applied sciences

MDPI

Article

Adaptive Binaural Cue-Based Amplitude Panning in Irregular Loudspeaker Configurations

Shang Zhao, Yunjia Zhou and Zhibin Lin *

Key Laboratory of Modern Acoustics, Institute of Acoustics, Nanjing University, Nanjing 210093, China; zhaoshang@smail.nju.edu.cn (S.Z.); 5020222220120@smail.nju.edu.cn (Y.Z.)
* Correspondence: zblin@nju.edu.cn

Featured Application: This method has significant potential for application in the field of spatial audio, providing listeners with precise spatial azimuth cues, thereby significantly enhancing the realism and immersiveness of the auditory experience within virtual and augmented reality applications.

Abstract: The amplitude panning method is a prevalent technique for controlling sound image directions in stereophonic and multichannel surround systems. However, conventional methods typically achieve accurate localization only with standardized loudspeaker configurations, limiting their applicability in diverse scenarios. This paper presents an adaptive binaural cue-based amplitude panning algorithm based on binaural cues tailored for accurate azimuthal sound image localization in irregular loudspeaker configurations. Our two-stage approach first employs inverse filtering to equalize loudspeaker magnitude responses. Subsequently, gains and inter-loudspeaker time delay are optimized based on interaural time difference and interaural cross-correlation values, derived from the measured binaural room impulse responses using a dummy head. Objective and subjective evaluations on a stereo system demonstrate that the proposed method significantly improves azimuth localization accuracy compared to existing techniques.

Keywords: amplitude panning; irregular loudspeaker configuration; binaural cue; spatial audio; inverse filtering



Academic Editor: Alexander Sutin

Received: 11 March 2025
Revised: 21 April 2025
Accepted: 22 April 2025
Published: 23 April 2025

Citation: Zhao, S.; Zhou, Y.; Lin, Z. Adaptive Binaural Cue-Based Amplitude Panning in Irregular Loudspeaker Configurations. *Appl. Sci.* **2025**, *15*, 4689. https://doi.org/10.3390/app15094689

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Spatial sound field reproduction seeks to establish an immersive acoustic environment within a predefined region, enabling the listener to experience a virtual yet realistic replication of the original sound field. Techniques that physically recreate an acoustic field are referred to as sound field synthesis [1]. Examples include Wave Field Synthesis (WFS) [2–6], Higher Order Ambisonics (HOA) [7–9], and the pressure matching method [10,11]. In contrast, amplitude panning methods attempt to create a plausible spatial perception by delivering the relevant psychoacoustic cues to the listener's ears. These perceptually motivated reproduction techniques distribute the source signal across multiple loudspeakers, assigning a gain to each to create a sound image or virtual sound source from the desired direction. Such methods are advantageous for practical applications, offering low computational complexity, absence of destructive interference in the 'sweet spot', high timbral quality, and gradual degradation of sound quality outside the 'sweet spot' [12]. Vector-based amplitude panning (VBAP) [13–16] is the most widely used perceptually motivated method for two- and three-dimensional multi-loudspeaker reproduction, forming the basis

of spatial and object-based audio standards, such as MPEG-H [17]. The multiple direction amplitude panning (MDAP) method [18] extends VBAP by introducing additional virtual sources around the intended source position, allowing for the synthesis of both uniform and variable source spreads. The distance-based amplitude panning (DBAP) method [19] is used for irregular setups, taking the actual positions of the loudspeakers in space as the point of departure. All-Round Ambisonic Panning (AllRAP) [20] creates phantom sources with stable loudness and adjustable source width. Additionally, a convex optimization-based method has been proposed to allow the precise control of source spread [12,21,22]. The Compensated Amplitude Panning (CAP) method [23–25] is proposed by considering the orientation of the listener's head. Frequency-dependent gain panning methods [26,27] have been proposed to maintain consistent loudness and localization.

Most amplitude panning methods assume that the listener is positioned at the 'sweet spot', with loudspeakers arranged equidistantly and symmetrically around them. However, achieving such an idealized listening environment can be challenging, particularly in domestic and cabin settings. In consumer systems, where loudspeaker positions may deviate from their canonical locations, the perceived direction of the virtual sound source may shift away from its intended location. To address this problem, techniques such as time-aligning and loudness-matching loudspeakers based on distance [28] have been proposed. A modified panning approach for non-equidistant loudspeakers, considering the composite contributions of direct sound from multiple loudspeakers, has also been suggested [29]. However, these compensation methods for amplitude panning assume each loudspeaker behaves as an omnidirectional point source in a free field, thereby neglecting frequency-dependent factors, including room modes, loudspeaker directivity, and inter-loudspeaker variations. Moreover, these methods frequently presume that, after compensation, the reproduction system approximates an ideal listening environment, enabling the application of panning methods predicated on idealized conditions. This assumption disregards the persistent effects of irregular setups and various listening environments, ultimately leading to inaccuracies in the perceived direction of virtual sound sources.

To enhance the listening experience in acoustic environments with irregular loudspeaker configurations, this paper introduces an adaptive binaural cue-based amplitude panning algorithm designed to accurately reproduce the azimuth angle of virtual sound sources. This algorithm leverages measured room impulse responses (RIRs) and binaural room impulse responses (BRIRs) to account for the characteristics of the listening environment when deriving loudspeaker gains. It consists of two stages: the compensation stage and the gain and delay optimization stage. In the compensation stage, the inverse filtering algorithm is employed to eliminate discrepancies in the magnitude frequency responses of RIRs, ensuring consistent sound output from each loudspeaker at the center of the listening area. During the gain and delay search optimization stage, the interaural time difference (ITD) and interaural cross-correlation (IACC) for different gains and time delays are predicted using BRIRs measured with a dummy head. By adjusting the gains and time delay between loudspeakers, the predicted ITD of the virtual source is matched to that of the real source in the free field, creating an accurate virtual sound source at the desired azimuth angle. To resolve multiple solutions during the search process, the optimal gain pair and time delay are determined by using IACC. The effectiveness of the proposed algorithm is validated through both subjective and objective evaluations on a stereo system.

The remainder of this paper is organized as follows: Section 2 introduces the proposed amplitude panning method, including the use of inverse filters to eliminate magnitude differences between loudspeakers and the determination of gains and time delay based on ITD and IACC. Section 3 presents the results of the simulation. Section 4 presents the results

of objective and subjective experiments conducted in a listening room. Finally, Section 5 concludes the paper.

2. Methods

Amplitude panning methods determine the gains for each loudspeaker based on the positions of both the loudspeakers and the virtual sound source, following a specified amplitude panning function. Each virtual source signal $s(t)$ is then scaled by the corresponding gain g_s and distributed across the S loudspeakers. As a result, the listener perceives the virtual sound source at the intended direction or location, as illustrated in Figure 1. The specific amplitude panning method varies depending on the acoustic environment and loudspeaker arrangement. In this section, we introduce a binaural cue-based amplitude panning method designed for irregular loudspeaker configurations within a room. The proposed algorithm comprises two distinct stages. The initial stage involves magnitude compensation, wherein an inverse filter is designed to equalize the magnitude frequency response of each loudspeaker. To ensure robustness, the magnitude is smoothed using one-third octave bands prior to computing the inverse filter. The subsequent stage explores the gain and delay combinations within a specified range. Using the acquired BRIRs, the ITD and IACC of the synthesized virtual sound source are calculated. The optimal gain and delay combination is then selected based on its ability to match the ITD of the real sound source at the intended azimuth. In instances where multiple combinations satisfy the target ITD criterion, the final gain and delay are determined by maximizing the IACC. We have focused our research on a stereo system, as the multichannel surround system can be realized by partitioning the multichannel setup into a combination of several stereo systems, akin to VBAP [13].

