Spatial Characteristics of Head-Related Transfer Function *

ZHONG Xiao-Li(钟小丽), XIE Bo-Sun(谢菠荪)**

Laboratory for Acoustics, School of Physics, South China University of Technology, Guangzhou 510641

(Received 19 January 2005)

The spatial characteristics of head-related transfer functions (HRTFs) are studied by a spatial Fourier analysis. A law of the HRTF spatial sampling in different elevation planes is obtained, and the corresponding spatial interpolating method used to recover the continuous HRTF is proposed. The method is valid, for the average error between the measured and interpolated HRTFs in the horizontal plane is about 0.5%. These results provide guidance for the discrete spatial measurement of HRTFs.

PACS: 43.60.+d, 43.66.+y

Head-related transfer functions (HRTFs), which are the acoustical transfer functions from sound sources to human eardrums, contain most localization cues. Therefore, they are important for understanding binaural hearing and psychoacoustics, and are widely used in virtual reality. HRTFs can be obtained by calculation.[1−3] However, limited by the physical models or computer capability, the results are unable to demonstrate the important characteristics of HRTFs above the frequency of 6.0 kHz. An HRTF of full frequency range can also be obtained by measurement.[4,5] However, because the measurement can only be carried out in finite discrete directions and because the HRTFs are continuous functions depending on spatial directions, the problems of how to sample in the space and how to recover the unmeasured HRTF by spatial interpolation arise. Evans[6] developed a continuous functional representation of discrete measured HRTFs by using surface spherical harmonics as basis functions; however, how to sample the continuous HRTF in space was not discussed. In Ref. [7], though a formula of calculating the measuring resolution in different elevation planes was given, it was derived from experience but not from theoretical analysis. Recently, some spatial interpolating methods,[8−10] including bilinear interpolation and spline interpolation, have been used to recover the unmeasured HRTF, but their validity has not yet been related to the sampling conditions, though in fact the recovering and the sampling are closely correlative. Here the problem of the HRTF spatial sampling by Fourier analysis in the spatial domain is investigated, and the corresponding spatial interpolating formula is given.

In the far field, the HRTF is a continuous function of (θ, ϕ, f), where $0^\circ \le \theta < 360^\circ$ and $-90^\circ \le \phi \le 90^\circ$ are the azimuth and elevation of the sound source respectively; $\phi = 0^\circ$ is the horizontal plane, $\phi = 90^\circ$ is the upside. In the horizontal plane, $\theta = 0^\circ$ and 90° are front and right respectively, f is the frequency. For a given ϕ_0, the HRTF is a continuous function of θ with a period of 2π. Therefore, it can be expanded as a Fourier series along θ, as expressed by the following equations. Provided that the HRTF is left–right symmetric, the following analysis is suitable for both ears, so the subscript denoting left or right ear can be omitted:

$$H(\theta, \phi_0, f) = \sum_{q=-\infty}^{+\infty} d_q(f) e^{iq\theta},$$
$$d_q(f) = \frac{1}{2\pi} \int_0^{2\pi} H(\theta, \phi_0, f) e^{-iq\theta} d\theta. \quad (1)$$

If an integer Q exists, for $|q| > Q$, we always have $d_q(f) = 0$, then Eq. (1) can be rewritten as

$$H(\theta, \phi_0, f) = \sum_{q=-Q}^{+Q} d_q(f) e^{iq\theta}. \quad (2)$$

Then H consists of $(2Q+1)$ weighted angle harmonics, and completely decided by $(2Q+1)$ weights $d_q(f)(|q| \le Q)$, which depend on frequency. Along $0 \le \theta < 2\pi$, sampling or measuring $H(\theta, \phi_0, f)$ in M azimuthal directions in equal increments, we obtain

$$H(\theta_m, \phi_0, f) = \sum_{q=-Q}^{+Q} d_q(f) e^{iq\theta_m},$$
$$\theta_m = \frac{2\pi m}{M}, \quad m = 0, 1, \cdots, (M-1). \quad (3)$$

