MADRID
inter.noise 2019
June 16 - 19
NOISE CONTROL FOR A BETTER ENVIRONMENT

Transaural reproduction of spatial surround sound using four actual loudspeakers

Xie Bosun[1]
Acoustic Lab, School of Physics and Optoeletronics, South China University of Technology, Guangzhou 510641, China;
State Key Laboratory of Subtropical Building Science, South China University of Technology, Guangzhou, China.
Liu Lulu[2]
Acoustic Lab, School of Physics and Optoeletronics, South China University of Technology, Guangzhou 510641, China;
Zhang, Chengyun [3]
School of Mechanical and Electrical Engineering, Guangzhou University, Guangzhou, 510006, China

ABSTRACT

Recently, multichannel sound has been evolving from horizontal surround to spatial surround with height. On the other hand, in some practical uses such as TV set, it is inconvenient to arrange multiple loudspeakers for multichannel sound reproduction. Transaural technique, which consists of HRTF-based binaural synthesis and cross-talk cancellation, enables to reproduce multichannel sound by using fewer actual loudspeakers. However, conventional transaural reproduction with two frontal loudspeakers is only able to recreate virtual source in the frontal-horizontal quadrants. In present work, a method for transaural reproduction of spatial surround sound using four actual loudspeakers is proposed. The four actual loudspeakers are arranged in the left-front and right-front directions in the horizontal plane, as well as left-front-up, right-front-up directions in a higher elevation plane, respectively. It is proved experimentally that, with transaural processing, this loudspeaker arrangement is able to recreate virtual source within the frontal-hemispherical directions. In practical use for TV set, the actual four-loudspeaker configuration can be realized by a pair of horizontal sound-bars (bar-shape loudspeaker boxes) arranged on and below the TV set, respectively; or realized by a pair of vertical sound-bars arranged on two sides of the TV set, respectively.

Keywords: Spatial sound, transaural reproduction, HRTF
I-INCE Classification of Subject Number: 61

1. INTRODUCTION

Conventionally, the multiple loudspeakers of horizontal surround sound are arranged in the horizontal plane. For example, the ITU-recommended loudspeakers configuration

[1] phbsxie@scut.edu.cn
[2] 348802659@ qq.com
[3] zhangcy@ gzhu.edu.cn

for 5.1 channel surround sound includes five loudspeakers with full audible bandwidth in the horizontal plane, namely left-front, centre, right-front, left-surround and right-surround loudspeakers. It also includes an optional subwoofer.

Recently, multichannel sound has been evolving from horizontal surround to spatial surround with height and increasingly applied to sound reproduction in cinema and home[1][2]. Compared with conventional horizontal surround sound, spatial surround sound greatly improves the spatial perceptual performance, but requires more complex loudspeakers configuration. Usually, layer-wise loudspeaker configurations are used in spatial surround sound. For example, the 9.1 channel spatial surround sound consists of a middle(horizontal)-layer and a high-layer loudspeaker configurations, as well as an optional subwoofer. The arrangement of five loudspeakers with full audible bandwidth in the middle layer is identical to that of 5.1 channel surround sound. And the four loudspeakers with full audible bandwidth in the upper layer are arranged above the left-front, right-front, left-surround and right-surround loudspeakers in the horizontal plane, respectively[3].

On the other hand, in some practical uses such as TV set, it is inconvenient to arrange multiple loudspeakers for multichannel sound reproduction. To address this problem, transaural technique, which consists of HRTF-based binaural synthesis and cross-talk cancellation, has been used to convert multichannel signals for reproduction with fewer loudspeakers. Such application of transaural technique is also called "virtual surround sound". The nature of "virtual surround sound" is first using transaural processing to create more virtual loudspeakers from fewer actual loudspeakers, and then reproducing multichannel signals through multiple virtual loudspeakers. Transaural technique has been applied to reproduce 5.1 channel and other multichannel surround sounds through two actual frontal loudspeakers[4][5]. Some commercial techniques, such as Dolby Virtual Surround, are available (see http://www.dolby.com).

