PAN
ARCHIVES OF ACOUSTICS
Vol. **46**, No. 2, pp. 213–228 (2021)
DOI: 10.24425/aoa.2021.136577

Research Paper

Analysis and Experiment on the Limitations of Static and Dynamic Transaural Reproduction with Two Frontal Loudspeakers

Lulu LIU, Bosun XIE*

Acoustic Lab, School of Physics and Optoelectronics
South China University of Technology
Guangzhou, 510641, China
*Corresponding Author e-mail: phbsxie@scut.edu.cn

(*received April 19, 2020; accepted December 19, 2020*)

By duplicating the binaural pressures of an actual source, transaural reproduction with two frontal loudspeakers is expected to recreate a virtual source in arbitrary direction. However, experiments indicated that in static transaural reproduction, the perceived virtual source is usually limited to the frontal-horizontal plane. The reasons for this limitation, as guessed, are that, in static reproduction, the dynamic cues for front-back and vertical localisation are incorrect, and the high-frequency spectral cues are unstable with head movement. To validate this hypothesis, the variations of ITD (interaural time difference) caused by head turning in both static and dynamic transaural reproductions are analysed. The results indicate that dynamic reproduction is able to create appropriate low-frequency ITD variations, and the static transaural reproduction is unable to do so. Psychoacoustic experiments are conducted to compare virtual source localisation in static and dynamic reproductions. The results indicate that dynamic reproduction is able to recreate the front, back, and vertical virtual source for low-pass stimuli below 3 kHz, while for full audible bandwidth stimulus, appropriate low-frequency dynamic cue and unstable high-frequency spectral cues in dynamic reproduction result in two splitting virtual sources. Therefore, the results of present work prove the limitations of static transaural reproduction.

Keywords: transaural reproduction; vertical localisation; dynamic cue.

1. Introduction

Binaural pressures include the major information for auditory localisation. Using head related transfer function (HRTF)-based binaural synthesis, a binaural reproduction or virtual auditory display (VAD) duplicates the binaural pressures caused by an actual sound source and then recreates the perception of virtual source localisation in three dimensional space through headphone reproduction (Xie, 2013).

It is also desired to reproduce binaural signals by loudspeakers, especially by two frontal loudspeakers for simplicity. By combining HRTF-based binaural synthesis and crosstalk cancellation, transaural reproduction with two (or more) frontal loudspeakers is theoretically able to duplicate the binaural pressures of an actual sound source, and therefore seems able to recreate the virtual source in an arbitrary direction of three-dimensional space (Schroeder, Atal, 1963; Schroeder, 1970; Cooper, Bauck, 1989). The transaural method has been applied to some commercial "virtual surround sound" in which multichannel sounds are reproduced by using two actual frontal loudspeakers (Kawano *et al.*, 1998; Toh, Gan, 1999).

To evaluate the performance of transaural reproduction, some authors have reported that under a series of critical conditions (e.g., individualised HRTF processing, restrictions of head movement, reproduction in anechoic rooms, etc.), a pair of frontal loudspeakers was able to recreate perceived virtual sources in all horizontal or even three-dimensional directions for listeners to some extent (Damaske, 1971; Takeuchi *et al.*, 1998). However, more other experimental results indicated that in static transaural reproduction with two frontal loudspeakers (especially with non-individualised HRTF processing), perceived virtual source positions are usually limited in the frontal-horizontal quadrants regions (Nelson *et al.*, 1996; Gardner, 1997). The virtual sources intended for rear-horizontal quadrants or high elevations are often perceived in the frontal-horizontal quadrants with the same cone of confusion.

In the case of an actual sound source, both high-frequency spectral cues included in binaural pressures and the dynamic variations in binaural pressures caused by head turning contribute to front-back and vertical localisation (Blauert, 1997; Wallach, 1940). The contributions of these two cues depend on frequency. In addition, the information providing by dynamic and spectral cues is somewhat redundant. When one cue is eliminated, another cue alone still enables vertical localisation to some extent (Jiang *et al.*, 2019). In static transaural reproduction, transaural processing is fixed regardless of head turning. In this case, head turning during the reproduction results in variations in binaural pressures that are related to the positions of actual loudspeakers and unmatched with those of the target virtual source. Therefore the dynamic cue is incorrect (rather than eliminated completely). Moreover, the high-frequency spectral cues in loudspeaker reproduction are unstable with head movement due to the short wavelength at high frequency. Therefore, it is reasonable to guess that the aforementioned limitation of static transaural reproduction with two frontal loudspeakers is caused by the unmatched dynamic cue and unstable high-frequency spectral cue.

Gardner (1997) proposed dynamic transaural reproduction with two frontal loudspeakers and band-limited crosstalk cancellation up to 6 kHz, in which the transaural processing was updated according to the temporary position of listener's head. Experimental results for pink noise indicated that dynamic processing reduced the percentage of back-front confusion in localisation to 50.7%, as compared with a percentage of 91.4% for conventional static processing. However the 50.7% back-front confusion in dynamic processing is still high. Moreover, this experiment excluded the vertical localisation in dynamic transaural reproduction.

Kurabayashi *et al.* (2014) also conducted a psychoacoustic experiment to compare the localisation performance of static and dynamic transaural reproduction with four frontal loudspeakers. The results indicated that dynamic cue caused by head rotation reduced back-front confusion obviously. The experiment also indicated that the dynamic cue seems improve vertical localisation (as asserted by the authors of that experiment). However, the results of vertical localisation experiment exhibit great dispersion across subjects. Actually, the contributions of different cues to vertical localisation may depend on frequency or frequency spectra of the stimuli. Only the white noise stimulus was used in that experiment. In addition, that experiment is for the case of transaural reproduction with four frontal loudspeakers rather than conventional two frontal loudspeakers, while the latter is more common in practical uses.

Overall, the influence of dynamic cues, especially the low-frequency dynamic cue to vertical localisation in transaural reproduction with two frontal loudspeakers has not been evaluated completely. Further analysis and experiment is required to test the guess that the limitations of static transaural reproduction with two frontal loudspeakers are caused by the unmatched dynamic cue and unstable high-frequency spectral cue. For this purpose, the dynamic variation of the interaural time difference (ITD) caused by head turning in transaural reproduction with two frontal loudspeakers is analysed in the present work. Virtual source localisation experiments involving both static and dynamic transaural reproductions through two frontal loudspeakers are conducted. The analysis and experimental results validate the hypothesis.

2. Principle of transaural reproduction

Various methods can be used to derive the loudspeaker signals for transaural reproduction with two frontal loudspeakers in the horizontal plane. The conventional method cascades the HRTF-based binaural synthesis and crosstalk processing (Schroeder, Atal, 1963; Bauck, Cooper, 1996). Here, an alternative but mathematically equivalent method is outlined (Sakamoto *et al.*, 1981).

For convenience in assessing vertical or elevation localisation, interaural polar coordinates are used in the present study. As shown in Fig. 1, the origin of the coordinates is located at the centre of the head. The source position is specified by (r, Θ, Φ), where $0 \leq r < \infty$ denotes the source distance and $-90° \leq \Theta \leq 90°$ denotes the interaural polar azimuth, that is, the angle between the directional vector of the sound source and the median plane, with $\Theta = -90°$, $0°$, and $90°$ being the left direction, median plane, and right direction, respectively. A constant Θ represents the cone of confusion. Additionally $-90° \leq \Phi < 270°$ denotes the interaural polar elevation, that is, the angle between the projection of the directional vector of the source to the median plane and frontal axis, with $\Phi = -90°$,

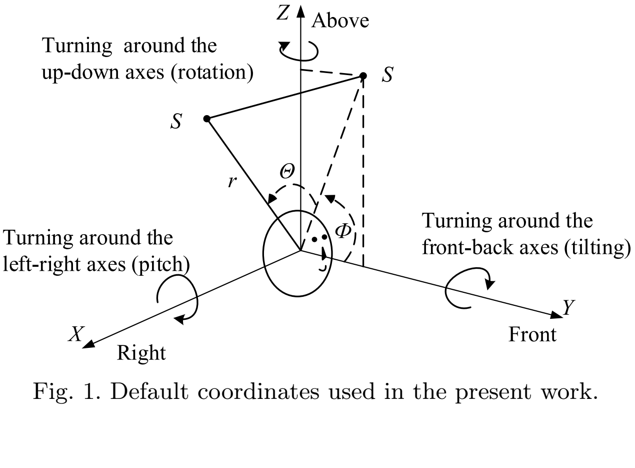
Z
Above
Turning around the up-down axes (rotation)
S
S
r
Θ
Φ
Turning around the left-right axes (pitch)
Turning around the front-back axes (tilting)
X
Right
Y
Front

Fig. 1. Default coordinates used in the present work.

0°, 90°, and 180° being the below, front, above, and back directions, respectively.