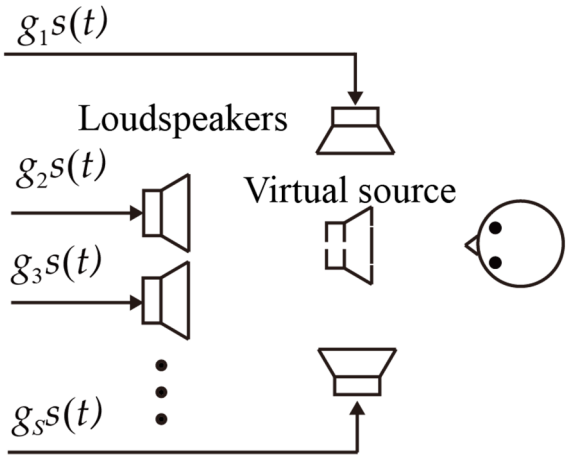

Figure 1. Schematic diagram of amplitude panning method.

2.1. Magnitude Compensation Stage

2.1.1. Inverse Filtering

The inverse filtering problem can be conceptualized from a model-matching perspective [30]. For a system with S loudspeakers, T program input channels, and R receiver points, it can be expressed as the following optimization problem:

$$\min_{\boldsymbol{C}} \|\boldsymbol{M} - \boldsymbol{H}\boldsymbol{C}\|_{\mathrm{F}}^2, \tag{1}$$

where F symbolizes the Frobenius norm, $\boldsymbol{H}$ is an $R \times S$ plant transfer matrix, $\boldsymbol{M}$ is an $R \times T$ matching model matrix, and $\boldsymbol{C}$ is an $S \times T$ inverse filter matrix. By utilizing Tikhonov regularization, the inverse filter matrix can be obtained as

$$\boldsymbol{C}=(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}+\beta\boldsymbol{I})^{-1}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{M}, \tag{2}$$

where β is the regularization factor. To simplify the system and reduce filter length, the inverse filter is calculated individually for each loudspeaker. For a single-input single-output system, the magnitude compensation algorithm is applied to each loudspeaker, expressed as

$$C(k) = \frac{|H^*(k)|}{|H(k)|^2 + \beta} \cdot M(k), \tag{3}$$

where k is the frequency index. By performing the inverse Fourier transform and applying circular shifts, a causal FIR filter is obtained.

2.1.2. Spectral Smoothing

Due to the inherent complexity and measurement inaccuracies of the RIR, a direct solution using Equation (3) may result in an unstable filter. To mitigate this issue, spectral smoothing is applied to the RIRs using the methodology from [31], which simultaneously smooths both the magnitude and phase spectra. When considering only the magnitude during the inverse filtering process, the smoothed magnitude $|H_{\mathrm{ts}}(k)|$ is obtained as

$$|H_{\mathrm{ts}}(k)| = \sqrt{\sum_{i=k-m(k)}^{k+m(k)} \left\{ [H_{\mathrm{R}}(i)]^2 + [H_{\mathrm{I}}(i)]^2 \right\} W_{\mathrm{sm}}[(i-k+m(k))]}, \tag{4}$$

where $H_{\mathrm{R}}(i)$ and $H_{\mathrm{I}}(i)$ are the real and imaginary parts of the RIR in the frequency domain, and $m(k)$ is the smoothing index corresponding to the length of the smoothing window W_{sm}. This method simulates the auditory system's property, where frequency resolution decreases at higher frequencies. By incorporating non-uniform resolution, the equalization becomes less precise at these frequencies. Moreover, this approach is less affected by positional changes and avoids the processing artifacts of traditional ideal inversion methods [32].

2.2. Gain and Time Delay Optimization Stage

Following the equalization of the magnitude frequency responses of the left and right loudspeakers using the inverse filter, the gain and delay need to be determined to create the virtual sound source. Given gains g_L and g_{R} applied to the left and right loudspeakers, respectively, the resulting sound pressures at the listener's left and right ears, denoted as $p_{\mathrm{L}}(t)$ and $p_{\mathrm{R}}(t)$, are calculated using the measured BRIRs as

$$p_{\mathrm{L}}(t) = g_L \cdot c_L * \mathrm{BRIR}_{\mathrm{LL}}(t-\Delta t) * s(t) + g_{\mathrm{R}} \cdot c_{\mathrm{R}} * \mathrm{BRIR}_{\mathrm{LR}}(t) * s(t) \tag{5}$$

and

$$p_{\mathrm{R}}(t) = g_L \cdot c_L * \mathrm{BRIR}_{\mathrm{RL}}(t-\Delta t) * s(t) + g_{\mathrm{R}} \cdot c_{\mathrm{R}} * \mathrm{BRIR}_{\mathrm{RR}}(t) * s(t), \tag{6}$$

where $*$ is defined as the convolution operation symbol. c_L and c_{R} are the inverse filters for the left and right loudspeakers, respectively. The terms $\mathrm{BRIR}_{\mathrm{LL}}$, $\mathrm{BRIR}_{\mathrm{LR}}$, $\mathrm{BRIR}_{\mathrm{RL}}$, and $\mathrm{BRIR}_{\mathrm{RR}}$ represent the BRIR from each loudspeaker to each ear. The first subscript indicates the ear (L for left ear, R for right ear), and the second subscript indicates the loudspeaker

(L for left loudspeaker, R for right loudspeaker). Additionally, Δt represents the time delay of the left loudspeaker. The ITD and the IACC are calculated as

$$\mathrm{ITD} = \underset{|\tau| \leq 1 \text{ ms}}{\operatorname{argmax}} \left(\int_{t_1}^{t_2} p_{\mathrm{L}}(t) p_{\mathrm{R}}(t+\tau) \mathrm{d}t \right) \tag{7}$$

and

$$\mathrm{IACC} = \max_{|\tau| \leq 1 \text{ ms}} \left(\frac{\int_{t_1}^{t_2} p_{\mathrm{L}}(t) p_{\mathrm{R}}(t+\tau) \mathrm{d}t}{\sqrt{\int_{t_1}^{t_2} (p_{\mathrm{L}}^2(t) \mathrm{d}t \int_{t_1}^{t_2} p_{\mathrm{R}}^2(t) \mathrm{d}t}} \right), \tag{8}$$

where integration limits t_1 and t_2 are the start and end moments of the signal. Prior to calculating the ITD and IACC, the sound pressure signals from both ears are subjected to low-pass filtering with a cutoff frequency of 1500 Hz. The ITD refers to the time difference between the arrival of a sound at each ear, serving as a key cue for sound localization. Low-frequency ITD cues, particularly those from content below 1500 Hz, are robust and dominate the directional cues at higher frequencies [23]. With a fixed time delay, the ITD depends solely on the gains of the left and right loudspeakers. By adjusting the gains, the ITD of the left and right ear signals can be aligned with the target ITD of the real sound source at the desired azimuth angle α, expressed as

$$g_{\mathrm{L}}^{\mathrm{opt}}(\alpha, \Delta t) = \underset{0 \leq g_{\mathrm{L}} \leq 1}{\operatorname{argmin}} |\mathrm{ITD} - \mathrm{ITD}_{\mathrm{Tar}}(\alpha)|. \tag{9}$$

To maintain a constant sound pressure level (SPL) across different angles, the gains of the left and right loudspeakers satisfy the constraint

$$\left(g_{\mathrm{L}}^{\mathrm{opt}}(\alpha, \Delta t) \right)^2 + \left(g_{\mathrm{R}}^{\mathrm{opt}}(\alpha, \Delta t) \right)^2 = 1. \tag{10}$$

In the experiment, multiple combinations of gains and time delays may minimize the ITD error. To determine the optimal gains and time delay, we propose maximizing the mean IACCs of N target azimuths as the selection criterion:

$$\Delta t^{\mathrm{opt}} = \underset{\Delta t}{\operatorname{argmax}} \frac{1}{N} \sum_{\alpha} \mathrm{IACC}^{\mathrm{opt}}(\alpha), \tag{11}$$

where $\mathrm{IACC}^{\mathrm{opt}}(\alpha)$ corresponds to the virtual sound source formed by the gains $g_{\mathrm{L}}^{\mathrm{opt}}(\alpha, \Delta t)$ and $g_{\mathrm{R}}^{\mathrm{opt}}(\alpha, \Delta t)$. The IACC was chosen as the criterion due to its strong, monotonic relationship with listener sensitivity to ITD, which increases as the IACC value rises [33]. Although Equations (9) and (11) cannot be solved directly, the optimal gains and time delay can be determined through parameter scanning within a defined value range.

