Exploring the Limits of Virtual Source Localization with Amplitude Panning on a Flat Panel with Actuator Array: Implications for Future Research

ZIYING YU,[1, a)] QIAOXI ZHU,[2] MING WU,[1] and JUN YANG[1, b)]

[1)] *Key Laboratory of Noise and Vibration Research, Institute of Acoustics, Chinese Academy of Sciences, Beijing, 100190, China*

[2)] *Centre for Audio, Acoustics and Vibration, Faculty of Engineering and IT, University of Technology Sydney, Sydney, NSW 2007, Australia*

(Dated: 9 August 2023)

This paper is part of a special issue on 3D Sound Reconstruction for Virtual Auditory Displays: Applications in Buildings.

Immersive and spatial sound reproduction has been widely studied using loudspeaker arrays. However, flat-panel loudspeakers that utilize thin flat panels with force actuators are a promising alternative to traditional coaxial loudspeakers for practical applications, with benefits in low-visual profiles and diffuse radiation. Literature has addressed flat-panel loudspeakers' sound quality and applications in 3D sound reproduction, such as wave field synthesis and sound zones. This paper revisits the spatial sound perception of flat-panel loudspeakers, specifically the localization mismatch between the perceived and desired sound directions when using amplitude panning. Subjective tests in an anechoic chamber with twenty-four subjects result in the mean azimuth direction mismatch within $\pm 6.0°$ and the mean elevation mismatch within $\pm 10.0°$. The experimental results show that the virtual source created by amplitude panning over a flat-panel loudspeaker still achieves spatial localization accuracy close to that of a real sound source, despite not using complex algorithms or acoustic transfer function information. The findings of this study establish a benchmark for virtual source localization in spatial sound reproduction using flat-panel loudspeakers, which can serve as a starting point for future research and optimization of algorithms.

[a] yuziying@mail.ioa.ac.cn

[b] jyang@mail.ioa.ac.cn; Also at: School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing, 100049, China

I. INTRODUCTION

Extensive research is dedicated to using loudspeaker arrays to reproduce spatial sound for creating immersive and realistic listening experiences[1]. For example, recreate the auditory sense of space[2] and the localization of perceived sound sources[3–5], essential for various applications such as augmented or mixed reality (AR/MR), multimedia content creation[1,4,6], and personalized sound zones[7,8]. However, developing this technology from laboratory prototypes to real-world settings, especially in complex acoustic environments like buildings, requires further exploration. Previous research has explored various approaches, including but not limited to equalization of room responses[9], optimization of robustness[6,10], optimization of loudspeaker placement[11,12], implementation simplification by reducing acoustics transfer function measurement[8,13] or using distributed systems[7]. However, loudspeakers are limited by the physical structure and spatial placement in sound reproduction. For example, coaxial loudspeakers can be impractical for certain applications due to their weight, cost, or other factors. Additionally, reproducing sound with sufficient spatial coverage requires loudspeakers to have an appropriate spatial span while maintaining a small enough spacing for controlling sound waves at high frequencies.

The flat-panel loudspeaker is a promising alternative to traditional coaxial loudspeakers with advantages in low-visual profile and wide sound dispersion[14]. It uses a thin, flat panel with force actuators at the rear side to generate acoustic radiation through the panel vibration. It can adapt to various indoor environments and even utilize existing displays, e.g., the organic light-emitting diode (OLED) screen, to generate spatialized audio[15]. On the other hand, the flat-panel loudspeaker, as the multi-actuator panel (MAP), has multiple exciters driven with different signals, respectively, using signal processing to allow dynamic control of the panel's spatial vibration profile with diffuse sound radiation characteristic[16]. This characteristic helps avoid the beaming properties of piston loudspeakers at high frequencies[17] and reduces modal excitation within rooms[18]. Compared to traditional loudspeakers, the sound quality could be challenging when applying the flat-panel loudspeaker. Though not the scope of this paper, existing literature has widely addressed flat-panel loudspeakers' sound quality improvement[19] and applications in 3D sound reproduction, such as wave field synthesis[14,20] and directional sound fields[21].

The main aim of sound reproduction is to provide listeners with a clear spatial perception by utilizing psychoacoustic cues that lead to perceptual satisfaction. However, the human auditory perception mechanism is complex, resulting in selective emphasis and suppression even under unfavorable conditions, such as the cocktail party effect. Therefore, considering subjective perception is crucial for evaluating, designing, and optimizing sound reproduction methods. The flat-panel loudspeaker has been perceptually evaluated, including loudness[22,23], sound localization, perception of sound distance by wave field synthesis[24], and sound quality enhancement[25]. Furthermore, it is compared with the electrodynamic loudspeaker on objective and subjective measures for wave field synthesis[26].

So far, vector-based amplitude panning (VBAP)[27] remains an effective and straightforward method to create virtual sound sources using traditional loudspeakers arbitrarily placed in space, with ongoing research and development[28,29]. VBAP has several practical advantages, including low computational complexity, no destructive interference within the sweet spot, superior timbral quality, and gradual sound quality degradation outside the sweet spot[29]. Moreover, it does not require precise information on the acoustic transfer function for implementation. This feature is essential for controlling flat-panel loudspeakers since their sound radiation is affected by several factors, such as material, boundary conditions, and coupling. While VBAP has been utilized with flat-panel loudspeakers[30], a complete and thorough evaluation of spatial sound panning with the flat-panel loudspeaker is still lacking[31].

This paper revisits the spatial sound perception of flat-panel loudspeakers, specifically the localization mismatch between the perceived and desired sound directions when using amplitude panning. An experiment involving subjective and objective tests utilized a vector-based amplitude panning algorithm to create eighty-one virtual sources with four actuators placed at the corners of a flat panel. Subjective listening tests included twenty-four normal-hearing subjects, while objective tests included results on the interaural time difference (ITD) and the interaural level differences (ILD) measured in an anechoic chamber. The study aims to determine the spatial localization accuracy of amplitude panning using a flat-panel loudspeaker. As amplitude panning does not rely on complex algorithms or acoustic transfer function information, the experiment results can be a benchmark for virtual source localization in spatial sound reproduction using flat-panel loudspeakers. This information can be useful for future research and optimization of algorithms in this area.