For an actual or target virtual source at a far-field distance ($r \geq 1.2$ m) in a given (Θ_S, Φ_S), the binaural pressures in the frequency domain can be calculated by filtering the input stimulus $E(f)$ with a pair of corresponding HRTFs $H_L(\Theta_S, \Phi_S, f)$ and $H_R(\Theta_S, \Phi_S, f)$

$$\begin{bmatrix} P_L \\ P_R \end{bmatrix} = \begin{bmatrix} H_L(\Theta_S, \Phi_S, f) \\ H_R(\Theta_S, \Phi_S, f) \end{bmatrix} E(f). \tag{1}$$

For transaural reproduction with two frontal loudspeakers, as shown in Fig. 2, two loudspeakers are arranged at a given far-field distance ($r \geq 1.2$ m) with specific directions (Θ_L, $\Phi_L = 0°$), (Θ_R, $\Phi_R = 0°$) respectively. Let $E_L(f)$ and $E_R(f)$ be the loudspeaker signals. $H_{LL}(\Theta_L, \Phi_L, f)$, $H_{RL}(\Theta_L, \Phi_L, f)$, $H_{LR}(\Theta_R, \Phi_R, f)$ and $H_{RR}(\Theta_R, \Phi_R, f)$ denote the four acoustic transfer functions from the two loudspeakers to two ears. The reproduced binaural sound pressures at the two ears are given as follows

$$\begin{bmatrix} P'_L \\ P'_R \end{bmatrix} = \begin{bmatrix} H_{LL}(\Theta_L, \Phi_L, f) & H_{LR}(\Theta_R, \Phi_R, f) \\ H_{RL}(\Theta_L, \Phi_L, f) & H_{RR}(\Theta_R, \Phi_R, f) \end{bmatrix} \begin{bmatrix} E'_L(f) \\ E'_R(f) \end{bmatrix}. \tag{2}$$

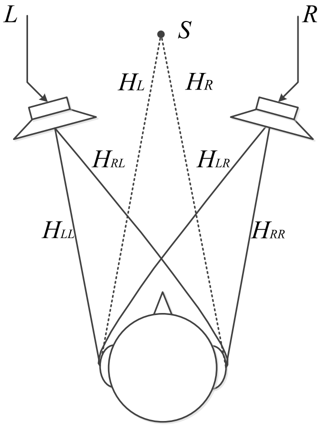

Fig. 2. Control of binaural pressures with a pair of loudspeakers.

Setting Eq. (2) equal to Eq. (1), that is, the binaural pressures in transaural reproduction are equal to those of the target, the loudspeaker signals can be found as follows:

$$\begin{bmatrix} E'_L(f) \\ E'_R(f) \end{bmatrix} = \begin{bmatrix} H_{LL}(\Theta_L, \Phi_L, f) & H_{LR}(\Theta_R, \Phi_R, f) \\ H_{RL}(\Theta_L, \Phi_L, f) & H_{RR}(\Theta_R, \Phi_R, f) \end{bmatrix}^{-1} \cdot \begin{bmatrix} H_L(\Theta_S, \Phi_S, f) \\ H_R(\Theta_S, \Phi_S, f) \end{bmatrix} E(f). \tag{3}$$

The second 2×1 matrix on the right side of Eq. (3) represents HRTF-based binaural synthesis. The inverse of the transfer matrix from the two loudspeakers to two ears represents the cross-talk cancellation matrix. If the transfer matrix from two loudspeakers to two ears is well-conditioned, the crosstalk cancellation matrix can be found from inverse manipulation. However, this is not always the case, especially at certain frequencies. In practice, a regularisation scheme is usually required to obtain the inverse of the transfer matrix (Kirkeby, Nelson, 1999; Ward, Elko, 1999; Papadopoulos, Nelson, 2010).

If the transfer matrix in Eq. (3) is invertible, Eq. (3) can be rewritten as

$$\begin{aligned} E'_L(f) &= G_L(\Theta_S, \Phi_S, f) E(f), \\ E'_R(f) &= G_R(\Theta_S, \Phi_S, f) E(f), \end{aligned} \tag{4}$$

where $G_L(\Theta_S, \Phi_S, f)$ and $G_R(\Theta_S, \Phi_S, f)$ are the responses of a pair of transaural filters which depend on the directions of the target source and loudspeakers with respect to the head,

$$\begin{aligned} G_L(b^{\circledast}) &= \frac{H_{RR}(a^{\circledast}) H_L(b^{\circledast}) - H_{LR}(a^{\circledast}) H_R(b^{\circledast})}{H_{LL}(c^{\circledast}) H_{RR}(a^{\circledast}) - H_{LR}(a^{\circledast}) H_{RL}(c^{\circledast})}, \\ G_R(b^{\circledast}) &= \frac{-H_{RL}(c^{\circledast}) H_L(b^{\circledast}) + H_{LL}(c^{\circledast}) H_R(b^{\circledast})}{H_{LL}(c^{\circledast}) H_{RR}(a^{\circledast}) - H_{LR}(a^{\circledast}) H_{RL}(c^{\circledast})}, \end{aligned} \tag{5}$$

where

$$a^{\circledast} = \Theta_R, \Phi_R, f,$$

$$b^{\circledast} = \Theta_S, \Phi_S, f,$$

$$c^{\circledast} = \Theta_L, \Phi_L, f.$$

Therefore, by filtering the input stimulus with a pair of transaural filters, transaural reproduction is able to control the binaural pressures caused by two loudspeakers so that they are equal to those caused by an actual source. This approach is the basic principle of transaural reproduction with two frontal loudspeakers.

In practice, an appropriate equalisation algorithm may be applied in transaural synthesis to reduce the perceived timbre colouration in reproduction. Timbre equalisation is based on the fact that in reproduction with two frontal loudspeakers, the perceived virtual source direction is dominated by interaural cues (especially ITD) and is limited to the frontal-horizontal quadrants. The interaural cues are controlled by the relative, rather than the absolute, magnitude, and phase of the left and right loudspeaker signals. Scaling both loudspeaker signals with identical frequency-dependent coefficients does not alter their relative magnitudes and phases and then the perceived virtual source azimuth. However, this manipulation alters the overall power spectra of the loudspeaker signals and therefore equalises the timbre. Of course, equalisation algorithms may alter the spectra of binaural pressures. There are various equalisation algorithms available (Hawksford, 2002). For constant power equalisation algorithms (Xie, 2013; Xie *et al.*, 2005), the responses of the transaural synthesis filters are equalised by their root mean square (RMS). That is, $G_L(\Theta_S,\Phi_S,f)$ and $G_R(\Theta_S,\Phi_S,f)$ in Eq. (4) are replaced by $G'_L(\Theta_S,\Phi_S,f)$ and $G'_R(\Theta_S,\Phi_S,f)$,

$$G'_L(\Theta_S,\Phi_S,f)=\frac{G_L(\Theta_S,\Phi_S,f)}{d^{\circledast}},\qquad G'_R(\Theta_S,\Phi_S,f)=\frac{G_R(\Theta_S,\Phi_S,f)}{d^{\circledast}}, \tag{6}$$

where

$$d^{\circledast}=\sqrt{|G_L(\Theta_S,\Phi_S,f)|^2+|G_R(\Theta_S,\Phi_S,f)|^2}$$

After equalisation, the loudspeaker signals given by Eq. (4) and Eq. (6) satisfy the following constant power spectral relationship,

$$|E'_L(f)|^2+|E'_R(f)|^2=|E(f)|^2. \tag{7}$$

Therefore, the overall power spectra of loudspeaker signals are equal to those of the input stimulus, thereby reducing reproduction colouration.

When the subject's head turns, the HRTFs from the target source and two loudspeakers to two ears change. In dynamic transaural reproduction, head turning is detected by a head tracker. According to the direction of the target virtual source relative to the temporary orientation of the subject's head, the HRTFs in the two transaural filters of Eq. (5) or (6) are updated constantly. The details of dynamic transaural reproduction are referred to Gardner (1997).

3. Analysis of dynamic localisation cues

To analyse the dynamic localisation cue, the ITD variation caused by head turning in transaural reproduction is analysed and compared with that of an actual source.

For an actual source in the direction (Θ_S, Φ_S), the binaural pressures are evaluated by Eq. (1). When the head turns, the binaural pressures are also evaluated by Eq. (1), but the HRTFs in the new direction with respect to the head are used.

For static transaural reproduction, the binaural pressures are evaluated by Eq. (2) with the loudspeaker signals given by Eqs (4) and (5) in the case without timbre equalisation, or by Eqs (4) and (6) in the case with timbre equalisation. When the head turns, the HRTFs from the two loudspeakers to the two ears should be replaced by those in new directions with respect to the head, but the loudspeaker signals are unchanged.

In dynamic transaural reproduction, the binaural pressures are evaluated in a manner similar to those in static reproduction. However when the head turns, both the HRTFs from the two loudspeakers to the two ears and the loudspeaker signals should be changed according to the new directions with respect to the head.

As shown in Fig. 1, the subject's head is able to turn with three degrees of freedom, including turning around the left-right axes (pitch), the front-back axes (tilting or rolling), and the up-down axes (rotation or yaw). The ITD variation caused by head rotation provides information for front-back and vertical localisation, and the ITD variation caused by head tilting also provides supplementary information for up-down discrimination (Wallach, 1940; Jiang *et al.*, 2019; Perrett, Noble, 1997). Therefore, the ITD variations caused by head rotation and tilting are evaluated.

There are various definitions and methods used for ITD calculation (Xie, 2013). Here, the ITDs are calculated by maximising the normalised cross-correlation function between the pressures at the two ears. In the frequency domain, the equations for ITD calculation are given by

$$\Psi(\tau)=\frac{\int P_L(f)\;P_R^*(f)\exp(j\,2\pi f\tau)\,\mathrm{d}f}{\left\{\left[\int|P_L(f)|^2\,\mathrm{d}f\right]\left[\int|P_R(f)|^2\,\mathrm{d}f\right]\right\}^{1/2}}, \tag{8}$$

and

$$\mathrm{ITD}=\tau_{\max}=\arg\max\Psi(\tau)\quad|\tau|\le 1\ \mathrm{ms}, \tag{9}$$

where the superscript "*" denotes complex conjugate. Because the ITD is an effective localisation cue at low frequency, the frequencies range for the integral in Eq. (8) is chosen up to 1.5 kHz.

The analysis scheme is as follows:

1) Calculate the binaural pressures for an actual source before and after head turning.
2) Calculate the ITD for an actual source before and after the head turning, and then evaluate the variation in the ITD.
3) Calculate the binaural pressures for transaural reproduction before and after head turning.
4) Calculate the ITD for transaural reproduction before and after head turning, and then evaluate the variation in the ITD.
5) Compare the ITD variations for the actual source and transaural reproduction.

