Chapter 6

Head-Related Transfer Functions and Virtual Auditory Display

Xiao-li Zhong and Bo-sun Xie

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/56907

1. Introduction

1.1. Sound source localization and HRTFs

In real environments, wave radiated by sound sources propagates to a listener by direct and reflected paths. The scattering, diffraction and reflection effect of the listener's anatomical structures (such as head, torso and pinnae) further disturb the sound field and thereby modify the sound pressures received by the two ears. Human hearing comprehensively utilizes the information encoded in binaural pressures and then forms various spatial auditory experiences, such as sound source localization and subjective perceptions of environmental reflections.

Psychoacoustic experiments have proved that the following cues encoded in the binaural pressures contribute to directional localization [1]:

1. The interaural time difference (ITD), i.e., the arrival time difference between the sound waves at left and right ears, is the dominant directional localization cue for frequencies approximately below 1.5 kHz.
2. The interaural level difference (ILD), i.e., the pressure level difference between left and right ears caused by scattering and diffraction of head etc., is the important directional localization cue for frequencies approximately above 1.5 kHz.
3. The spectral cues encoded in the pressure spectra at ears, which are caused by the scattering, diffraction, and reflection of anatomical structures. In particular, the pinna-caused high-frequency spectral cue above 5 to 6 kHz is crucial to front-back disambiguity and vertical localization.

4. The dynamic cue, i.e., the change in binaural pressures (thus ITD and ILD) introduced by head movement, also contributes significantly to front-back disambiguity and vertical localization.

In this chapter, the sound source position is specified by a spherical coordinate (r, θ, ϕ), where r denotes the source distance relative to the head center (i.e., the origin). Elevation ϕ varies from –90° to 90° with –90°, 0°, 90° denoting below, horizontal and above, respectively. Azimuth θ varies from 0° to 360° with θ = 0°, 90°, 180°, and 270° denoting front, right, behind, and left in the horizontal plane, respectively.

When both sound source and listener are fixed, the acoustical transmission from a point source to the two ears can be regarded as a linear-time-invariable (LTI) process (see Figure 1). Head-related transfer functions (HRTFs) are defined as the acoustical transfer function of this LTI system:

$$H_L(r,\theta,\phi,f,a)=\frac{P_L(r,\theta,\phi,f,a)}{P_0(r,f)},\quad H_R(r,\theta,\phi,f,a)=\frac{P_R(r,\theta,\phi,f,a)}{P_0(r,f)}. \tag{1}$$

where P_L and P_R represent sound pressures at left and right ears, respectively; P_0 represents the free-field sound pressure at head center with the head absent. Generally, HRTFs vary as functions of frequency f and source position (r, θ, ϕ) (distance and direction) as well as individual a. For r > 1.0 − 1.2 m, HRTFs are approximately independent of source distance and called far-field HRTFs. For r < 1.0 m, however, HRTFs are relevant to source distance and called near-field HRTFs.

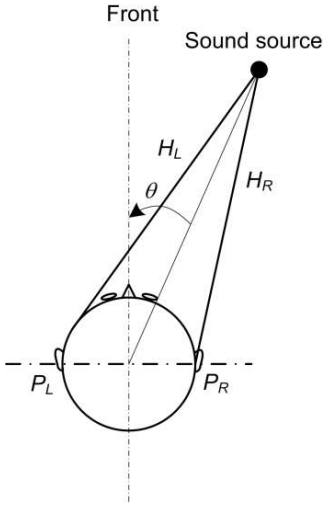

Figure 1. Acoustic transmission from a point sound source to the two ears

http://dx.doi.org/10.5772/56907

The measurement point for P_L and P_R in Eq. (1) varies across studies, among which the eardrum is a most natural choice. Since the external ear canal is proved to be a direction-independent one-dimensional transmission line below 10 kHz, the binaural pressures can be measured at an arbitrary point from the blocked or open entrance of ear canal to the eardrum [2]. Although the pressures differ at different reference points, they all capture the directional information of sound source.

The time-domain counterparts of HRTFs are known as head-related impulse responses (HRIRs), which relate to HRTFs by Fourier transform. HRIRs are the impulse responses from a point sound source to two ears in the free-field. More generally, in reflective environments such as a room, the impulse responses from a source to two ears are called binaural room impulse responses (BRIRs). BRIRs can be regarded as generalized HRIRs from a free-field without reflections to a sound field with reflections.

HRTFs or HRIRs contains most of above-mentioned source localization cues, except the dynamic cue caused by head movement. Therefore, they are vital to the study of binaural localization [3]. One important application of HRTFs is the binaural synthesis in virtual auditory display (VAD). These are the major contents of this chapter.

2. Obtainment of HRTF

2.1. Measurement

Measurement is a conventional and accurate way to obtain HRTFs, especially for human individuals. The principle and methods for HRTF measurement are similar to those for measuring the response of an acoustical LTI system. Figure 2 shows a typical block diagram of HRTF measurement. The measuring signal generated by a computer is rendered to a loudspeaker after passing through a D/A converter and a power amplifier. Resultant signals are recorded by a pair of microphones positioned at subject's two ears, and then delivered to the computer after amplification and A/D conversion. Finally, HRTFs or HRIRs are obtained after some necessary signal processing.

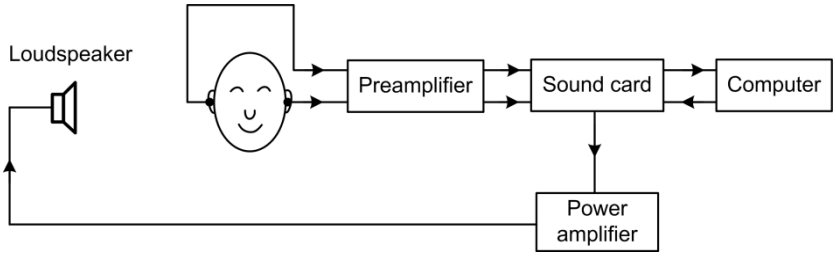

Figure 2. The block diagram of HRTF measurement

To avoid the influence of environment, measurements are usually undertaken in an anechoic chamber, or in a reflective room with a succeeding time-domain windowing so as to rule out reflections. Post-equalization is usually supplemented to correct the non-ideal transmission response in the measurement chain in Figure 2.

Due to the one-dimensional transmission characteristic from the entrance of ear canal to the eardrum, the binaural pressures can be recorded at an arbitrary point from entrance of ear canal to eardrum. In practice, recording binaural pressures with miniature microphones at the blocked ear cannel entrance is the most convenient method for HRTF measurements of human subjects, see Figure 3.

Figure 3. Blocked-ear-canal measurement with miniature microphone

Various signals, such as impulse signals, exponential sweep signals, and pseudo-random noise signals, have been used in HRTF measurement, among which the bipolar maximal length sequence (MLS) is often used. The MLS is a pseudo-random noise sequence (signal) with a deterministic and periodic structure, but possesses characteristics similar to a random noise. In particular, it possesses the lowest crest factor and pulse-like autocorrelation function (equivalent to a nearly uniform power spectrum). For a long N-point MLS, its HRIR $h(n)$ is related to the circle cross-correlation calculation between the recorded signal y and MLS signal x as:

$$R_{xy}(n) \approx h(n) - \frac{1}{N}\sum_{n=0}^{N-1} h(n). \tag{2}$$

One advantage of the measurement using the MLS-like pseudo-random noise sequence is its noise immunity. The deterministic and periodic characteristics of the MLS allow a high signal-

http://dx.doi.org/10.5772/56907

to-noise ratio in measurement by means of averaging. In addition, the low cross-correlation among the time-order-reversed MLS also allows for a fast measurement of HRTFs at different directions using multiple sources simultaneously [4].

Figure 4 is the photo of a set of computer-controlled HRTF measurement apparatus in our laboratory [5]. Multiple sound sources (i.e., small loudspeakers) are arranged in different elevations. A computer-controlled horizontal turntable is adopted, on which a rod is installed to support the artificial head or a seat for a human subject. The source distance relative to the head center is adjustable with a maximum distance of 1.2 m.

Thus far, some research groups have constructed databases for measured far-field HRTFs from artificial heads or human subjects [6-13]. Some databases are available on the internet. Foremost of these are the HRTFs of Knowles Electronic Manikin for Acoustic Research (KEMAR), an artificial head-and-torso model for the research of binaural hearing, see Figure 4. The KEMAR HRTF database constructed by the MIT Media laboratory has been widely used in numerous studies. The database contains 512-point far-field (r = 1.4 m) HRIRs of 710 spatial directions from elevation −40° to 90°. In the measurements, the binaural pressures were recorded at the ends of the occluded-ear simulator, i.e., at eardrums.

Figure 4. Photo of HRTF measurement apparatus in our lab.

However, the HRTFs of an artificial head merely represent the mean characteristics of a certain population, based on which the artificial head was designed, rather than the individual characteristics of humans. For human HRTFs, the CIPIC database consists of 43 subjects mainly from western population [10]. There are statistically significant differences in anatomical dimensions and shapes as well as resulting HRTFs among different populations. Thus, our group measured and established a far-field HRTF database with 52 Chinese subjects (half males and half females) in 2005 [13]. This database includes far-field 512-point HRTFs at 493 source directions per subject with 44.1 kHz sampling frequency and 16-bit quantization. The database also includes 17 anthropometric parameters relating to dimensions of head and pinna, and so on.

Near-field HRTF measurement is relatively difficult. First, a near-field point sound source is urgently needed. In the case of near-field, an ordinary small-size loudspeaker system is no long approximately being as a point sound source due to its size, directivity, and multiple scattering between source and subject. Second, near-field HRTF measurement is much more time-consuming because measurements at various distances are required due to the distance dependency of near-field HRTF. Such tedious measurement process is particularly unbearable for human subjects. Till now, only a few research groups have measured near-field HRTFs for artificial heads, and no public database is available [14-16]. Based on a spherical dodecahedron sound source, Yu et al. measured the near-field HRTF for KEMAR with DB 60/61 small pinnae [17]. The binaural pressures were recorded at the ends of a pair of Zwislocki occluded-ear simulators. The resultant database includes HRIRs at 10 source distances of 0.20, 0.25, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, and 1.00 m, and 493 directions at each source distance. Each HRIR is 512-point length with 44.1 kHz sampling frequency and 32-bit (float) quantization.

2.2. Computation

Computation is an alternative method for obtaining HRTFs. From mathematical and physical perspectives, calculating HRTFs pertains to solving the scattering problem caused by the human anatomical structures; that is, solving the wave or Helmholtz equation subject to certain boundary conditions.

The analytical solution of HRTFs can be solved from some simplified human anatomical geometry. The spherical-head model is the simplest model for HRTF calculation. As shown in Figure 5, the head is simplified as a rigid sphere with radius a, and the ears as two opposite points on the sphere. For an incident plane wave or a sinusoidal point source that is infinitely distant from the sphere center, the far-field HRTF can be calculated by Rayleigh's solution for pressure at the sphere surface, as [18]

$$P(\Gamma, f) = -\frac{P_0}{(ka)^2}\sum_{l=0}^{\infty}\frac{(2l+1)\,j^{l+1}P_l(\cos\Gamma)}{d\,h_l(ka)/d(ka)}, \tag{3}$$

where Γ is the angle between incident direction and received point (ear) on the sphere surface; $k = 2\pi f/c$ is the wave number; $P_l(\cos\Gamma)$ is the Legendre polynomial of degree l; $h_l(ka)$ is the lth-order spherical Hankel function of the second kind. The calculation of spherical-head HRTF can be extended to the case of an arbitrary (finite) source distance [19].