II. THEORY

A. Flat Panel with Actuator Array

The motion for a thin flat panel can be expressed as[31]

$$D\nabla^4 u\left(y,z,t\right) + \rho h \frac{\partial^2 u\left(y,z,t\right)}{\partial^2 t} = -f\left(y,z,t\right), \tag{1}$$

where $u\left(y,z,t\right)$ is the out-of-plane displacement of time t for point $\left(y,z\right)$ on the panel. The coordinate system is defined in Sec. III. $f\left(y,z,t\right)$ is the external forcing function applied to the panel, h and ρ represent the thickness and density of the panel, respectively, and D is the bending stiffness per unit width given by Young's modulus E and Poisson's ratio υ as[19]

$$D = \frac{Eh^3}{12\left(1-\upsilon^2\right)}. \tag{2}$$

The forced response of a rectangular panel with dimensions $L_y \times L_z \times h$ and simply supported edges can be expressed as a sum of modes of the panel's free response as

$$\begin{aligned} u\left(y,z,t\right) &= \sum_{r=1}^{\infty} \alpha_r \Phi_r\left(y,z\right) e^{j\omega_r t} \\ &= \sum_{r=1}^{\infty} \alpha_r \sin\left(\frac{m_r \pi}{L_y} y\right) \sin\left(\frac{n_r \pi}{L_z} z\right) e^{j\omega_r t}, \end{aligned} \tag{3}$$

where α_r is the amplitude of the mode $\Phi_r\left(y,z\right)$, ω_r is the resonant frequency of each mode, j denotes the imaginary unit as the square root of -1, m_r and n_r represent the number of sinusoidal half-wavelengths for each mode along the y and z axes, respectively.

Using the Rayleigh integral, the surface area S of the flat panel can be divided into several sub-regions ds, each of which is treated as a point source that radiates sound waves outward. The total acoustic response in space is the superposition of these point sources. The sound pressure at $\boldsymbol{r}$ with origin at the center of the panel is

$$p\left(\mathbf{r}\right)=\int_{S}\frac{-j\omega\rho_{0}\dot{u}\left(\boldsymbol{r}_{s}\right)\exp\left(-jkR\right)}{2\pi R}ds, \tag{4}$$

where $\dot{u}(\boldsymbol{r}_s)$ is the complex transverse velocity at any point $\boldsymbol{r}_s$ on the surface, $R=|\boldsymbol{r}-\boldsymbol{r}_s|$, ρ_0 is the density of air, and k is the wave number. The complex velocity $\dot{u}(\boldsymbol{r}_s)$ is determined by the first time derivative of the panel response.

Please note that the derivation presented here may need to be more precise for the near-field condition, where the sound pressure radiated by the source is considerably more complex and has intricate oscillatory features that cannot be accurately approximated. However, near-field scenarios can be quite demanding, such as when the user is situated within one meter of the display screen. Furthermore, the practical boundary conditions can differ from those assumed in the derivation, making it challenging to obtain an analytical solution.

Though the sound field produced by a flat-panel loudspeaker is highly dependent on frequency and the relative observation position to the sound source[21], it is found that each actuator element can vibrate independently without being affected by neighboring exciters and panel edges, as confirmed by laser Doppler vibrometer measurements[32]. It implies that individual exciters can be considered as independent sources for spatial sound reproduction. So that we can create a virtual sound using amplitude panning without accurate sound pressure value estimation or measurement.

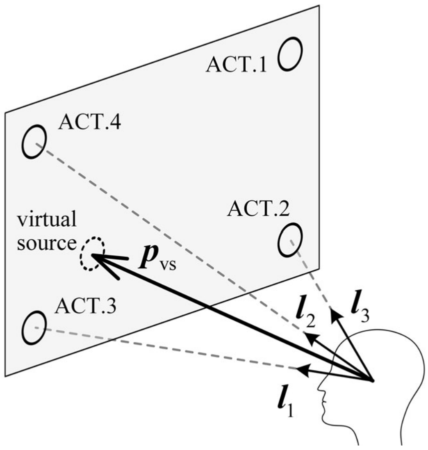


FIG. 1. Three-dimensional amplitude panning using a flat-panel loudspeaker with four actuators "ACT.1∼4". For example, actuators 1, 2, and 4 are the activated triplets to create a virtual sound source in the direction of $\boldsymbol{p}_{\mathrm{vs}}$. Unit vectors $\boldsymbol{l}_1$, $\boldsymbol{l}_2$, and $\boldsymbol{l}_3$ represent the directions from the listener to each actuator in Cartesian coordinates.

B. Amplitude Panning Using the Flat-Panel Loudspeaker

Figure 1 illustrates the vector-based amplitude panning of actuator triplets on a flat panel to create a virtual sound source. For a given virtual source direction, the three closest actuators are activated simultaneously with respective signal gains as a triplet[27]. The direction to the virtual source is defined as[33,34]

$$\boldsymbol{p}_{\mathrm{vs}} = \boldsymbol{g}\mathbf{L}_{123} = g_1\boldsymbol{l}_1 + g_2\boldsymbol{l}_2 + g_3\boldsymbol{l}_3, \tag{5}$$

where unit vectors $\boldsymbol{l}_1$, $\boldsymbol{l}_2$, and $\boldsymbol{l}_3$ represent the directions from the listener to each actuator in Cartesian coordinates, $\mathbf{L}_{123} = [\, \boldsymbol{l}_1 \;\; \boldsymbol{l}_2 \;\; \boldsymbol{l}_3 \,]$ and the normalized gains $\boldsymbol{g} = [\, g_1 \;\; g_2 \;\; g_3 \,]^{\mathrm{T}}$ is

$$\boldsymbol{g} = \boldsymbol{p}_{\mathrm{vs}}\mathbf{L}_{123}^{-1} / \left\| \boldsymbol{p}_{\mathrm{vs}}\mathbf{L}_{123}^{-1} \right\| , \tag{6}$$

where $(\cdot)^{\mathrm{T}}$ denotes matrix transposition, $\|\boldsymbol{g}\| = 1$, and the inverse matrix $\mathbf{L}_{123}^{-1}$ satisfies $\mathbf{L}_{123}^{-1}\mathbf{L}_{123} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. This work implemented the vector-based amplitude panning based on codes from Ref. 35 and Ref. 36.

VBAP includes a geometric determination of the triangle of active loudspeakers and an algebraic solution to compute the panning gains such that the velocity vector of the synthesized sound field matches the direction of the virtual source. Though it does not require acoustic transfer function information, the spatial localization accuracy generated by conventional loudspeakers with VBAP is within $\pm 8^\circ$ and $\pm 18^\circ$ in azimuth and elevation, respectively[34]. In comparison, perceiving a real sound source has the *mean* azimuth mismatch ranges from 1° to 3°, and the mean elevation mismatch in the median plane ranges from 4° for white noise to 17° for speech[3]. Conventional loudspeakers are discrete sound sources in terms of spatial distribution. On the other hand, a flat panel with multiple actuators is a continuous sound source, e.g. in Eq. (3). The sound received from the flat-panel loudspeaker is contingent upon the plate's vibration, i.e. $\dot{u}(\boldsymbol{r}_s)$ in Eq. (4). Since actuators have different gains but the same phase under VBAP, the higher $\dot{u}(\boldsymbol{r}_s)$ value aligns with the vicinity of the actuators. Thus, due to spatial masking, the perceived sound of VBAP using the flat panel may be comparable to that of conventional loudspeakers. The following section will present experimental characteristics of the virtual source localization with amplitude panning on a flat panel with actuators.