However, conventional transaural reproduction with two frontal loudspeakers is only able to recreate virtual source in the frontal-horizontal quadrants. In order to recreate virtual source in vertical directions, a method of transaural reproduction of spatial surround sound using four actual loudspeakers is proposed. It is proved experimentally that the method is able to recreate virtual source in both frontal-horizontal quadrants and vertical directions.

2. PRINCIPLE AND ANALYSIS

In the following, the spatial direction is specified by elevation -90° ≤ ϕ ≤ 90° and azimuth -180°< θ ≤ 180°. Where ϕ = -90°, 0° and 90° denote below, horizontal and above directions, respectively. In the horizontal plane, azimuth θ = 0°, 90°, 180° and -90° denote the front, left, back and right directions, respectively.

As shown in Fig.1, four actual loudspeakers, namely L1, R1, L2 and R2, are used in reproduction. They are arranged in the directions of left-front, right-front, left-front-up and right-front-up, respectively. The left-front and right-front loudspeakers are arranged in the horizontal plane with an elevation

$$\phi_{L1} = \phi_{R1} = 0° \quad (1)$$

Of course, the left-front and right-front loudspeakers can also be arranged in a little bit lower elevation than the horizontal plane. For practical uses in TV set, the span azimuth between a pair of left-front and right-front loudspeakers are smaller than the standard of 60º. It usually varies between 20º and 30º. Then, the azimuths of left-front and right-front loudspeakers are:

$$\theta_{L1} = 10° \sim 15° \qquad \theta_{R1} = \text{-}10° \sim \text{-}15° \quad (2)$$

The left-front-up and right-front-up loudspeakers are arranged in the position above the horizontal plane with elevation

$$\phi_{L2} = \phi_{R2} = 30° \pm 15° \tag{3}$$

And their azimuths are

$$\theta_{L2} = 10° \sim 15° \qquad \theta_{R2} = \text{-}10° \sim \text{-}15° \tag{4}$$

In practical use, this four-loudspeaker arrangement can be realized by a pair of horizontal sound-bar (there are two loudspeakers in each sound sound-bar), with one sound-bar being above and another below the TV set; or realized by a pair of vertical sound-bar, with one on the left and another on the right of the TV set.

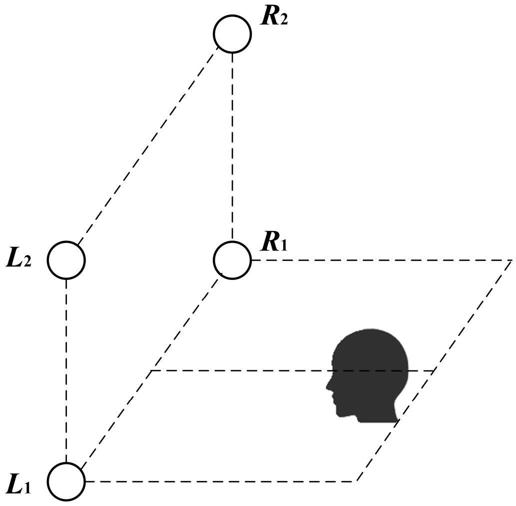

Fig.1 The arrangement of four actual loudspeakers

There are various formats of multichannel spatial surround sound with height. Usually, the loudspeakers of multichannel spatial surround sound with height are arranged in two elevation layers, including a middle (horizontal) layer and an upper layer. An additional low layers is also possible. Suppose that there are $M + 2$ loudspeakers in the middle layer, their elevations are $\phi_m = 0^\circ$ and signals are E_m, with $m = 1, 2, \ldots, (M + 2)$. The azimuths of M loudspeakers which are not located in the front or back are denoted by θ_m, with $m = 1, 2, \ldots, M$; and the azimuths of frontal and back loudspeakers (if they exist) are $\theta_{M+1}= 0^\circ$ and $\theta_{M+2}= 180^\circ$, respectively. Also suppose that there are $M' + 2$ loudspeakers in the upper layer, their elevations are $\phi'_{m'} = \phi_H$, and signals are $E'_{m'}$, with $m' = 1, 2, \ldots, (M' + 2)$. The azimuths of M' loudspeakers which are not located in the front or back are denoted by $\theta'_{m'}$, with $m' = 1, 2, \ldots, M'$; and the azimuths of frontal and back loudspeakers (if they exist) are $\theta'_{M'}+1 = 0^\circ$ and $\theta'_{M'}+2 = 180^\circ$, respectively.