If $M \ge (2Q+1)$, using the orthogonality and integrity of $\{\exp(iq\theta_m)\}$, we can obtain

$$d_q(f) = \frac{1}{M} \sum_{m=0}^{M-1} H(\theta_m, \phi_0, f) e^{-iq\theta_m}, \ |q| \le Q,$$
$$Q < |q| \le (M-1)/2, \ M \text{ is odd};$$
$$d_q(f) = 0, \quad -(M/2) \le q < -Q, \text{ and},$$
$$Q < q \le (M/2) - 1, \ M \text{ is even}. \quad (4)$$

* Supported by the National Natural Science Foundation of China under Grant No 10374031, and the Natural Science Foundation of the South China University of Technology under Grant No 123-E5040600.

** Email: phbsxie@scut.edu.cn

©2005 Chinese Physical Society and IOP Publishing Ltd

Substituting Eq. (4) into Eq. (3) yields

$$H(\theta,\phi_0,f)=\frac{1}{M}\sum_{m=0}^{M-1}H(\theta_m,\phi_0,f)\cdot\frac{\sin\left[\left(Q+\frac{1}{2}\right)(\theta-\theta_m)\right]}{\sin\left(\frac{\theta-\theta_m}{2}\right)}. \quad (5)$$

Therefore, by using Eq. (5), the continuous $H(\theta,\phi_0,f)$ can be recovered from M discrete azimuthal samplings $H(\theta_m,\phi_0,f)$ if $M\geq(2Q+1)$. This is the law of the HRTF spatial sampling, and it presents a new interpolating method related to the sampling conditions for recovering the continuous HRTF in different elevation planes.

The above proposition of spatial sampling and recovering is based on Eq. (2). In Eq. (2), the infinite angle harmonics can be truncated at Q degree. Hence, Q is the highest degree harmonic that should be taken into account. Theoretically, in order to determine Q, the continuous $H(\theta,\phi_0,f)$ is needed as shown in Eq. (1). However, the continuous $H(\theta,\phi_0,f)$ of the full frequency range is not achievable, because only $H(\theta,\phi_0,f)$ under 6.0 kHz is available by calculation and only the discrete $H(\theta_m,\phi_0,f)$ is available by measurement. Therefore, as an approximation, if a database contains enough azimuthal samplings satisfying $M\geq(2Q+1)$, then

$$\sum_{q=-\infty}^{+\infty}d_q(f)e^{iq\theta}=\sum_{q=-Q_1}^{Q_2}d_q(f)e^{iq\theta},\ (Q_1+Q_2+1)=M. \quad (6)$$

The database used is measured on a KEMAR dummy head by MIT Media Lab.[4] It contains HRTFs of far field in 710 different spatial directions. For each direction, a 512-point HRIR (HRTF in the time domain, and can be transformed into frequency domain by fast Fourier transform) at a 44.1 kHz sampling rate is given. Details are listed in Table 1. Because the samplings at $\phi=\pm40^\circ$ are not uniform, they are omitted here.

Table 1. Numbers of the azimuthal sampling in different elevation planes.

Elevation	-30°	-20°	-10°	0°	10°	20°	30°	50°	60°	70°	80°	90°
Numbers of Azimuthal Sampling	60	72	72	72	72	72	60	45	36	24	12	1

To show the relative contribution of each spatial angle harmonic, $\eta_s(f)$ is defined as

$$\eta_s(f)=\frac{|d_s(f)|^2+|d_{-s}(f)|^2}{\sum_{q=-Q_1}^{+Q_2}|d_q(f)|^2},$$
$$s=1,2,3,\cdots\min(Q_1,Q_2),$$
$$Q_1=Q_2=(M-1)/2,\ M \text{ is odd},$$
$$Q_1=M/2, Q_2=(M/2)-1, M \text{ is even}. \quad (7)$$