FIG. 2. (color online) Experimental setup using the KEMAR Head and Torso simulator at the listener location with a distance of 70 cm from the panel.

III. OBJECTIVE EVALUATION OF VIRTUAL SOURCE DIRECTION

The experiment was designed to objectively and subjectively evaluate the spatial sound perception of flat-panel loudspeakers, specifically the localization mismatch between the perceived and desired sound directions when using amplitude panning. As illustrated in Fig. 2, we consider the indoor displays scenario and the listener in the near-field with a distance $L = 70$ cm from the center of the subject's head O to the center of the panel. The flat panel is a 0.2 mm thick aluminum stencil of dimensions 60.5×57.5 cm with fixed edges. Virtual sound sources were created respectively at eighty-one locations within the square region of sizes 50.0×50.0 cm. The four actuators are at the vertexes $(L, -a, a)$, $(L, -a, -a)$, $(L, a, -a)$, and (L, a, a), respectively, with $a = 25$ cm, as shown in Fig. 3. Virtual sources are denoted as Sij, where $i = 1, 2, \ldots, 9$, and $j = 1, 2, \ldots, 9$, represent the row and column numbers, respectively. So the available azimuth and elevation ranges of virtual sound sources were within $\pm 19.65°$.

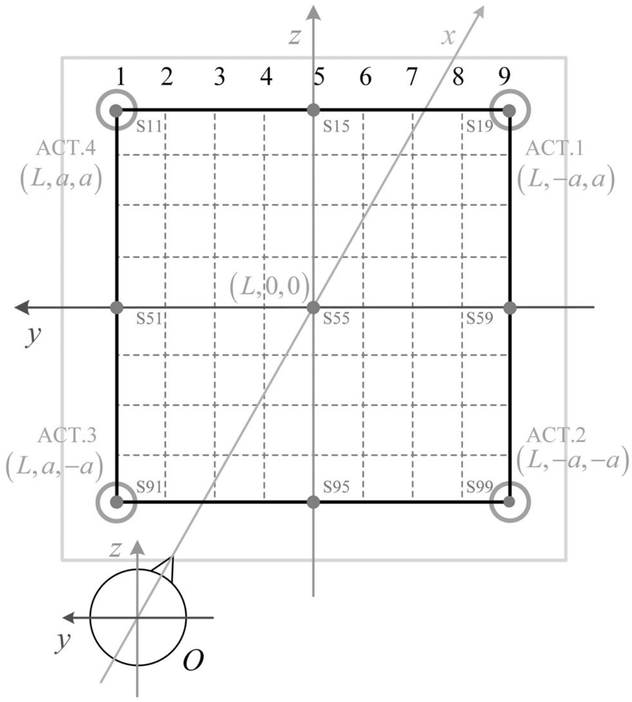


FIG. 3. (color online) The distance from the center of the listener's head to that of the panel was $L = 70$ cm. Four actuators ACT.1 4 were at the rear side of the panel with $a = 25$ cm. Eighty-one virtual sources Sij, $i = 1, 2, \ldots, 9$ and $j = 1, 2, \ldots, 9$, were within the square region (contoured in black) and the four actuators at the vertexes.

Stimuli were generated at a sampling rate of 48 kHz. VBAP gains were calculated based on codes from Ref. 35 and Ref. 36. Output signals were played as multi-channel .flac files, with a pink noise signal in each channel for the corresponding actuator. The computer was equipped with a Fireface UC audio interface for digital to analog conversion. Separate power amplifiers drove four actuators. Stepped sweep signals tested channel distortion to determine the effective volume range resulting in total harmonic distortion (THD) of less than 10%. All measurements were conducted in an anechoic chamber.

Equalization employed an inverse filter through linear predictive coding (LPC)[37] of the impulse response measured by a microphone at the origin O to have flat and uniform frequency responses from every actuator. Inter-channel calibration was also applied using Gaussian white noise pulses with a duration of 120 s as the test signal to minimize amplitude discrepancies among actuators. So the received frequency responses using different actuators were flattened and aligned over 100 Hz to 20 kHz.

The ITD and ILD are widely used as auditory cues for the localization of a single source in psychoacoustics[3,5]. ITD depends on the different durations that sounds travel towards two ears[38]. It is calculated based on the position of the interaural cross-correlation peak within a maximum interaural delay time of 1 ms for frequencies below 1.5 kHz that[39]

$$\mathrm{ITD}\left(\theta\right) = \arg\max\left\{\underset{\tau}{E}\left[s_L\left(t\right) s_R\left(t+\tau\right)\right]\right\}, \tag{7}$$

where $s_L\left(t\right)$ and $s_R\left(t\right)$ are the sound signals the left and right ear receives, respectively. ITD reflects the shading effect of a human head, that sound pressure degrades in the ear furthest away from the source and increases at the other, when the sound source deviates from the median plane[40]. ILD is defined as.

$$\mathrm{ILD}\left(x_s, y_s, z_s, f_0\right) = 20\lg\left|\frac{P_{\mathrm{R}}\left(x_s, y_s, z_s, f_0\right)}{P_{\mathrm{L}}\left(x_s, y_s, z_s, f_0\right)}\right|\left(\mathrm{dB}\right), \tag{8}$$

where $P_{\mathrm{L}}\left(x_s, y_s, z_s, f_0\right)$ and $P_{\mathrm{R}}\left(x_s, y_s, z_s, f_0\right)$ are left and right frequency-domain sound pressures at the ear canals generated by the sound source at location $\left(x_s, y_s, z_s\right)$ with frequency f_0, respectively.

In the experiment, ITD and ILD for different virtual sources are calculated from recordings using the GRAS 45BE KEMAR Head and Torso simulator and 5s pink noise as the test signal. Then, we obtained the corresponding perceptual virtual source directions by referencing a high-resolution lookup table with simulated ITD and ILD values based on the CIPIC database on head-related transfer functions[41] and interpolation.

Figure 4 presents the localization mismatch between the perceptual virtual source direction (obtained from the ITD measurement) and the desired sound direction for each virtual source. The tested values are smoothed for visualization. For virtual sources at most locations on the panel, the azimuth and elevation mismatch values are relatively small. A few azimuth mismatches of negative values but no worse than -8.0° appear in the lower right area ($y < 100,\ z < 0$), while a few elevation mismatches of positive values but no worse than 5.0° locate in the lower right area ($y < 0,\ z < 0$).

