ICSV 21
13-17 July, 2014, Beijing/China

The 21st International Congress on Sound and Vibration

13-17 July, 2014, Beijing/China

EFFECT OF SPATIAL SAMPLING APPROACHES ON VIRTUAL HIGH ORDER AMBISONICS

Guangzheng Yu

Physics Department, South China University of Technology, Guangzhou, China, 510641

Joshua D. Reiss

School of Electronic Engineering and Computer Science, Queen Mary University of London, UK, E1 4NS

e-mail: scgzyu@scut.edu.cn

Ambisonics is a sound spatialisation technology that aims at physically reconstructing the sound field. When the ambisonics signals are not reproduced by loudspeakers, but instead synthesized to binaural signals suitable for headphone reproduction, the concept of virtual ambisonics or binaural synthesis is introduced. Virtual high order ambisonics (HOA) may be applied since there is no difficulty in arranging a large number of loudspeakers. In order to use head-related transfer functions (HRTFs) in binaural synthesis, the sampling of virtual loudspeaker of HOA should be set up to match the sampling distribution of HRTFs. However, the existing sampling approaches of measured HRTFs may be inefficient for HOA. In this work, based on the calculated HRTFs of a scanned KEMAR artificial head, the effects of sampling approaches of HRTFs on binaural synthesis are analysed. Results show that the nearly uniform sampling is more efficient in virtual HOA and binaural synthesis, but the HRTFs should be obtained first via spatial interpolation or numerical simulation.

1. Introduction

The human hearing system is very effective at acquiring auditory experiences from a spatial sound field.[1, 2] The desired sound field in a space can be either artificial synthesized or recorded using various microphone arrays, and then reproduced via an array of loudspeakers. Binaural synthesis is the method of converting multichannel spatial sound signals into two signals through convolution with head-related transfer functions (HRTFs), so that the complicated rendering system assembled by many loudspeakers can be simplified into reproduction based on only a pair of headphones.[3, 4]

A sound field can be accurately reconstructed by ambisonics based on spherical harmonics functions expansion [5] and by wavefield synthesis (WFS) based on the Kirchhoff–Helmholtz (KH) integral formula.[6] In binaural synthesis, the used HRTFs are generally defined under a concentric array of sound sources with respect to the coordinate origin, which matches the concentric reconstructed sound field of high order ambisonics (HOA).[7–9] Therefore, virtual HOA is suitable for binaural synthesis and adopted in this paper.

In order to use the HRTFs in the binaural synthesis, we could intuitively design the spatial sampling of HOA as the same sampling scheme used by the HRTFs. However, the samplings of HRTFs may be inefficient for the HOA and binaural synthesis. The HOA based on spherical har-

monics functions expansion has better performance when the virtual loudspeakers are distributed throuthout the whole space, but the sampling schemes of measured HRTFs are usually incomplete in the low elevations due to the difficulty in measurements. For example, HRTFs of (Knowles Electronic Manikin for Acoustic Research) KEMAR were measured from elevations –40° to 90° in MIT,[10] and HRTFs of 52 human subjects were obtained from elevations from –30° to 90° in SCUT[11]. Then, the HRTFs were usually measured with the subject being on a servo turntable,[12] which supports the equiangular intervals of azimuths at each elevation and thereby the denser distributions as the elevations deviate from the horizontal plane. The inhomogeneous spatial sampling may cause wasteful computation and decrease the effectiveness of virtual HOA.

In order to improve the binaural synthesis based on the virtual HOA, the nearly uniform sampling approach [3] for virtual loudspeakers was adopted. The HRTFs used in binaural synthesis are obtained by numerical simulation because the HRTFs at any directions can be easily calculated.[13] In addition, the distance encoding (or distance compensation) [9, 10] for virtual HOA is also introduced.

2. Method

2.1 Model and coordinate

Binaural synthesis techniques can be classified as data-based or model-based synthesis depending on a captured sound field or a number of virtual sound sources derived from analytical spatial source models, respectively. In comparison, the parametric analysis of binaural synthesis is easily done through the model-based binaural synthesis. Therefore, only model-based binaural synthesis is considered in this paper.