In transaural reproduction, the original multichannel signals are reproduced by a set of virtual loudspeakers. The virtual loudspeakers are created by HRTF-based binaural synthesis and cross-talk cancellation, or equivalently, created by controlling the signals of actual loudspeakers so that the binaural pressures equal to those in multichannel reproduction. Of course, some additional equalization is needed to avoid timbre artifacts in reproduction.

For transaural reproduction by aforementioned four actual loudspeakers, the virtual loudspeakers in the middle layer are created by the two actual loudspeakers L1 and R1 in left-front and right-front directions, respectively. To create a virtual loudspeaker with azimuth θ_m at the non-frontal and non-back directions, the original input signal E_m is filtered by a pair of transaural filters $G_{L1}(\theta_m, f)$, $G_{R1}(\theta_m, f)$

$$E_{L1,m} = G_{L1}(\theta_m, f) E_m \qquad E_{R1,m} = G_{R1}(\theta_m, f) E_m \tag{5}$$

Then, the binaural pressures created by the two actual loudspeakers L1 and R1 are given by

$$P'_L = \alpha_1 E_{L1,m} + \beta_1 E_{R1,m} \qquad P'_R = \beta_1 E_{L1,m} + \alpha_1 E_{R1,m} \tag{6}$$

Because two actual loudspeakers in the horizontal plane are arranged left-right symmetrically, $\alpha_1 = \alpha_1(f)$ and $\beta_1 = \beta_1(f)$ are the HRTFs from actual loudspeaker at left-front (or right-front) to the ipsilateral and contralateral ear, respectively; f denotes frequency.

For a loudspeaker at azimuth θ_m and with signal E_m in multichannel reproduction, the binaural pressures are given by

$$P_L = H_L(\theta_m, f) E_m \qquad P_R = H_R(\theta_m, f) E_m \tag{7}$$

Where $H_L(\theta_m, f)$ and $H_R(\theta_m, f)$ are a pairs of HRTFs for a loudspeaker at horizontal direction θ_m.

The responses of two transaural filers can be directly derived by letting Eq.(6) be equal to Eq.(7). A constant-power equalization is applied to the responses of two transaural filers to improve the timbre. The principle of timbre equalization in two-loudspeaker reproduction is explained as follows. Given the difficulty in robustly rendering the fine high-frequency spectral cues to listeners' ears in loudspeaker reproduction, the perceived virtual source direction is dominated by interaural cues (especially ITD) and is limited to frontal-horizontal quadrants. The interaural cues are controlled by the relative, rather than the absolute, magnitude and phase of left and right loudspeaker signals. Scaling both loudspeaker signals with identical frequency-dependent coefficients does not alter their relative magnitude and phase, or the perceived virtual source direction. However, this manipulation alters the overall power spectra of loudspeaker signals and therefore equalizes timbre. For a constant-power equalization, the responses of the transaural synthesis filters $G_{L1}(\theta_m, f)$ and $G_{R1}(\theta_m, f)$ in Eqs.(5) and (6) are equalized by their RMS so that the total power of the outputs of two transaural filters is independent of frequency.

Incorporating constant-power equalization, the responses of two transaural synthesis filters are given by [4][6]

$$\begin{aligned} G_{L1}(\theta_m, f) &= \frac{\alpha_1 H_L(\theta_m, f) - \beta_1 H_R(\theta_m, f)}{\sqrt{|\alpha_1 H_L(\theta_m, f) - \beta_1 H_R(\theta_m, f)|^2 + |-\beta_1 H_L(\theta_m, f) + \alpha_1 H_R(\theta_m, f)|^2}} \frac{|\alpha_1^2 - \beta_1^2|}{\alpha_1^2 - \beta_1^2} \\ G_{R1}(\theta_m, f) &= \frac{-\beta_1 H_L(\theta_m, f) + \alpha_1 H_R(\theta_m, f)}{\sqrt{|\alpha_1 H_L(\theta_m, f) - \beta_1 H_R(\theta_m, f)|^2 + |-\beta_1 H_L(\theta_m, f) + \alpha_1 H_R(\theta_m, f)|^2}} \frac{|\alpha_1^2 - \beta_1^2|}{\alpha_1^2 - \beta_1^2} \end{aligned} \tag{8}$$