As an example, suppose that two loudspeakers are arranged in the horizontal plane. The distance between the loudspeakers and the head centre of the subject is 1.5 m, and the directions are as follows:

$$\Theta_L=-15^\circ,\qquad \Theta_R=15^\circ,\qquad \Phi_L=\Phi_R=0^\circ. \tag{10}$$

The actual or target source is located in two vertical planes, including the median plane $\Theta=0^\circ$ and the sagittal plane $\Theta=45^\circ$. It should be noted that

Kirkeby *et al.* (1998) suggested a ±5° loudspeaker arrangement (stereo dipole) to improve the stability of virtual sources in transaural reproduction, but this arrangement is established at the cost of reducing low-frequency performance. Some later works suggested that a ±15° loudspeaker arrangement yields the best overall performance (Xie *et al.*, 2005; Lopez, Gonzalez, 2001). Therefore, a ±15° rather than conventional ±30° loudspeaker arrangement is chosen in this analysis and experimental work.

The HRTFs of a KEMAR artificial head (with DB-060/061 small pinnae but no torso) are used in the analysis. The HRTFs were obtained by first scanning the images of the KEMAR using a laser scanner and then performing calculations with the fast boundary element method (Gumerov, Duraiswami, 2009; Rui *et al.*, 2013). The sample rate of HRTFs is 44.1 kS/s and the length is 512 points.

Figure 3 plots the results of ITD before head turning for actual source and transaural reproduction. Figures 3a and 3b show the cases of actual or target source polar elevations in the median plane $\Theta = 0°$ and sagittal plane $\Theta = 45°$, respectively. Because the results for transaural reproduction with and without timbre equalisation are almost identical, Fig. 3 only plots the results for transaural reproduction with timbre equalisation. In addition, the ITDs for static and dynamic transaural reproduction are identical before head turning. Therefore, these ITDs are plotted as a curve and marked as "transaural". The ITDs for transaural reproduction effectively match these of the actual sources. In the median plane $\Theta = 0°$, as expected, the ITDs are approximately zero, and in the sagittal plane of $\Theta = 45°$, the ITDs vary from 439 μs to 533 μs.

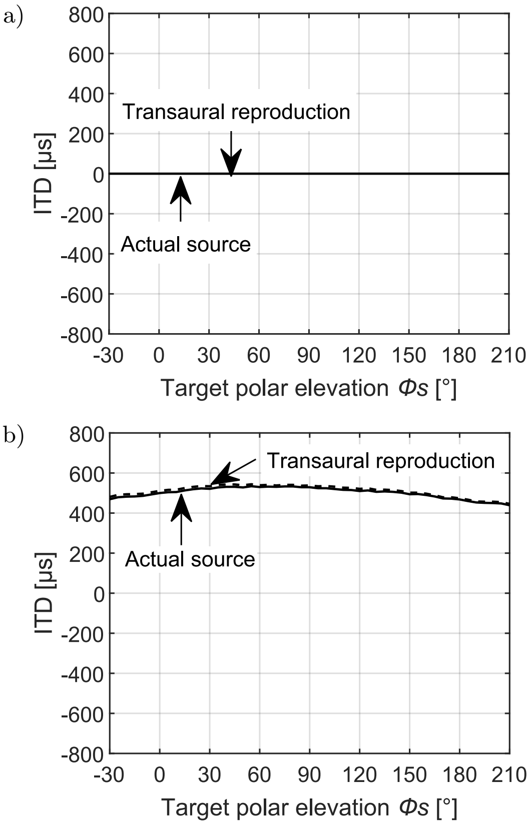

Fig. 3. ITD results for a target source located in the median plane $\Theta = 0°$ (a) and in the sagittal plane $\Theta = 45°$ (b).

The ITD changes after head turning. Because the ITD variations for transaural reproduction with and without timbre equalisation are almost identical, only the results for transaural reproduction with timbre equalisation are given. Moreover, the results for the actual source, static and dynamic transaural reproduction are plotted in the same figure.

Figure 4 shows the ITD variations (ΔITD) after head rotation to the left with an azimuth of 10°. Figure 4a shows ΔITD for the actual or target source in the median plane of $\Theta = 0°$. In the case of the actual source, ΔITD varies with elevation. In the frontal-median plane of $-30° \le \Phi_S < 90°$, ΔITD > 0; and in the backward-median plane, ΔITD < 0. Therefore, the sign of ΔITD provides information for front-back discrimination. In addition, the magnitude of ΔITD reaches a maximum in the frontal-horizontal and backward directions. As the source elevation departs from the horizontal plane, the magnitude of ΔITD decreases. At $\Phi_S = 90°$ (or $\Phi_S = -90°$, not shown in the figure), ΔITD

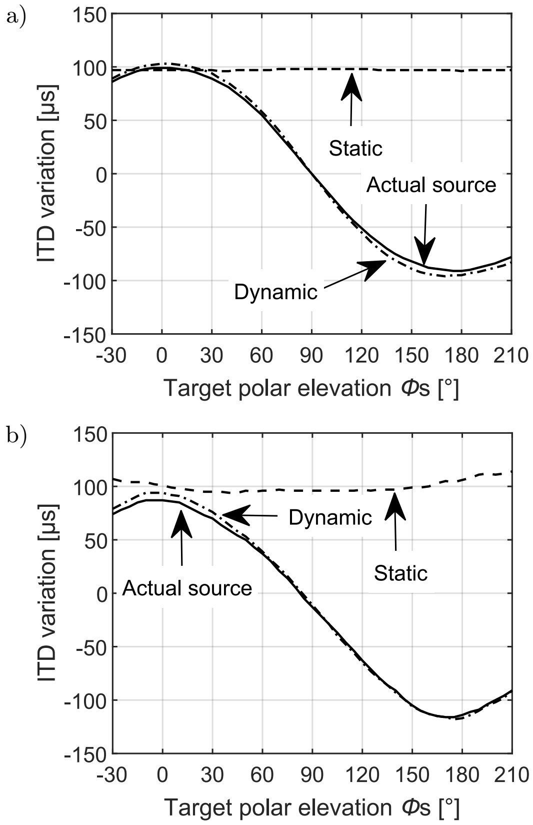

Fig. 4. ITD variation results for target source polar elevations with head rotation to the left with an azimuth of 10°in the median plane $\Theta = 0°$ (a) and the sagittal plane $\Theta = 45°$ (b).

is zero. Therefore, ΔITD with head rotation provides information for vertical displacement from the horizontal plane. However, ΔITD with head rotation is approximately up-down symmetric. For example, the ΔITD values for $\Phi_S = 30°$ and $-30°$ are almost identical. In the case of static transaural reproduction, the ΔITD values with head rotation are almost invariant to the target elevation (96 μs to 97 μs) and largely consistent with those of an actual source in the horizontal direction ($\Theta_S = 0°$, $\Phi_S = 0°$). In the case of dynamic transaural reproduction, ΔITD with head rotation is almost consistent with that of the actual source at various elevations.

Figure 4b shows ΔITD for the actual or target source in the sagittal plane of $\Theta = 45°$. Similar to the case in the median plane, in static transaural reproduction, the ΔITD values with head rotation vary from 96 μs to 114 μs, which are very close to those of an actual source in the near-horizontal direction ($\Theta_S = 0°$, $\Phi_S = 45°$). These variations are almost insignificant in auditory perception. In dynamic transaural reproduction, the ΔITD value with head rotation is basically consistent with that of the actual source at various polar elevations.

Figure 5 plots the ITD variations (ΔITD) after head tilting to the left with an azimuth of 10°. Figure 5a shows the actual or target source in the median plane $\Theta = 0°$. In the case of the actual source, ΔITD varies with elevation and is approximately up-down asymmetric. Therefore, the ITD variation caused by head tilting provides further information for up-down discrimination (of course, scattering by the torso also provides information for up-down discrimination (Kirkeby, Nelson, 1999). In the case of static reproduction, the ΔITD value with head tilting is almost invariant to the target elevation (0 μs) and basically consistent with that of an actual source in certain direction ($\Theta_S = 0°$, $\Phi_S = 0°$). In the case of dynamic transaural reproduction, the ΔITD value with head tilting is almost consistent with that of the actual source at various elevations. Figure 5b shows the results of actual or target source in the sagittal plane of $\Theta = 45°$. The results are similar to those in the median plane.

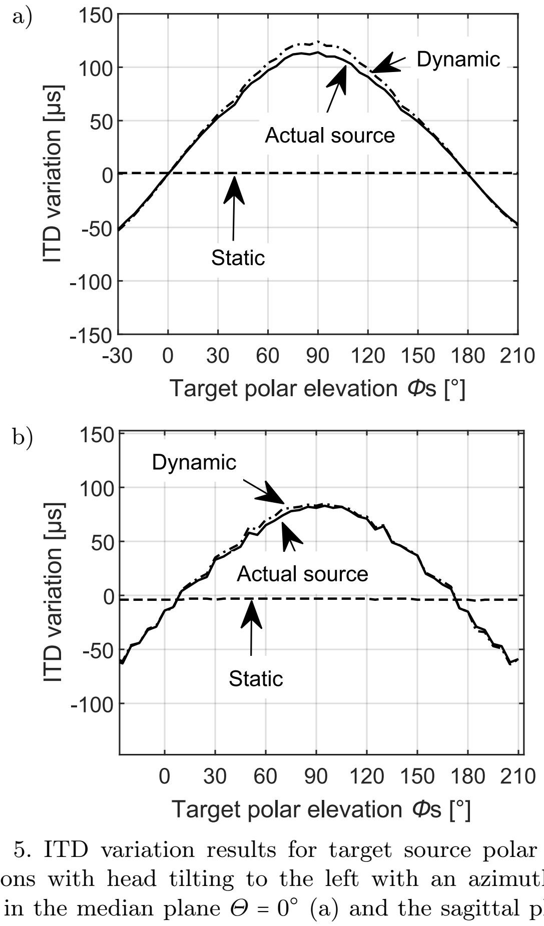

Fig. 5. ITD variation results for target source polar elevations with head tilting to the left with an azimuth of 10° in the median plane $\Theta = 0°$ (a) and the sagittal plane $\Theta = 45°$ (b).