To investigate the torso effect on HRTFs, a simplified head-and-torso model called the snowman model was used for HRTF calculation [20]. The model consists of a spherical head located above a spherical torso, and the HRTFs of the model can be solved using the method of multi-scattering or multipole re-expansion [21].

The calculation from the simplified head-and-torso model reflects some basic features of HRTFs, but it is roughly valid at low and mid frequencies below 3 kHz. The geometry of a real human head is more complex than a sphere and the contribution of pinnae to high-frequency HRTFs is significant. To improve HRTF calculation accuracy, some numerical methods such

http://dx.doi.org/10.5772/56907

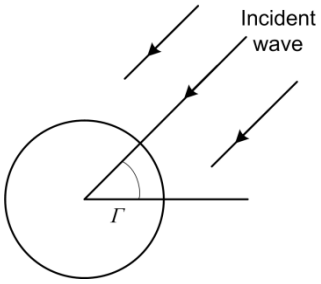

Figure 5. The spherical-head model for HRTF calculation

as boundary element method (BEM) have been developed [22-24]. In BEM calculation, the solution to the scattering problem of human anatomical structures can be expressed as a Kirchhoff–Helmholtz integral equation. The geometrical surfaces of a human or artificial head (such as head and pinnae) are first acquired by a laser 3D scanner or other scanning devices, and then discretized into a mesh of triangular elements. The largest length of the elements is closely related to the maximal frequency to be analyzed and should not exceed 1/4 to 1/6 of the shortest wavelength concerned. Consequently, the Kirchhoff –Helmholtz integral equation is converted into a set of linear algebra equations. Currently, the BEM calculation yields HRTFs with reasonable accuracy up to or near 20 kHz.

However, numerical methods are also time-consuming. It usually takes dozens to hundreds of hours for a typical personal computer to calculate a set of HRTFs at various source directions by conventional BEM (depending on computational power, the number of elements, frequency, and spatial resolution, etc.). High computational costs make calculation difficult. To reduce the computational cost, the acoustic principle of reciprocity can be incorporated in HRTF calculation. According to the acoustic principle of reciprocity, interchanging the source/receiver positions results in identical pressures. In HRTF calculation, therefore, source position can be fixed at the two ears and receiver points are selected at various spatial directions outside the body. There is still some calculation due to each receiver, but these calculations are much faster than the conventional calculation [23]. Moreover, some researches proposed a fast multipole accelerated boundary element method (FMM BEM) for HRTF calculation [25].

2.3. Customization

Aside from measurement and calculation, in practical use, individualized HRTFs can also be approximately obtained by customization. Generally, HRTFs can be customized using anthropometry-based or subjective-selection-based methods.

The anthropometry-based methods hypothesize that there exists a strong relationship between individual HRTFs and individual anatomical features, because HRTFs characterize the interaction between incident sound waves and human anatomical structures. Accordingly, the individualized HRTFs can be approximately estimated or matched from appropriate anatomical measurements and a baseline database of HRTFs. Practical

customization methods include selecting the best-matched HRTFs from a baseline database in terms of the similarity on the measured anatomical parameters among the subject and those in the baseline database [26]; scaling the logarithmic HRTF magnitude from a generic HRTF using anthropometry-predicted scale factor [27]; establishing statistical relationship between the parameterized representation of HRTFs and anatomical parameters, and then predicting the parameters for HRTF representation by anthropometric measurements [28]. The subjective-selection-based methods approximately evaluate the individual HRTFs by appropriate subjective evaluation schemes so as to achieve improved perceived performance, such as localization performance in VAD [29, 30].

Customization of individual HRTFs usually necessitates a baseline database with adequate subjects so as to adapt to the diversity in individualized HRTFs. Customization is simpler than measurement or calculation and yields moderate results, but its accuracy is inferior to measurement and calculation.

3. Physical characters of HRTF

3.1. Time- and frequency-domain characteristics

Although HRIRs or HRTFs vary across individual, some common characteristics in time- and frequency-domain are observed. Figure 6 shows far-field HRIRs of KEMAR with small pinnae at horizontal azimuths 30° and 90°[8]. At azimuth 30°, the HRIR magnitude at preceding 30 to 58 samples is approximately zero, corresponding to the propagation delay from sound source to ears. In practice, a time window is usually applied to raw HRIRs, and thus the initial delay only has relative significance. The main body of the HRIRs, which reflects the complicated interactions between incident sound waves and anatomical structures, persists for about 50 to 60 samples. Subsequently, the HRIR magnitude returns to nearly zero. When the sound source deviates from directly front and back directions, the initial delay difference in the left- and right-ear HRIRs reflects the propagation time difference from the sound source to the left and right ears, i.e., ITD. At azimuth 90°, for instance, the left-ear HRIR lags to the right-ear HRIR with a relative delay of 28 samples (approximately 635 μs at a sampling frequency of 44.1 kHz). Moreover, when the sound source is located contralateral to the concerned ear, for example, at an azimuth of 90° for the left ear, the HRIR magnitude is visibly attenuated because of the head shadow effect. As elevation deviates from the horizontal plane, the difference in initial delay and magnitude between left and right HRTFs at lateral directions reduces.

Figure 7 shows the magnitudes of HRTFs corresponding to the HRIRs in Figure 6. At low frequencies below 0.4 to 0.5 kHz, the normalized log-magnitudes of HRTFs approach 0 dB and are roughly frequency-independent because of the negligible scattering and shadow effect of the head. The decrease in magnitude below 150 Hz is caused by the low-frequency limit of loudspeaker response used in HRTF measurement, rather than by the HRTF itself. Because of the finite source distance relative to the head center (r = 1.4 m) in HRTF measurement, a 2 to 4 dB difference between the left- and right-ear HRTF magnitudes is observed at a lateral azimuth of 90° even at low frequencies. As frequency increases, the normalized log-magni-

http://dx.doi.org/10.5772/56907

tudes of HRTFs vary with frequency and azimuth in a complex manner, due to the overall filtering effects of the head, pinna, torso, and ear canal. The apparent peak in HRTF magnitude at 2 to 3 kHz results from the resonance of the occluded-ear simulator of KEMAR. Above 4 kHz, the contralateral HRTF magnitudes (for example, the left ear at an azimuth of 90°) are visibly attenuated because of the low-pass filtering properties of the head shadow. The ipsilateral HRTF magnitudes (for example, the right ear at an azimuth of 90°) increase to a certain extent, although some notches occur. This phenomenon is partially attributed to the approximate mirror-reflection effect of the head on ipsilateral incidence at high frequencies, thereby leading to increased pressure for ipsilateral sound sources.

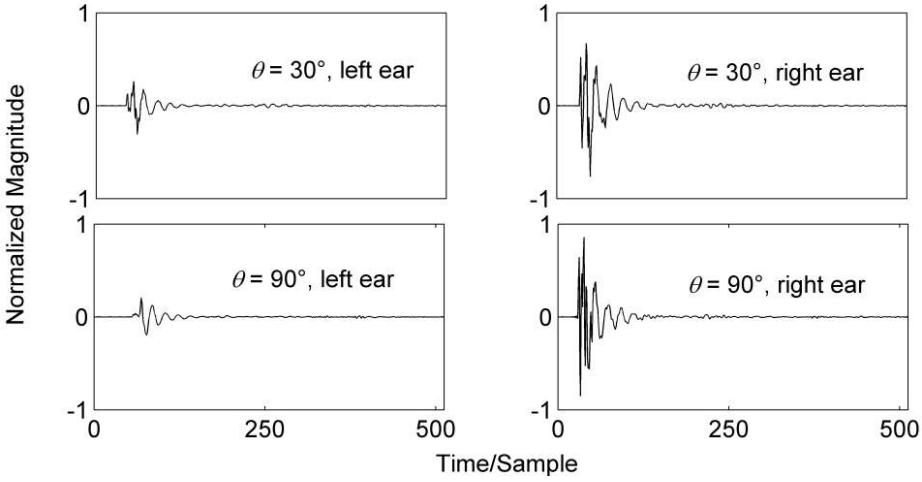

Figure 6. KEMAR far-field HRIRs at azimuths of 30° and 90° in the horizontal plane

To demonstrate the individuality of HRTFs, Figure 8 shows the normalized magnitudes of left-ear HRTFs at ($\theta = 0°$, $\phi = 0°$) for 10 subjects randomly selected from the Chinese subject HRTF database. Considerable inter-subject differences in HRTF magnitudes are observed above 6 to 7 kHz.

3.2. Localization cues in HRTFs

Various localization cues stated in Section 1 can be evaluated from measured HRTFs. ITD is a dominant azimuthal localization cue below 1.5 kHz. There are various evaluation methods for ITD, among which ITD_p calculated from interaural phase delay difference is directly related to low-frequency localization,

$$ITD_p(\theta,\phi,f) = \frac{\Delta\psi}{2\pi f} = -\frac{\psi_L - \psi_R}{2\pi f} \tag{4}$$

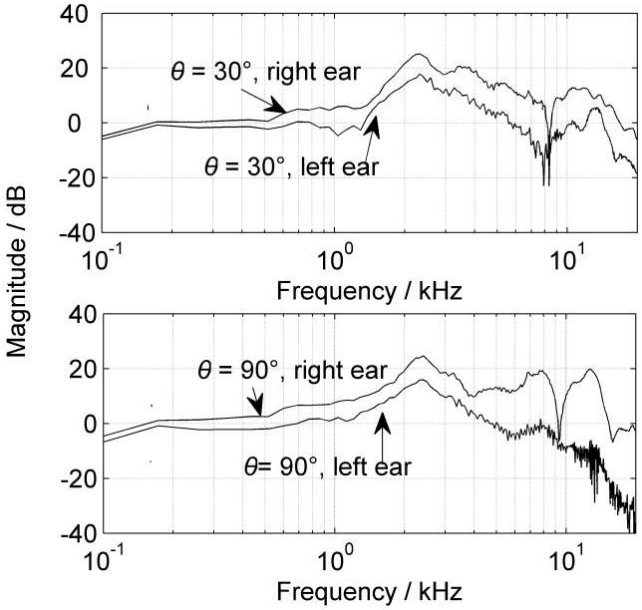

Figure 7. Magnitudes of KEMAR HRTFs at azimuths of 30° and 90° in the horizontal plane

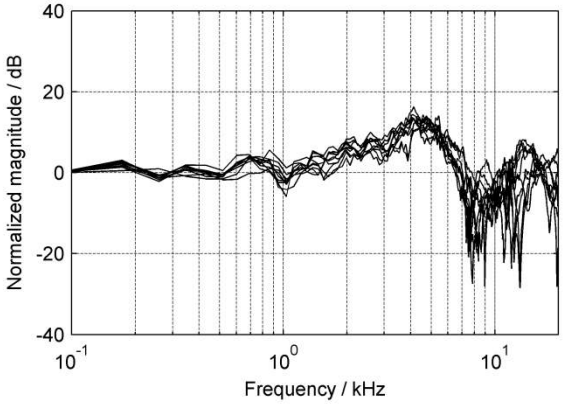

Figure 8. Left-ear HRTF magnitudes for 10 subjects at azimuth 0° in the horizontal plane

where ψ_L and ψ_R denote the unwrapped phases of left- and right-ear HRTFs, respectively. Besides, ITD can be evaluated as τ_{max} at which the normalized interaural cross-correlation function of a left- and right-ear HRIR pair maximizes.

http://dx.doi.org/10.5772/56907

$$\max\{\Phi_{LR}(\tau)\} = \max\left\{\frac{\int_{-\infty}^{+\infty} h_L(t+\tau)h_R(t)dt}{\left\{\left[\int_{-\infty}^{+\infty} h_L^2(t)dt\right]\left[\int_{-\infty}^{+\infty} h_R^2(t)dt\right]\right\}^{1/2}}\right\} \quad with\ |\tau| \le 1ms \tag{5}$$

$$ITD_{corre}(\theta,\phi) = \tau_{\max} \tag{6}$$

In some studies, ITD is usually evaluated by leading-edge detection, i.e., detecting instants $t_{L,\eta}$ and $t_{R,\eta}$ at which the HRIRs first reach a certain percentage η (e.g., 10%) of maximum peak amplitudes. Then, ITD_{lead} is calculated by

$$ITD_{lead}(\theta,\phi) = t_{L,\eta} - t_{R,\eta}\ . \tag{7}$$

The ITD_{corre} and ITD_{lead} are relevant to source direction but independent of frequency.