3. Simulations

The simulations were conducted using measured RIRs and BRIRs obtained in the listening room of the Acoustics Institute at Nanjing University, with a reverberation time of approximately 0.3 s. The spatial arrangement of the loudspeakers and listening area is shown in Figure 2a. The distance between the left and right loudspeakers is 1.5 m, and the distance from the center of the listening position to the loudspeaker axis is 1.3 m. The center of the listening area is laterally offset by 0.4 m from the loudspeaker centerline, with its height aligned with the loudspeakers. To acquire the BRIRs, the MegaSig AH 262 desktop dummy head is placed at the listening location, oriented directly forward as shown in Figure 2b. This dummy head is interfaced with a Fireface UC sound card connected to a

laptop, which serves to control the loudspeakers and record signals from the left and right ear microphones of the dummy head. The sampling rate is set to 44.1 kHz.

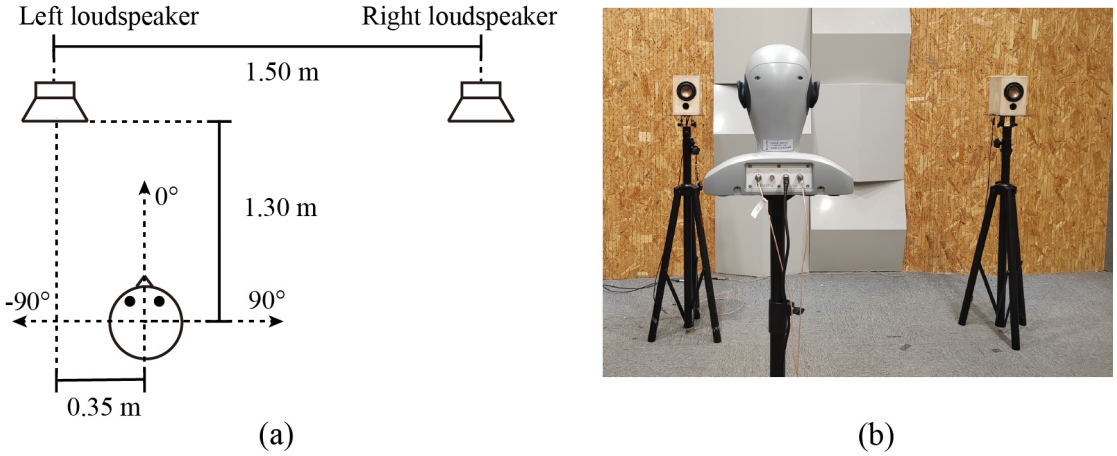

Figure 2. (**a**) Relative positions of the loudspeakers and listening area and (**b**) illustration of the BRIR measurement setup.

The RIRs of both loudspeakers, measured with the microphone placed at the listening position, are shown in Figure 3. Irregular loudspeaker placement introduces variations in the distances between each loudspeaker and the measurement position, causing differences in the direct sound components of the RIRs, both in time and amplitude, compared to standard configurations. Additionally, room modes, loudspeaker directivity, and their differences lead to magnitude variations across frequencies. Therefore, the inverse filtering algorithm is applied to mitigate the magnitude frequency response differences between the left and right loudspeakers, with the target response set to a constant. The frequency range of the inverse filter is limited to 150 Hz to 9000 Hz by adjusting the frequency-dependent regularization factor based on the loudspeakers' working frequency bands. This prevents excessive compensation outside the loudspeakers' working range. The regularization factor β is shown in Figure 3c. For the frequency range from 150 Hz to 9000 Hz, β is set to 1×10^{-4}, and for frequencies below 75 Hz and above 18,000 Hz, it is set to 1×10^{-3}. The factor for the remaining frequencies is determined through logarithmic interpolation. To improve the robustness of inverse filtering, one-third octave smoothing is applied to the RIRs before calculating the inverse filters. The resulting filtered magnitude frequency responses are shown in Figure 3d. Compared to the raw RIRs shown in Figure 3b, the magnitude differences between the loudspeakers are substantially reduced, with their magnitude frequency curves exhibiting a near-flat characteristic within the designated frequency range.

During the gain and delay optimization phase, the gains and inter-loudspeaker delay are determined to accurately reproduce the virtual sound source at various azimuth angles. In this stage, a range of gains is systematically evaluated, with the corresponding ITD of the virtual sound source calculated. The gain matching the target ITD is selected to reproduce the virtual sound source at the desired azimuth angle. In the simulation, the azimuth angle search range extends from $-15°$ to $40°$, with $5°$ intervals in the front horizontal plane. The target ITD values are derived from measurements of a real source acquired in the anechoic chamber at Nanjing University, corresponding to the relevant azimuth angles. During the optimization stage, the left loudspeaker's gain is incrementally increased from 0 to 1 with discrete steps of 0.001, and the right loudspeaker's gain is calculated using Equation (10). The ITD for each gain pair is calculated using Equations (5)–(7), with the input signal set as the unit impulse signal. Prior to the gain optimization process, a preliminary time

delay is applied to the left loudspeaker's signal to ensure the proper combination of direct sounds from both loudspeakers, preventing excessive time differences that could lead to auditory dominance by the leading loudspeaker. Figure 4a shows the variation of ITD with gain for the left loudspeaker time delays of 0.88 ms, 0.93 ms, and 0.98 ms. A positive correlation exists between the left loudspeaker gain and the ITD. Specifically, when the left loudspeaker gain is set to 0, only the right loudspeaker emits sound, resulting in the smallest ITD. Conversely, when the left loudspeaker gain reaches its maximum value of 1, the ITD reaches its maximum, corresponding to the left loudspeaker's ITD. Therefore, as the left loudspeaker gain increases from 0 to 1, the ITD of the virtual sound source gradually shifts from the right loudspeaker's ITD to the left loudspeaker's ITD. Furthermore, the precise time alignment significantly impacts the algorithm's performance. When employing time delays of 0.93 ms or 0.98 ms, the ITD changes uniformly with varying gain, with the difference between adjacent ITD values corresponding to the duration of a single time sample. As the gain increases from 0 to 1, every ITD value between the left and right loudspeakers' ITDs is covered, enabling the acquisition of the gain for each desired azimuth angle. When multiple gains correspond to the same ITD value, the arithmetic mean gain is selected to prevent minor gain changes from affecting the ITD, thereby enhancing robustness. However, with a time delay of 0.88 ms, a notable discontinuity arises as the ITD abruptly shifts from -0.36 ms to 0 ms as the left loudspeaker gain transitions from 0.472 to 0.473. This discontinuity causes the gain search method to fail in accurately reproducing the target ITDs within this range. The gain results for virtual sound source angles ranging from -15° to 40° are shown in Figure 4b. The ITD of the virtual sound source is influenced by both time delay and gains. For a given azimuth angle, varying time delays correspond to different gains. To ensure a continuous ITD variation during the gain search process, the time delay must be judiciously selected. Otherwise, an abrupt discontinuity in the ITD during the gain and delay search stage will prevent the virtual sound source's ITD from matching the target ITD through gain adjustment alone.