By definition, an HRTF describes the acoustic effects of the head and torso on the sound signals impinging upon listener's ears, and is related to the sound source position, including the distance r, azimuth θ, and elevation ϕ, and also the anatomical parameters of subjects. As shown in Fig. 1(a), $\phi = 90°$ is the vertical direction; in the horizontal plane of $\phi = 0°$, $\theta = 0°$ and 90° are the front and right of listener, respectively. Other directions can be similarly deduced.

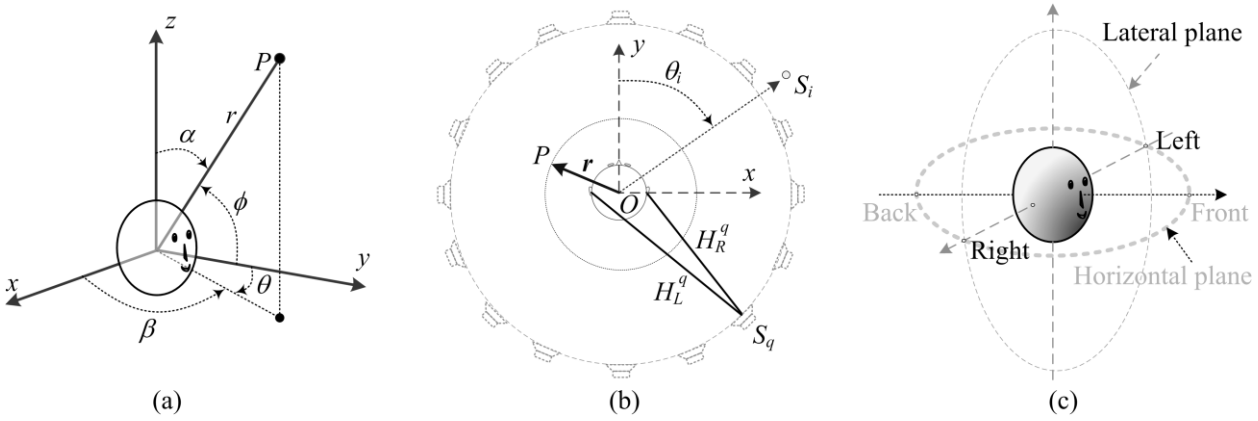

Figure 1. (a) Spherical coordinate system; (b) Diagrams of near-field virtual ambisonics; (c) Illustration of the horizontal and lateral plane.

For a point sound source S_i with unknown HRTFs, it can be decoded by an array of virtual loudspeakers [S_q] of HOA as shown in Fig. 1(b), and then a reconstructed spherical wave field within radius r can be reproduced. If the reconstructed region is larger than the subject's head, binaural signals can be synthesized via convolving the decoding signals [S_q] of virtual loudspeakers with the corresponding HRTFs [H_L^q, H_R^q]. The encoding and decoding of near-field HOA generally is defined in a spherical coordinate system (r, α, β) as shown in Fig. 1(a),[7, 8] and the near-field HRTFs are usually described by coordinate (r, θ, ϕ).[12, 13] Two coordinate systems are related with $\theta = 90° - \beta$

and $\phi = 90° - \alpha$. In Fig. 1(c), the horizontal and lateral planes are also described.

2.2 HOA and binaural synthesis

As shown in Fig. 1(a), the pressure $P(\boldsymbol{r}, \boldsymbol{r}_i, k)$ from a point source S_i is denoted by,

$$P(\boldsymbol{r}, \boldsymbol{r}_i, k) = \mathrm{j}k \frac{A_0}{4\pi|\boldsymbol{r} - \boldsymbol{r}_i|} e^{-\mathrm{j}k|\boldsymbol{r} - \boldsymbol{r}_i|}, \tag{1}$$

where $\boldsymbol{r}$ and $\boldsymbol{r}_i$ are the distance vectors of field point P and sound source S_i, respectively; k is wave number, $k = 2\pi f / c$, f is frequency, c is sound speed in air; and A_0 is the intensity of the point source or the known sound signal. In near-field, the sound pressure $P(\boldsymbol{r}, \boldsymbol{r}_i, k)$ can be expanded by, [7]