Figure 5 presents the localization mismatch between the perceptual virtual source direction (obtained from the ILD measurement) and the desired sound direction for each virtual source. The tested values are smoothed for visualization. The corresponding rounded frequencies are $f = 2500$ and 5000 Hz, respectively. Similar to Fig. 4 for virtual sources at most locations on the panel, the azimuth and elevation mismatch values are relatively small. ILD values are frequency dependent. Mismatch values associated with the lower-frequency at 2500 Hz range exhibit more significant deviations among different virtual source heights. At a frequency of 2500 Hz, the lower region near the center of the panel ($y < 100,\ z < -50$) exhibits larger azimuth localization mismatch values. Meanwhile, the consistency of different locations across various locations increases for elevation mismatch, and elevation mismatch

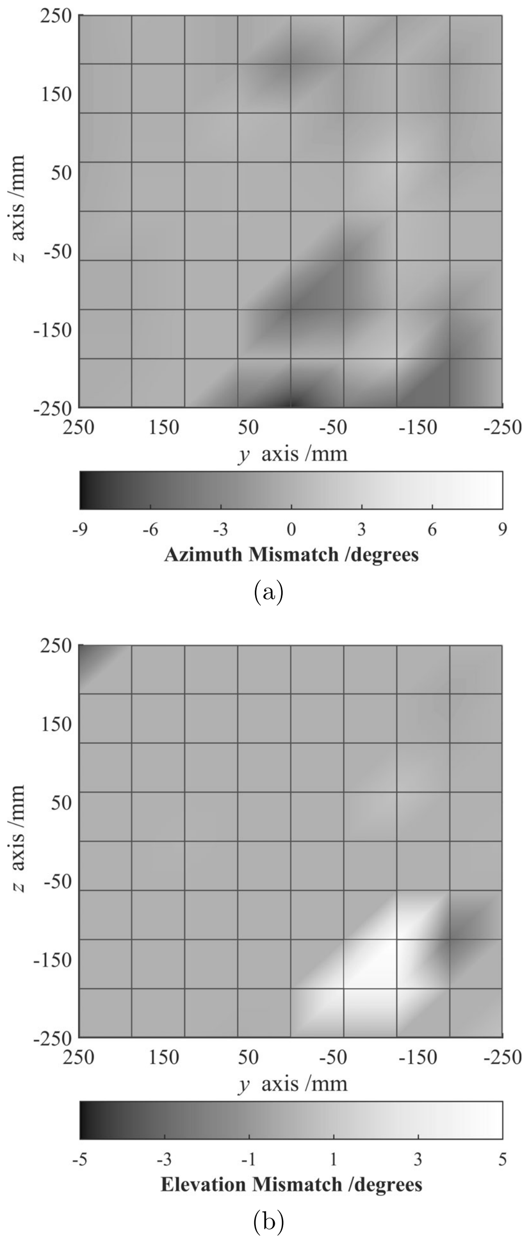


FIG. 4. (color online) Localization mismatch between the perceptual virtual source direction (obtained from the ITD measurement) and the desired sound direction for each virtual source in Fig. 3. (a) Azimuth mismatch and (b) elevation mismatch.

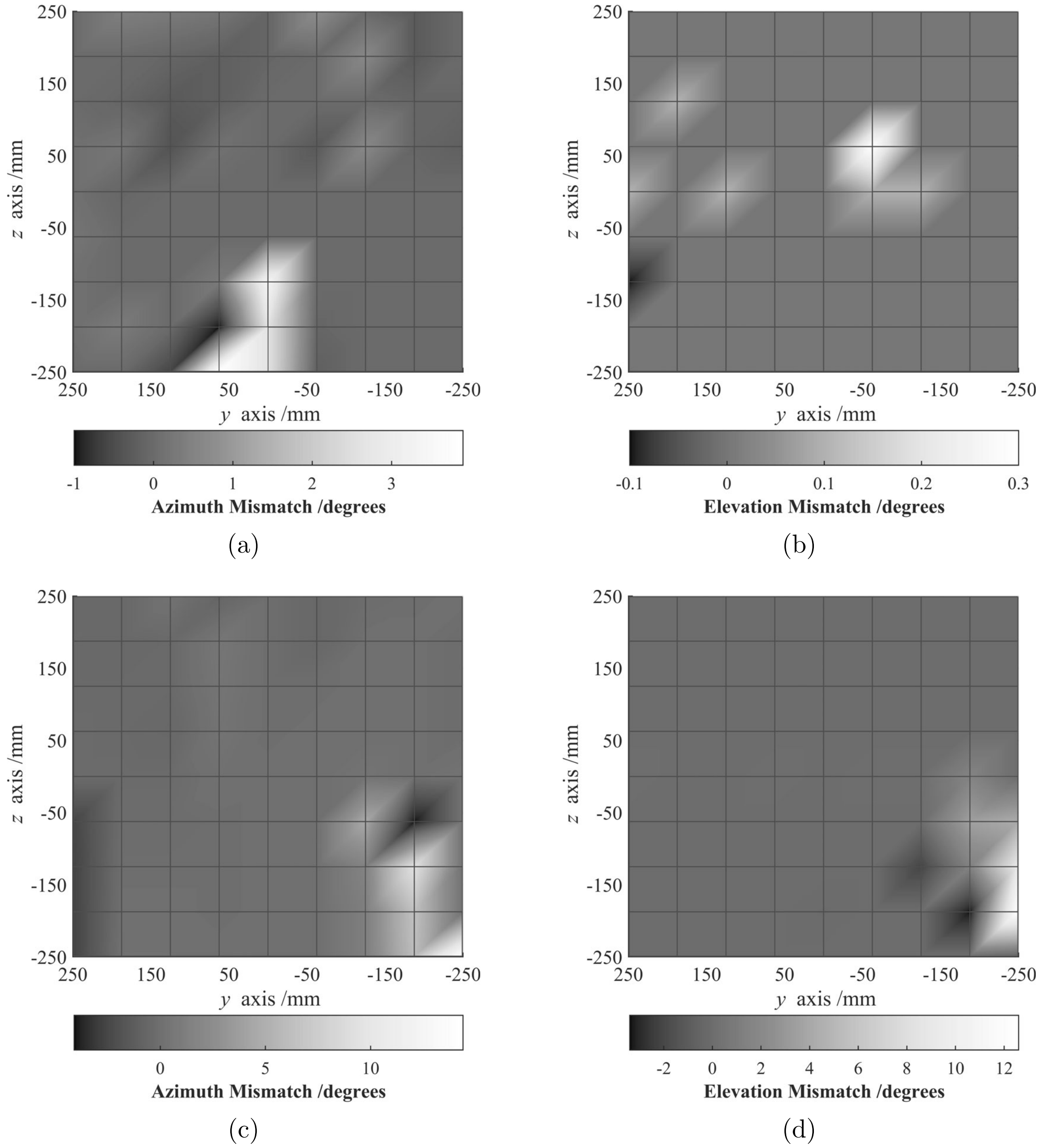


FIG. 5. (color online) Localization mismatch between the perceptual virtual source direction (obtained from the ILD measurement) and the desired sound direction for each virtual source in Fig. 3. (a) Azimuth mismatch at 2500 Hz, (b) elevation mismatch at 2500 Hz, (c) azimuth mismatch at 5000 Hz, and (d) elevation mismatch at 5000 Hz.

values fall within a range of 0.3. At a frequency of 5000 Hz, larger mismatch values occur when the virtual source is positioned on the panel's lower right area ($y < -50,\ z < 0$).