For original signals corresponding to the frontal and back loudspeakers in multichannel sound reproduction, they can be fed to the actual loudspeakers L1 and R1 at the same time (after an 3 dB attenuation). That is, the virtual loudspeakers are created by traditional phantom image method. The finial signals for two actual loudspeakers L1 and R1 are the combination of the signals for all virtual loudspeakers,

$$\begin{aligned} E_{L1} &= \sum_{m=1}^{M} E_{L1,m} + 0.7E_{M+1} + 0.7E_{M+2} = \sum_{m=1}^{M} G_{L1}(\theta_m, f) E_m + 0.7E_{M+1} + 0.7E_{M+2} \\ E_{R1} &= \sum_{m=1}^{M} E_{R1,m} + 0.7E_{M+1} + 0.7E_{M+2} = \sum_{m=1}^{M} G_{R1}(\theta_m, f) E_m + 0.7E_{M+1} + 0.7E_{M+2} \end{aligned} \tag{9}$$

Similarly, virtual loudspeakers in the upper layer are created by the two actual loudspeakers L2 and R2 in left-front-up and right-front-up directions, respectively. The finial signals for two actual loudspeakers L2 and R2 are the combination of the signals for all virtual loudspeakers in the upper layer.

$$\begin{aligned} E_{L2} &= \sum_{m'=1}^{M'} G_{L2}(\theta'_{m'}, f) E'_{m'} + 0.7E'_{M'+1} + 0.7E'_{M'+2} \\ E_{R2} &= \sum_{m'=1}^{M'} G_{R2}(\theta'_{m'}, f) E'_{m'} + 0.7E'_{M'+1} + 0.7E'_{M'+2} \end{aligned} \tag{10}$$

The responses of two transaural filers are derived by the same method as that of Eq.(8),

$$\begin{aligned} & G_{L2}(\theta'_{m'}, f) \\ & = \frac{\alpha_2 H_L(\theta'_{m'}, f) - \beta_2 H_R(\theta'_{m'}, f)}{\sqrt{|\alpha_2 H_L(\theta'_{m'}, f) - \beta_2 H_R(\theta'_{m'}, f)|^2 + |-\beta_2 H_L(\theta'_{m'}, f) + \alpha_2 H_R(\theta'_{m'}, f)|^2}} \frac{|\alpha_2^2 - \beta_2^2|}{\alpha_2^2 - \beta_2^2} \\ & G_{R2}(\theta'_{m'}, f) \\ & = \frac{-\beta_2 H_L(\theta'_{m'}, f) + \alpha_2 H_R(\theta'_{m'}, f)}{\sqrt{|\alpha_2 H_L(\theta'_{m'}, f) - \beta_2 H_R(\theta'_{m'}, f)|^2 + |-\beta_2 H_L(\theta'_{m'}, f) + \alpha_2 H_R(\theta'_{m'}, f)|^2}} \frac{|\alpha_2^2 - \beta_2^2|}{\alpha_2^2 - \beta_2^2} \end{aligned} \tag{11}$$

Where, $H_L(\theta'_{m'}, f)$ and $H_R(\theta'_{m'}, f)$ are a pairs of HRTFs for virtual loudspeakers at azimuth $\theta'_{m'}$ in the upper layer. Because two actual loudspeakers at the left-front-up and right-front-up directions are arranged left-right symmetrically, $\alpha_2 = \alpha_2(f)$ and $\beta_2 = \beta_2(f)$ are the HRTFs from actual loudspeaker at left-front–up (or right-front-up) to the ipsilateral and contralateral ear, respectively.