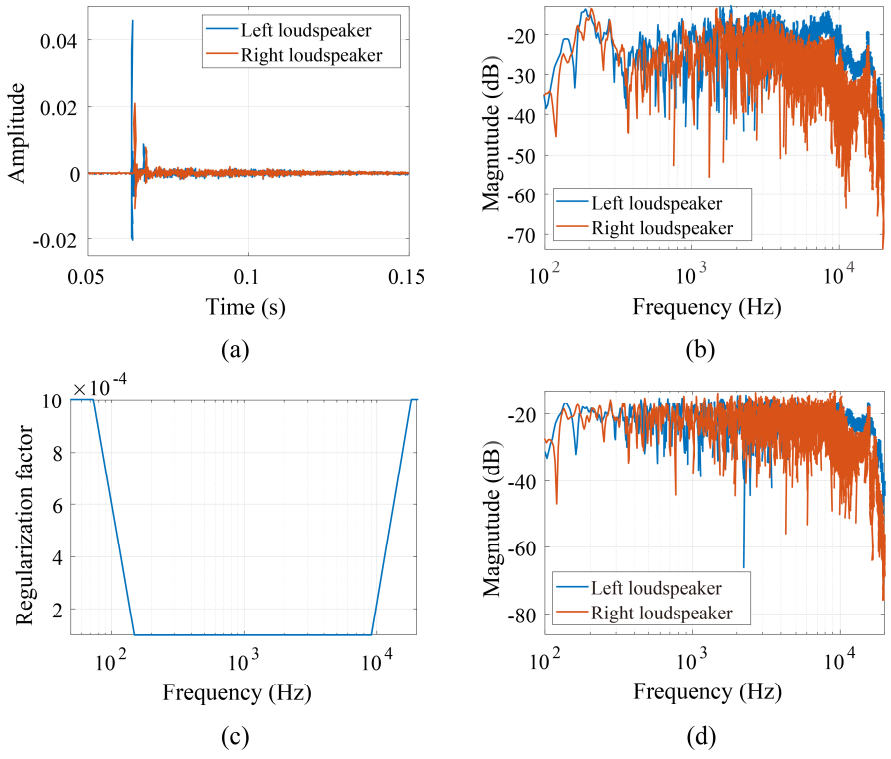

Figure 3. (**a**) RIRs; (**b**) magnitude frequency responses; (**c**) frequency-dependent regularization factor; and (**d**) filtered magnitude frequency responses.

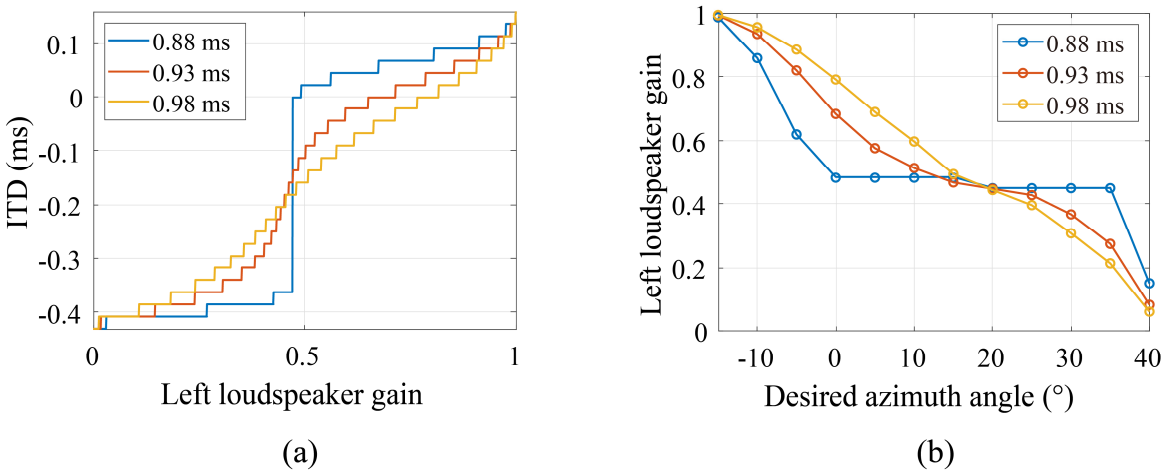

Figure 4. (**a**) ITD corresponding to different gains and (**b**) optimal gains obtained at various desired azimuth angles for three different time delays.

To determine the appropriate time delay, we investigate the influence of varying time delays on the virtual sound source ITD obtained by the gain optimization algorithm. Given that the inherent time difference for the direct sound position between the two loudspeakers is 1.00 ms, the search range for time delays is set from 0.77 ms to 1.22 ms, centered around this value, covering 20 adjacent time sampling points. The results of the loudspeaker gain search for each discrete time delay are shown in Figure 5. When the time delay is between 0.93 ms and 1.16 ms, the loudspeaker gain pair can be adjusted for each desired azimuth angle to align the virtual sound source's ITD with that of the real sound source in the free field. However, outside this range, the ITD fails to accurately match the real sound source's ITD at certain angles. As the deviation from the ideal time delay increases, an increasing number of angles exhibit ITD errors.

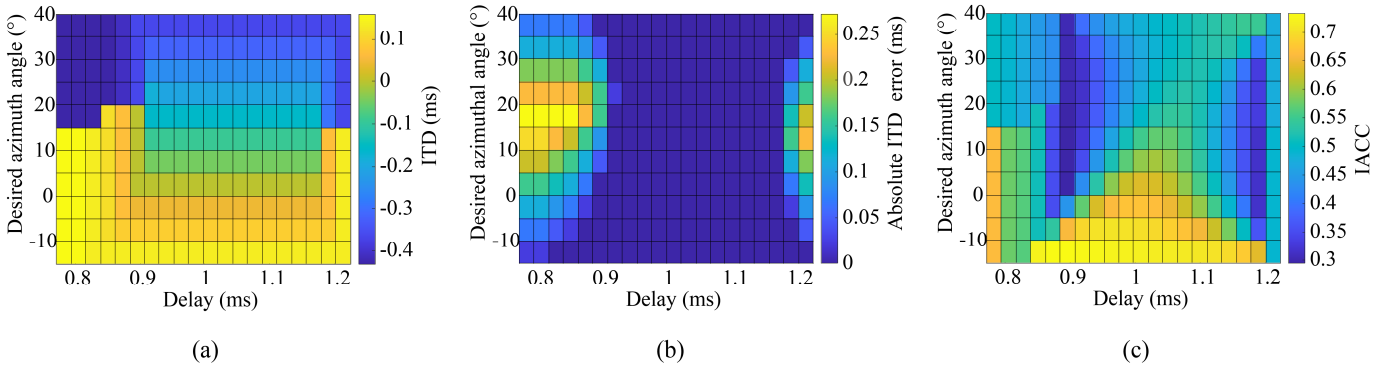

Figure 5. (**a**) ITD, (**b**) absolute ITD error, and (**c**) IACC of virtual source under different time delays.