For head rotation or tilting with other small azimuths (for example, 20°) or in other sagittal planes, the ITD variation exhibits a similar trend. That is, in the case of static transaural reproduction, the ITD variations with head turning are basically consistent with those of the actual source near the horizontal plane in the same cone of confusion. In the case of dynamic transaural reproduction, the ITD variations with head turning basically match those of an actual source in certain target directions.

Overall, the analysis on aforementioned cases indicates that dynamic transaural reproduction with two frontal loudspeakers provides ITD variations with head rotation and tilting that match with those of the target source. Therefore dynamic transaural reproduction is able to provide appropriate dynamic cue for front-back and vertical localisation. In contrast, the static transaural reproduction with two frontal loudspeakers is unable to do so. The ITD variations with head rotation and tilting in static reproduction always provide (sometimes misleading) localisation information that virtual source approximately locates in frontal-horizontal quadrants with the same cone of confusion as that of target source. This is consistent with some previous experimental results (Nelson *et al.*, 1996; Gardner, 1997).

4. Experimental method

A series of virtual source localisation experiments were conducted to examine the perceived virtual source directions of static and dynamic transaural reproduction with two frontal loudspeakers.

The experiment was conducted in a listening room with a reverberation time of 0.15 s. Static and dynamic transaural reproduction with timbre equalisation was examined. Two loudspeakers (GENELEC 8010AP-5) were arranged in the horizontal plane at a distance of 1.5 m and azimuths given by Eq. (10). Target virtual sources were located in two vertical planes, including the median plane $\Theta = 0°$ and sagittal plane $\Theta = 45°$. In each vertical plane, nine target elevations from $\Phi_S = -30°$ to 210° at an interval of 30° were chosen. The HRTFs used in transaural synthesis were identical to those in Sec. 3.

The aim of present work is to explore the role of dynamic cue caused by head turning on front-back and vertical localisation in transaural reproduction. Continuous stimuli with sufficient length were used in the experiment so that subjects had enough time to turn their head during the stimulus presentation. Two types of stimuli, pink noise with a full audible bandwidth and 3 kHz low-pass pink noises, were used. A FIR-based filter was used to create the low-pass stimuli. The cut-off frequency of pass-band was 3 kHz, the maximum attenuation in the pass-band is 1 dB. The cut-off frequency of stop-band is 3.3 kHz. The minimum attenuation in pass-band is −50 dB. The length of the stimuli was 10 s. The low-pass pink noise stimulus is for examining the influence of low-frequency dynamic cue to vertical localisation.

Signal processing was performed using a PC-based dynamic VAD with crosstalk cancellation. In dynamic reproduction, an electromagnetic head tracker (Polhemus FASTRAK) was used to detect the orientation of the subject's head. This tracker was able to detect head turning with three degrees of freedom. The update rate and system latency time of dynamic reproduction were 60 Hz and 25.4 ms, respectively.

The subjects judged the perceived virtual source direction and reported this direction using an electromagnetic tracker (Polhemus FASTRAK). The tracker included two receivers. One receiver was fixed on the subject's head surface to monitor the position and orientation of the head. Another receiver was fixed at one end of a 1.0 m wooden rod. The subject pointed the rod at the position of the perceived virtual source and a computer recorded the result. The direction of the virtual source was measured relative to the head centre because the data relative to the receiver on the head surface were transformed into relative to the head centre. The subjects made the judgment and pointed at the perceived direction during the stimulus presentation. In addition, the subjects were told that two virtual sources may be perceived during the reproduction with full audible bandwidth stimulus, and when this phenomenon occurs, the directions of the two virtual sound sources need to be reported separately. Before and after a presentation, the subjects were allowed to turn their heads to recognise the direction.

For static reproduction, the subject's heads were restricted during the judgment. The data from the head tracker indicated that the angle of head rotation and tilting was less than 2°. In dynamic reproduction, the subjects were encouraged to turn their heads. The angle of head rotation ranged from ±5° to ±15°, and the angle of head tilting ranged from ±10° to ±25°.

Eight subjects participated in the experiment. The subjects were from 22 to 30 years old and had normal hearing. For each condition and target direction, each subject repeatedly judged three times. Therefore, there were 3 repetitions × 8 subjects = 24 judgments under each condition. Statistical analysis was applied to the 24 judgments.

5. Statistical method and experimental results

5.1. Statistical method

The Kruskal-Wallis H test at a significant level of $\alpha = 0.05$ was used for the homogeneity tests. The results showed that there were no significant differences for all the tests, i.e. the localisation results for all of the subjects and repetitions were consistent and therefore reliable and stable. The preliminary results from the subjects indicated that, for static reproduction with the stimuli of two bandwidths and for dynamic reproduction with 3 kHz low-pass pink noise, a single virtual source was perceived. In this case, statistical analysis was applied to the perceived directions of this virtual source. However, for dynamic reproduction with pink noise of a full audible bandwidth, two splitting virtual sources were perceived. The high-frequency one was always perceived near the horizontal plane in the $\Theta \approx 0°$ or $\Theta \approx 45°$ direction. The low-frequency one was perceived at various elevations in the median plane $\Theta = 0°$ or in the sagittal plane $\Theta = 45°$ depending on the target elevation and target azimuth. In this case, subjects judged the perceived directions of two virtual sources separately, and statistical analysis was individually applied to the perceived direction of each virtual source.

If the overall tendency of the localisation results from all subjects and repetitions was similar except some reversal errors, the mean unsigned polar azimuth error and the mean unsigned polar elevation error were calculated to evaluate the overall localisation performance:

$$\begin{aligned} \Delta\Theta &= \frac{1}{N}\sum_{n=1}^{N} |\Theta_I(n) - \Theta_S(n)|, \\ \Delta\Phi &= \frac{1}{N}\sum_{n=1}^{N} |\Phi_I(n) - \Phi_S(n)|, \end{aligned} \tag{11}$$

where $\Theta_S(n)$ is the target polar azimuth of the n-th judgment, $\Theta_I(n)$ is the perceived polar azimuth of the n-th judgment, $\Phi_S(n)$ is the target polar elevation of the n-th judgment, $\Phi_I(n)$ is the perceived or reported polar elevation of the n-th judgment, and N is the total number of judgments. The mean was calculated across 3 repetitions, 8 subjects, and all target elevations in each target sagittal plane. Prior to calculating the mean unsigned error, the judged directions for the cases with reversal errors (the front-back and up-down confusion) were resolved.

5.2. Preliminary statistics on experimental results

Table 1 lists the mean unsigned polar azimuth error and the mean unsigned polar elevation error. As shown, in the median plane, the perceived virtual source was located near the median plane with mean unsigned polar azimuth errors less than 5.5° in all cases; in the sagittal plane, the perceived virtual source was located near the sagittal plane $\Theta = 45°$, with mean unsigned polar azimuth errors less than 13.1° in all cases. In addition, high-frequency virtual source of dynamic transaural reproduction exhibited almost the same unsigned polar elevation errors as those of static reproduction. However, low-frequency virtual source of dynamic transaural reproduction exhibited smaller unsigned polar elevation errors than those of static reproduction.

A multi-way ANOVA (with $\alpha = 0.05$) is conducted to check the effect of vertical plane ($\Theta = 0°$ and 45°), splitting virtual sources (high-frequency and low-frequency), reproducing manner (static and dynamic), bandwidth of stimuli (full and low pass filtered), on unsigned polar elevation error and unsigned polar azimuth error. The interactions between reproduction manners and bandwidth of stimulus are also checked. The results are listed in Table 2.

The results of the multi-way ANOVA indicated that the effect of vertical planes on unsigned polar elevation error was insignificant, unlike the effect of vertical planes on unsigned azimuth error which was significant. In addition, concerned with both unsigned polar elevation error and unsigned azimuth error, the difference between the high-frequency virtual source of dynamic reproduction with stimulus of full audible bandwidth and static reproduction of stimuli with two bandwidths was insignificant; the difference between the low-frequency virtual source of dynamic reproduction with full audible bandwidth stimuli and dynamic reproduction with low pass-filtered stimuli was insignificant either. However, the differences between static reproduction of stimuli with two bandwidths and the low-frequency virtual source of dynamic reproduction with full audible bandwidth stimulus were significant.

Table 1. Mean and standard deviation of the unsigned errors.

Reproduction manner	Stimulus bandwidths	Splitting into two virtual sources	Vertical plane Θ [°]	Mean/standard deviation $\Delta\Theta$ [°]	Mean/standard deviation $\Delta\Phi$ [°]
Static	Low-pass	No	0	2.9/0.2	33.2/26.0
		No	45	10.6/0.8	33.6/25.4
	Full band	No	0	2.7/0.4	33.7/25.9
		No	45	10.6/0.9	32.7/25.1
Dynamic	Low-pass	No	0	5.5/1.6	22.1/9.6
		No	45	11.5/3.3	20.4/7.7
	Full band	low-frequency	0	5.3/1.7	23.3/10.6
		high-frequency		1.9/0.2	33.5/25.9
		low-frequency	45	13.1/3.3	21.9/7.7
		high-frequency		10.4/0.6	32.8/25.6

Table 2. The results of the multi-way ANOVA on the experimental results.

