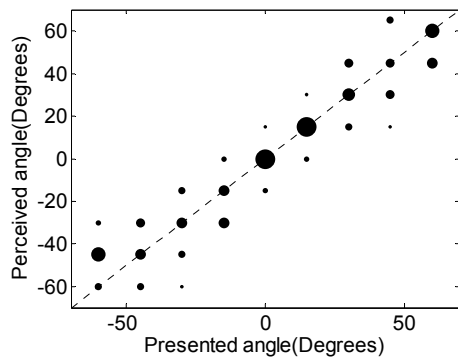
Figure 4. Left ear HRTF (*N*=6).

Figure 5. Right ear HRTF (*N*=6).

Figure 6. Left ear HRTF (*N*=12).

Figure 7. Right ear HRTF (*N*=12).

Figure 8. The results of the subjective listening experiment using approximated HRTF
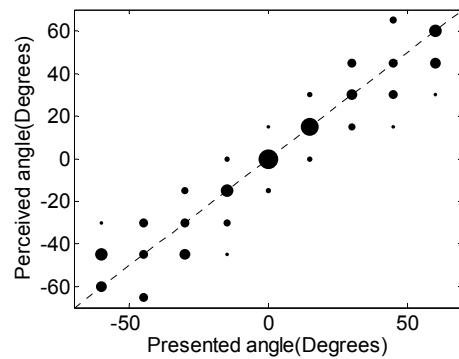
Figure 9. The results of the subjective listening experiment using measured HRTF
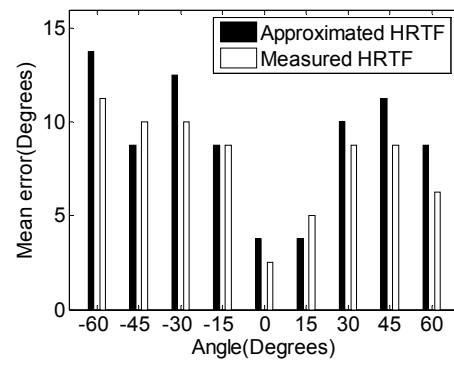
Figure 10. The localization errors for the presented angles