When the height of the sound source is level with the ear, the trend of flat-panel loudspeakers is more consistent with the positioning rules of traditional coaxial loudspeakers. However, when the sound source is lower than the ear level, ILD localization mismatch values become more pronounced.

IV. SUBJECTIVE EVALUATION OF VIRTUAL SOURCE DIRECTION

Due to the variations in principles and reproduction methods among different spatial sound techniques, there has yet to be an established standard for assessing spatial sounds, including evaluation criteria, methodologies, experimental conditions, and data processing. The International Telecommunication Union (ITU) has set standards for subjectively assessing spatial sound, which can be referenced in specific evaluations of spatial sound[42,43].

A. Setup

Twenty-four normal-hearing listeners aged 22 to 49 were involved in subjective tests. Amongst them, thirteen are male, and eleven are female. The subjects sat naturally on a chair in an anechoic chamber with a flat panel located 0.7 m in front of them. During the listening test, the subjects were instructed to maintain their head orientation toward the direction of the unknown perceptual virtual source. They were also asked to maintain a stable body position while slightly rotating their head to keep their eyes and ears at approximately the same height as the center of the panel.

A pre-study with three groups of training was conducted. Firstly, virtual sound sources of different azimuth angles in the same height were played from sequence left to right, namely 'S51', 'S55', and 'S59' in Fig. 3, respectively. Then, virtual sound sources of different elevation angles in the median plane were played sequentially from top to bottom, namely 'S15', 'S55', and 'S95'. Finally, the virtual sound sources in four corners were played, namely 'S11', 'S19', 'S99', and 'S91'. Finally, the virtual sound sources in four corners were played. During the pre-study test, subjects were instructed to figure out the perceived direction of the virtual source once heard before being told the correct direction. Then the subjects proceeded to the formal testing phase after a five-minute rest interval. An equalized VBAP pink noise sample with a duration of five seconds and a three-second pause between each pulse was used and presented randomly to simulate the perception of sound sources with varying locations.

During this experiment step, subjects were instructed to orient their head toward the unknown direction of the perceptual virtual sound source. A slight head rotation was recommended to reduce potential confusion. During the localization experiment, the selection of the reporting method is of utmost importance. It should demonstrate an accuracy level at least as high as the human localization accuracy, which is approximately one degree for frontal sound incidence. Therefore, we opted for absolute evaluation over auditory comparison and discrimination experiments. Once the direction of the virtual source was determined, subjects were instructed to point a laser pointer in that direction. After confirming the direction, the subjects were asked to hit the controller's button connected to the phone with Bluetooth. Meanwhile, a mobile phone positioned behind the subject was used to take a picture and record the localization of the laser mark. The laser pointer marks should be confined within the black frame indicated as the target region on the panel. The controlling computer was used to verify the results to avoid any omissions. If there was any error in operation, the subjects were allowed to revise their answers before the end of the audio playback. After a three-second rest, the subsequent trial began immediately. If the virtual source was found to be indeterminate, the subject was asked to point the laser marker to a random position. The listening test had a total length of fourteen minutes for each subject to avoid fatigue.

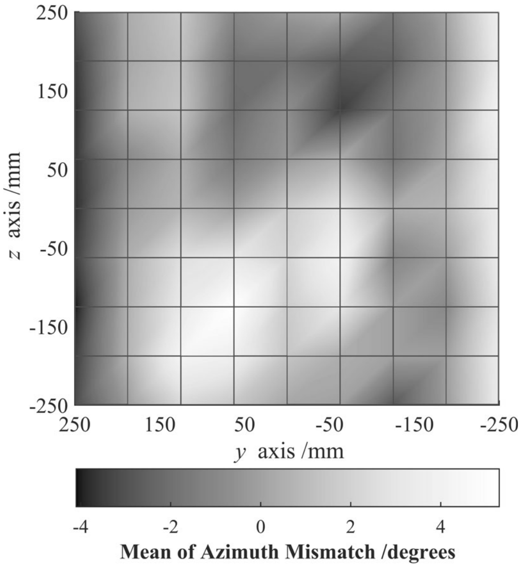
Azimuth and elevation angles can be achieved through the laser marks in the aforementioned mobile phone. Given that the mobile phone may have shifted slightly due to ground shaking caused by walking on the steel net while changing subjects in the anechoic chamber, we must carefully reposition the data points and select precise values for the coordinates of the four corner marks to correct any deviation.

B. Result

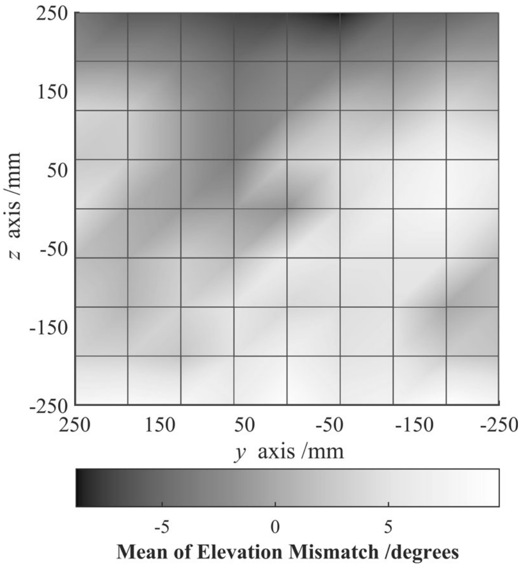
The mean and standard deviation of subjective localization mismatch are shown in Fig. 6 and Fig. 7. The mismatch of subjective experiments presents a more complex distribution than the results of objective experiments. Larger regions exhibit obvious mismatch values. When the virtual source is located on both sides of the panel, the perceptual azimuth angle demonstrates a trend generally consistent with traditional coaxial loudspeakers, indicating relatively low accuracy in localization on sides. However, the flat-panel loudspeaker localization accuracy for the center position is relatively lower than that of a coaxial loudspeaker, which has a 1° to 3° accuracy range[34]. When the virtual source is at the same height as the human ear, the accuracy of flat-panel loudspeakers' center localization azimuth angle is only moderate. This is in contrast to coaxial loudspeaker results, which showed an accuracy error about three times higher on both sides than the center. The localization accuracy of the upper part is slightly better than that of the lower part, indicating a relatively concentrated and fuzzy localization area in the lower part. The results presented here can be discussed from the standpoint that all the subjects are right-handed, so horizontal localization on the left does not perform as well as on the right.