			Significance (α = 0.05)	
Conditions			Unsigned polar elevation error	Unsigned polar azimuth error
Vertical plane Θ	$\Theta = 0°$ *vs* $\Theta = 45°$		No (0.498)	Yes (0.000)
Splitting sources	High-frequency *vs* low-frequency		Yes (0.000)	Yes (0.000)
Reproduction manner	Low-pass	Static *vs* dynamic	Yes (0.000)	Yes (0.001)
	Full band	Static *vs* dynamic of low-frequency	Yes (0.000)	Yes (0.000)
		Static *vs* dynamic of high-frequency	No (0.975)	No (0.301)
Bandwidths	Static	Low-pass *vs* full band	No (0.928)	No (0.789)
	Dynamic	Low-pass *vs* full band of low-frequency	No (0.575)	No (0.215)
		Low-pass *vs* full band of high-frequency	Yes (0.000)	Yes (0.000)
Reproduction manner × bandwidths	Static low-pass *vs* dynamic full of low-frequency		Yes (0.000)	Yes (0.000)
	Static low-pass *vs* dynamic full of high-frequency		No (0.903)	No (0.176)
	Static full *vs* dynamic low-pass		Yes (0.000)	Yes (0.000)

5.3. Vertical localization results for the median plane of Θ = 0°

To explore the localisation performance at different target elevations in detail, Fig. 6 shows the scatterplots of perceived elevations from all subjects and repetitions at the median plane of Θ = 0° and for:

a) static reproduction of low-pass stimulus;
b) static reproduction of full audible bandwidth stimulus;
c) dynamic reproduction of low-pass stimulus;
d) low-frequency virtual source for dynamic reproduction of full audible bandwidth stimulus;
e) high-frequency virtual source for dynamic reproduction of full audible bandwidth stimulus.

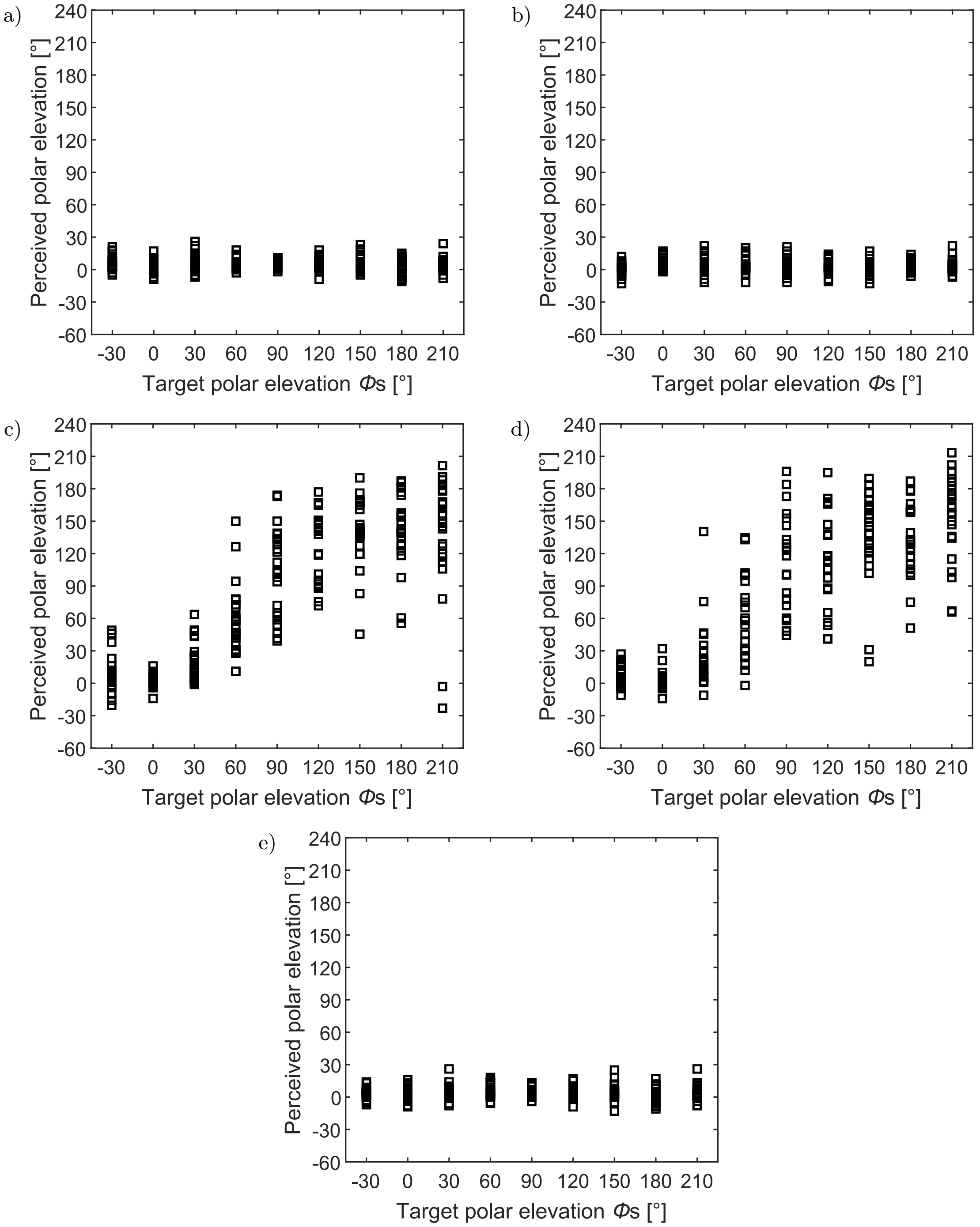

Fig. 6. Scatter plots of perceived polar elevations in the median plane of Θ = 0°: a) static reproduction of low-pass stimulus, b) static reproduction of full audible bandwidth stimulus, c) dynamic reproduction of low-pass stimulus, d) low-frequency virtual source for dynamic reproduction of full audible bandwidth stimulus, e) high-frequency virtual source for dynamic reproduction of full audible bandwidth stimulus.

It is observed from Fig. 6 that:

1) For static reproduction, the results of the low-pass stimulus and full audible bandwidth stimulus are similar. A single virtual source is perceived. The perceived virtual sources always locate near the frontal-horizontal direction of ($\Theta = 0°$, $\Phi = 0°$), in spite of the target virtual source elevation.
2) For dynamic reproduction of the low-pass stimulus, a single virtual source is perceived near the median plane. The perceived virtual source elevations roughly follow the target elevation, although a few front-back confusions may occasionally occur and lager up-down confusions occur for the cases of target elevations at $\Phi_S = -30°$ and 210°.
3) For dynamic reproduction of full audible stimulus, two splitting virtual sources are perceived near the median plane. The low-frequency virtual sources locate at the elevations similar to above (2). However, the high-frequency virtual sources locate similar to above (1). That is, the high-frequency virtual sources always locate near the horizontal direction of ($\Theta = 0°$, $\Phi = 0°$), in spite of the target virtual source elevation.

To see the results more clearly, Fig. 7 plots the mean perceived polar elevation and standard deviation corresponding to various cases in Fig. 6. The means were calculated across the 24 judgments for 3 repetitions and 8 subjects at each elevation in the median plane. Reversal was also resolved for the raw localisation results with front-back (F-B) and up-down (U-D) confusion. Table 3 lists the percentage of confusion for each target virtual source directions.

Figures 7a and 7b show the results of static reproduction for stimuli with two bandwidths, respectively. The results for the stimuli with two different bandwidths are similar. For target source at a front median plane and in the top direction with $-30° \leq \Phi_S \leq 90°$, the mean perceived direction is near $\Phi_I = 0°$, and Table 3 indicates that no front-back confusion occurs. For a target source at the back median plane, with $120° \leq \Phi_S \leq 210°$, Figs 7a and 7b indicate that the mean perceived direction is near $\Phi_I = 180°$. However, Table 3 indicates that the percentage of front-back confusion is 100%. Therefore, for a target source at $-30° \leq \Phi_S \leq 210°$, the actual perceived directions are all at the horizontal plane with $\Phi_I = 0°$, in spite of the target direction.

Figure 7c shows the results of dynamic reproduction with the low-pass stimulus. The mean perceived elevations largely match with the target elevations. Table 3 indicates that no or little front-back and up-down confusions occur in most cases. Two exceptions are for the cases of low-elevation target directions of $\Phi_S = -30°$ and 210°, at which the up-down confusions reach 79.2% and 75%.

Figure 7d shows the results of a low-frequency virtual source for dynamic reproduction with the full

Table 3. Percentage of confusion for each target with virtual source reproduction.