Figure 9 plots the variation of horizontal ITDs with azimuths from 0° to 180°. The ITDs are calculated from MIT KEMAR (far-field) HRTFs, and left-right symmetric HRTFs are assumed. The ITDs evaluated by four different methods, including ITD_p at 0.35 and 2.0 kHz, ITD_{lead} with η = 10%, and ITD_{corre}, are shown in the figure. Before the ITD_{corre} is calculated, a pair of HRIRs is subjected to low-pass filtering below 2.0 kHz to avoid the influence of resonance from the occluded-ear simulator. The ITDs derived by different methods generally vary with azimuth in a similar manner. The ITDs are zero at azimuths of 0° and 180°, then gradually increase as the source deviates from the median line and maximizes at directions close to the lateral. For example, the maximal ITD_{corre} is 710 μs at azimuth 90°. At a given azimuth, however, some differences in ITD value exist among the ITDs derived from different methods, with the ITD_p at 0.35 kHz being the largest and ITD_{lead} being the smallest. The range of ITD variation decreases as elevation deviates from the horizontal plane.

ILD defined in Eq. (8) is another localization cue at high frequency.

$$ILD(r,\theta,\phi,f) = 20\log_{10}\left|\frac{H_R(r,\theta,\phi,f)}{H_L(r,\theta,\phi,f)}\right| (dB). \tag{8}$$

According to Eq. (8), ILD depends on both source direction and frequency. Figure 10 shows ILD varying with azimuth at different frequencies. This ILD is calculated using the MIT-KEMAR (far-field) HRTFs associated with the DB-061 small pinna. At low frequency of 0.35 kHz, ILD is small (within 4.5 dB) and almost invariable with source azimuth. The non-zero ILD at low frequency is partly due to the finite source distance (1.4 m) in the MIT-KEMAR HRTF measurement. For an infinitely distant source, the ILD at low frequency trends to zero.

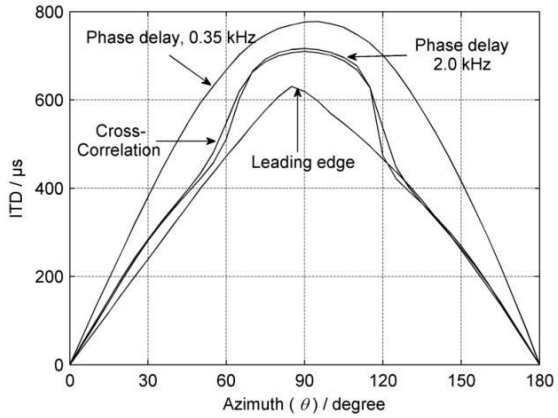

Figure 9. Horizontal ITDs of KEMAR evaluated by various methods.

As frequency increase, ILD increases and exhibits a complex variation manner with azimuth and frequency, with the value at the front (0°) and back (180°) always being zero. The range of ILD variation decreases as elevation deviates from the horizontal plane.

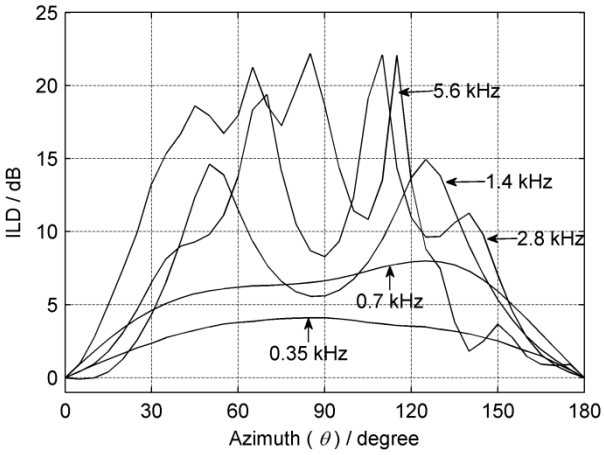

Figure 10. KEAMR ILDs in the horizontal plane for five frequencies.

The spectral cues provided by HRTFs at high frequency are vital for front-back and elevation localization. Among various spectral cues, the first (lowest) frequency notch in HRTF magnitude caused by the out-of-phase interference of pinna reflection/diffraction and direct sound wave in the ear canal is of importance. The elevation dependence of the central frequency of

http://dx.doi.org/10.5772/56907

the pinna notch is regarded as an important vertical localization cue. Figure 11 shows the HRTF magnitude spectra of a typical Chinese subject in the median plane with $\theta = 0°$ and elevation $\phi = -30°$, 0°, and 30°[13]. The pinna notch at 6 to 9 kHz is observed in the spectra. The central frequency of the pinna notch at $\phi = -30°$, 0°, and 30° are 6.5 (6.2), 8.1(7.9), and 8.8 (8.7) kHz for the right (left) ear, respectively. At high elevations with $\phi \geq 60°$, the pinna notch gradually vanishes. Considerable inter-individual differences exist in the central frequency of the pinna notch and other high-frequency spectral features of HRTFs. Therefore, HRTFs are highly individual dependent. Actually, statistical results indicate that HRTFs are left-right asymmetric above 5 − 6 kHz [31].

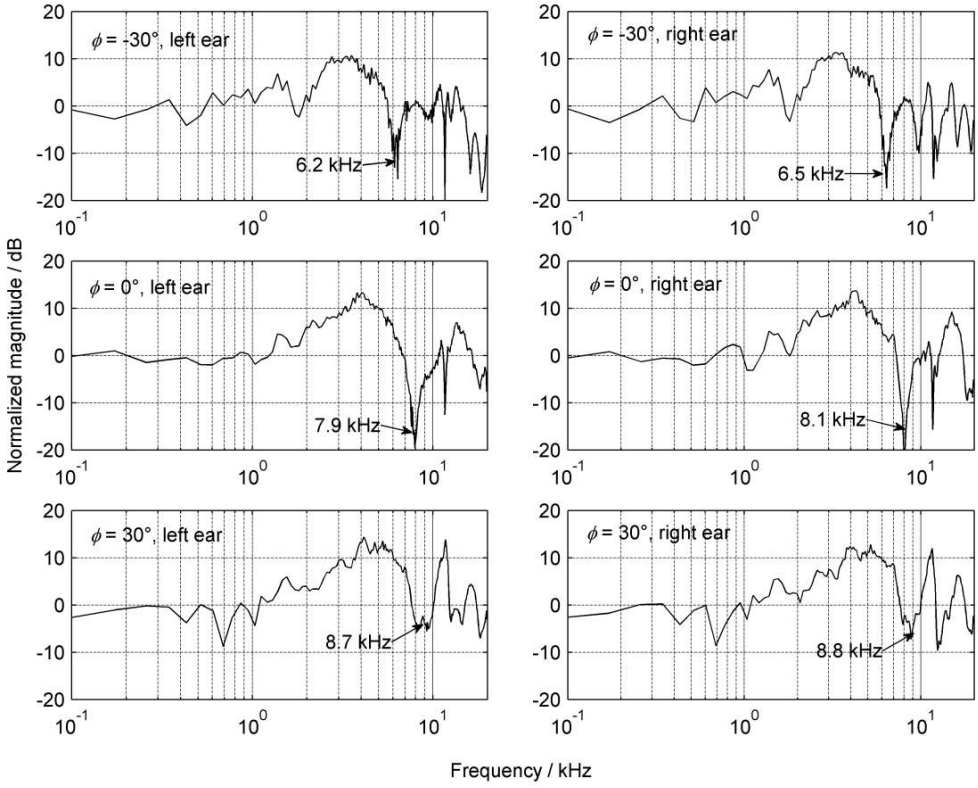

Figure 11. HRTF magnitude spectra for a typical Chinese subject at elevations –30°, 0°, and 30°

3.3. The minimum-phase characteristics of HRTFs

At a given source direction, HRTF is a complex-valued function of frequency and can be decomposed by the product of a minimum-phase function $H_{min}(\theta, \phi, f)$, an all-pass function $\exp[\, j\psi_{all}(\theta, \phi, f)]$, and a linear-phase function $\exp[- j2\pi f T(\theta, \phi)]$:

$$H(\theta,\phi,f) = H_{\min}(\theta,\phi,f)\exp\left[j\psi_{all}(\theta,\phi,f)\right]\exp\left[-j2\pi fT(\theta,\phi)\right] \tag{9}$$

The phase of the minimum-phase function is related to the logarithmic HRTF magnitude by Hilbert transform:

$$\psi_{\min}(\theta,\phi,f) = -\frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{\ln|H(\theta,\phi,x)|}{f-x}dx\,. \tag{10}$$

If the contribution of the all-pass phase component is negligible, Eq. (9) can be approximated as

$$H(\theta,\phi,f) \approx H_{\min}(\theta,\phi,f)\exp\left[-j2\pi fT(\theta,\phi)\right]. \tag{11}$$

Eq. (11) is known as the minimum-phase approximation of HRTFs, in which an HRTF is approximated by its minimum-phase function cascaded with a linear phase or a pure delay. Studies have proved that, in most cases, HRTF is of minimum-phase below 10 − 12 kHz[32]. This conclusion is greatly convenient to the HRTF-related signal processing.

Excluding the all-pass phase component from the overall ITD calculation may cause errors when the contribution of this component is non-negligible. Minnaar et al. investigated the all-pass phase of the HRTFs of 40 subjects with 97 spatial directions per subject, and found that below 1.5 kHz the contribution of the all-pass phase component to interaural group delay difference is nearly independent of frequency[33]. If the interaural group delay difference caused by the all-pass phase component is replaced by its value at 0 Hz, the error caused by approximation is less than 30 μs and is inaudible [34].