From previous work on VBAP localization of coaxial loudspeakers, elevation angle localization is only accurate when the virtual sound source and loudspeakers are at the same height. Localization is not precise in other situations for coaxial loudspeakers. Here, the localization phenomenon for flat-panel loudspeakers does not conform to the abovementioned rules. Localization on the right side is worse than on the left for horizontal localization. In comparison, localization on the bottom side is worse than on the top side for vertical localization.

For the entire panel, the mean of azimuth mismatch values is within ±6.0°, which is generally better than that of elevation mismatch within ±10.0°. This indicates that horizontal localization is more accurate than vertical localization. This result is consistent with previous research that suggests the human ear has a lower vertical resolution than horizontal resolution. Fig. 6 also shows that when the virtual source is located precisely on the edge of the target region, the mismatch values of azimuth and elevation angles are large. However, the judgment of virtual sources near the edge is much more precise. For azimuth angle, if the



(a)



(b)

FIG. 6. (color online) Localization mismatch between the perceptual virtual source direction (averaged among 24 subjects in the listening test) and the desired sound direction for each virtual source in Fig. 3. (a) Azimuth mismatch and (b) elevation mismatch.

virtual sound source is positioned near the border frame but not directly on edge on either side of the subject ($y = \pm 187.5$, as the cross line in the second column from the left and right marked as numbers 2 and 8 in Fig. 3), the mismatch values are relatively small within $\pm 1.0°$. But when virtual source is located on the left and right borders $y = \pm 250$, mean mismatch values are rather large within $\pm 6.0°$. There is a similar pattern for the elevation angle as well. We call this phenomenon the "edge-deterioration effect". This interesting phenomenon suggests that subjects may subconsciously shift the laser mark towards the center to avoid exceeding the control area. Relatively large edge deviation may be caused by the limitations of boundary conditions and the distribution of actuators.

The mismatch observed in our experiment is related to various factors, such as the limit directional resolution of human hearing, the limitation of VBAP or pair-wise amplitude panning itself, and the use of flat-panel loudspeakers as sources for reproduction. First, human hearing of a real sound source has the mean azimuth mismatch ranges from 1° to 3°, and the mean elevation mismatch in the median plane ranges from 4° for white noise to 17° for speech. Second, multichannel sound with conventional loudspeakers when using the VBAP algorithm has a similar localization mismatch pattern, that the azimuth mismatch is much better that the elevation mismatch. Specifically, the azimuth mismatches of the flat-panel loudspeaker are comparable to those of conventional loudspeakers using VBAP[34] over the desired virtual sources at (0°, 0°), (0°, 15°) and (10°, 0°), where the difference in the median value, the interquartile range, or the data range, is within 2°, except that the flat-panel loudspeaker's data range at (10°, 0°) are 5° larger. For elevation mismatches, the difference in the median value is within 2°, while the interquartile range and the data range at (0°, 15°) and at (10°, 0°) are 5° to 13° larger, probably due to disparities in array configurations, as the triplets are placed differently in the flat-panel loudspeaker in our experiment and conventional loudspeakers[34].

Thus, the perceived location mismatch can be largely caused by the limitation of VBAP itself, the limited directional resolution of human hearing, and the edge effect of the flat panel when the desired virtual source is geometrically close to the panel edge.

Standard deviations of azimuth and elevation mismatch values are illustrated in Fig. 7. Azimuth angle exhibits a relatively uniform distribution. Standard deviations near the real sound source actuators at four vertices are small, while those at other localization places are slightly larger. The standard deviation is smaller in the upper part of the central area around the panel, rather than precisely in the center. This phenomenon may be because the slight rotation of the subject's head improves the localization accuracy and thus results in smaller inconsistencies. For elevation angle, the standard deviation is large around the central height area but small at the top and bottom. This indicates that the best localization is not the same height as the human ear. Slight head rotation may improve the localization effect, which could also explain this phenomenon.

We also analyzed individual differences by the standard deviation of the subjects as shown in Fig. 8. Subjects 5, 8, 15, 21, and 22 had large standard deviations in horizontal localization. Subjects 1, 7, 11, 15, and 19 had large deviations in vertical localization. The overall standard deviation of subject No. 15 is relatively large. Subject No. 8 had a large standard deviation of azimuth mismatch with a rather small standard deviation of elevation mismatch. The subject's statistical standard deviation fluctuated less horizontally while

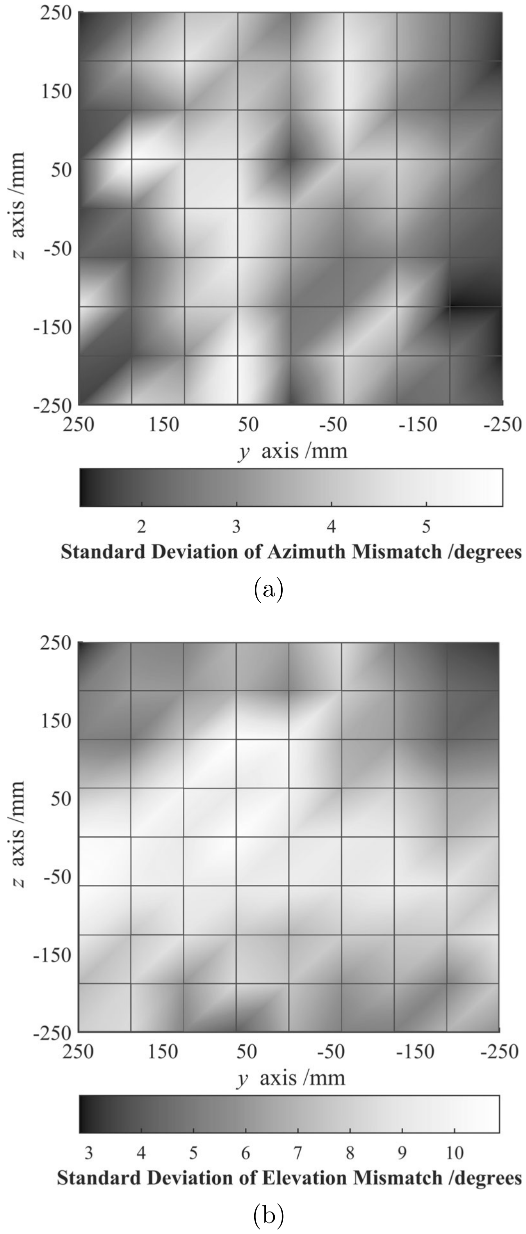


FIG. 7. (color online) Standard deviation of the localization mismatch in (a) azimuth and (b) elevation among 24 subjects in the listening test for each virtual source in Fig. 3.

vertically fluctuated greatly. The result coincides with the conclusion that human vertical perception is less accurate than horizontal localization. Spearman's ρ test was performed between the perceptual and ideal localization angles. The output of this test indicates a correlation within the range of 88% to 98% ($p < 0.05$). There is a strong correlation between the ideal and test values.