Since multiple time delays can result in zero absolute ITD error, we calculate the mean IACC and mean absolute ITD error for each desired sound source azimuth angle at each delay to determine the optimal time delay and corresponding gains, as shown in Figure 6a. Within the time delay range where the mean absolute ITD error approaches zero, the mean IACC initially increases and subsequently decreases, reaching a maximum value at 1.04 ms. Outside this range, the highest mean IACC occurs at 0.77 ms. At this particular time delay, the ITDs for each azimuth angle are restricted to either 0.16 ms or −0.43 ms, corresponding to the ITDs when the left or right loudspeakers work individually. This indicates that during the search process, the sound image abruptly shifts from the left loudspeaker to the right, failing to synthesize the virtual sound source between loudspeakers. Consequently,

compared to the virtual sound source reproduced by two loudspeakers, the IACC value for a single loudspeaker is higher. Therefore, a time delay of 1.04 ms is selected as the optimal setting, and the corresponding gains obtained during the search are applied to reproduce the virtual sound source at the desired azimuth angles, as shown in Figure 6b. As the desired azimuth angle increases, the gain of the left loudspeaker decreases, and that of the right loudspeaker increases. The proposed algorithm uses a uniform time delay for all azimuth angles, thereby simplifying the creation of continuously moving virtual sound sources compared to methods with variable time delays. For instance, when the virtual sound source is intended to move continuously from −15° to −10°, interpolation techniques can be used to achieve the desired spatial trajectory. When the delays for −15° and −10° are consistent, the loudspeaker gains can be easily interpolated. However, if the time delays for −15° and −10° differ, both the time delay and gain would need simultaneous adjustments when moving the virtual sound source. Since time delay adjustments occur in discrete steps (at least a single time sample), maintaining continuous changes becomes challenging and complicates the interpolation process.

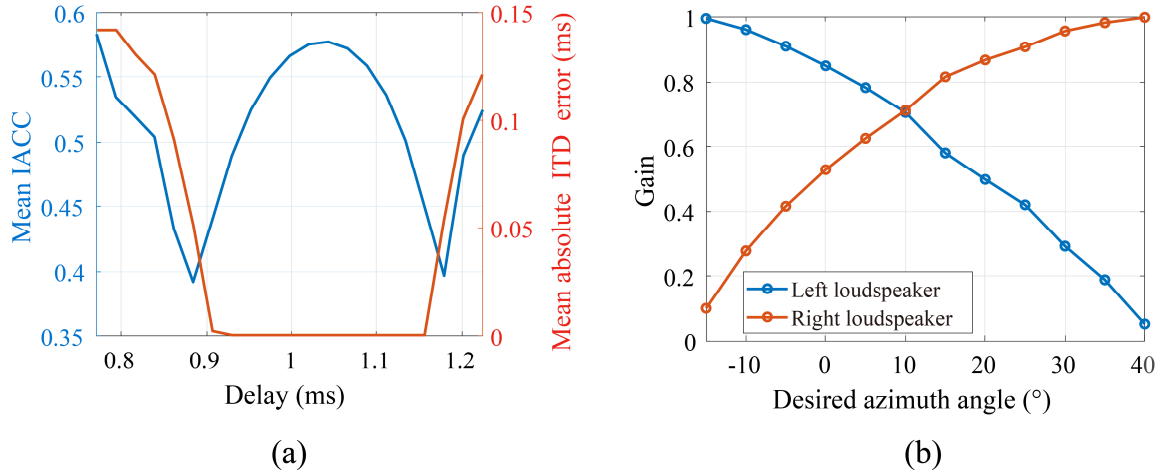

Figure 6. (**a**) Mean IACC and mean absolute ITD error at different time delays, and (**b**) optimal gains for the left and right loudspeakers at different desired azimuth angles.

The performance of the proposed algorithm was evaluated in comparison with the VBAP [13] and DBAP [19] algorithms. As the VBAP algorithm is inherently designed for loudspeaker arrangements with equidistant placement, amplitude and time delay compensation are required when applying it to setups with unequal loudspeaker distances [28]. In the simulation, measured RIRs were used to implement delay compensation, aligning the direct sound peaks of the two loudspeakers. Additionally, the amplitude was adjusted to equalize the peak values of the direct sound. Regarding the DBAP algorithm, which simultaneously controls both the azimuth angle and distance of the virtual sound source, the virtual source was restricted to the loudspeaker axis. Figure 7a presents the ITD simulation results for the three evaluated methods. The VBAP algorithm accurately reproduces virtual sound sources with ITD errors within 0.02 ms (equivalent to one sample delay) and azimuth angle deviations within 5° for desired angles between −15° and 10°, as well as at 40°. However, in the range of 15° to 35°, the reproduced ITDs are consistently lower than the target values. In particular, at 20°, the ITD error reaches 0.14 ms (equivalent to six sample delays), with corresponding azimuth errors approaching 15°. For the DBAP algorithm, the ITDs remain nearly constant at approximately 0.14 ms for azimuth angles between −15° and 15°. However, at 20° and 25°, the ITDs exhibit a sudden increase to 0.91 ms, while for angles above 30°, they drop sharply to −0.41 ms. In contrast, the

proposed method yields ITD values that closely match the target ITDs across the entire azimuth range, indicating superior and consistent performance. These results demonstrate that the proposed algorithm achieves reliable ITD reproduction, while the VBAP algorithm maintains accuracy only at azimuths near the loudspeakers and exhibits significant bias at mid-range angles. The DBAP algorithm fails to synthesize coherent virtual sources, primarily due to the excessive time delay between sound arrival from the two loudspeakers. At 20° and 25°, the calculated ITDs reflect only the geometric path length difference, rather than corresponding to a perceivable auditory image. Figure 7**b** presents the IACC simulation results of virtual sound sources reproduced by the three methods across various azimuth angles. In a free field, the IACC of a real sound source approaches unity. However, in a listening room, acoustic reflections reduce the IACC values for all virtual sources. The overall differences in IACC among the three methods are relatively small. Compared to the other two methods, the proposed algorithm yields slightly lower IACC values at azimuth angles near the loudspeakers, but higher values at mid-range angles.

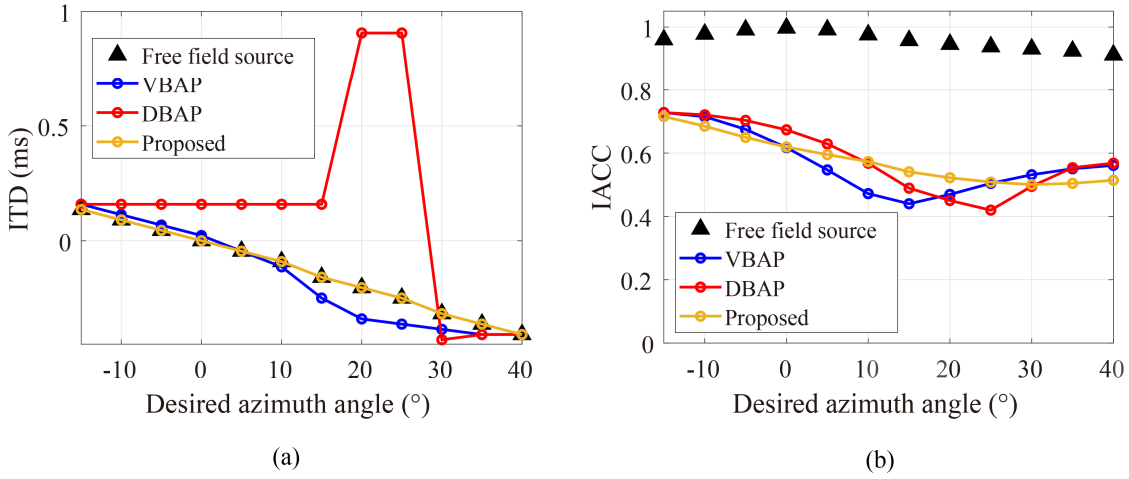

Figure 7. (**a**) ITD and (**b**) IACC of the virtual sound sources reproduced by three methods at various desired azimuth angles in simulation.