$$P(\boldsymbol{r}, \boldsymbol{r}_i, k) = \mathrm{j}kA_0 \left\{ \boldsymbol{\Lambda} \left[h_l(kr_i)\, j_l(kr) \right] \mathbf{Y}_i^* \right\}^{\mathrm{T}} \mathbf{Y}, \quad r \le r_i \tag{2}$$

where j_l is the l^{th} order spherical Bessel function, h_l is the l^{th} order spherical Hankel function. $\boldsymbol{\Lambda}$ is a $N \times N$ diagonal matrix with $N = (L+1)^2$, L is the truncated number of order l; and $\mathbf{Y}_i$ and $\mathbf{Y}$ are $N \times 1$ matrixes of spherical harmonics functions $[Y_{lm}(\alpha_i, \beta_i)]$ and $[Y_{lm}(\alpha, \beta)]$, respectively. The superscript "*" denotes the complex conjugate, the superscript "T" means the transpose of a matrix.

If input signals of all virtual loudspeakers consist of a $Q \times 1$ matrix $\mathbf{S}$ with $Q \ge (L+1)^2$, the total pressure $P(\boldsymbol{r}, [\boldsymbol{r}_q], k)$ at sound field P can be summed by the $Q \times 1$ weight coefficient matrix $\mathbf{C}$,

$$P(\boldsymbol{r}, [\boldsymbol{r}_q], k) = \mathbf{C}^{\mathrm{T}}\, \mathbf{S} \tag{3}$$

where $[\boldsymbol{r}_q]$ is the assembly of vector $\boldsymbol{r}_q$ of virtual loudspeakers. $\mathbf{C}$ is the re-encoding matrix,

$$\mathbf{C} = \mathrm{j}k \left\{ \boldsymbol{\Lambda} \left[h_l(kr_q)\, j_l(kr) \right] \left[\mathbf{Y}_q^* \right] \right\}^{\mathrm{T}} \mathbf{Y}, \quad r \le r_q \tag{4}$$

where $[\mathbf{Y}_q]$ is a $N \times Q$ matrix; $\mathbf{Y}_{\boldsymbol{q}}$ is $N \times 1$ matrix of spherical harmonics functions $[Y_{lm}(\alpha_q, \beta_q)]$.

If the pressure $P(\boldsymbol{r}, [\boldsymbol{r}_q], k)$ given by Eq. (3) is equal to the pressure $P(\boldsymbol{r}, \boldsymbol{r}_i, k)$ given by Eq. (2), the input signals $\mathbf{S}$ of all virtual loudspeakers can be solved by,

$$\mathbf{S} = A_0 \; pinv\left(\left[\mathbf{Y}_q^* \right] \right) \left\{ \boldsymbol{\Lambda} \left[\frac{h_l(kr_i)}{h_l(kr_q)} \right] \mathbf{Y}_i^* \right\}, \quad r \le r_i \text{ and } r \le r_q. \tag{5}$$

Eq. (2) is effective within $r \le L/k$, Eq. (5) is only satisfied under the conditions of $r \le r_i$ and $r \le r_q$, and the effective reconstruction region should be larger than the radius of the subject' head in binaural synthesis. Therefore, the effective region of a reconstructed field should be evaluated by,

$$a \le r \le \min\{cL/(2\pi f), r_i, r_q\} \tag{6}$$

Furthermore, the binaural signals can be generated by convolution between the multichannel signals $\mathbf{S}$ and the corresponding HRTFs as the following,

$$\mathbf{B} = conv(\mathbf{S}, \mathbf{H}). \tag{7}$$

where the abbreviation "*conv*" means the convolution operation; $\mathbf{B} = [\mathbf{B}_L\ \mathbf{B}_R]$ is a synthesized binaural signal; $\mathbf{S}$ is Q channels of input signals for virtual loudspeakers; $\mathbf{H} = [\mathbf{H}_L\ \mathbf{H}_R]$ is Q pairs of near-field HRTFs corresponding to the virtual loudspeakers.

2.3 Spatial sampling approaches

In order to compare the effect of spatial sampling on virtual HOA and binaural synthesis, three spatial sampling approaches are adopted, as shown in Fig. 2.