For instance, $\eta_s(f)$ in the horizontal plane is illustrated in Fig. 1. In Fig. 1, $\eta_s(f)$ of low degree harmonics are distinct in all the frequency range, though it seems to decrease with the increasing frequency. However, $\eta_s(f)$ of high degree harmonics are nearly zero until at high frequency. For instance, $\eta_s=20$ ($f=172$ Hz) is 3.2×10^{-5}, while $\eta_{s=20}$ ($f=15.4$ kHz) is 0.12. There are some physical facts behind those phenomena. The angle harmonics reveal the distribution of the HRTF caused by wave scattering around head, torso, and pinna. The low degree harmonics mainly caused by the nearly regular head shape reveal a smooth change of the HRTF. In contrast, the high degree harmonics originating from the irregularity of head and pinna demonstrate the fast change of HRTF with azimuth. It is known that the sphere approximation is a good one of head under 3 kHz. Therefore at low frequency, low degree harmonics are so significant that their combination is enough to represent the HRTF. With the increasing frequency, the irregularities of head and pinna begin to play roles, therefore the relative contributions of higher harmonics become significant. Hence, analysing the angle harmonics can provide a new means of detecting the scattering of head and pinna. The results at other elevations are similar.

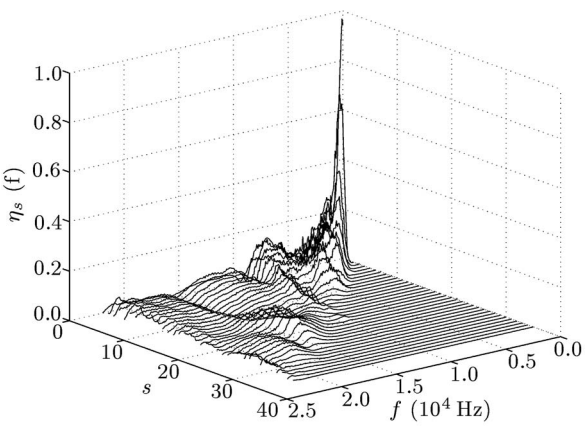

Fig. 1. Contributions of each spatial angle harmonics at $\phi_0=0^\circ$.

The relative energy ratio η of preceding S degree harmonics is defined by

$$\eta=\frac{\sum_{q=-S}^{+S}|d_q(f)|^2}{\sum_{q=-Q_1}^{+Q_2}|d_q(f)|^2}, \quad (8)$$

where Q_1 and Q_2 are the same as those in Eq. (7), determined by the actual azimuthal sampling number in the database.

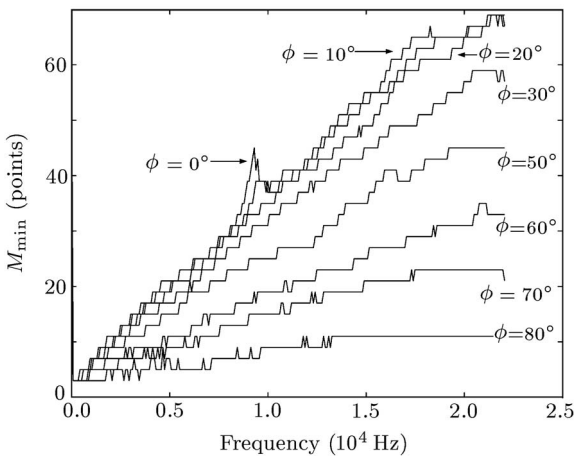

Fig. 2. $M_{\min}$ calculated for each elevation planes.

The steps of determining the highest degree Q of harmonics and the minimal number of azimuthal sampling $M_{\min} = (2Q + 1)$ for each elevation are shown as follows. Firstly, calculate Eq. (8) at $S = 1, 2, 3 \cdots$, $\min(Q_1, Q_2)$ for each discrete frequency band f_n until $S = S_1$ which makes $\eta \geq 0.99$. This means that the ratio of the energy contributed by the preceding S_1 harmonics to the energy contributed by all the harmonics is larger than 0.99, which is set according to error. Secondly, analyse two different conditions about the value of S_1. If $S_1 < \min(Q_1, Q_2)$, then the sampling number analysed is of redundancy. This is because S_1 can be regarded as the highest degree Q of harmonic under $\eta \geq 0.99$ under such a condition. Thus, according to the proposition of spatial sampling and recovering, $M_{\min} = (2S_1 + 1)$ is enough to recover the continuous HRTF, while the actual number of sampling points $M = (Q_1 + Q_2 + 1)$ is larger than $M_{\min}$. If $S_1 = \min(Q_1, Q_2)$, then the sampling number analysed is considered as insufficient conservatively, for it is difficult to judge whether the sampling number is just enough or not.