3.4. Spatial-domain characteristics

Far-field HRTFs are continuous functions of source direction. As stated in Section 2.1, HRTFs are usually measured at discrete and finite directions, i.e., sampled at directions around a spatial spherical surface. Under certain conditions, the HRTFs at unmeasured directions (θ, ϕ) can be estimated from measured data by following linear interpolation method:

$$\hat{H}(\theta,\phi,f) \approx \sum_{i=0}^{M-1} A_i H(\theta_i,\phi_i,f)\,, \tag{12}$$

where $H(\theta_i, \phi_i, f)$ with (θ_i, ϕ_i) $(i = 0, 1, \ldots M–1)$ denotes the measured HRTFs at a constant source distance $r = r_0$ and M appropriate spatial directions; A_i are a set of weights related to the target direction (θ, ϕ).

http://dx.doi.org/10.5772/56907

There are various HRTF interpolation schemes, leading to different selection of measured directions and weights. The bilinear interpolation scheme shown in Figure 12 is commonly used. Let θ_{grid} and ϕ_{grid} denote the measured intervals of azimuth and elevation, respectively. The four adjacent measured directions (θ_1, ϕ_1), $(\theta_1+\theta_{grid}, \phi_1)$, $(\theta_1+\theta_{grid}, \phi_1+\phi_{grid})$ and $(\theta_1, \phi_1+\phi_{grid})$ are denoted by number 1, 2, 3 and 4, respectively. Then the HRTF at a target direction $(\theta, \phi) = (\theta_1+\Delta\theta, \phi_1+\Delta\phi)$ within the grid is estimated as

$$\hat{H}(\theta,\phi,f) \approx A_1 H(1,f) + A_2 H(2,f) + A_3 H(3,f) + A_4 H(4,f), \tag{13}$$

where $A_\theta = \Delta\theta/\theta_{grid}$, $A_\phi = \Delta\phi/\phi_{grid}$, $A_1=(1-A_\theta)(1-A_\phi)$, $A_2=A_\theta(1-A_\phi)$, $A_3=A_\theta A_\phi$, $A_4=(1-A_\theta)A_\phi$.

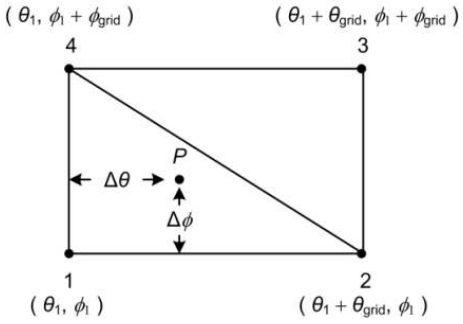

Figure 12. The bilinear interpolation

The HRTF spatial interpolation is closely related to the basis functions linear decomposition of HRTFs. HRTF linear decomposition is categorized into two basic types: spectral shape basis function decomposition and spatial basis function decomposition. Generally, the basis function decomposition representation of an HRTF for a given ear can be written as

$$H(\theta,\phi,f) = \sum_q w_q(\theta,\phi) d_q(f). \tag{14}$$

For spectral shape basis function decomposition, $d_q(f)$ are a series of frequency-dependent spectral shape basis functions; $w_q(\theta, \phi)$ are source direction-dependent weights which may also depend on individual. When the basis functions $d_q(f)$ are specified, $H(\theta, \phi, f)$ is completely determined by weights $w_q(\theta, \phi)$.

Various methods for deriving the spectral shape basis functions $d_q(f)$ are available, and appropriate selection of basis functions depends on situation. There usually exist some correlations among the HRTFs at different directions. If these correlations are completely removed so that the HRTF can be represented by a small set of spectral shape basis functions, data dimensionality is efficiently reduced. Principal components analysis (PCA) is a statistical

algorithm for deriving a small set of orthonormal spectral shape basis functions and then decomposing HRTFs. For example, Kistler et al. found that five spectral shape basis functions derived from PCA accounted for more than 90% variation of logarithmic binaural HRTF magnitudes for $S = 10$ human subjects at 256 source directions[35].

In contrast, in spatial basis function decomposition, $w_q(\theta, \phi)$ in Eq. (14) denote a set of source direction-dependent spatial basis functions; $d_q(f)$ are frequency-dependent weights which may also depend on individual. There are various selections for spatial basis functions, among which azimuthal Fourier series and spatial spherical harmonic functions are two sets of pre-determined and orthonormal spatial basis functions. In the former, HRTF at each elevation plane is decomposed into a weighted sum of azimuthal harmonics. While in the latter, HRTF at arbitrary direction is decomposed into a weighted sum of spherical harmonic functions.

The spatial sampling (Shannon–Nyquist) theorem for HRTF measurement can be derived from the spatial harmonics representation of HRTF. Suppose that the spatial basis functions $w_q(\theta, \phi)$ in Eq. (14) are specified, and the basis functions up to order Q are sufficient for accurately representing HRTF. Given the measured HRTFs at M appropriate, Eq. (14) yields

$$H(\theta_i, \phi_i, f) = \sum_{q=1}^{Q} d_q(f) w_q(\theta_i, \phi_i) \qquad i = 0, 1, 2 \ldots (M-1). \tag{15}$$

At each frequency f, Eq. (15) is a set of M linear equations, with the number of unknown $d_q(f)$ equal to the number of basis functions Q. Selecting M appropriate measurement directions and providing $M \geq Q$, the exact or approximate solution of $d_q(f)$ can be obtained from Eq. (15). The spatial basis functions representation of $H(\theta, \phi, f)$ can then be realized by substituting the resultant $d_q(f)$ into Eq. (14). Given a set of directionally continuous basis functions, HRTF at arbitrary unmeasured direction can be recovered from M directional measurements. Therefore, spatial basis functions decomposition of HRTFs can also be regarded as spatial interpolation or fitting algorithm for HRTFs. Using the azimuthal Fourier series representation of HRTF, Zhong and Xie proved that continuous HRTF in horizontal plane can be recovered from 72 azimuth measurements [36]. When extended to three-dimensional space, recovering spatial continuous HRTF using spherical harmonic functions representation requires $M = 2209$ directional measurements at least [37].

The number of directional measurements required for recovering HRTF is related to the total number of spatial basis functions (i.e., Q) for HRTF representation with $M \geq Q$. Aside from the azimuthal Fourier series and spatial spherical harmonic functions representation, if we can find another small set of spatial basis functions to efficiently represent HRTF, HRTF at unmeasured direction can be recovered from a small set of directional measurements. Xie applied spatial principal components analysis (SPCA) to a baseline HRTF dataset with high directional resolution to derive the small set of spatial basis functions[38]. SPCA is applied to spatial domain rather than frequency (or time) domain in conventional PCA. Using the resultant spatial basis functions, HRTF magnitudes at 493 directions can be recovered from 73 directional measurements. This method is applicable to simplifying HRTF measurement.

http://dx.doi.org/10.5772/56907

3.5. Characteristics of near-field HRTFs

When $r < 1.0$ m, the near-field HRTFs vary with source distance, and exhibit some characteristics that are remarkably distinguished from the far-field HRTFs [14, 39]. The distance dependence of near-field HRTFs is regarded as a distance perception cue. Figure 13 shows KEMAR HRTF magnitudes at $r = 0.2$ m, 0.5 m, 1.0 m and $(\theta, \phi) = (90°, 0°)$ [40]. The magnitudes vary obviously with source distance from $r = 0.2$ m to 0.5 m, and vary less with source distance from $r = 0.5$ m to 1.0 m. The ipsilateral (right) HRTF magnitude increases with decreasing r when a direct propagation path from source to concerned ear exists; the contralateral HRTF magnitude decreases with decreasing r because of the enhancement of the head shadow when a direct propagation path is missing. The variations in HRTF magnitude with r increase the ILD associated with decreasing r. This phenomenon is particularly prominent at low frequencies, thereby relatively increases low-frequency magnitude and therefore causes a perceptible change in timbre.

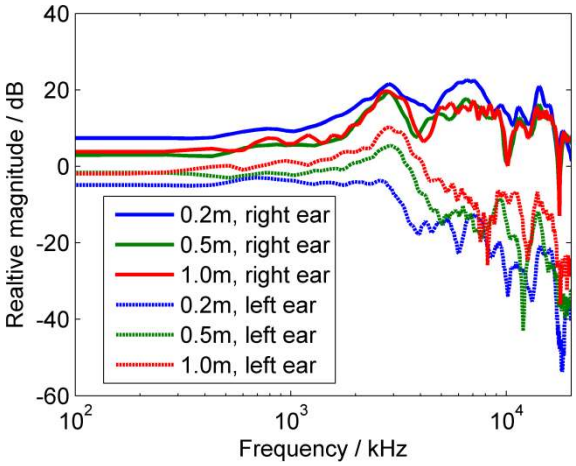

Figure 13. KEMAR HRTF magnitudes at $r = 0.2$ m, 0.5 m, 1.0 m and $(\theta, \phi) = (90°, 0°)$

4. Virtual auditory display

4.1. Basic principles

The binaural sound pressures recorded in the ear canals of a human subject or an artificial head contain the main spatial information of sound events [41]. If the eardrum pressures generated by a real sound event are replicated by sound reproduction, the same spatial auditory event or experience as the real sound event is recreated. This is the basic principle of binaural reproduction. The most straightforward method for binaural reproduction is

recording the binaural signals through a pair of microphones placed in the ear canal of an artificial head or human subject and then presenting the resultant signals via a pair of headphones. This is the basic principle of the binaural recording and playback technique. Another method is synthesizing the binaural signals by signals processing and then presenting via a pair of headphone. This is the core idea of virtual auditory display (VAD).

As stated in Section 1, in the static state, the acoustical transmission from a sound source to two ears is a linear time-invariable course. In the case of free-field sound source, the binaural pressures are related to HRTFs by Eq. (1). If a mono stimulus $E_0(f)$ is filtered with a pair of HRTFs at source direction (θ, ϕ) and the resultant signals are presented via headphone, i.e.,

$$E_L(\theta,\phi,f) = H_L(\theta,\phi,f)E_0(f), \quad E_R(\theta,\phi,f) = H_R(\theta,\phi,f)E_0(f), \tag{16}$$

then the binaural pressures in reproduction is equal to or directly proportional to those created by a real source at direction (θ, ϕ), resulting in a perceived virtual source at corresponding direction. Replacing the HRTFs with different directions in Eq. (16) yields virtual sources at various directions. Note that HRTFs are individual dependent, thus an ideal VAD should use individualized HRTFs in binaural synthesis [42]. Eq.(16) can be equally expressed in the time domain as

$$e_L(\theta,\phi,t) = h_L(\theta,\phi,t) * e_0(t), \quad e_R(\theta,\phi,t) = h_R(\theta,\phi,t) * e_0(t). \tag{17}$$

That is, convoluting the mono stimulus $e_0(t)$ with a pair of HRIRs yields binaural sound signals.

4.2. Signal processing

Direct implementation of binaural synthesis in VAD by Eq. (16) or Eq. (17) usually suffers from low computational efficiency. Alternatively, various HRTF filter model and structure are often designed for binaural synthesis processing. The commonly used HRTF filter models are classified into two catalogs: the moving average (MA) model and autoregressive moving-average (ARMA) model.