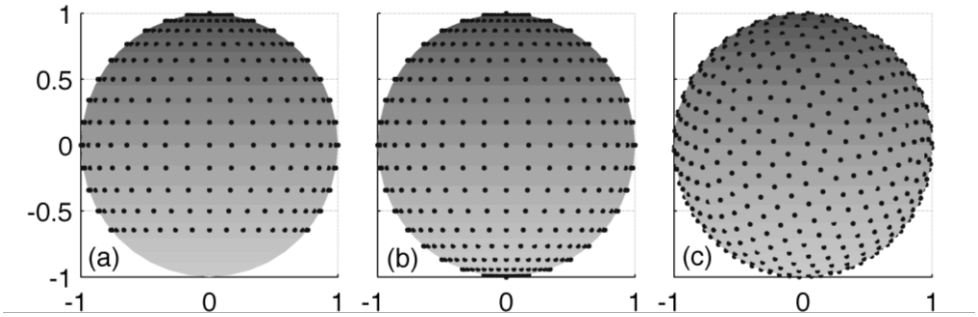

Figure 2. Three approaches of spatial sampling: (a) incomplete equiangular interval sampling with Q = 469; (b) complete equiangular interval sampling with Q = 614; (c) nearly uniform sampling with Q = 614.

3. Results

The effect of spatial sampling approaches on HOA and binaural synthesis can be analysed separately. In section 3.1, the reconstructions of HOA are discussed firstly. Then, the influence of sampling methods on binaural synthesis is investigated in section 3.2.

3.1 Reconstruction

For model-based binaural synthesis via virtual HOA in section 2, the multichannel feeds **S** of virtual loudspeakers in Eq. (5) will result in an effective sound field within a region limited by Eq. (6). The errors of the reconstructed sound field can be given by,

$$E(\boldsymbol{r},\boldsymbol{r}_i,k)=\frac{P(\boldsymbol{r},[\boldsymbol{r}_q],k)-P(\boldsymbol{r},\boldsymbol{r}_i,k)}{\left|P(\boldsymbol{r},\boldsymbol{r}_i,k)\right|}, \tag{8}$$

where "| |" is the absolute value sign, $P(\boldsymbol{r}, \boldsymbol{r}_i, k)$ and $P(\boldsymbol{r}, [\boldsymbol{r}_q], k)$ are the original and reconstructed sound pressures, respectively. The error E should be within [-1, 1] when the sound field is reconstructed properly. The original sound field, reconstructed sound fields, and their errors in the horizontal plane and lateral plane are shown in Fig. 3 and Fig. 4, respectively.