Reproduction manner	Bandwidths	Confusion [%]	Vertical plane Θ [°]	−30	0	30	60	90	120	150	180	210
Static	Low-pass	front-back	0	0	0	0	0	No	100	100	100	100
			45	0	0	0	0	No	100	100	100	100
		up-down	0	No	No	No	No	No	No	No	No	No
			45	No	No	No	No	No	No	No	No	No
	Full band	front-back	0	0	0	0	0	No	100	100	100	100
			45	0	0	0	0	No	100	100	100	100
		up-down	0	No	No	No	No	No	No	No	No	No
			45	No	No	No	No	No	No	No	No	No
Dynamic	Low-pass	front-back	0	0	0	0	16.7	No	16.7	8.3	8.3	12.5
			45	0	0	4.2	12.5	No	16.7	8.3	0	16.7
		up-down	0	79.2	No	4.2	0	0	0	4.2	No	75
			45	62.5	No	8.3	0	0	0	0	No	79.2
	Low-frequency virtual source for full band stimulus	front-back	0	0	0	4.2	25	No	29.1	8.3	8.3	8.3
			45	0	0	4.2	4.2	No	12.5	4.2	12.5	20.8
		up-down	0	70.8	No	4.2	4.2	8.3	4.2	8.3	No	70.8
			45	45.8	No	4.2	0	0	0	0	No	83.3
	High-frequency virtual source for full band stimulus	front-back	0	0	0	0	0	No	100	100	100	100
			45	0	0	0	0	No	100	100	100	100
		up-down	0	No	No	No	No	No	No	No	No	No
			45	No	No	No	No	No	No	No	No	No

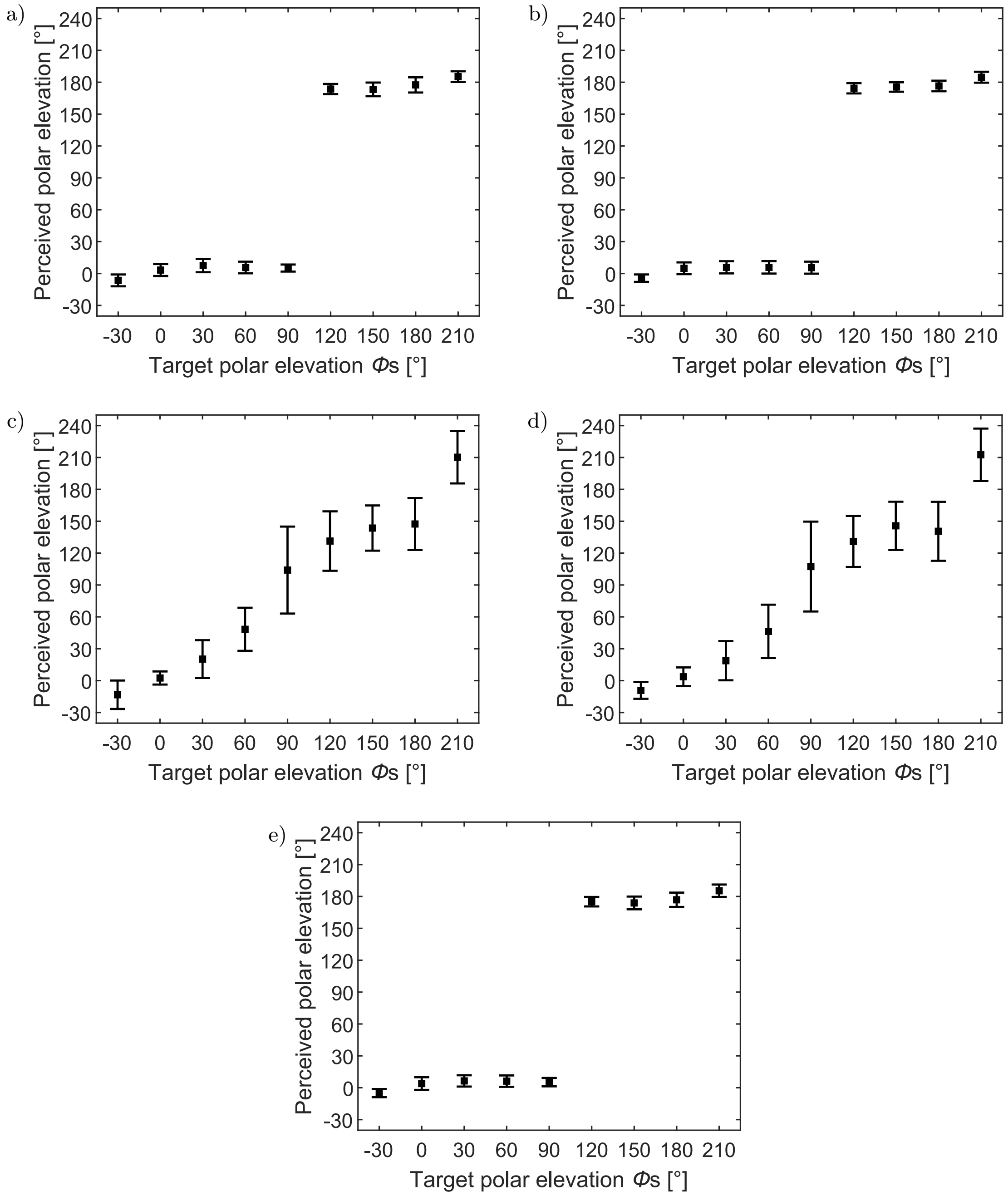

Fig. 7. Mean and standard deviation of perceived polar elevations in the median plane of Θ = 0°. Symbols of square are mean experimental results, respectively, the error bars are corresponding standard deviations: a) static reproduction of the low-pass stimulus, b) static reproduction of the full bandwidth stimulus, c) dynamic reproduction of the low-pass stimulus, d) low frequency virtual source for dynamic reproduction of the full band stimulus, e) high-frequency virtual source for dynamic reproduction of the full band stimulus.

audible stimulus, which are similar to the results of dynamic reproduction with the low-pass stimulus in Fig. 7c. The mean perceived elevations largely match the target elevations. Table 3 indicates that no or little front-back and up-down confusions occur in most cases. Two exceptions are for the case of target directions of Φ_S = 60° and 120°, at which the front-back confusion reaches 29.1%. Other exceptions are for the cases of low-elevation target directions of $\Phi_S = -30°$ and 210°, at which the up-down confusions reach 70.8%.

Figure 7e shows the results of a high-frequency virtual source for dynamic reproduction with the full audible stimulus, which are similar to the results of static reproduction for stimuli with two bandwidths in Figs 7a and 7b. For a target source at $-30° \le \Phi_S \le 210°$, the actual perceived directions are all near the frontal-horizontal directions of ($\Theta = 0°$, $\Phi = 0°$), in spite of the target direction.

It is observed that, for dynamic reproduction with full audible bandwidth stimulus, the percentages of front-back confusion for the low-frequency virtual sources at the target directions of $\Phi_S = 60°$ and 120° are a little bit higher than those at other target directions. This may be due to the fact that the magnitudes of ITD variation caused by head rotation are small in these directions (see Fig. 4a), which may cause errors in perceived directions easily.

It is also observed that, for dynamic reproduction with the low-pass stimulus and full audible stimulus, large up-down confusions occur for the low-frequency virtual sources at the target directions of low-elevation

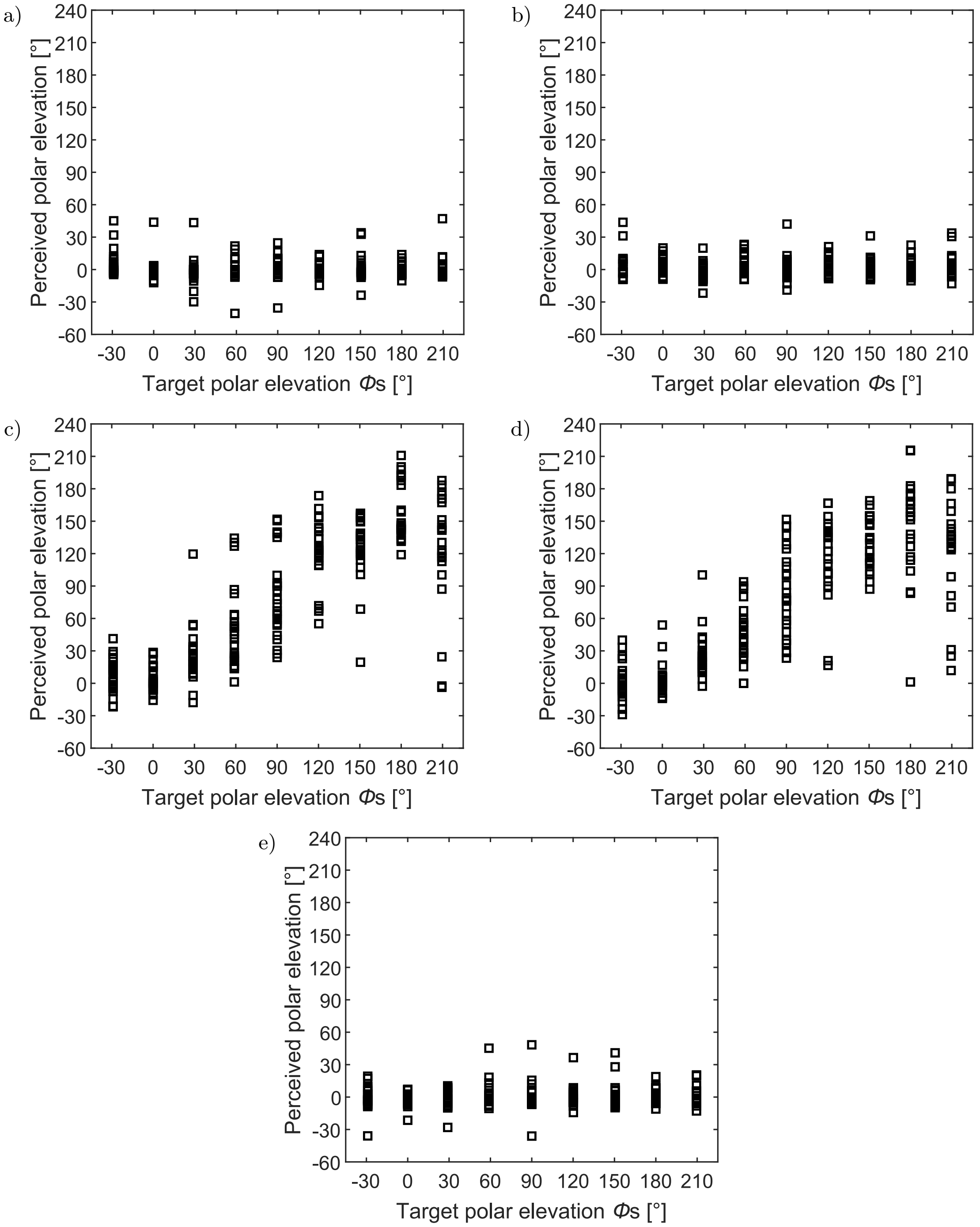

Fig. 8. Scatter plots of perceived polar elevations in the sagittal plane of $\Theta = 45°$: a) static reproduction of the low-pass stimulus, b) static reproduction of the full bandwidth stimulus, c) dynamic reproduction of the low-pass stimulus, d) low frequency virtual source for dynamic reproduction of the full band stimulus, e) high-frequency virtual source for dynamic reproduction of the full band stimulus.

of Φ_S = −30° and 210°. This may be due to the HRTFs of KEMAR without torso used in the present experiment. The torso-related spectral cues at low frequencies contribute to up-down discrimination, although head tilting also provides supplementary information for up-down discrimination (Perrett, Noble, 1997).