In the complex-Z domain, the system function of a Q-order MA model can be written as

$$H(z) = b_0 + b_1 z^{-1} + ... + b_Q z^{-Q}, \tag{18}$$

where $b_0, b_1, \ldots b_Q$ are filter coefficients. In the discrete time domain, the impulse response length of a MA model is $N = Q + 1$, therefore MA is a finite impulse response (FIR) filter model.While the system function of a (Q, P)-order ARMA model can be written as

http://dx.doi.org/10.5772/56907

$$H(z) = \frac{b_0 + b_1 z^{-1} + ... + b_Q z^{-Q}}{1 + a_1 z^{-1} + ... + a_P z^{-P}}, \tag{19}$$

where $a_1,\ldots, a_P$ and $b_0, b_1,\ldots, b_Q$ are filter coefficients. The impulse response length of an ARMA model is infinite, therefore ARMA is an infinite impulse response (IIR) filter model.

HRTF filter design is to appropriately select the coefficients in Eq.(18) or Eq.(19) so that the filter response exactly or approximately matches the target HRTF in some mathematical or perceptual senses. Prior to filter design, some pre-processing schemes are often applied to raw HRTFs so as to simplify the resultant filters. The common simplifications include truncation by a time window so as to reduce the response length, smooth by auditory bandwidth to discard the spectral details of HRTF insignificant to auditory perception, among others. Minimum-phase approximation of HRTF is also beneficial to reduce the filter length.

Various conventional filter design methods, such as windowing or frequency sampling method for FIR filter, and Prony or the Yule–Walker method for IIR filter, have been used in HRTF filter design. Some other sophisticated methods for IIR filter design, such as balanced model truncation (BMT) [43], method using logarithmic error criterion [44] and method of common-acoustical-pole and zero [45], have also been suggested. Frequency-warped filter for HRTFs based on non-uniform frequency resolution of human hearing was also proposed [46]. Those filters can be implemented by various structures and yield reasonable physical and auditory perception performance in VAD. Reference [47] gives a review of HRTF filter design. Aside from above methods, the methods of basis functions linear decomposition of HRTFs (such as PCA) have been applied to binaural synthesis processing. The basis function decomposition-based methods allow for synthesizing multiple virtual sources with a parallel bank of common filters, and then improve the efficiency in multiple virtual source synthesis [48].

4.3. Headphone presentation

As stated in Section 2.1, the binaural signals or HRTFs can be recorded at an arbitrary reference point along the entrance of ear canal to the eardrum, or even at the blocked entrance of ear canal. Therefore, directly rendering the recorded or synthesized binaural signals via headphone without accounting for the measurement position may lead to incorrect eardrum pressures. Moreover, the non-ideal transfer characteristics of the recording and playback chain, which originates from the non-flat frequency responses of the recording microphone and reproducing headphone as well as the unwanted coupling between headphone and external ear, will inevitably cause linear distortions in both magnitude and phase of the reproduced sound pressures at the eardrums. The overall non-ideal transfer characteristics of the recording and playback chain can be represented by a pair of transfer functions, $H_{pL}(f)$ and $H_{pR}(f)$, one for each ear. Ideally, if the recorded binaural signals is equalized by the inverse of $H_{pL}(f)$ and $H_{pR}(f)$ prior to rendering to headphone, the linear frequency distortion in the signal chain can then be eliminated or at least reduced as minimally as possible.

$$F_L(f) = \frac{1}{H_{pL}(f)} \quad and \quad F_R(f) = \frac{1}{H_{pR}(f)}. \tag{20}$$

Figure 14 is the blocked diagram of binaural synthesis along with headphone equalization in a VAD.

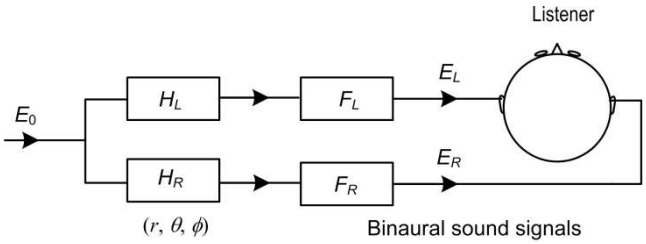

Figure 14. The blocked diagram of binaural synthesis along with headphone equalization

In particular, the transmission character from the electric input signal of headphone to the reference point in the ear canal is described by headphone-to-ear-canal transfer function (HpTF). If the reference point of HpTFs are identical to that of HRTFs and microphones for binaural recording or HRTF measurement have an ideal transmission response, the $H_{pL}(f)$ and $H_{pR}(f)$ in Eq. (20) can be replaced by HpTF, i.e., the binaural signals are equalized by the inverse of HpTFs. For microphone with non-ideal transmission response, providing that the microphones used in HpTFs measurement are identical to those in binaural recording or HRTF measurement, the effect of the non-ideal transmission response of microphone is cancelled in equalization [2]. Note that, for a blocked ear canal measurement, the above equalization method is not always valid unless a headphone with free-air equivalent coupling to the ear (FEC-hedphone) is used. The transmission response on 14 types of headphones were measured [49], and results indicated that the responses of all the headphones (except one) deviated from that of ideal FEC-headphone on the order of 2 to 4 dB above 2 kHz. Moreover, the measurements above 7 kHz were unreliable. In practical uses, whether a headphone can be considered as an FEC-headphone depends on acceptable error.

Similar to the case of HRTFs, HpTFs is individual dependent because of the difference in structures and dimensions of the external ear. Ideally, individualized HpTFs should be incorporated into equalization processing. Moreover, the measured HpTFs for some types of headphone exhibit poor repeatability above 5 to 6 kHz due to the variation of compressive deformation of pinna caused by headphone. This phenomenon makes the equalization difficult.

In headphone presentation, an accurate virtual source can be rendered if the sound pressures for a real sound source are exactly replicated at eardrums. Results of some psychoacoustic experiments with careful individualized HRTFs processing and HpTFs equalization indicate that headphone-rendered virtual source could achieve the equivalent localization performance

http://dx.doi.org/10.5772/56907

as that of free-field real source [50]. However, numerous experimental results indicate that subject-dependent errors in perceived virtual source position are generally existed such as

1. Reversal Error (i.e., front-back or back-front confusion). That is, a virtual source intended in the front hemisphere is perceived at a mirror position in the rear hemisphere, or, less frequently, the reverse. Sometimes, there is confusion with up and down source positions termed up-down or down-up confusion.

2. Elevation error. For example, the direction of a virtual source in the front median plane is usually elevated.

3. In-head localization (i.e., intracranial lateralization). The virtual source or auditory event is perceived inside the head rather than outside headphone, leading to an unnatural hearing experience.

As stated in Section 1, the interaural cues such as ITD and ILD only determine a confusion cone rather than a well-defined spatial position of sound source. The dynamic cue caused by head movement and high-frequency spectral cue introduced by pinnae etc. response for resolving reversal ambiguation and vertical localization. However, conventional static VAD is lack of dynamic cues, so that front-back and vertical localization depend more on high-frequency spectral cue. Unfortunately, the high-frequency spectral cue is elaborate and highly individual-dependent. Errors in binaural recording/synthesis and playback chain, such as non-individualized HRTFs processing, incorrect or lack of headphone equalization, are possible sources responsible for perceived position errors in headphone presentation. Using individual HRTFs and HpTFs processing reduces localization errors. In addition, modeling room reflections in binaural synthesis effectively eliminates in-head localization.

4.4. Loudspeaker presentation

Binaural signals from either binaural recording or synthesis, are originally intended for headphone presentation. When binaural signals are reproduced through a pair of left and right loudspeakers arranged in front of the listener, an unwanted cross-talk from each loudspeaker to the opposite ear occurs. Cross-talk impairs the directional information encoded in the binaural signals. Therefore, cross-talk cancellation should be introduced for binaural reproduction through loudspeakers [51]. That is, prior to loudspeaker reproduction, binaural signals should be pre-corrected or filtered so as to cancel the transmission from each loudspeaker to the opposite ear.

Let $E_L(f)$ and $E_R(f)$, or simply E_L and E_R, denote frequency-domain binaural signals. As illustrated in Figure 15, binaural signals are pre-filtered by a 2×2 cross-talk cancellation matrix and then reproduced through the loudspeakers. The loudspeaker signals are given by

$$\begin{bmatrix} L' \\ R' \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} E_L \\ E_R \end{bmatrix} \tag{21}$$

where A_{11}, A_{12}, A_{21} and A_{22} are the four transfer functions or filters forming the cross-talk cancellation matrix.

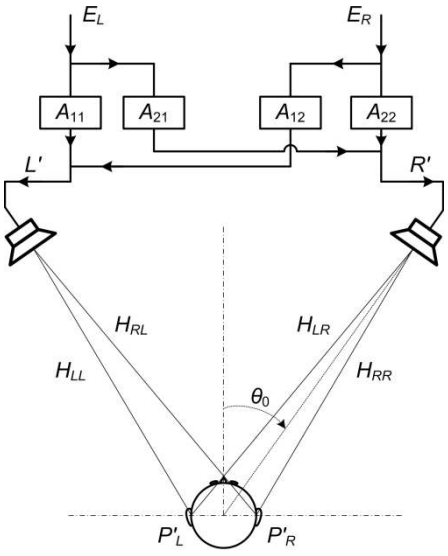

Figure 15. Binaural reproduction through loudspeakers

Let H_{LL}, H_{RL}, H_{LR} and H_{RR} denote the four acoustic transfer functions (HRTFs) from two loudspeakers to two ears, respectively. These four transfer functions are determined by the loudspeaker configuration and listener's location. Then the reproduced pressures at two ears are given by

$$\begin{bmatrix} P_L' \\ P_R' \end{bmatrix} = \begin{bmatrix} H_{LL} & H_{LR} \\ H_{RL} & H_{RR} \end{bmatrix} \begin{bmatrix} L' \\ R' \end{bmatrix} = \begin{bmatrix} H_{LL} & H_{LR} \\ H_{RL} & H_{RR} \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} E_L \\ E_R \end{bmatrix} \tag{22}$$

with Eq. (21) substituted. The transfer characters of cross-talk cancellation matrix are properly selected so that the product of two 2×2 matrixes in Eq. (22) equals to an identity matrix, then the cross-talk is completely cancelled out and the desired binaural signals are exactly delivered to listener's two ears. Therefore, the cross-talk cancellation matrix is obtained from the inverse of the acoustic transfer matrix. In the left-right symmetrical case, $H_{LL} = H_{RR} = H_\alpha$ and $H_{LR} = H_{RL} = H_\beta$, the element of cross-talk cancellation matrix is

$$A_{11} = A_{22} = \frac{H_\alpha}{H_\alpha^2 - H_\beta^2}, \qquad A_{12} = A_{21} = \frac{-H_\beta}{H_\alpha^2 - H_\beta^2}. \tag{23}$$

http://dx.doi.org/10.5772/56907

If the signal processing initially aims to create appropriate loudspeakers signals, the two stages of binaural synthesis and cross-talk cancellation can be merged together, yielding

$$L' = G_L(\theta, f)E_0(f), \qquad R' = G_R(\theta, f)E_0(f). \tag{24}$$

with

$$G_L(\theta, f) = \frac{H_\alpha H_L(\theta, f) - H_\beta H_R(\theta, f)}{H_\alpha^2 - H_\beta^2}, G_R(\theta, f) = \frac{-H_\beta H_L(\theta, f) + H_\alpha H_R(\theta, f)}{H_\alpha^2 - H_\beta^2}. \tag{25}$$

Eq. (24) demonstrates that loudspeaker signals L' and R' for target virtual source at direction θ can be directly synthesized by filtering a mono stimulus $E_0(f)$ with a pair of filters $G_L(\theta, f)$ and $G_R(\theta, f)$. This is the basic principle of loudspeaker-based binaural reproduction or transaural synthesis. The cross-talk cancellation and transaural synthesis can be generalized to the case of binaural reproduction through more than two loudspeakers and with more than one listener [52]. In practice, the acoustic transfer matrix in Eq. (24) may be singular and thus non-invertible at some frequencies. To address this problem, some approximate methods for solving cross-talk cancellation matrix have been proposed [53].