Figure 2 shows the calculated $M_{\min}$ for each elevation plane and frequency based on the MIT database. It is obvious that $M_{\min}$ tends to increase with the increasing frequency in a given elevation plane, and for a given frequency the higher the elevation is, the less the calculated $M_{\min}$ is. Within $f \leq 20\,\text{kHz}$, a comparison of $M_{\min}$ with M is shown in Table 2. The numbers of the azimuthal sampling in the MIT database are enough, except for $\phi = 50°, 70°, 80°$.

Table 2. Calculated $M_{\min}$ and M used in the MIT measurements.

ϕ	$-30°$	$-20°$	$-10°$	$0°$	$10°$	$20°$	$30°$	$50°$	$60°$	$70°$	$80°$
M of MIT	60	72	72	72	72	72	60	45	36	24	12
$M_{\min}$ calculated	57	63	63	65	65	65	57	45	31	23	11

According to Table 2, the relationship between $M_{\min}(\phi)$ at elevation ϕ and $M_{\min}(0)$ at elevation $0°$ is evaluated approximately (with a correlation coefficient as high as 0.996):

$$M_{\min}(\phi) = M_{\min}(0)\cos\phi. \tag{9}$$

A similar relationship is found by half-empirical experiment in Ref. [7]. Actually, a similar relationship can be found in the perimeter $L(\phi)$ for each elevation plane of a sphere and that for the horizontal plane:

$$L(\phi) = L(0)\cos\phi. \tag{10}$$

Such a result implies that the spatial sampling ratios relative to the perimeter of each elevation plane are similar. Then, we can infer that the variations of the HRTF in different elevation planes are very similar.

As expressed in Eq. (5), if $M \geq M_{\min}$, the continuous HRTF can be recovered by M discrete azimuthal samplings. To test this proposition, we recover the continuous HRTF of the horizontal plane for example. From $0°$ to $350°$ in $10°$ increments, 36 azimuthal samplings in total are substituted into Eq. (5) to recover the continuous HRTF. The remainders, whose azimuths are from $5°$ to $355°$ in $10°$ increments, are used as references for evaluation. In the horizontal plane, the azimuthal sampling ratio of 36 points is only sufficient for frequency below 10 kHz as shown in Fig. 2. Hence, the comparison of measured and interpolated HRTFs, which are extracted from the recovered HRTF, should be carried out in such a frequency range. The comparison in the time domain at $\theta = 215°$ for the left ear is shown in Fig. 3. We can see that Eq. (5) is valid, for the two overlaid are very similar with the correlation coefficient as high as 0.917 (after 10 kHz low-pass filter, it is 0.999) and no relative time delay.

The signal-to-deviation ratio (SDR or R_{SDR}) is defined to evaluate the results quantitatively:

$$R_{\text{SDR}}(\theta, \phi_0) = 10\log_{10}\Big(\frac{\sum_f |H(\theta, \phi_0, f)|^2}{\sum_f |H(\theta, \phi_0, f) - \hat{H}(\theta, \phi_0, f)|^2}\Big)(\text{dB}), \tag{11}$$

where $\hat{H}(\theta, \phi_0, f)$ is the interpolated HRTF, $H(\theta, \phi_0, f)$ is the measured HRTF. The summations are taken over all the frequencies evaluated. Figure 4 shows the distribution of SDR along the azimuth in the horizontal plane for the left ear.