5.4. Vertical localisation results for the sagittal plane of Θ = 45°

Figure 8 shows the scatterplots of perceived elevations from all subjects and repetitions in the sagittal plane of Θ = 45° and for: a) static reproduction of the low-pass stimulus; b) static reproduction of the full audible bandwidth stimulus; c) dynamic reproduction of the low-pass stimulus; d) low-frequency virtual source for dynamic reproduction of the full audible bandwidth stimulus; e) high-frequency virtual source for dynamic reproduction of the full audible bandwidth stimulus.

Figure 9 plots the mean perceived polar elevation and standard deviation corresponding to various cases in Fig. 8. Reversal was resolved for the raw localisation results with front-back (F-B) and up-down (U-D) confusion. Table 3 also lists the percentages of the front-back and up-down confusion. Overall, the results for

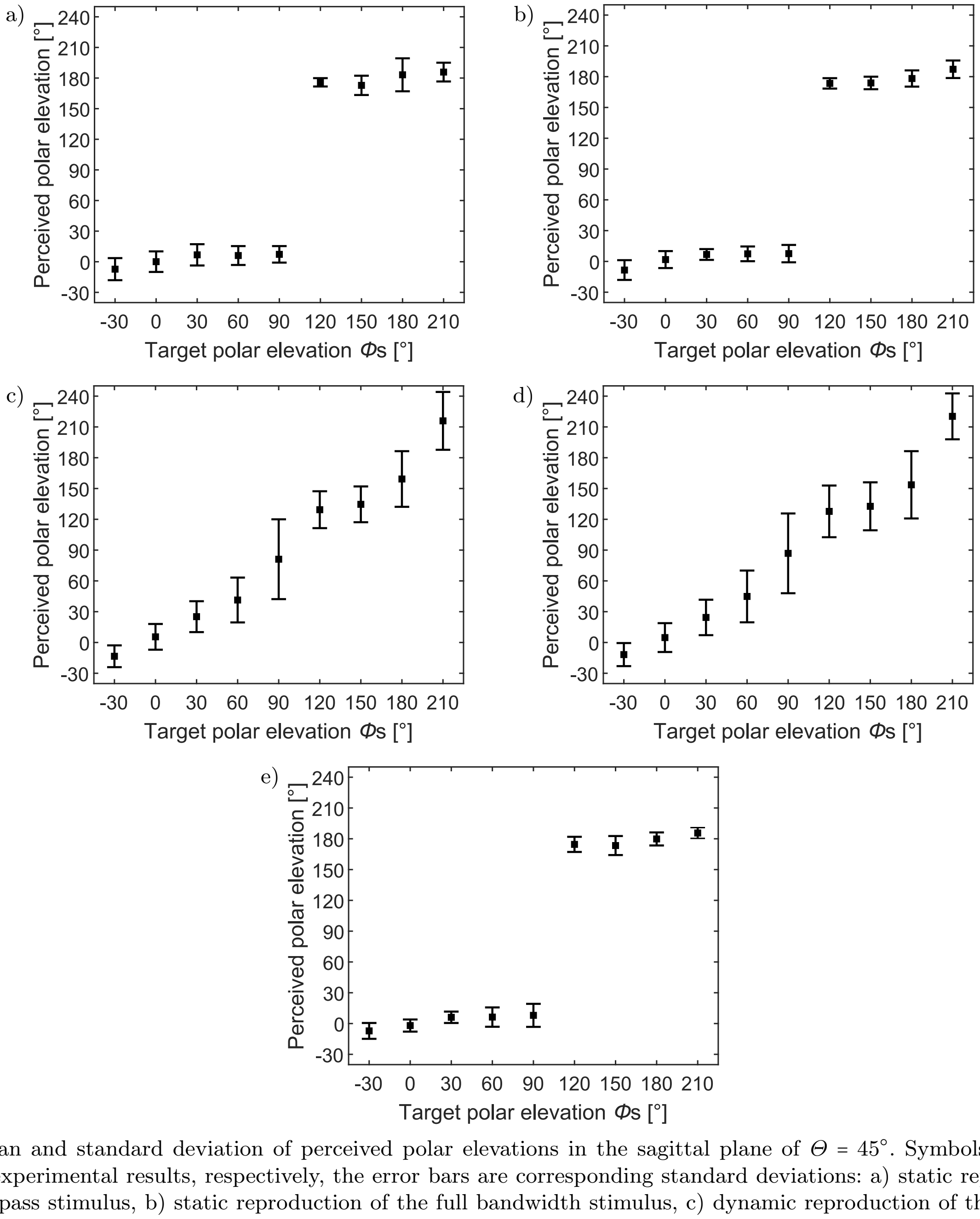

Fig. 9. Mean and standard deviation of perceived polar elevations in the sagittal plane of Θ = 45°. Symbols of square are mean experimental results, respectively, the error bars are corresponding standard deviations: a) static reproduction of the low-pass stimulus, b) static reproduction of the full bandwidth stimulus, c) dynamic reproduction of the low-pass stimulus, d) low frequency virtual source for dynamic reproduction of the full band stimulus, e) high-frequency virtual source for dynamic reproduction of the full band stimulus.

the sagittal plane of Θ = 45° are similar to those for the median plane of Θ = 0° and can be summarised as follows:

1) For static reproduction, the results of the low-pass stimulus and full audible bandwidth stimulus are similar. A single virtual source is perceived. The perceived virtual sources always locate near the horizontal directions of (Θ = 45°, Φ = 0°), in spite of the target virtual source elevation.
2) For dynamic reproduction of the low-pass stimulus, a single virtual source is perceived near the azimuth of Θ = 45°. The perceived virtual source elevations roughly follow the target elevation, although a few front-back confusions may occasionally occur. In addition, lager up-down confusions occur for the cases of target elevations at $\Phi_S = -30°$ and 210°.
3) For dynamic reproduction of the full audible stimulus, two splitting virtual sources are perceived near the azimuth of Θ = 45°. The low-frequency virtual sources locate at the elevations similar to above (2). However, the high-frequency virtual sources locate similar to above (1). That is, they always locate near the horizontal direction of (Θ = 45°, Φ = 0°), in spite of the target virtual source elevation.

6. Discussion

The experimental results in Sec. 5 indicate that static transaural reproduction with two horizontal frontal loudspeakers is unable to recreate a virtual source behind a subject and at high or low elevations in the median plane of Θ = 0° and the sagittal plane of Θ = 45°. The target virtual sources at all elevations in the median plane of Θ = 0° are perceived near the horizontal frontal direction (Θ = 0°, Φ = 0°), and the target virtual sources at all elevations in the sagittal plane of Θ = 45° are perceived near the horizontal direction (Θ = 45°, Φ = 0°).

In contrast, dynamic transaural reproduction with two horizontal frontal loudspeakers is able to recreate virtual sources behind a subject or at various high elevations in the median plane of Θ = 0° and the sagittal plane of Θ = 45°, at least for low-frequency stimuli below 3 kHz. Under the situation that the subjects are informed before the listening tests that there may be two virtual sound images, for the full-audible bandwidth stimulus, dynamic transaural reproduction creates two splitting virtual sources. The perceived directions of the low-frequency virtual source are similar to those of the low-pass stimulus, while the directions of the high-frequency virtual source basically locate at horizontal directions of Θ = 0° or Θ = 45°. In other words, dynamic transaural reproduction is still unable to recreate high-frequency virtual sources behind a subject at high or low elevations in the median plane of Θ = 0° and the sagittal plane of Θ = 45°.

By analysing the experimental results of static and dynamic transaural reproduction, it is possible to explore the origin of limitations in static transaural reproduction with two frontal loudspeakers. Both the spectral and dynamic cues contribute to front-back and vertical localisation. The spectral cue introduced by the pinnae is effective at high frequencies above 5 kHz and is individual dependent (Xie, 2013). A carefully designed transaural reproduction method could theoretically be possible to create a high-frequency spectral cue at the ideal listening position. Due to the short wavelength at high frequencies, however, the spectral cue is very sensitive to small deviations in the listening position and other errors in the reproduction chain such as unmatched HRTFs and loudspeaker's position. Even if dynamic HRTF-based binaural synthesis and cross-talk cancellation is included in dynamic transaural reproduction, it is still difficult to accurately reproduce the fine high-frequency spectral cues. In other words, both static and dynamic transaural reproductions are unable to create stable high-frequency spectral cues. If timbre equalisation is applied, the spectral cue is further distorted. Previous research (Nelson *et al.*, 1996; Gardner, 1997) has shown that even if there is no timbre equalisation in transaural reproduction, the front-back and vertical localisation cannot be achieved. Therefore, the front-back and vertical localisation in transaural reproduction rely on dynamic cues.