4. Experiments

4.1. Objective Experiments

The objective experiments were conducted to validate the effectiveness of the proposed algorithm, using pink noise low-pass filtered at 9 kHz as the input signal. Signal processing was performed using a laptop, and the stimuli were delivered to the loudspeakers through the Fireface UC sound card and a power amplifier. Binaural signals were recorded using a dummy head for analysis. During the audio generation process, the VBAP algorithm first applies a gain to the signal sent to the left loudspeaker and performs a time shift to compensate for the direct sound. The gain pair corresponding to the desired azimuth angle is then applied to the monophonic audio signal to produce a two-channel output. Similarly, the DBAP algorithm directly applies the computed gain pair to the monophonic signal to generate the two-channel audio. The proposed method first filters the monophonic audio signal using inverse filters designed for the left and right loudspeakers, respectively. A time shift is then applied to the left channel, followed by the application of the gain pair to produce the final two-channel output. Compared to the VBAP and DBAP algorithms, the main additional computational cost of the proposed method lies in the two-channel filtering process. In this experiment, each inverse filter is implemented with 1000 taps. For real-time processing applications, the filter length can be reduced through central truncation or the use of higher-order octave-band smoothing.

Figure 8a shows the ITD results for the three methods. After compensating for direct sound, the VBAP algorithm reproduces virtual sound sources with minimal ITD errors (0.02 ms, equivalent to one sample time delay) and corresponding azimuth angle errors within 5° for desired azimuth angles ranging from −15° to 10° and at 40°. However, within the range of 15° to 35°, the ITDs of the virtual sound source are demonstrably below the target ITDs. Specifically, at 20° and 25°, the ITD error measures 0.09 ms (four time samples), with associated azimuth errors approaching 10°. In the case of the DBAP algorithm, the ITDs remain relatively constant at 0.14 ms for azimuth angles spanning from −15° to 20°. However, at 25°, the ITD undergoes an abrupt transition, and for azimuth angles of 30° and beyond, the ITDs register at −0.41 ms. In contrast, the proposed algorithm reproduces virtual sound sources with ITD values that closely approximate the target ITD. For azimuth angles below 10° and at 40°, the ITDs of the virtual sound source align favorably with those of the real sound source. At other azimuth angles, the ITD errors are minimal at 0.02 ms (a single time sample), with angle errors within 5°. Compared to the other two methods, the ITDs generated by the proposed algorithm exhibit superior agreement with those of the real sound source. These experimental findings suggest that, notwithstanding direct sound compensation, the VBAP algorithm still produces ITD deviations in irregular loudspeaker placements, with errors exceeding 5° at certain azimuth angles. This phenomenon arises because of the VBAP algorithm's inherent limitation of disregarding the unique acoustic environment and treating the loudspeaker configuration as a standard setup. The DBAP algorithm, lacking time delay adjustments, yields two distinct ITD values. The ITD of the virtual sound source shifts directly between these two values within a small angle range, failing to produce intermediate angles characterized by matching ITDs. This behavior mirrors the simulation in the proposed algorithm when time delay values fall outside a reasonable range. This implies that, for irregular loudspeaker placements, the time delay must be carefully adjusted to ensure that sounds from both loudspeakers arrive at the listening area nearly simultaneously, thereby enabling virtual sound source creation through amplitude panning. The proposed method can accurately reproduce virtual sources by incorporating ITD into the gain calculation using measured BRIRs.

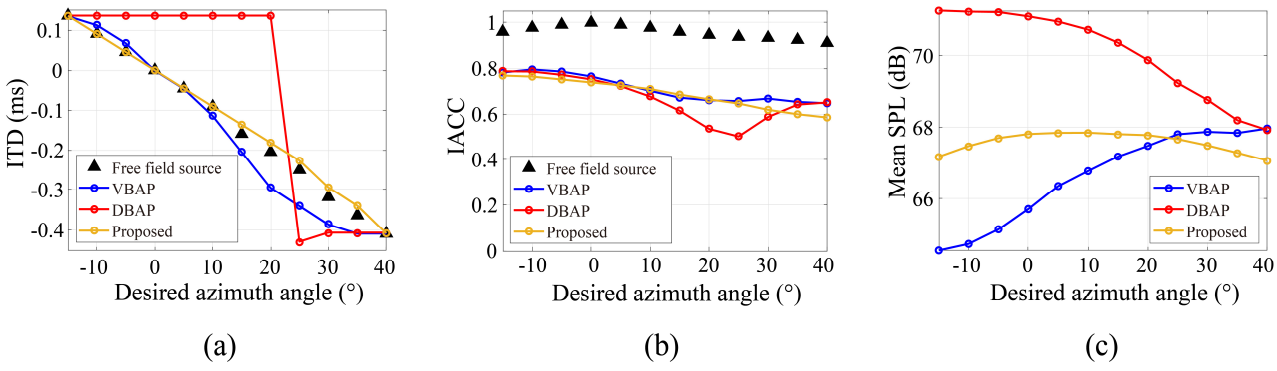

Figure 8. (**a**) ITD, (**b**) IACC, and (**c**) mean SPL of the virtual sound sources reproduced by three methods at various desired azimuth angles in the experiment.

Figure 8b shows the IACC values of virtual sound sources reproduced by the three methods at various azimuth angles. The IACC of the real sound source in a free field approaches unity. In the listening room, acoustic reflections induce a reduction in IACC for all virtual sound sources. The IACC differences among the three methods are minimal, except for the DBAP algorithm, which exhibits lower values at azimuth angles of 20° and 25°. The comparison of mean SPL at the left and right ears for different virtual sound source azimuth angles across the three methods is shown in Figure 8c. In the VBAP algorithm, the mean SPL tends to increase with the azimuth angle because the left

loudspeaker, being closer to the listener, generates a higher direct sound pressure level than the right loudspeaker. However, the reflected sound and reverberation levels from both loudspeakers are similar. After compensation for direct sound, the reflected and reverberated pressure levels fall below those of the right loudspeaker, thereby reducing the total SPL from the left. Consequently, as the azimuth angle increases, the contribution from the right loudspeaker grows, leading to an overall increase in the SPL. In the DBAP algorithm, the left loudspeaker produces a higher mean SPL than the right loudspeaker due to disparities in distance. As the virtual sound source moves from left to right, the distance to the listening position increases, causing the mean SPL to decrease with the increasing azimuth angle. In the proposed method, the mean SPLs of both loudspeakers are equalized during the compensation stage. As a result, the mean SPL remains relatively stable across different azimuth angles, ensuring perceptually consistent loudness as the virtual sound source moves.

The VBAP and DBAP methods assume that each loudspeaker behaves as an omnidirectional point source in a free field, thereby overlooking frequency-dependent factors such as room modes, loudspeaker directivity, and inter-loudspeaker variability. These methods rely on direction- or distance-based approximations, which are insufficient to ensure accurate ITD reproduction across arbitrary virtual source azimuths and irregular loudspeaker configurations. In contrast, the proposed method predicts the ITD corresponding to each gain pair and incorporates the acoustic characteristics of the listening environment by leveraging measured RIRs and BRIRs in the gain computation. The algorithm adaptively adjusts the inter-loudspeaker gains based on the resulting ITD and IACC values. Furthermore, inverse filtering of the RIRs, which capture both direct and reflected components, including reverberation, ensures that the two loudspeakers exhibit matched magnitude frequency responses at the listening position. As a result, the virtual sound source maintains a consistent mean SPL across a wide range of desired azimuth angles.