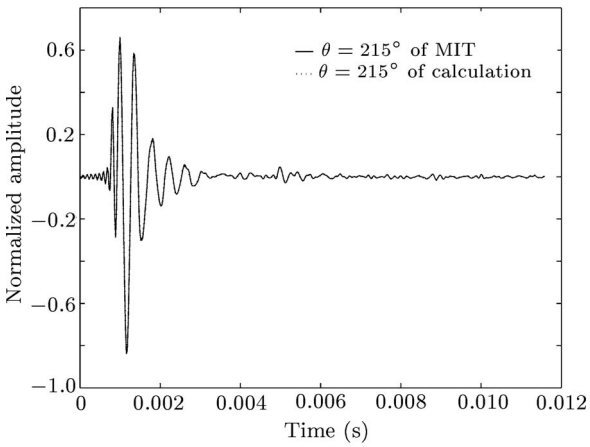

Fig. 3. Comparison of measured HRTF and interpolated HRTF at $\theta = 215°$.

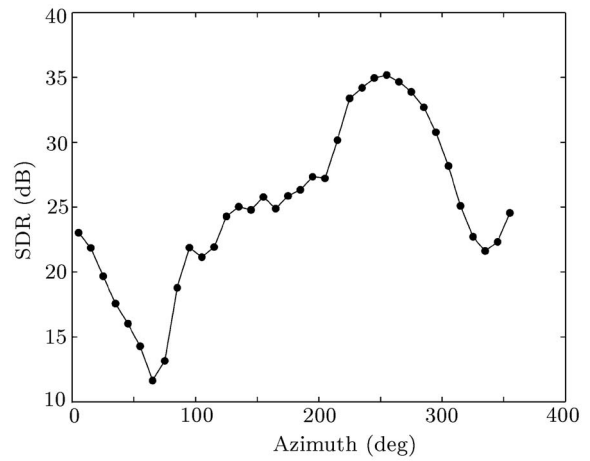

Fig. 4. The signal-to-deviation ratio (SDR or R_{SDR}) in the horizontal plane.

Generally, the SDR is higher when the source is ipsilateral, i.e. $\theta = 180°$–$360°$ for the left ear. In contrast, when the source is contralateral to the left ear, the SDR is relatively lower. The reason is that the original measured HRTF at the contralateral side has a low signal–noise ratio because of the head shadow effect. However, the average SDR is around 25 dB, which means that the error caused by the new spatial interpolation of Eq. (5) is just about 0.5%. Thus, the proposition of the HRTF spatial sampling and recovering is proven to be correct.

In summary, we have applied spatial Fourier analysis in studying HRTFs in each elevation planes. The distribution of the HRTF along the azimuth is determined by the spatial angle harmonics. At a given elevation, only if the actual uniform sampling number M is not less than the minimal sampling number $M_{\mathrm{min}} = (2Q + 1)$, where Q is denoted as the highest spatial angle harmonic, can the continuous HRTF of azimuth be recovered from M discrete azimuthal samplings of the HRTF.

References

[1] Gumerov N A and Duraiswami R 2002 *J. Acoust. Soc. Am.* **112** 2688
[2] Tian X and Liu Q H 2003 *J. Acoust. Soc. Am.* **113** 2434
[3] Katz B F G 2001 *J. Acoust. Soc. Am.* **110** 2449
[4] Gardner W G and Martin K D 1995 *J. Acoust. Soc. Am.* **97** 3907
[5] Algazi V R and Duda R O 2001 *IEEE Proc. Workshop on Applications of Signal Processing to Audio and Acoustics* (New York: Mohonk House) p 99
[6] Evans M J 1998 *J. Acoust. Soc. Am.* **104** p 2400
[7] Bovbjerg B P et al 2000 *109th AES Convention Preprint* p 5264
[8] Freeland F P et al 2002 *22nd AES International Conference on Virtual Synthetic and Entertainment Audio* (Espoo, Finland) p 000232
[9] Nishino T et al 2000 *Proc. 7th Western Pacific Regional Acoustics Conference* (Kumamoto, Japan) p 086
[10] Biscainho L W P 2002 *IEEE Proc. International Conference on Acoustics, Speech, and Signal Processing* (Orlando, FL) vol 2 p 1961