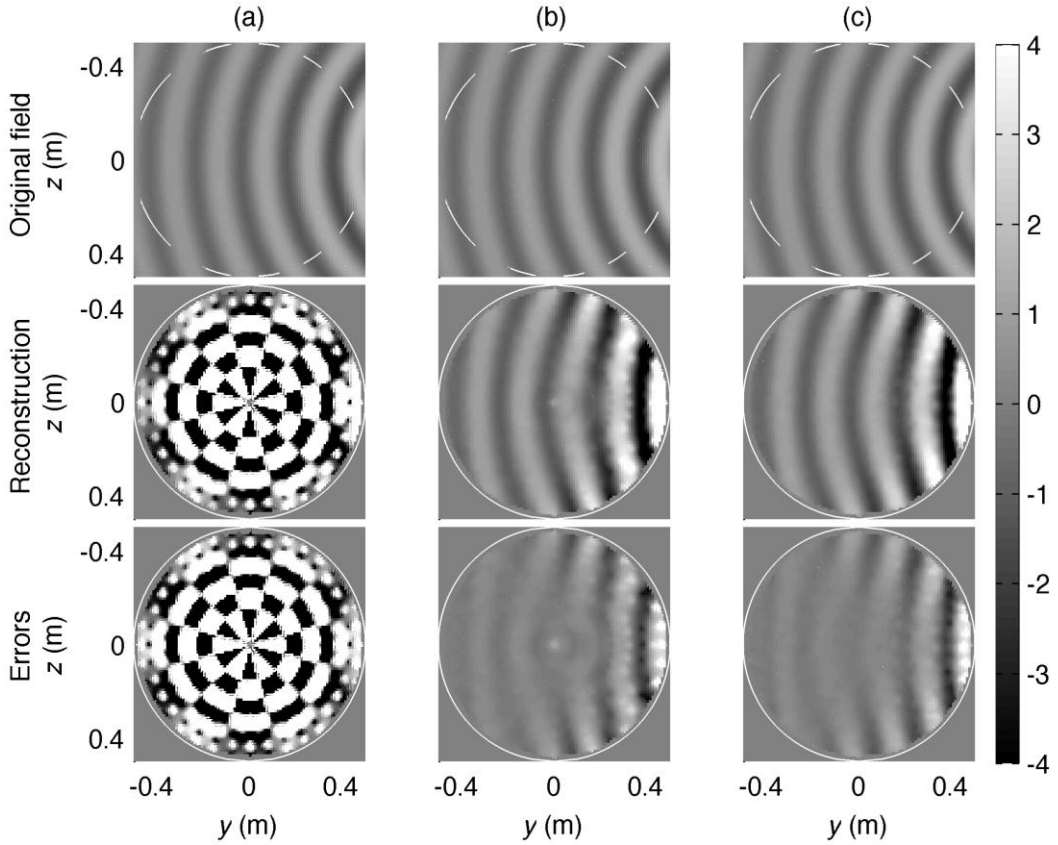

Figure 3. In the horizontal plane, the original and reconstructed sound fields are shown in the first and second rows; the errors are presented in the third row, under order L = 18, frequency f = 1.7 kHz, distance r_i = 1.0 m and r_q = 0.5 m, and the spatial sampling approaches in Fig. 2(a), (b) and (c).

For the incomplete spatial sampling approaches shown in Fig. 2(a), the sound field in the horizontal plane in Fig. 3(a) cannot be reconstructed properly, the errors exceed the bounds [−1, 1] in the whole region of $r \leq 0.5$ m; but the results in Fig. 4(a) indicate that the errors in the lateral plane gradually become acceptable as the elevation increases. For the complete equiangular interval sam-

pling and nearly uniform sampling in Fig. 2(b) and 2(c), the reconstructions round the centric regions in Fig. 3 and Fig. 4 are acceptable, although the reconstructed field under nearly uniform sampling seems a little better.