4.2. Subjective Experiments

Subjective localization tests were conducted to evaluate the perceived azimuth angles of virtual sources. Stereo audio was generated using three distinct methods and four types of audio content: pink noise low-pass filtered at 1.5 kHz (10 s), pink noise low-pass filtered at 9 kHz (10 s), male and female conversation (6 s), and a song clip from 'The Show' by Lenka [34] (10 s). The mean SPL of each method was calibrated to 70 dB. Based on the findings from the objective experiments, the experimental azimuth angles were set at $-10°$, $0°$, $10°$, $20°$, $25°$, $30°$, $35°$, and $40°$, presented in random order. A total of 12 participants with clinically normal hearing (10 males, 2 females, aged 22 to 27 years) took part in the subjective experiments. Participants were instructed to sit in a chair at the listening position, close their eyes prior to audio playback to mitigate visual cues from the loudspeakers, and then identify the location of the virtual sound source after the audio played in loops. During the experiments, participants were permitted to rotate their heads to improve localization of the virtual sound source. Once the azimuth was determined, they opened their eyes and used a laser pointer to indicate the perceived direction of the virtual sound source. Subsequently, the experimenter used the scales in front of the participants to calculate the azimuth angle based on the position indicated by the laser pointer using geometric relationships. Prior to the formal test, participants underwent training to familiarize themselves with the procedure.

To assess the efficacy of the proposed method, the experimental results are evaluated using statistical analysis. The manipulated variables comprise the desired azimuth angles, audio content types and processing methods, with the perceived azimuth angle serving as the dependent variable. As initial assessments revealed that the data did not conform

to a normal distribution, a non-parametric method is employed. Specifically, Friedman's tests are conducted to examine the effects of method and audio content. The Friedman test shows a statistically significant difference between the three methods (χ^2 = 394.4, $p < 0.0001$). Furthermore, a significant difference is also found when analyzing the effects of audio content as a variable (χ^2 = 61.8, $p < 0.0001$). These results, stratified by azimuth angles and audio content types, are shown in Figure 9, offering a comprehensive overview of the localization performance of the three methods. The black dots in the figure represent the desired angle values corresponding to each azimuth angle. The results for the pink noise filtered at 1.5 kHz are shown in Figure 9a. The VBAP method yields perceived angles close to the desired values at −10°, 0°, 35°, and 40°, with a median azimuth error within 5°. However, at 10°, 20°, 25°, and 30°, the perceived angles deviate by more than 5°, with the largest error of 9° occurring at 25°. The perceived direction of the virtual sound source reproduced by the DBAP method significantly deviates from the desired direction. Although the perceived angles are within 5° of the desired angles at −10°, 35°, and 40°, the errors exceed 14° at other angles. The perceived azimuth angle in the DBAP method tends to bias toward the directions of the two loudspeakers. Below 20°, it collapses near the left loudspeaker, and above 35°, it collapses near the right loudspeaker. The virtual sound source is successfully reproduced only at 25° and 30°, with significant perceptual differences observed among subjects at 30°. Additionally, some participants also reported difficulty in identifying the direction of the virtual sound source during the experiments. The perceived angle results for the proposed method show that the median perceived azimuth angles closely aligned with the desired angles, with errors consistently remaining within 3°, thereby outperforming both the VBAP and DBAP methods. The results obtained with the other three audio content types are qualitatively similar to those observed with the 1.5 kHz low-pass filtered pink noise. For the VBAP method at 20°, 25°, and 30°, the perceived angles deviate by more than 5° from the desired angles, except for the conversation at 20° and 25°. In the case of the 9 kHz low-pass filtered pink noise, the largest error occurs at 20°, with an error of 11°. The perception of virtual sound sources using DBAP tends to shift toward the two loudspeakers, with a significant variation observed among subjects at 30°. The proposed method consistently results in perceived azimuth angle errors within 5° across all audio content types and azimuth angles, with the largest error being 5° at the 10° azimuth angle for the song clip. Wilcoxon signed-rank tests are conducted between the proposed method and the other methods. Significant differences are observed between the proposed method and VBAP in the 10° to 30° range for all four audio content types ($p < 0.05$ for the song at 20° and $p < 0.01$ for the other cases), except for the 9 kHz low-pass filtered pink noise at 10°. Significant differences are also found between the proposed method and DBAP at all azimuth angles for all four audio content types ($p < 0.01$), except at 35° and 40°. These results suggest that the proposed method improves localization accuracy compared with both VBAP and DBAP.

The subjective test results for azimuth angles are consistent with the ITD results from the objective experiments. For the VBAP method, the virtual sound sources in the 20° to 30° azimuth range exhibit both ITDs and perceived azimuth angles that deviate from the desired values. The virtual sound sources generated by the DBAP method are biased toward the left or right loudspeaker, failing to reproduce stable sound images in the region between the loudspeakers, which aligns with the segmented characteristics of the ITD. The proposed method consistently ensures a close correspondence between the perceived and desired angles. This result suggests that, although the ITD is calculated based on the received signals below 1500 Hz, the proposed method still achieves accurate virtual sound source localization for wideband audio. This indicates that the ITD parameters exert a dominant influence on the subjective localization process.

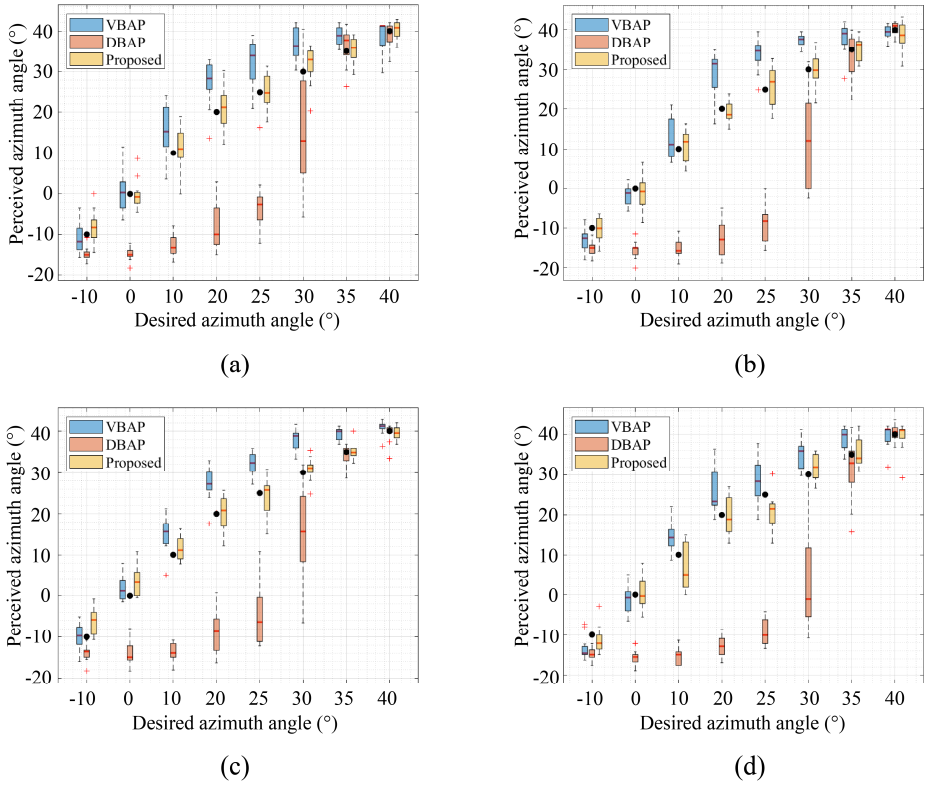

Figure 9. Perceived azimuth angles of virtual sources reproduced by three methods of (**a**) pink noise low-pass filtered at 1.5 kHz, (**b**) pink noise low-pass filtered at 9 kHz, (**c**) male and female conversation, and (**d**) song clip.