Usually, the original arrangement of loudspeakers in multichannel spatial sound are left-right symmetric. The signal processing in Eq.(9) can be simplified by considering the symmetry[4][5]. The M original channels (loudspeakers) at the non-frontal and non-back directions in the middle layer are numbered so that the odd numbers denote the channels at left and the even numbers denote the symmetric channels at the rights. Then the responses of transaural filters in Eq.(9) satisfy following symmetric relationship:

$$\begin{aligned} & G_{L1}(\theta_1, f) = G_{R1}(\theta_2, f) \qquad G_{L1}(\theta_2, f) = G_{R1}(\theta_1, f) \\ & G_{L1}(\theta_3, f) = G_{R1}(\theta_4, f) \qquad G_{L1}(\theta_4, f) = G_{R1}(\theta_3, f) \\ & \vdots \\ & G_{L1}(\theta_{M-1}, f) = G_{R1}(\theta_M, f) \qquad G_{L1}(\theta_M, f) = G_{R1}(\theta_{M-1}, f) \end{aligned} \tag{12}$$

Then Eq.(9) is equivalent to following Eq.(13)

$$\begin{bmatrix} E_{L1} \\ E_{R1} \end{bmatrix} = 0.7 \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \left\{ \sum_{m=odd}^{M-1} \begin{bmatrix} \Sigma_{m,m+1} & 0 \\ 0 & \Delta_{m,m+1} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} E_m \\ E_{m+1} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} (E_{M+1} + E_{M+2}) \right\} \tag{13}$$

The sum in Eq.(13) is over all odd numbers m， and

$$\begin{aligned} \Sigma_{m,m+1} &= 0.707[G_{L1}(\theta_m, f) + G_{L1}(\theta_{m+1}, f)] \\ \Delta_{m,m+1} &= 0.707[G_{L1}(\theta_m, f) - G_{L1}(\theta_{m+1}, f)] \end{aligned} \tag{14}$$

Similarly, the M' original channels (loudspeakers) at the non-frontal and non-back directions in the upper layer are numbered so that the odd numbers denote the channels at left and the even numbers denote the symmetric channels at the rights. The responses of transaural filters in Eq.(10) satisfy following symmetric relationship:

$$
\begin{aligned}
&G_{L2}(\theta'_1,f)=G_{R2}(\theta'_2,f) \qquad G_{L2}(\theta'_2,f)=G_{R2}(\theta'_1,f)\\
&G_{L2}(\theta'_3,f)=G_{R2}(\theta'_4,f) \qquad G_{L2}(\theta'_4,f)=G_{R2}(\theta'_3,f)\\
&\vdots\\
&G_{L2}(\theta'_{M'-1},f)=G_{R2}(\theta'_{M'},f) \qquad G_{L2}(\theta'_{M'},f)=G_{R2}(\theta'_{M'-1},f)
\end{aligned} \tag{15}
$$

Then the signals processing in Eq.(10) is equivalent to following Eq.(16)

$$
\begin{bmatrix} E_{L2} \\ E_{R2} \end{bmatrix} = 0.7 \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \left\{ \sum_{m'=odd}^{M'-1} \begin{bmatrix} \Sigma'_{m',m'+1} & 0 \\ 0 & \Delta'_{m',m'+1} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} E'_{m'} \\ E'_{m'+1} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} (E'_{M'+1} + E'_{M'+2}) \right\} \tag{16}
$$

The sum in Eq.(16) is over all odd numbers m', and

$$
\begin{aligned}
\Sigma'_{m',m'+1} &= 0.707[G_{L2}(\theta'_{m'},f)+G_{L2}(\theta'_{m'+1},f)]\\
\Delta'_{m',m'+1} &= 0.707[G_{L2}(\theta'_{m'},f)-G_{L2}(\theta'_{m'+1},f)]
\end{aligned} \tag{17}
$$

The transaural processing in Eq.(13) and Eq.(16) includes (M + M') filters in total, which is just half of 2(M + M') filters in Eq.(9) and Eq.(10). Therefore, the efficiency of transaural processing is improved. Fig.2 is the block diagram of the signal processing for the middle layer input. The block diagram of the signal processing for the upper layer input is similar to this.