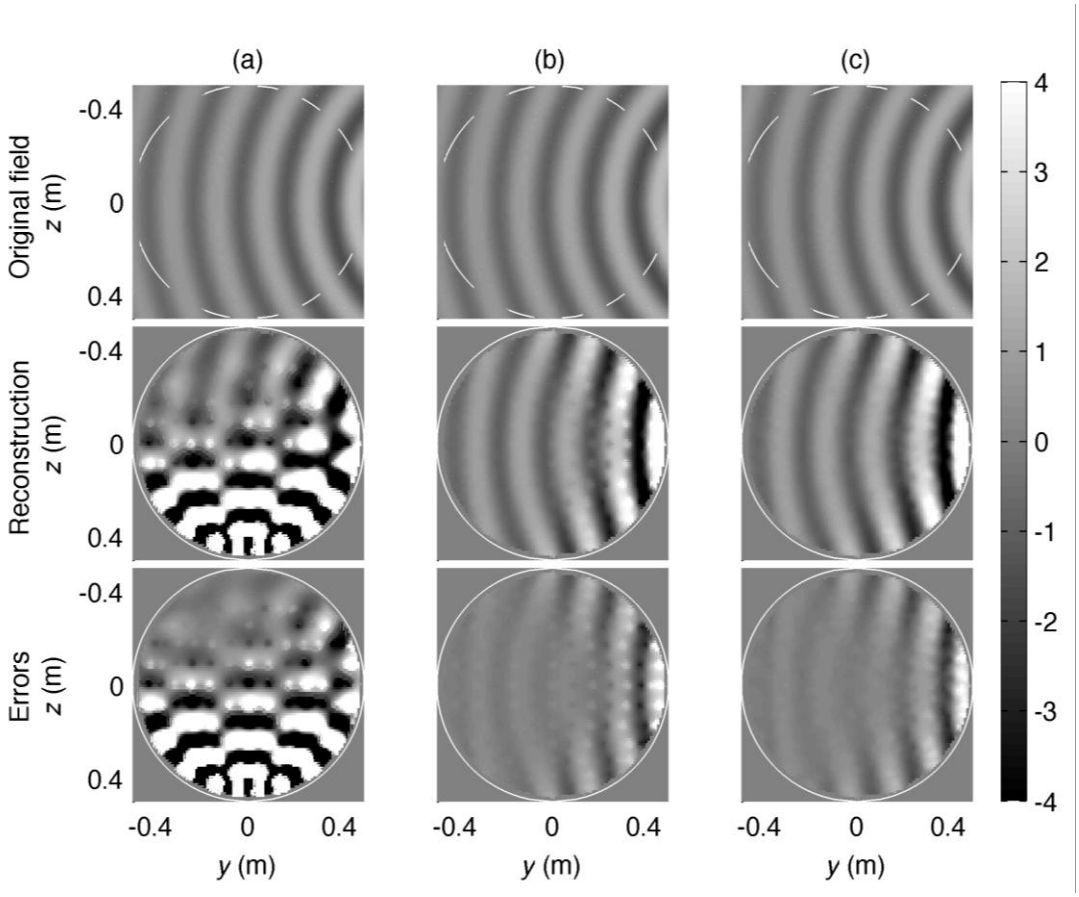

Figure 4. In the lateral plane, the original and reconstructed sound fields are shown in the first and second rows; the errors are presented in the third row, under order L = 18, frequency f = 1.7 kHz, distance r_i = 1.0 m and r_q = 0.5 m, and the spatial sampling approaches in Fig. 2(a), 2(b) and 2(c).

3.2 Binaural synthesis

Because the incomplete sampling results in Q = 469 directional samplings, the maximum order L = 20 is adopted in binaural synthesis (L = 20 results in $(L+1)^2$ = 441).

In the horizontal and lateral plane, the synthesized signals at the right ear are shown in Fig. 5 and Fig. 6. Overall, the errors of synthesized signals under the incomplete sampling approach are the most significant. At high frequencies, for all sampling approaches, the errors of ipsilateral synthesis (azimuth from 0 °to 180 °in Fig. 5, and various elevations in the right side in Fig. 6) are less than that of the contralateral synthesis (azimuth from 180 °to 360 °in Fig. 5, and various elevations in the left side in Fig. 6).

In Fig. 5(a), an unwanted peak of synthesis and errors around about 1 kHz goes through all azimuths and reaches about 20 dB. Except for the ipsilateral synthesized signals between about 5 kHz and 10 kHz, almost all synthesized signals are distorted. In Fig. 6(a), only the synthesis at high elevations above about 45 °are acceptable.

At high frequencies, the ipsilateral errors in Fig. 6(c) are less than that in Fig. 6(b). The aliasing phenomenon of synthesized signal at high frequencies in Fig. 5(c) is weaker than that in Fig. 5(a) and 5(b), which means that the nearly uniform sampling results in better synthesis and wider frequency limit. For Eq. (6), presumed the radius a = 0.0875m of KEMAR, the effective frequency limit should be not more than 12.4 kHz, which is nearly satisfied only in Fig. 5(c) and Fig. 6(c) under the nearly uniform sampling. Therefore, the nearly uniform sampling is a better choice in com-

parison.