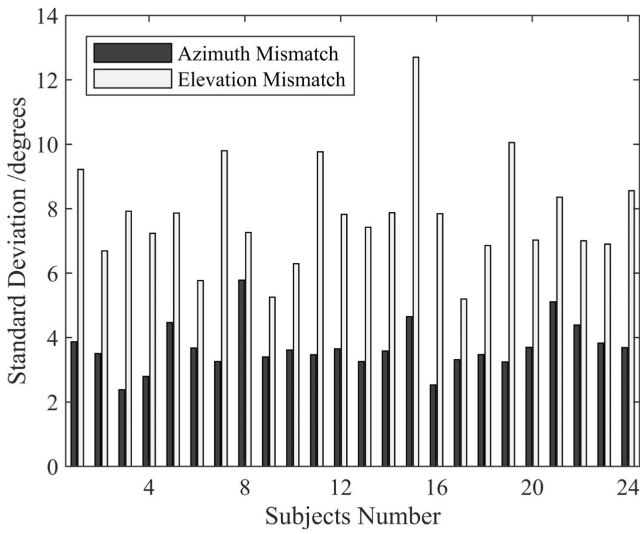


FIG. 8. (color online) The analysis of individual differences through the standard deviation of the subjects. The abscissa represents the subject number, ranging from 1 to 24. The dark filled bar represents the standard deviation of azimuth mismatch, while the light filled bar represents the standard deviation of elevation mismatch.

We classified the test results into two groups based on the virtual sound source positioned at different horizontal and vertical locations. Then we analyzed the influence of these two factors on the accuracy of perceptual azimuth and elevation angles. Two variables combined with two factors result in four datasets. The Lilliefors test determines whether each subjective dataset conforms to a normal distribution. We also conducted Bartlett's test to determine whether the data are derived from normal distributions with equal variances. The results suggest that the Kruskal-Wallis test is used rather than the analysis of variance (ANOVA) as not all tests accept the null hypothesis at the 5% significance level. According to the results, the exact placement of the virtual source on the panel substantially impacts both the perceptual azimuth and elevation localization accuracy ($p < 0.05$). Despite the statistical significance of all interactions, discernible differences in the magnitude of their impacts exist among the four scenarios. There is a statistically significant difference in perceptual azimuth angle among various horizontal angles where the virtual sound source is located ($p < 0.001$). The nine positions where the virtual source position is located at various heights also exhibit a statistically significant difference in perceptual elevation angle ($p < 0.001$). The effect of virtual source position at different heights on azimuth angle perception shows less variability than elevation angle, but the associated p-value is still less than 0.001. Although the significance is very high for the three scenarios mentioned above, virtual source position at different horizontal angles has the least impact on elevation angle perception, with a factor that is more than 1000 times lower ($p = 0.03$).

C. Discussion

Though the experiment was conducted over a chosen panel with corner-positioned actuators, the findings may have broader implications. In application, the flat-panel loudspeakers can be used solely or in multi-panel setups for immersive sound with extended spatial coverage. Under VBAP, each panel can operate independently, so we assessed a corner-actuated single-panel scenario as a representative module. Since VBAP doesn't require the exact sound source or propagation, and considering auditory perception and masking effects, the result may hold for other flat-panel loudspeakers with comparable geometry.

This experiment assesses the performance of VBAP on the flat-panel loudspeaker, and the results can serve as a baseline for perceptual evaluation in current and future research on flat-panel loudspeakers, since VBAP is a simple but effective approach for spatial sound reproduction that does not require acoustic transfer function information. This study also further enhances the understanding of the transition of sound reproduction using the conventional loudspeaker to the flat panel from a perceptual aspect. Those findings extend the existing theory and practical value on flat-panel loudspeakers, especially for the auditory display in buildings.

V. CONCLUSIONS

This paper explored the localization mismatch between desired and perceived sound directions using amplitude panning with flat-panel loudspeakers. The study involved creating virtual sound sources of various locations and evaluating the perceptual source direction through both objective and subjective tests. The subjective tests resulted in a mean azimuth direction mismatch within $\pm 6.0°$ and a mean elevation mismatch within $\pm 10.0°$. Additionally, the objective tests using the head and torso simulator and auditory localization cues indicated a good match. These findings suggest that the virtual source created by amplitude panning over a flat-panel loudspeaker can achieve spatial localization accuracy comparable to that of a real sound source without the need for complex algorithms or acoustic transfer function information. Future research will focus on optimizing algorithms for virtual source localization in spatial sound reproduction using flat-panel loudspeakers, along with perceptual evaluation.

ACKNOWLEDGMENTS

This work was supported by the Beijing Natural Science Foundation (L223032). The authors would like to express gratitude to all the participants who took part in the listening test, Professor Feiran Yang for insightful discussions on binaural modeling, Dr. Yuzhen Yang for engaging discussions on acoustic modeling, and Cong Wang, Yong Chen, Kai Wang, and Xuyang Zhu for their assistance during the experiment.

REFERENCES

[1]S. Spors, H. Wierstorf, A. Raake, F. Melchior, M. Frank, and F. Zotter, "Spatial sound with loudspeakers and its perception: A review of the current state," Proc. IEEE **101**(9), 1920–1938 (2013).

[2]P. Coleman, A. Franck, P. J. Jackson, R. J. Hughes, L. Remaggi, and F. Melchior, "Object-based reverberation for spatial audio," J. Audio Eng. Soc. **65**(1/2), 66–77 (2017).

[3]J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, Massachusetts, 1997), pp. 137–191.

[4]V. Erbes and S. Spors, "Localisation properties of wave field synthesis in a listening room," IEEE/ACM Trans. Audio, Speech, Language Process. **28**, 1016–1024 (2020).

[5]B. Xie, *Head-Related Transfer Function and Virtual Auditory Display*, 2nd ed. (J. Ross Publishing, Plantation, FL, 2013), pp. 8–12.

[6]C. Kyriakakis, P. Tsakalides, and T. Holman, "Surrounded by sound," IEEE Signal Process. Mag. **16**(1), 55–66 (1999).

[7]T. Betlehem, W. Zhang, M. A. Poletti, and T. D. Abhayapala, "Personal sound zones: Delivering interface-free audio to multiple listeners," IEEE Signal Process. Mag. **32**(2), 81–91 (2015).

[8]Q. Zhu, P. Coleman, M. Wu, and J. Yang, "Robust acoustic contrast control with reduced in-situ measurement by acoustic modeling," J. Audio Eng. Soc. **65**(6), 460–473 (2017).

[9]S. Cecchi, A. Carini, and S. Spors, "Room response equalization—a review," Appl. Sci. **8**(1), 16 (2017).

[10]M. Poletti, "Robust two-dimensional surround sound reproduction for nonuniform loudspeaker layouts," J. Audio Eng. Soc. **55**(7/8), 598–610 (2007).

[11]S. Koyama, G. Chardon, and L. Daudet, "Optimizing source and sensor placement for sound field control: An overview," IEEE/ACM Trans. Audio, Speech, Language Process. **28**, 696–714 (2020).