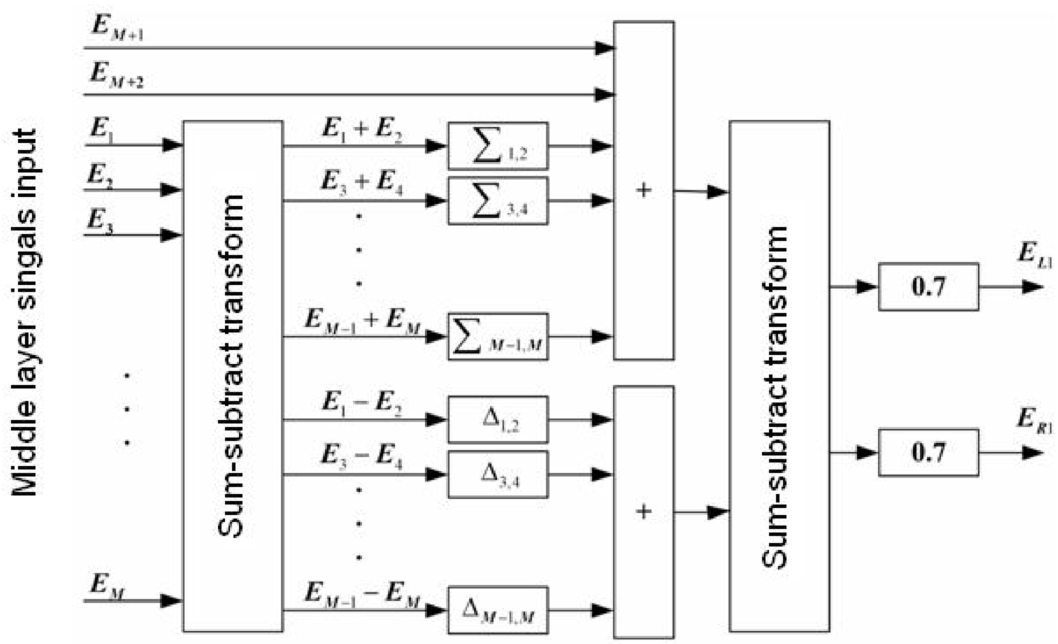

Fig.2 The block diagram of the signal processing for the middle layer input.

3. TRANSAURAL REPRODUCTION OF 9.1 CHANNEL SPATIAL SOUND

As an example of the method discussed in Section 2, transaural reproduction of 9.1 channel spatial sound is discussed in this section. 9.1 channel system is a typical spatial sound system with two-layer loudspeaker arrangement. There are nine independent channels with full audible bandwidth in 9.1 channel system. As shown in Fig.3, the middle layer includes M = 5 channels (loudspeakers), namely left (L), centre (C), right (R), left surround (LS) and right surround (RS). It also includes a low frequency effect channel and subwoofer, but the low frequency effect channel is omitted here. Four non-frontal and non-back channels in the horizontal plane, i.e., the L, R, LS and RS channels are left-right symmetric, the corresponding signals are E_L, E_C, E_{LS}, and E_{RS}, respectively. According to the notation in Section 2, the five channel signals are numbered as:

$$E_1 = E_L \quad E_2 = E_R \quad E_3 = E_{LS} \quad E_4 = E_{RS} \quad E_5 = E_C \tag{18}$$

The elevations of all five loudspeakers in the middle layer are 0º, and azimuths are

$$\theta_1 = \theta_L = 30° \quad \theta_2 = \theta_R = -30° \quad \theta_3 = \theta_{LS} = 110° \quad \theta_4 = \theta_{RS} = -110° \quad \theta_5 = \theta_C = 0° \tag{19}$$