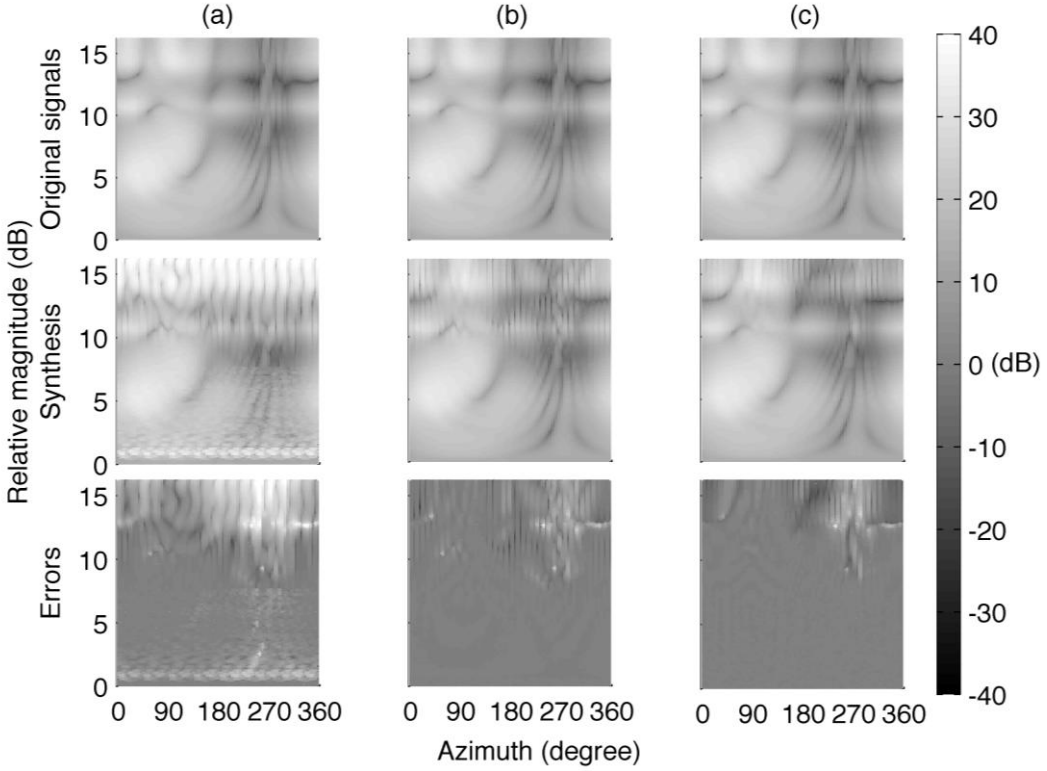

Figure 5. In the horizontal plane, the original and synthesized pressures at right ear are shown in the first and second rows; the errors are presented in the third row, under order L = 20, distance r_i = 1.2 m and r_q = 1.0 m, and the spatial sampling approaches (a), (b) and (c) as shown in Fig. 2.