Wallach (1940) hypothesised that the ITD variations caused by head rotation provide information for identifying the front-back location and vertical displacement from the horizontal plane. Wallach's hypothesis has been validated by some modern experiments (Jiang *et al.*, 2019; Perrett, Noble, 1997). As shown in Sec. 3, in the case of static reproduction, the information provided by the low-frequency ITD variations caused by head rotation is consistent with that of a source near the frontal-horizontal direction. The inappropriate information provided by low-frequency ITD variations along with unstable (inappropriate) high-frequency spectral cue makes the perceived virtual source locate in the frontal-horizontal plane, the horizontal azimuth of the perceived virtual source is determined by the ITD itself. In the case of dynamic reproduction, the information provided by the low-frequency ITD variations matches with those of a target or actual source. Therefore, it is able to create low-frequency virtual source in front, back, and vertical directions. For the full audible bandwidth stimuli, of course, appropriate low-frequency ITD variations and unstable or incorrect high-frequency spectral cue in dynamic reproduction will cause two splitting virtual sources.

The previous psychoacoustic experiment by Gardner (1997) indicated that compared with static transaural reproduction with two frontal loudspeakers, dynamic transaural reproduction moderately reduces the percentages of back-front confusion for the pink noise stimulus and target virtual source at horizontal-back. The results in present work are basically consistent with those of Gardener's. However, vertical localisation was not included in Gardner's experiment.

Kurabayashi *et al.* (2014) conducted similar experiments which included both front-back and vertical localisation. The results reconfirmed that dynamic cues reduced back-front confusion obviously. However, the results of vertical localisation for the full audible bandwidth stimulus in that experiment exhibit great dispersion across subjects. In addition, no splitting perceived virtual sources were reported for the full audible bandwidth stimulus and localisation of the low-pass stimulus was not evaluated separately. Moreover, there are basically differences between the present experiment and the experiment of Kurabayashi *et al.* (2014).

1) The experiment of Kurabayashi *et al.* (2014) was based on transaural reproduction with four frontal loudspeakers. The present experiment focuses on transaural reproduction with two frontal loudspeakers, which is much common in practical uses. The high-frequency spectra cues in binaural pressures are unstable in transaural reproduction. The variations of high-frequency spectra cues are different for transaural reproduction with two and four loudspeakers.
2) White noise stimulus was used in the experiment of Kurabayashi *et al.* (2014). White noise stimulus includes more high frequency components than it is in the case of the pink noise which was used in the present experiment; this makes the localisation rely more on the spectral cue.
3) The system latency time of dynamic transaural reproduction in the experiment of Kurabayashi *et al.* (2014) was 100 ms, while the system latency time of the instrument in the present work is 25.4 ms. A lager system latency time may influence the use of dynamic cue for localisation (Sandvad, 1996).

7. Conclusions

Transaural reproduction is unable to provide stable high-frequency spectral cues, therefore front-back and vertical localisation relies on dynamic cues. Dynamic transaural reproduction with two frontal-horizontal loudspeakers is able to provide the correct low-frequency dynamic cues and then create virtual sources behind a subject and at various elevations below 3 kHz. Static reproduction is unable to do so, and the perceived virtual source is usually limited to the frontal-horizontal plane. In dynamic transaural reproduction with full audible bandwidth stimuli, appropriate low-frequency ITD variations and unstable or incorrect high-frequency spectral cue will cause two splitting virtual sources. Therefore, the present analysis and experiment validate the hypothesis of the reasons for the limitations of conventional static transaural reproduction with two frontal loudspeakers.

Recently, multichannel sound with height is developed rapidly (Herre *et al.*, 2015; ITU-R Report BS.2159-7, 2015). In some cases, multichannel sound signals must be downmixed for reproduction with few loudspeakers. Transaural processing has been suggested for downmixing. The limitations discussed in the present work should be considered if a downmixing scheme is designed.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (11674105) and State Key Lab of Subtropical Building Science, South China University of Technology. The funders had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

References

1. Bauck J., Cooper D.H. (1996), Generalized transaural stereo and applications, *Journal of the Audio Engineering Society*, **44**(9): 683–705.
2. Blauert J. (1997), Spatial hearing, [in:] *The Psychophysics of Human Sound Localization*, MIT Press, Cambridge.
3. Cooper D.H., Bauck J.L. (1989), Prospects for transaural recording, *Journal of the Audio Engineering Society*, **37**(1/2): 3–19.
4. Damaske P. (1971), Head-related two-channel stereophony with loudspeaker reproduction, *The Journal of the Acoustical Society of America*, **50**(4B): 1109–1115, doi: 10.1121/1.1912742.
5. Gardner W.G. (1997), *3-D audio using loudspeakers*, Ph.D. Thesis, University of Massachusetts Institute of Technology.
6. Gumerov N.A., Duraiswami R. (2009), A broadband fast multipole accelerated boundary element method for the three dimensional Helmholtz equation, *The Journal of the Acoustical Society of America*, **125**(1): 191–205, doi: 10.1121/1.3021297.
7. Hawksford M.J. (2002), Scalable multichannel coding with HRTF enhancement for DVD and virtual sound systems, *Journal of the Audio Engineering Society*, **50**(11): 894–913.
8. Herre J., Hilpert J., Kuntz A., Plogsties J. (2015), MPEG-H audio – The new standard for coding of immersive spatial audio, *IEEE Journal of Se-*

lected Topics in Signal Processing, **9**(5): 770–779, doi: 10.1109/JSTSP.2015.2411578.

9. ITU-R Report BS.2159-7 (2015), *Multichannel sound technology in home and broadcasting applications*, International Telecommunication Union, Geneva.

10. Jiang J.L., Xie B.S., Mai H.M., Liu L.L., Yi K.L., Zhang C.Y. (2019), The role of dynamic cue in auditory vertical localisation, *Applied Acoustics*, **146**: 398–408, doi: 10.1016/j.apacoust.2018.12.002.

11. Kawano S., Taira M., Matsudaira M., Abe Y. (1998), Development of the virtual sound algorithm, *IEEE Transactions on Consumer Electronics*, **44**(3): 1189–1194, doi: 10.1109/30.713254.

12. Kirkeby O., Nelson P.A., Hamada H. (1998), The 'stereo dipole': A virtual source imaging system using two closely spaced loudspeakers, *Journal of the Audio Engineering Society*, **46**(5): 387–395.

13. Kirkeby O., Nelson P.A. (1999), Digital filter design for inversion problems in sound reproduction, *Journal of the Audio Engineering Society*, **47**(7/8): 583–595.

14. Kurabayashi H., Otani M., Itoh K., Hashimoto M., Kayama M. (2014), Sound image localization using dynamic transaural reproduction with noncontact head tracking, *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, **97**(9): 1849–1858, doi: 10.1587/transfun.E97.A.1849.

15. Lopez J.J., Gonzalez A. (2001), Experimental evaluation of cross-talk cancellation regarding loudspeakers' angle of listening, *IEEE Signal Processing Letters*, **8**(1): 13–15, doi: 10.1109/97.889637.

16. Nelson P.A., Orduna-Bustamante F., Engler D., Hamada H. (1996), Experiments on a system for the synthesis of virtual acoustic sources, *Journal of the Audio Engineering Society*, **44**(11): 990–1007.

17. Papadopoulos T., Nelson P.A. (2010), Choice of inverse filter design parameters in virtual acoustic imaging systems, *Journal of the Audio Engineering Society*, **58**(1/2): 22–35.

18. Perrett S., Noble W. (1997), The effect of head rotations on vertical plane sound localization, *The Journal of the Acoustical Society of America*, **102**(4): 2325–2332, doi: 10.1121/1.419642.

19. Rui Y.Q., Yu G.Z., Xie B.S., Liu Y. (2013), Calculation of individualized near-field head-related transfer function database using boundary element method, *Audio Engineering Society 134th Convention*, Paper No. 8901, Rome.

20. Sakamoto N., Gotoh T., Kogure T., Shimbo M., Clegg A.H. (1981), Controlling sound-image localization in stereophonic reproduction, *Journal of the Audio Engineering Society*, **29**(11): 794–799.

21. Sandvad J. (1996), Dynamic aspects of auditory virtual environments, *Audio Engineering Society 100th Convention*, Preprint No. 4226, Copenhagen.

22. Schroeder M.R., Atal B.S. (1963), Computer simulation of sound transmission in rooms, *Proceedings of the IEEE*, **51**(3): 536–537, doi: 10.1109/PROC.1963.2180.

23. Schroeder M.R. (1970), Digital simulation of sound transmission in reverberant spaces, *The Journal of the Acoustical Society of America*, **47**(2A): 424–431, doi: 10.1121/1.1911541.

24. Takeuchi T., Nelson P.A., Kirkeby O., Hamada H. (1998), Influence of individual head-related transfer function on the performance of virtual acoustic imaging systems, *Audio Engineering Society 104th Convention*, Tokyo.

25. Toh C.W., Gan W.S. (1999), A real-time virtual surround sound system with bass enhancement, *Audio Engineering Society 107th Convention*, Preprint No. 5052, New York.

26. Wallach H. (1940), The role of head movement and vestibular and visual cue in sound localization, *Journal of Experimental Psychology*, **27**(4): 339–368, doi: 10.1037/h0054629.

27. Ward D.B., Elko G.W. (1999), Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation, *IEEE Signal Processing Letters*, **6**(5): 106–108, doi: 10.1109/97.755428.

28. Xie B.S. (2013), *Head-Related Transfer Function and Virtual Auditory Display*, J. Ross Publishing – USA.

29. Xie B.S., Shi Y., Xie Z.W., Guan S. (2005), Virtual reproducing system for 5.1 channel surround sound, *Journal of South China University of Technology*, **24**(1): 76–88.