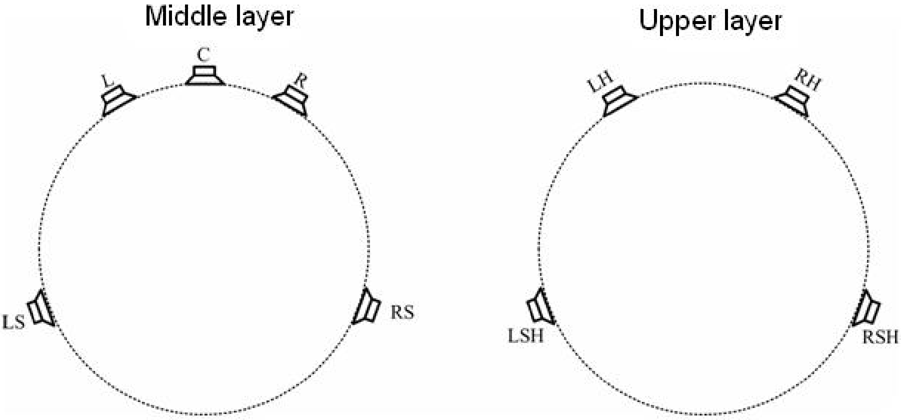

Fig.3 Middle layer and upper layer loudspeaker arrangement in 9.1 channel system

The upper layer of 9.1 channel system includes $M' = 4$ left-right symmetric channels, namely, left-high (LH), right-high (RH), left-surround-high (LSH) and right-surround-high (RSH), the corresponding signals are E'_{LH}, E'_{RH}, E'_{LSH} and E'_{RSH}, respectively. There are no centre-upper and back-upper channels. According to the notation in Section 2, the four channel signals are numbered as:

$$E'_1 = E'_{LH} \quad E'_2 = E'_{RH} \quad E'_3 = E'_{LSH} \quad E'_4 = E'_{RSH} \tag{20}$$

The elevations of all four loudspeakers in the upper layer are 30º, and azimuths are

$$\theta'_1 = \theta'_{LH} = 30° \quad \theta'_2 = \theta'_{RH} = -30° \quad \theta'_3 = \theta'_{LSH} = 110° \quad \theta'_4 = \theta'_{RSH} = -110° \tag{21}$$

Given the loudspeaker arrangement of 9.1 channel system and those of the four actual loudspeakers in transaural reproduction [Eq.(1) ~ Eq.(4)], the transaural processing can be implemented according to Eq.(13) and Eq.(16). Because it is difficult to create virtual loudspeakers in the back-hemispherical space by using the actual loudspeakers in the frontal-hemispherical space, in the transaural processing, the virtual surround loudspeakers in the middle and upper layers are moved to the lateral directions with azimuths,

$$\theta_3 = \theta_{LS} = 90° \quad \theta_4 = \theta_{RS} = -90° \qquad \theta'_3 = \theta'_{LS} = 90° \quad \theta'_4 = \theta'_{RS} = -90° \tag{22}$$

KEMAR-HRTFs are used in transaural processing. The HRTFs are obtained by measurement. The sampling frequency of HRTF is 44.1 kHz, and the length of filter response is 128 points.

4. EXPERIMENT

A virtual source localization experiment was conducted to validate the proposed method. Transaural reproduction of 9.1 channel spatial sound is taken as an example. Experiment was conducted in a listening room with reverberation time of 0.15 s. The elevations and azimuths of the four actual loudspeakers are $\phi_{L1} = \phi_{R1} = 0º$, $\phi_{L2} = \phi_{R2} = 30º$; $\theta_{L1} = \theta_{L2} = 15º$, $\theta_{R1} = \theta_{R2} = -15º$, respectively. The actual loudspeakers are located at a distance of 1.5 m relative to the head centre of subject. Two kind of stimuli, including speech (Chinese male voice) and orchestral music (segment from J.Strauss, On The Beautiful Blue Danube), were used in the experiment. The length of each stimulus was 10 s. The raw stimuli were processed according to the method described in Section 2 so as to create each of nine virtual loudspeakers in 9.1 channel reproduction. The target directions of virtual loudspeakers were chosen according to Eq.(19) and Eq.(21), except that the azimuths of four virtual surround loudspeakers were revised according to Eq.(22). After transaural processing, the resultant signals were fed to the four actual loudspeakers. Eight subjects participated in the experiment. The subjects were from 22 to 30 years old and had normal hearing.

In the experiment, subject judged the perceived direction (including elevation and azimuth) of each of nine virtual loudspeakers. On each condition (each virtual loudspeaker, each kind of stimulus), each subject repeatedly judged three times. Therefore, there were 3 repetitions × 8 subjects = 24 judgments under each condition.

The performance of localization was evaluated by four statistical measures on the experimental results, including the percentage of front-back (F-B) confusion, the percentage of up-down(U-D) confusion, the mean unsigned azimuthal error, and the mean unsigned elevation error[4]. Table 1 lists the results.