[12]Q. Zhu, P. Coleman, X. Qiu, M. Wu, J. Yang, and I. Burnett, "Robust personal audio geometry optimization in the SVD-based modal domain," IEEE/ACM Trans. Audio, Speech, Language Process. **27**(3), 610–620 (2019).

[13]Q. Zhu, X. Qiu, P. Coleman, and I. Burnett, "An experimental study on transfer function estimation using acoustic modelling and singular value decomposition," J. Acoust. Soc. Am. **150**(5), 3557–3568 (2021).

[14]B. Pueo, J. J. López, J. Escolano, and L. Hörchens, "Multiactuator panels for wave field synthesis: Evolution and present developments," J. Audio Eng. Soc. **58**(12), 1045–1063 (2011).

[15]Z. Li, P. Luo, C. Zheng, and X. Li, "Vibrational contrast control for local sound source rendering on flat panel loudspeakers," in *Proceedings of the 145th Audio Engineering Society Convention*, New York, NY (October 17–20, 2018).

[16]M. C. Heilemann, D. Anderson, and M. F. Bocko, "Sound-source localization on flat-panel loudspeakers," J. Audio Eng. Soc. **65**(3), 168–177 (2017).

[17]V. P. Gontcharov and N. P. Hill, "Diffusivity properties of distributed mode loudspeakers," in *Proceedings of the 108th Audio Engineering Society Convention*, Paris, France (February 19-22, 2000).

[18]N. Harris, "Modelling acoustic room interaction for pistonic and distributed-mode loudspeakers in the correlation domain," in *Proceedings of the 117th Audio Engineering Society Convention*, San Francisco, CA (October 28-31, 2004).

[19]M. C. Heilemann, D. A. Anderson, S. Roessner, and M. F. Bocko, "The evolution and design of flat-panel loudspeakers for audio reproduction," J. Audio Eng. Soc. **69**(1/2), 27–39 (2021).

[20]M. M. Boone, "Multi-actuator panels (MAPs) as loudspeaker arrays for wave field synthesis," J. Audio Eng. Soc. **52**(7/8), 712–723 (2004).

[21]N. Kournoutos and J. Cheer, "A system for controlling the directivity of sound radiated from a structure," J. Acoust. Soc. Am. **147**(1), 231–241 (2020).

[22]S. Flanagan and N. Harris, "Loudness: A study of the subjective difference between DML and conventional loudspeaker," in *Proceedings of the 106th Audio Engineering Society Convention*, Munich, Germany (May 8-11, 1999).

[23]N. J. Harris, "The acoustics and psychoacoustics of the distributed-mode loudspeaker (DML)," Ph.D. dissertation, University of Essex, Essex, United Kingdom, 2001.

[24]M. Rébillat, E. Corteel, B. F. Katz, and X. Boutillon, "From vibration to perception: Using large multi-actuator panels (LaMAPs) to create coherent audio-visual environments," in *Acoustics 2012*, Nantes, France (April 23-27, 2012).

[25]M. Heilemann, D. Anderson, S. Roessner, and M. F. Bocko, "Quantifying listener preference of flat-panel loudspeakers," in *Proceedings of the 145th Audio Engineering Society Convention*, New York, NY (October 17–20, 2018).

[26]E. Corteel, K.-V. Nguyen, O. Warusfel, T. Caulkins, and R. Pellegrini, "Objective and subjective comparison of electrodynamic and MAP loudspeakers for wave field synthesis," in *Proceedings of the Audio Engineering Society Conference: 30th International Conference: Intelligent Audio Environments*, Saariselka, Finland (March 15–17, 2007).

[27]V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," J. Audio Eng. Soc. **45**(6), 456–466 (1997).

[28]E. Erdem, Z. Cvetkovic, and H. Hacihabiboglu, "3D perceptual soundfield reconstruction via virtual microphone synthesis," IEEE/ACM Trans. Audio, Speech, Language Process. **31**, 1305–1317 (2023).

[29]A. Franck, W. Wang, and F. M. Fazi, "Sparse ℓ_1-optimal multiloudspeaker panning and its relation to vector base amplitude panning," IEEE/ACM Trans. Audio, Speech, Language Process. **25**(5), 996–1010 (2017).

[30]M. C. Heilemann, D. A. Anderson, and M. F. Bocko, "Near-field object-based audio rendering on flat-panel displays," J. Audio Eng. Soc. **67**(7/8), 531–539 (2019).

[31]M. C. Heilemann, "Spatial audio rendering with flat-panel loudspeakers," Ph.D. dissertation, University of Rochester, Rochester, New York, United States, 2018.

[32]M. Kuster, D. D. Vries, D. Beer, and S. Brix, "Structural and acoustic analysis of multi-actuator panels," J. Audio Eng. Soc. **54**(11), 1065–1076 (2006).

[33]V. Pulkki, "Generic panning tools for MAX/MSP," in *Proceedings of the 2000 International Computer Music Conference*, Berlin, Germany (August 27 - September 1, 2000), pp. 304–307.

[34]V. Pulkki, "Localization of amplitude-panned virtual sources II: Two-and three-dimensional panning," J. Audio Eng. Soc. **49**(9), 753–767 (2001).

[35]A. Politis, "Microphone array processing for parametric spatial audio techniques," Ph.D. dissertation, Aalto University, Otaniemi, Espoo, Finland, 2016.

[36]H. Wierstorf and S. Spors, "Sound field synthesis toolbox," in *Proceedings of the 132nd Audio Engineering Society Convention*, Budapest, Hungary (April 26-29, 2012).

[37]U. Zölzer, *DAFX: Digital Audio Effects*, 2nd ed. (John Wiley & Sons, West Sussex, United Kingdom, 2002), pp. 283–289.

[38]D. Howard and J. Angus, *Acoustics and Psychoacoustics*, 4th ed. (Focal Press, Oxford, UK, 2009), pp. 107–111.

[39]J. Zheng, T. Zhu, J. Lu, and X. Qiu, "A linear robust binaural sound reproduction system with optimal source distribution strategy," J. Audio Eng. Soc. **63**(9), 725–735 (2015).

[40]B. Xie, H. Mai, D. Rao, and X. Zhong, "Analysis of and experiments on vertical summing localization of multichannel sound reproduction with amplitude panning," J. Audio Eng. Soc. **67**(6), 382–399 (2019).

[41]R. Algazi, C. Avendano, and R. O. Duda, "Estimation of a spherical-head model from anthropometry," J. Audio Eng. Soc. **49**(6), 472–479 (2001).

[42]ITU-R BS.2159-9, *Multichannel Sound Technology in Home and Broadcasting Applications* (International Telecommunications Union, 2022).

[43]B. Xie, *Spatial Sound: Principles and Applications* (CRC Press, Boca Raton, FL, 2023), pp. 704–710.