The first problem with binaural reproduction through two frontal loudspeakers is reversal and elevation errors. High-frequency spectral cues is vital to front-back and vertical localization. But these cues cannot be stably replicated in loudspeaker reproduction because of the short wavelength at high frequency. A slight change in listening position causes an obvious variation in binaural pressures. Incorrect dynamic cues often causes back-front confusion in static binaural reproduction through a pair of frontal loudspeakers. In contrast to headphone reproduction, two-front loudspeaker reproduction can recreate only stable perceived virtual sources in frontal-horizontal quadrants rather than in full three-dimensional directions.

The second problem with loudspeaker reproduction is a limited listening region or sweet points. For a given loudspeaker configuration, the cross-talk in Eq. (23) or transaural synthesis in Eq. (25) is designed according to a default (optimal) listening position. Head deviation from the default position (including translation and tuning) spoils the cross-talk cancellation, and thus alters the binaural pressures. Therefore, the performance of cross-talk cancellation is position-dependent. There have been a lot of works on the stability of loudspeaker-based binaural reproduction against head movement [54-57]. Kirkeby *et al*. proved that two frontal loudspeakers configuration with narrow span angle is beneficial to the stability of virtual source [54, 55]. Kirkeby further proposed using a pair of frontal loudspeakers with 10° span (in contrast to 60° span in conventional stereo) for binaural or transaural reproduction, which is known as "stereo dipole". A stereo dipole improves the stability of virtual source at mid-frequency at the cost of making low-frequency signal processing difficult because a large low-frequency boost is required.

The third problem with loudspeaker reproduction is timbre coloration. Ideally, a perfect cross-talk cancellation yields the same binaural pressures as those with a real source. Nevertheless, as stated above, it is difficult to cancel out cross-talk completely within a full audible frequency range. In practice, some reasons, such as slight movement, unmatched HRTFs, and room reflection etc., inevitably lead to incomplete cross-talk cancellation so that the binaural pressures at reproduction deviate from those of a real source. This in turn leads to perceived coloration, especially at high frequency and for an off-center listener. Therefore additional timbre equalization is required.

The principle of timbre equalization in two frontal loudspeakers reproduction can be explained as follows. Due to the difficulty in robust rendering the fine high-frequency spectral cues to listener's ears in loudspeaker reproduction, the perceived virtual source direction is dominated by the interaural cues (especially ITD) and limited to the frontal horizontal quadrant. While the interaural cues are controlled by the relative rather than the absolute magnitude and phase between left and right loudspeaker signals. Scaling both loudspeaker signals with identical frequency-dependent coefficient does not alter their relative magnitude and phase and thus the perceived virtual source direction. However, this manipulation alters the overall power spectra of the loudspeaker signals and thus impairs the timbre. Xie *et al*. proposed a constant-power equalization algorithm, in which the responses of transaural synthesis filters $G_L(\theta, f)$ and $G_R(\theta, f)$ in Eq. (24) were equalized by their root-mean-square [58]. As a result, the $G_L(\theta, f)$ and $G_R(\theta, f)$ in Eq. (25) are replaced by

$$\begin{aligned} G_L'(\theta, f) &= \frac{G_L(\theta, f)}{\sqrt{|G_L(\theta, f)|^2 + |G_R(\theta, f)|^2}} \\ &= \frac{H_\alpha H_L - H_\beta H_R}{\sqrt{|H_\alpha H_L - H_\beta H_R|^2 + |-H_\beta H_L + H_\alpha H_R|^2}} \frac{|H_\alpha{}^2 - H_\beta{}^2|}{H_\alpha{}^2 - H_\beta{}^2}; \\ G_R'(\theta, f) &= \frac{G_R(\theta, f)}{\sqrt{|G_L(\theta, f)|^2 + |G_R(\theta, f)|^2}} \\ &= \frac{-H_\beta H_L + H_\alpha H_R}{\sqrt{|H_\alpha H_L - H_\beta H_R|^2 + |-H_\beta H_L + H_\alpha H_R|^2}} \frac{|H_\alpha{}^2 - H_\beta{}^2|}{H_\alpha{}^2 - H_\beta{}^2}. \end{aligned} \tag{26}$$

With $G_L'(\theta, f)$ and $G_R'(\theta, f)$, it can be proved that the loudspeaker signals given by Eq. (24) satisfy following relationship of constant power spectra:

$$|L'|^2 + |R'|^2 = E_0^2 \tag{27}$$

Therefore, the overall power spectra of loudspeaker signals is equal to that of the input stimulus, so reproduction coloration reduces.

http://dx.doi.org/10.5772/56907

4.5. Simulation of reflections

Free-field virtual source synthesis and rendering are discussed above, in which room or environment reflections are ignored. However, reflections exist in most real rooms and are vital to spatial auditory perception. Therefore, a complete VAD should include reflection modeling, and thereby is called virtual auditory or acoustic environment (VAE). Incorporating reflections into VAE processing brings following advantages: (1) recreating the spatial auditory perception in a room or reflective environment; (2) eliminating or reducing the in-head localization in headphone presentation; (3) controlling perceived virtual source distance.

Usually, there are two basic methods for room or environment reflection rendering. The physics-based method simulates the physical propagation of sound from source to receiver inside a room, or equally, the binaural room impulse responses (BRIRs), and then synthesizes the binaural signals by convoluting the input stimulus with BRIRs. The perception-based method recreates desired auditory perception of reflections by some signal processing algorithms from perceptual rather than physical viewpoint.

A complete physical modeling of BRIRs consists of source modeling (such as source radiation pattern), transmission or room acoustics modeling (such as frequency-dependent surface reflection, scattering and absorption, air absorption, etc.), listener modeling (scattering and diffraction by human anatomical structures). The room acoustics modeling methods are divided into two categories according to physical principle, i.e., geometrical acoustics-based method and wave acoustics-based method. Geometrical acoustics neglects most wave nature of sound, yielding the approximate solutions of room acoustic field. This approximation is reasonable for high frequency and smooth boundary surface. The image-source method and ray-tracing method are two commonly used geometrical acoustics-based methods. The former decomposes the reflection sound field into the radiations of multiple image sources in free space. While the later treats sound radiation like a number of rays, which propagate and then are reflected and absorbed by boundary surface according to certain rule.

When the wave nature of sound is taken into account, wave acoustics-based methods should be used. These methods solve the wave equation for pressure inside the room and yield more accurate results. Various numerical methods, such as the finite element method, boundary element method, finite-difference time domain method and digital waveguide mesh method, have been suggested to solve the acoustic field in rooms with complex geometries. Limited to the extensive computational workload, however, these numerical methods are merely suitable for low-frequency and small room modeling.

Room acoustic field modeling yields time, direction, magnitude (or energy) as well as the spectra of each reflection arriving at a received point. Each reflection is filtered with a pair of corresponding HRTFs and the contribution of all reflections are combined to form complete BRIRs. In actual VAD or VAE, convolution of the input stimulus with HRIRs can be implemented by some decomposed structures.

The calculation for modeling and convoluting with a pair of complete BRIRs is complex. In some practical uses, the physics-based methods mentioned above are used to simulate and render the early room reflections in VAEs. To simplify processing, the late and diffuse room

reflections are often simulated by some perception-based methods, such as various artificial delay and reverberation algorithms [59]. These algorithms are based on the pre-measured or pre-calculated room acoustic attributes or parameters (such as reverberation time) and render the reflections from the perceptual rather than physical point of view.

4.6. Dynamic VAD

In static VAD or VAE discussed above, both virtual sources and listeners are assumed to be fixed and real-time processing is not always required. In a real acoustic environment, however, either source or listener movement alters the binaural pressures and brings dynamic acoustic information. This dynamic information should be incorporated into VAD or VAE processing, because it is significant for both source localization and recreating convincing auditory perceptions of acoustic environment. Therefore, in addition to modeling the sound source, room (environment) and listener, a sophisticated VAD should be able to constantly detect the position and orientation of listener's head, based on which the signal processing is updated in real-time. In other words, a faithful VAE should be an interactive, dynamic and real-time rendering system, and thus called dynamic and real-time VAD system.

Figure 16 shows the basic structure of a dynamic VAD system, which consists of three parts:

1. Information input and definition

This part inputs the prior information and data for dynamic VAD through a user interface. These information and data are classified into three categories: source information, environment information and listener information. The source information includes type of source stimuli, the number, spatial positions, orientation, directivities (radiation pattern) and level of sources, or predetermined trajectory for a moving source, etc. The environment information includes room or environment geometry, absorption coefficients of surface material and air, etc. The listener information includes the initial spatial position, orientation and individual data of listener (such as HRTFs). A head-tracking device detects the position and orientation of listener's head and then provides those information to the system.

2. Dynamic VAD signal processing

According to prior information and data in part 1, this part simulates sound source as well as both direct and reflected/scattered propagation from sound sources to two ears using certain physical algorithms. Based on the temporary position of the head detected by head-tracking devices, the HRTFs for binaural synthesis are constantly updated so as to obtain dynamic binaural signals.

3. Reproduction

The resultant binaural signals are reproduced through headphone after headphone equalization, or through loudspeakers after cross-talk cancellation.

Ideally, the binaural signals or auditory scenario created by a dynamic VAD should synchronously vary with head movement just as in the real environment. Therefore, an ideal dynamic VAD should be a linear time-variable system. However, the signal processing schemes in

http://dx.doi.org/10.5772/56907

dynamic VAD are deduced from the static scheme, in which a series of short "static state" are used for approximating the transient. Therefore, the dynamic behaviors of VAD should be considered.

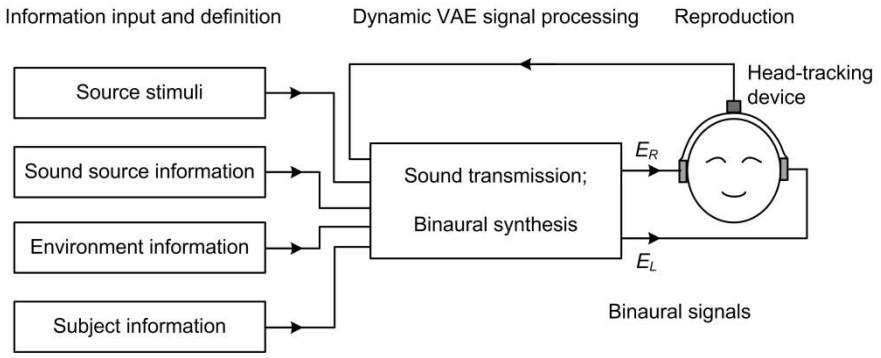

Figure 16. Structure of a typical dynamic VAD system

One problem concerned with the dynamic behaviors is the scenario update rate. A dynamic VAD updates the binaural signals and thereby auditory scenario at certain time interval. The scenario update rate of a VAD refers to the number of update scenario manipulations per second. Another problem concerned with dynamic behaviors is the system latency time. When the head moves, the synthesized binaural signals in existing VAD do not change synchronously but with a delay. The system latency time refers to the time from the listener's head movement to corresponding change in the synthesized binaural signals output, which is contributed by the hardware (such as head tracker) and software structures, as well as the data transmission and communication of the system. Usually, a high scenario update rate and low system latency time are preferred for improving the performance of a dynamic VAD system. Limited by the available system capacity, however, some tradeoffs in system performance should be made in practical dynamic VAD on basis of psychoacoustic rules. In addition, the auditory continuity of scenario update should also be taken in account.