Table 1 Statistical analysis of the experimental results

Virtual loudspeakers	Target direction(°) (θ,ϕ)	Stimuli	Confusion		Mean/standard deviation	
			F-B (%)	U-D (%)	Unsigned azimuth error(°)	Unsigned elevation error(°)
C	(0,0)	speech	/	/	0.4/1.3	0.9/1.6
		music	/	/	0.5/1.5	1.1/1.4
L	(30,0)	speech	0	0	4.1/2.6	1.2/1.8
		music	0	0	3.8/4.9	1.8/1.5
R	(-30,0)	speech	0	0	2.5/2.4	1.5/1.7
		music	0	0	4.5/2.7	1.4/1.4
LS	(90,0)	speech	0	0	30.2/8.0	2.8/2.6
		music	0	0	25.8/12.1	2.6/1.9
RS	(-90,0)	speech	0	0	36.4/7.1	1.8/2.4
		music	0	0	36.7/6.8	1.9/1.8
LH	(30,30)	speech	0	0	6.3/3.1	2.4/2.0
		music	0	0	4.8/2.7	2.1/1.5
RH	(-30,30)	speech	0	0	4.2/3.8	2.1/1.3
		music	0	0	4.4/3.6	2.9/1.3
LSH	(90,30)	speech	0	0	26.7/7.5	3.4/1.9
		music	0	0	30.3/9.1	2.9/1.6
RSH	(-90,30)	speech	0	0	27.2/9.4	2.6/1.4
		music	0	0	33.3/13.8	3.3/2.1

It can be seen that there are no front-back and up-down confusion, all virtual loudspeakers are perceived within the frontal and up-hemispherical space (including the frontal-horizontal plane). The mean unsigned elevation errors are within the reasonable bound. Therefore, the proposed method is able to recreate virtual loudspeakers in the vertical directions. The mean unsigned azimuthal errors for the four virtual surround loudspeakers LS, RS, LSH and RSH are a little bit large. The mean perceived azimuths are about $\theta = \pm 60^{\circ}$, rather than at the target direction of $\theta = \pm 90^{\circ}$. Therefore, the four virtual surround loudspeakers at lateral directions move towards the front. This is an inherent defect of transaural reproduction by actual loudspeakers arranged in the front[4].

5. CONCLUSIONS

For the applications of multichannel spatial sounds to some practical cases, such as TV set, a method of transaural reproduction using four actual loudspeakers is proposed. The four actual loudspeakers are arranged in the left-front and right-front directions in the horizontal plane, as well as left-front-up, right-front-up directions in a higher elevation plane, respectively. It is proved experimentally that, with transaural processing, this method is able to recreate virtual loudspeakers within the frontal-hemispherical directions, including vertical directions. Of course, it is unable to create virtual loudspeakers or virtual sources in the back-hemispherical directions. In addition, the virtual loudspeakers move towards the front. To further overcome these defects, more actual loudspeakers are needed. This is our future work.

6. ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (11674105) and the State Key Lab of Subtropical Building Science, South China University of Technology.

7. REFERENCES

1. ITU-R Report BS.2159-7, *"Multichannel sound technology in home and roadcasting applications."* International Telecommunication Union, Geneva, Switzerland, (2015).
2. Herre J., Hilpert J., Kuntz A., et al., *"MPEG-H audio—the new standard for universal spatial/3D audio coding."* Journal of the Audio Engineering Society 62.12 (2015): 821-830.
3. Theile, G., Wittek H., *"Principles in surround recordings with height."* Audio Engineering Society Convention 130. Audio Engineering Society, (2011).
4. Xie, B.S., *"Head-related transfer function and virtual auditory display."* J. Ross Publishing, (2013).
5. Bauck J., Cooper D. H., *"Generalized transaural stereo and applications."* Journal of the Audio Engineering Society 44.9 (1996): 683-705.
6. Xie B.S.，Zhong X.L., Yu G.Z. et al., *"Report on Research Projects on Head-Related Transfer Functions and Virtual Auditory Displays in China."* Journal of the Audio Engineering Society 61.5 (2013): 314-326.