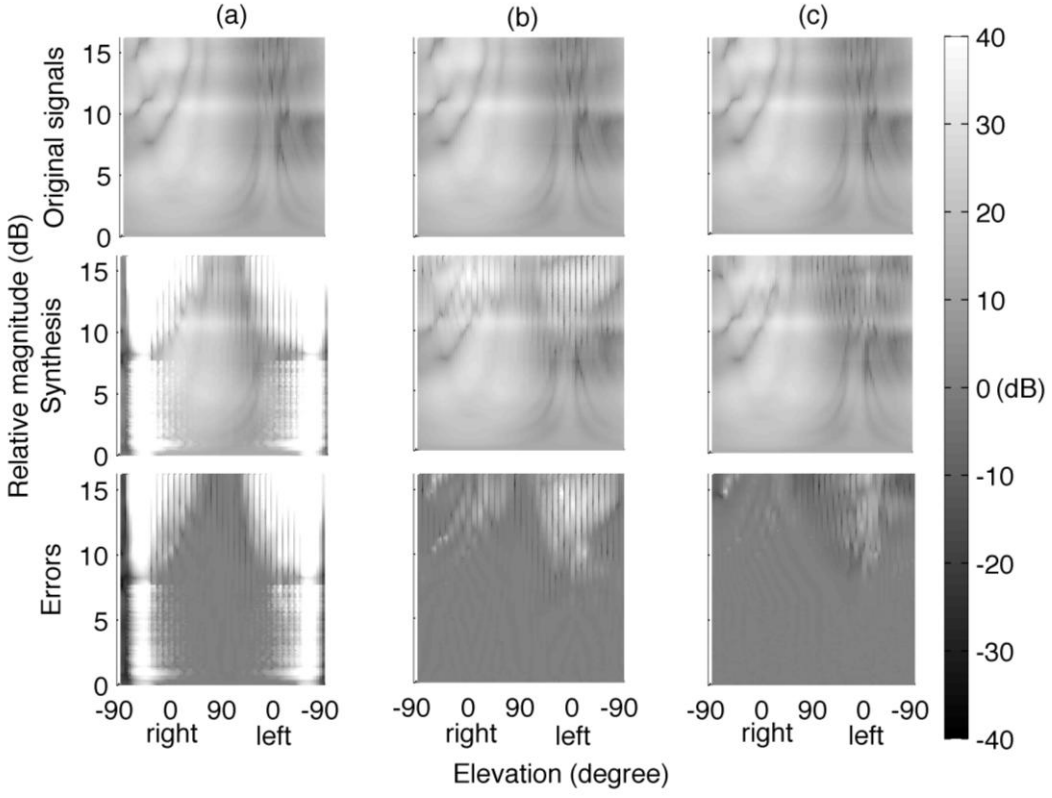

Figure 6. In the lateral plane, the original and synthesize pressures at right ear are shown in the first and second rows, respectively; the errors are presented in the third row, under order L = 20, distance r_i = 1.2 m

and r_q = 1.0 m, and the spatial sampling approaches (a), (b) and (c) as shown in Fig. 2.

4. Conclusions

Based on the calculated HRTFs of the scanned KEMAR artificial head, the effects of three sampling approaches of HRTFs on model-based binaural synthesis are analysed. Results show that the incomplete spatial sampling causes significant errors of sound field reconstructions and binaural synthesis, especially in the directions without samplings. Therefore, the complete sampling should be preferred in applications of virtual HOA. In comparison of the complete sampling methods, the nearly uniform sampling is more efficient than the equiangular interval sampling in virtual HOA and binaural synthesis. Since the HRTFs at low elevations are difficult to be measured, the interpolation or numerical method can be used to obtain the HRTFs at low elevations in practice.

Acknowledgement

The work is supported by EPSRC EP/K007491/1, Multisource audio-visual production from user-generated content, National Natural Science Foundation of China for Young Scholars (No 11104082), Natural Science Foundation of Guangdong Province (S2012010010015), and Fundamental Research Funds for the Central Universities (2013ZZ0087).

References

1 Xie, B. Head-related transfer function and virtual auditory display, J. Ross Publishing, New York, 20–22, 234–241, (2013).

2 Blauert, J. *Spatial Hearing*, *Revised edition*, MIT Press, Cambridge, UK, (1997).

3 Rafaely, B. Analysis and design of spherical microphone arrays, IEEE Trans. Speech Audio Process, **13**(1), 135–143, (2005).

4 Duraiswami, R., Zotkin, D. N., Li, Z., Grassi, E., Gumerov, N. A., and Davis, L. S. High order spatial audio capture and its binaural head-tracked playback over headphones with HRTF cues, 119th AES Convention, New York, USA, preprint 6540, October 7–10, (2005).

5 Gerzon, M. A. Ambisonics in multichannel broadcasting and video, J. Audio Eng. Soc. **33**(11), 859–871, (1985).

6 Berkhout, A. J., Vries, D. de, and Vogel, P. Acoustic control by wave field synthesis, J. Acoust. Soc. Am. **93**(5), 2764–2778, (1993).

7 Daniel, J. Spatial sound encoding including near field effect: Introducing distance coding filters and a viable new ambisonics format, 23rd AES International Conference, Copenhagen, Denmark, May 23–25, (2003).

8 Daniel, J., Nicol, R., and Moreau, S. Further investigations of high order ambisonics and wavefield synthesis for holophonic sound imaging, 114th AES Convention, Amsterdam, Netherlands, preprint 5788, March 22–25, (2003).

9 Menzies, D., Al-Akaidi, M. Nearfield binaural synthesis and ambisonics. J Acoust Soc Am, **121**(3), 1559–1563, (2007).

10 Gardner, W. G., MARTIN, K D. HRTF measurements of A KEMAR, J. Acoust. Soc. Am ., **97** (6), 3907–3908, (1995).

11 Xie, B. S., Zhong X. L., Rao D., and Liang, Z. Q. Head-related transfer function database and its analyses, Sci. China, Ser. G: Physics, Mechanics & Astronomy, **36** (5), 464–479, (2006).

12 Yu, G. Z., Xie, B. S., and Rao, D. Characteristics of Near-field head- related transfer function for KEMAR. AES 40th Convention, Tokyo in Japan, October 8–10, (2010).

[13] Rui, Y. Q., Yu, G. Z., Xie B. S., and Liu, Y. Calculated individualized near-field head-related transfer function database using boundary element method, the 134th Convention, Rome, Italy, May 4–7, (2013).