Some psychoacoustic experiments have been carried out to investigate the required scenario update rate and system latency time in a dynamic VAD. The experiment conducted by Sandvad indicated that a scenario update rate of 10 Hz or less degraded the speed of localization; and a scenario update rate of 20 Hz almost would not degrade the speed of localization, although audible artifacts may occur for moderate to fast head movement [60]. While the results for system latency time vary among different studies [61-63]. A general accepted conclusion is that a system latency time lower than 60 ms is adequate for most applications [63].

Some dynamic VAD systems have been developed for various purposes and applications [64-71]. For example, the SLAB (Sound Laboratory) developed by NASA in U. S. was intended

to provide a platform for psychoacoustic study [64-66]. It is a software-based system written with VC++ language, and is implemented by a PC or server under Microsoft Windows operating system. Through application programming interface (API), the SLAB provides access to different psychoacoustic researches. In the SLAB, the dynamic virtual auditory environment caused by moving sources in real-time can be rendered, including the simulation of source radiation pattern, sound propagation, environment reflection and absorption, air absorption etc. Six of 1st-order reflections were modeled by image source method. The maximum achievable number of virtual sources depends on the available computational ability of system (typical 4 CPUs). The typical scenario update rate is 120 Hz. Excluding the external latency caused by head tracker, the internal system latency time is 24 ms. The binaural signals are reproduced through headphone. The SLAB system has been updated for several times. The latest version also supports using individualized HRTFs.

5. Applications of VADs

5.1. Psychoacoustic experiment and hearing training

Psychoacoustic acoustics investigates the relationship between acoustics-related physical factors and resulting subjective perceptions. By means of VADs, the complete and precise controlling over some physical characteristics of binaural signals is allowed, and corresponding subjective perceptions can be created. Therefore, VADs have become an important experimental tool for psychoacoustic acoustics, such as auditory localization mechanism [72] and masking [73]. VADs also benefit to hearing training for musicians and sound engineers.

5.2. Virtual reality and multimedia

Virtual reality is a kind of human-computer interface technology that provides users the feeling of being presence by including various perceptual cues such as visual, auditory, tactile sense [68]. The interaction and complementary of multiple information on above aspects strengthen the sense of reality and immerse. By means of VADs, various auditory perceptions to source localization or acoustical environments can be generated. Therefore, VADs are important to virtual reality in regard to auditory simulation. A typical example is driving training simulation [74], which can be realized by the dynamic VADs presented in Section 4.6. Similar methods can also be applied to some special environment trainings such as virtual aviation, aerospace, submarine environments.

VADs have been widely applied on the entertainment functions of multimedia PC. At present, game softwares under Windows platform possess the functions of VADs. In such kind of consumer electronics, simplified signal processing in VADs is needed in consideration of cost and computer capacity.

http://dx.doi.org/10.5772/56907

5.3. Speech communication

Psychoacoustic research indicates that the target detection ability of the binaural hearing is prior to the monoaural hearing in presence of background interference. In daily life, a listener can detect the target speech information even in a noisy environment with negative signal-to-noise ratio, suggesting the high speech intelligibility of the target. This is so-called the cocktail party effect [75]. However, mono signal transmission is dominant in currently available communication systems, resulting in low speech intelligibility. This condition can be improved by using binaural signal transmission, in which spatial separation between target and competing sources is realized through VADs. This method can be applied to teleconference and other speech communication systems. VADs are also helpful in aeronautical communication on aspects of improving speech intelligibility and reducing the react time of the pilot in the case of accident hazard [76].

5.4. Binaural auralization

On-site listening is the most straightforward way for subjective assessment of room acoustic quality. However, this is difficult in practical use. One reason is the impossibility of accurately compare among halls at different areas due to human short-term memory and expansive travelling cost. Moreover, it is difficult to orginize the same band to play the same music at different halls, which is needed in accurate subjective assessment.

As mentioned, BRIRs contain the main information of direct sounds and reflections. Binaural auralization is achieved by convoluting the mono "dry" signal with mathematically or physically-obtained BRIRs (see Section 4.5) and reproducing the synthesized binaural signals through headphone or louderspeakers with proper crosstalk cancelling. In past decades, binaural auralization has become an important tool in the research and design of room acousitc quality [77]. Especially, binaural auralization is helpful to detect acoustic defects in regard to subjective properties on the stage of room design. This function has been included in some softwares for room acoustic design such as Odeon. Besides, binaural auralization has been generally used in subjective assessment such as noise evaluation [78], subjective assessment of sound reproduction systems [79], and virtual sound recording.

5.5. Virtual reproduction for multi-channel surround sound

Multi-channel surround sound reproduction, such as 5.1 channel surround sound, requires multiple loudspeakers, which is complex and inconvenient in some practical applications such as TV or multimedia computer. Hence, some HRTF-based virtual loudspeaker-based approaches (i.e., HRTF-based binaural synthesis in Section 4.4) for multi-channel surround sound reproduction have been introduced to reduce the number of loudspeaker needed. For example, some commercial products have been introduced for the virtual reproduction of 5.1 channel surround sound, that is, simulating 5.1 channel surround sound through a pair of actual stereophonic loudspeakers, see Figure 17. Signals *L* and *R* are directly fed to left and right loudspeakers respectively so as to create summing virtual source within the span of two loudspeakers. Signal *C* is attenuated 3 dB and then fed to the left and right loudspeakers to

create a summing virtual source at the front $\theta = 0°$. Two surround signals *LS* and *RS* are filtered by transaural synthesis filters and then fed to the loudspeakers.

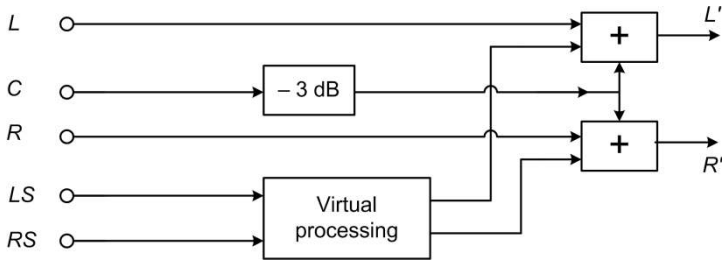

Figure 17. The block diagram of virtual 5.1 channel surround sound through loudspeakers

HRTF-based binaural synthesis can also be applied into multi-channel surround sound virtual reproduction through headphone. When directly rendering multi-channel surround sound signals to headphone, unnatural perceptions such as in-head localization occur. Using HRTF-based binaural synthesis, multiple loudspeakers can be virtually generated by headphone. Correspondingly, subjective perception of headphone-based multi-channel surround sound reproduction can be improved. Some related products have been introduced, such as Dolby headphone (http://www.dolby.com).

In virtual reproduction for multi-channel surround sound, defects presented in Sections 4.3 and 4.4 for VADs also exist, such as timbre coloration, limited listening area, and directional distortions. Our group has proposed some patents on the improvement of those defects.

6. Summary

HRTFs capture most localization information in binaural pressures and exhibit important physical and auditory characteristics. One major application of HRTFs is VAD or VAE, in which HRTF-based signal processing is used to recreate virtual source and other spatial auditory events in headphone presentation or loudspeaker presentation with proper equalization. Great developments have been achieved in the field of HRTFs and VADs, but many issues need further research. VADs have currently been applied to various fields in scientific research, engineering, entertainment and consumer electronic products, among others.

Acknowledgements

This work is supported by the National Nature Science Foundation of China (Nos. 11174087, 11004064), and State Key Lab of Subtropical Building Science, South China University of Technology.

http://dx.doi.org/10.5772/56907

Author details

Xiao-li Zhong and Bo-sun Xie

Acoustic Lab, Physics Dept., School of Science, South China University of Technology, Guangzhou, China

References

[1] Blauert J. Spatial Hearing (Revised edition), MIT Press, Cambridge, MA, England, 1997

[2] Møller H. Fundamentals of binaural technology. Applied Acoustics, 1992; 36(3/4), 171-218

[3] Xie B S. Head Related Transfer Function and Virtual Auditory Display. USA: J.Ross Publishing, 2013

[4] Xiang N, Schroeder M R. Reciprocal maximum-length sequence pairs for acoustical dual source measurements. J. Acoust. Soc. Am., 2003; 113 (5), 2754-2761

[5] Yu G Z, Liu Y, Xie B S. Fast measurement system and super high directional resolution head-related transfer function database. J. Acoust. Soc. Am.,2012; 131(4), 3304

[6] Wightman F L, Kistler D J. Measurement and validation of human HRTFs for use in hearing research. Acta Acoustica United with Acoustica, 2005; 91 (3), 429–439

[7] Blauert J, Brueggen M, Bronkhorst A W, et al. The AUDIS catalog of human HRTFs. J. Acoust. Soc. Am., 1998; 103 (5), 3082

[8] Gardner W G, Martin K D. HRTF measurements of a KEMAR. J. Acoust. Soc. Am., 1995; 97 (6), 3907-3908

[9] Møller H, Sørensen M F, Hammershøi D, et al. Head-related transfer functions of human subjects. J.Audio Eng. Soc., 1995; 43(5), 300-321

[10] Algazi V R, Duda R O, Thompson D M, et al. The CIPIC HRTF database, Proceeding of 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics, 99 -102, 2001

[11] IRCAM Lab (2003): Listen HRTF database, http://recherche.ircam.fr/equipes/salles/listen/ (accessed 25 Aug. 2013)

[12] Genuit K, Xiang N. Measurements of artificial head transfer functions for auralization and virtual auditory environment. Proceeding of 15th International Congress on Acoustics (invited paper), Trondheim, Norway, II 469-472 ; 1995

[13] Xie B S, Zhong X L, Rao D, et al. Head-related transfer function database and analyses. Science in China Series G, Physics, Mechanics & Astronomy. 2007; 50(3), 267-280

[14] Brungart D S, Rabinowitz W M. Auditory localization of nearby sources. Head-related transfer functions. J. Acoust. Soc. Am., 1999; 106 (3), 1465-1479

[15] Hosoe S, Nishino T, Itou K, et al. Measurement of Head-related transfer function in the proximal region. Proceeding of Forum Acusticum 2005, Budapest, Hungary, 2539-2542,2005

[16] Gong M, Xiao Z, Qu T S, et al. Measurement and analysis of near-field head-related transfer function, Applied Acoustics (in Chinese). 2007; 26, 326—334

[17] Yu G Z, Xie B S, Rao D. Characteristics of Near-field head-related transfer function for KEMAR. AES 40th Conference. Japan, Tokyo; 2010

[18] Morse P M, Ingrad K U. Theoretical Acoustics, McGraw-Hill, New York, USA, 1968

[19] Duda R O, Martens W L. Range dependence of the response of a spherical head model. J.Acous.Soc.Am.,1998; 104(5), 3048-3058

[20] Algazi V R, Duda R O, Duraiswami R, et al. Approximating the head-related transfer function using simple geometric models of the head and torso. J.Acoust.Soc.Am., 2002; 112 (5), 2053-2064

[21] Gumerov N A, Duraiswami R. Computation of scattering from N spheres using multipole reexpansion. J. Acoust. Soc. Am., 2002; 112 (6), 2688-2701

[22] Kahana Y, Nelson P A. Boundary element simulations of the transfer function of human heads and baffled pinnae using accurate geometric models. J. Sound and Vibration, 2007; 300(3/5), 552-579

[23] Katz B F G. Boundary element method calculation of individual head-related transfer function.I. Rigid model calculation. J.Acoust.Soc.Am., 2001; 110(5), 2440-2448

[24] Otani M, Ise S. Fast calculation system specialized for head-related transfer function based on boundary element method. J.Acoust.Soc.Am.,2006; 119(5), 2589-2598

[25] Gumerov N A, O'Donovan A E, Duraiswami R, et al. Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation. J. Acoust. Soc. Am., 2010; 127(1), 370-386

[26] Zotkin D N, Hwang J, Duraiswami R, et al. HRTF personalization using anthropometric measurements, Proceedings of the 2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 157-160; 2003

[27] Middlebrooks J C. Individual differences in external-ear transfer functions reduced by scaling in frequency, J. Acoust. Soc. Am., 1999; 106 (3), 1480-1492

[28] Nishino T, Inoue N, Takeda K, et al. Estimation of HRTFs on the horizontal plane using physical features, Applied Acoustics, 2007; 68(8), 897-908

[29] Seeber B, Fastl H. Subjective selection of non-individual head-related transfer functions, Proc. ICAD 2003, 259–262; 2003

[30] Yairi S, Iwaya Y, Suzuki Y. Individualization feature of head-related transfer functions based on subjective evaluation, Proceedings of the 14 International Conference on Auditory Display, Paris, France June 24 – 27; 2008

[31] Zhong X L, Xie B S. Spatial symmetry of head-related transfer function, Chinese Journal of Acoustics, 2007; 26, 73—84

[32] Kulkarni A， Isabelle S K, Colburn H. S. Sensitivity of human subjects to head-related transfer-function phase spectra. J. Acoust. Soc. Am., 1999; 105,2821—2840

[33] Minnaar P, Christensen F, and Møller H, et al. Audibility of all-pass components in binaural synthesis. AES 106th Convention, Munich, Germany, Preprint: 4911, 1999.

[34] Plogsties J, Minnaar P, and Olesen S, et al. Audibility of all-pass components in head-related transfer functions. AES 108th Convention, Paris, France, Preprint: 5132, 2000

[35] Kistler D J, Wightman F L. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. J. Acoust. Soc. Am., 1992; 91,1637-1647

[36] Zhong X L, Xie B S. Maximal azimuthal resolution needed in measurements of head-related transfer functions. J. Acoust. Soc. Am.,2009; 125,2209—2220

[37] Zhang W, Zhang M Q, Kennedy R A, et al. On high-resolution head-related transfer function measurements: an efficient sampling scheme. IEEE Transactions on Audio, Speech, and Language Processing, 2012; 20(2):575—584

[38] Xie B S. Recovery of individual head-related transfer functions from a small set of measurements, J. Acoust. Soc. Am., 2012; 132(1): 282—294

[39] Brungart D.S., Durlach N.I., Rabinowitz W.M., Auditory localization of nearby sources. II. Localization of a broadband source, J. Acoust. Soc. Am., 1999; 106 (4), 1956–1968.

[40] Yu G Z, Xie B S, Rao D. Near-field head-related transfer functions of a artificial head and its characteristics. Acta Acousticca (in Chinese), 2012;37(4), 378-385

[41] Møller H, Jensen C B, Hammershøi D, et al. Using a typical human subject for binaural recording, AES 100th Convention, Copenhagen, Denmark, Preprint: 4157; 1996

[42] Wenzel E M, Arruda M, Kistler D J, et al. Localization using nonindividualized head-related transfer functions. J. Acoust. Soc. Am., 1993; 94,111-123

[43] Mackenzie J, Huopaniemi J, Valimaki V, et al. Low-order modeling of head-related transfer functions using balanced model truncation. IEEE Signal Processing Letter, 1997; 4 (2), 39-41

[44] Blommer M A, Wakefield G H. Pole-zero approximations for head-related transfer functions using a logarithmic error criterion. IEEE Trans. on Speech and audio processing, 1997; 5 (3), 278-287

[45] Haneda Y, Makino S, Kaneda Y, et al. Common-acoustical-pole and zero modeling of head-related transfer functions. IEEE Trans. on Speech and Audio Processing, 1999; 7 (2), 188-196

[46] Harma A, Karjalainen M, Savioja L, et al. Frequency-warped signal processing for audio applications. J. Audio Eng. Soc.,2000; 48 (11), 1011-1031

[47] Huopaniemi J, Zacharov N. Objective and subjective evaluation of head-related transfer function filter design. J. Audio. Eng. Soc., 1999; 47(4), 218-239

[48] Jot. J M, Walsh M, Philp A. Binaural Simulation of Complex Acoustic Scenes for Interactive Audio, AES 121st Convention, San Francisco, U.S.A., Preprint: 6950; 2006

[49] Møller H, Hammershøi D, Jensen C B, et al. Transfer characteristics of headphones measured on human ears. J. Audio. Eng. Soc., 1995; 43(4), 203-217

[50] Wightman F L, Kistler D J. Headphone simulation of free-field listening, I: stimulus synthesis. J. Acoust. Soc. Am., 1989; 85 (2), 858-867

[51] Schroeder M R, Atal B S. Computer simulation of sound transmission in rooms, Proceeding of IEEE, 1963; 51(3), 536-537

[52] Bauck J, Cooper D H. Generalization transaural stereo and applications. J.Audio.Eng.Soc.,1996; 44(9), 683-705

[53] Bai M R, Tung C W, Lee C C. Optimal design of loudspeaker arrays for robust cross-talk cancellation using the Taguchi method and the genetic algorithm. J.Acoust.Soc.Am,2005; 117(5), 2802-1813

[54] Kirkeby O, Nelson P A, Hamada H. The "Stereo Dipole"—a virtual source imaging system using two closely spaced loudspeakers, J. Audio Eng. Soc., 1998; 46(5), 387-395

[55] Kirkeby O, Nelson P A, Hamada H. Local sound field reproduction using two closely spaced loudspeakers. J. Acoust. Soc. Am., 1998; 104(4), 1973–1981

[56] Takeuchi T, Nelson P A, Hamada H. Robustness to head misalignment of virtual sound image system. J.Acoust.Soc.Am.,2001; 109(3), 958-971

[57] Ward D B, Elko G W. Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation. IEEE Signal Processing Letters, 1999; 6(5), 106-108

[58] Xie B S, Shi Y, Xie Z W, et al. Virtual reproduction system for 5.1 channel surround sound. Chinese Journal of Acoustics, 2005; 24, 76—88

[59] Gardner W G. Reverberation algorithms, In Applications of Signal Processing to Audio and Acoustics (Edited by Kahrs M, Brandenburg K.), Kluwer Academic Publishers, USA, 1998

[60] Sandvad J. Dynamic aspects of Auditory virtual environments, AES 100th Convention, Copenhaqen, Denmark, Preprint 4226; 1996

[61] Bronkhorst A W. Localization of real and virtual sound sources. J. Acoust. Soc. Am., 1995; 98 (5), 2542-2553

[62] Wenzel E M. Effect of increasing system latency on localization of virtual sounds, AES 16th International Conference: Spatial Sound Reproduction, Rovaniemi, Finland; 1999

[63] Brungart D S, Kordik A J, Simpson B D. Effects of headtracker latency in virtual audio displays. J. Audio Eng. Soc., 2006; 54 (1/2), 32-44

[64] Wenzel E M, Miller D J, Abel J S. Sound Lab: a real-time, software-based system for the Study of Spatial hearing, AES 108 th Convention, Paris, France, Preprint: 5140, 2000

[65] Miller J D, Wenzel E M. Recent developments in SLAB: A software-based system for interactive spatial sound synthesis, Proceedings of the 2002 International Conference on Auditory Display, Kyoto, Japan; 2002

[66] Begault D R, Wenzel E M, Godfroy M, et al., Applying spatial audio to human interfaces: 25 years of NASA experience. AES 40th Conference, Tokyo, Japan; 2010

[67] Saviojia L, Lokki T, Huopaniemi J. Auralization applying the parametric room acoustic modeling technique-The DIVA Auralization system, Proceedings of the 2002 International Conference on Auditory Display, Kyoto, Japan; 2002

[68] Blauert J, Lehnert H, Sahrhage J, et al. An interactive virtual-environment generator for psychoacoustic research I: architecture and implementation. Acta Acustica united with Acustica, 2000; 86 (1), 94–102

[69] Silzle A, Novo P, Strauss H. IKA-SIM: A system to generate auditory virtual environments, AES 116th Convention, Berlin, Germany, Preprint: 6016; 2004

[70] Lentz T, Assenmacher I, Vorländer M, et al. Precise near-to-head acoustics with binaural synthesis, Journal of Virtual Reality and Broadcasting, 2006; 3(2)

[71] Zhang C Y, Xie B S. Platform for virtual auditory environment real time rendering system, ACOUSTICS 2012 HONG KONG Conference and Exhibition; 2012

[72] Langendijk E H A, Bronkhorst A W. Contribution of spectral cues to human sound localization. J. Acoust. Soc. Am., 2002; 112 (4), 1583-1596

[73] Kopco N, Shinn-Cunningham B G. Spatial unmasking of nearby pure-tone targets in a simulated anechoic environment, J. Acoust. Soc. Am., 2003; 114 (5), 2856-2870

[74] Krebber W, Gierlich H W, Genuit K. Auditory virtual environments: basics and applications for interactive simulations, Signal Processing, 2000; 80 (11), 2307-2322

[75] Bronkhorst A W, The cocktail party phenomenon: a review of research on speech intelligibility in multiple-talker conditions, Acta Acustica united with Acustica,2000; 86(1), 117-128.

[76] Begault, D R. Virtual acoustics, aeronautics, and communications. J. Audio Eng. Soc., 1998; 46(6), 520-53

[77] Kleiner M, Dalenback B I, Svensson P. Auralization-an overview. J. Audio Eng. Soc., 1993; 41(11), 861-875

[78] Song W, Ellermeier W, Hald J. Using beamforming and binaural synthesis for the psychoacoustical evaluation of target sources in noise. J. Acoust. Soc. Am., 2008; 123(2),910-924

[79] Toole F E. Binaural record/reproduction systems and their use in psychoacoustic investigation, AES 91st Convention, New York, USA, Preprint:3179; 1991