5. Conclusions

This paper proposes an adaptive amplitude panning algorithm designed to accurately reproduce the virtual sound source at the desired azimuth angle in irregular loudspeaker configurations and then implements it within a stereo system. The method leverages measured RIRs and BRIRs to effectively incorporate environmental information. Inverse filters are applied to the loudspeakers to improve their similarity. Subsequently, ITD and IACC are employed to optimize gain and time delay parameters. The method was validated through both objective and subjective experiments. The objective evaluation confirmed the accurate reproduction of the virtual sound source with the target ITD, achieving an error margin within 0.02 ms. Correspondingly, the subjective results demonstrated that the proposed method can reproduce virtual sound sources with a maximum angular deviation of 5°.

Author Contributions: Conceptualization, S.Z. and Y.Z.; resources, Z.L.; writing—original draft preparation, S.Z. and Y.Z.; writing—review and editing, Z.L.; supervision, Z.L.; funding acquisition, Z.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China Grant No.12274221.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: We would like to thank Jingru Yi for her assistance with the experimental design.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Spors, S.; Wierstorf, H.; Raake, A.; Melchior, F.; Frank, M.; Zotter, F. Spatial sound with loudspeakers and its perception: A review of the current state. *Proc. IEEE* **2013**, *101*, 1920–1938. [CrossRef]
2. Berkhout, A.J. A holographic approach to acoustic control. *J. Audio Eng. Soc.* **1988**, *36*, 977–995.
3. Berkhout, A.J.; de Vries, D.; Vogel, P. Acoustic control by wave field synthesis. *J. Acoust. Soc. Am.* **1993**, *93*, 2764–2778. [CrossRef]
4. Rabenstein, R.; Spors, S. Spatial aliasing artifacts produced by linear and circular loudspeaker arrays used for wave field synthesis. In Proceedings of the Audio Engineering Society Convention 120, Paris France, 20–23 May 2006.
5. Spors, S.; Rabenstein, R.; Ahrens, J. The theory of wave field synthesis revisited. In Proceedings of the 124th AES convention, Amsterdam, The Netherlands, 17–20 May 2008; pp. 17–20.
6. Firtha, G.; Fiala, P.; Schultz, F.; Spors, S. Improved referencing schemes for 2.5 D wave field synthesis driving functions. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2017**, *25*, 1117–1127. [CrossRef]
7. Ahrens, J.; Spors, S. Analytical driving functions for higher order ambisonics. In Proceedings of the 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, USA, 31 March–4 April 2008; pp. 373–376.
8. Ward, D.B.; Abhayapala, T.D. Reproduction of a plane-wave sound field using an array of loudspeakers. *IEEE Trans. Speech Audio Process.* **2001**, *9*, 697–707. [CrossRef]
9. Betlehem, T.; Abhayapala, T.D. Theory and design of sound field reproduction in reverberant rooms. *J. Acoust. Soc. Am.* **2005**, *117*, 2100–2111. [CrossRef]
10. Kirkeby, O.; Nelson, P.A. Reproduction of plane wave sound fields. *J. Acoust. Soc. Am.* **1993**, *94*, 2992–3000. [CrossRef]
11. Poletti, M. Robust two-dimensional surround sound reproduction for nonuniform loudspeaker layouts. *J. Audio Eng. Soc.* **2007**, *55*, 598–610.
12. Franck, A.; Wang, W.; Fazi, F.M. Sparse l_1-Optimal Multiloudspeaker Panning and Its Relation to Vector Base Amplitude Panning. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2017**, *25*, 996–1010. [CrossRef]
13. Pulkki, V.; Karjalainen, M. Localization of amplitude-panned virtual sources I: Stereophonic panning. *J. Audio Eng. Soc.* **2001**, *49*, 739–752.
14. Pulkki, V. Localization of amplitude-panned virtual sources II: Two-and three-dimensional panning. *J. Audio Eng. Soc.* **2001**, *49*, 753–767.
15. Pulkki, V. Spatial Sound Generation and Perception by Amplitude Panning Techniques. Ph.D. Thesis, Helsinki University of Technology, Espoo, Finland, 2001.
16. Yu, Z.; Zhu, Q.; Wu, M.; Yang, J. Exploring the limits of virtual source localization with amplitude panning on a flat panel with actuator array: Implications for future research. *J. Acoust. Soc. Am.* **2023**, *154*, 1362–1371. [CrossRef]
17. Herre, J.; Hilpert, J.; Kuntz, A.; Plogsties, J. MPEG-H audio—The new standard for universal spatial/3D audio coding. *J. Audio Eng. Soc.* **2015**, *62*, 821–830. [CrossRef]
18. Pulkki, V. Uniform spreading of amplitude panned virtual sources. In Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. WASPAA'99 (Cat. No. 99TH8452), New Paltz, NY, USA, 20–20 October 1999; pp. 187–190.
19. Lossius, T.; Baltazar, P.; de la Hogue, T. DBAP–distance-based amplitude panning. In Proceedings of the ICMC, Montreal, QC, Canada, 16–21 August 2009.
20. Zotter, F.; Frank, M. All-round ambisonic panning and decoding. *J. Audio Eng. Soc.* **2012**, *60*, 807–820.
21. Franck, A.; Fazi, F.M.; Hamdan, E. An optimization approach to control sound source spread with multichannel amplitude panning. In Proceedings of the 24th International Congress on Sound and Vibration, London, UK, 23–27 July 2017.
22. Franck, A.; Fazi, F.M. Towards constant-spread panning for multi-loudspeaker configurations. In Proceedings of the Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio, York, UK, 27–29 March 2019.
23. Menzies, D.; Galvez, M.F.S.; Fazi, F.M. A low-frequency panning method with compensation for head rotation. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2017**, *26*, 304–317. [CrossRef]
24. Menzies, D.; Fazi, F.M. A complex panning method for near-field imaging. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2018**, *26*, 1539–1548. [CrossRef]
25. Menzies, D.; Fazi, F.M. Multichannel Compensated Amplitude Panning, An Adaptive Object-Based Reproduction Method. *J. Audio Eng. Soc.* **2019**, *67*, 549–556. [CrossRef]
26. Laitinen, M.-V.; Vilkamo, J.; Jussila, K.; Politis, A.; Pulkki, V. Gain normalization in amplitude panning as a function of frequency and room reverberance. In Proceedings of the Audio Engineering Society Conference: 55th International Conference: Spatial Audio, Helsinki, Finland, 27–29 August 2014.

27. Hughes, R.J.; Franck, A.; Cox, T.J.; Shirley, B.G.; Fazi, F.M. Dual frequency band amplitude panning for multichannel audio systems. In Proceedings of the Audio Engineering Society Conference: 2018 AES International Conference on Spatial Reproduction-Aesthetics and Science, Tokyo, Japan, 7–9 August 2018.
28. Galvez, M.F.S.; Menzies, D.; Mason, R.; Fazi, F.M. Object-based audio reproduction using a listener-position adaptive stereo system. *J. Audio Eng. Soc.* **2016**, *64*, 740–751. [CrossRef]
29. Hanschke, J.-H.; Arteaga, D.; Cengarle, G.; Lando, J.; Thomas, M.R.; Seefeldt, A. Improved Panning on Non-Equidistant Loudspeakers with Direct Sound Level Compensation. *arXiv* **2023**, arXiv:2310.17004.
30. Bai, M.R.; Lee, C.-C. Comparative study of design and implementation strategies of automotive virtual surround audio systems. *J. Audio Eng. Soc.* **2010**, *58*, 141–159.
31. Hatziantoniou, P.D.; Mourjopoulos, J.N. Generalized fractional-octave smoothing of audio and acoustic responses. *J. Audio Eng. Soc.* **2000**, *48*, 259–280.
32. Mourjopoulos, J.N.; Hatziantoniou, P.D. Real-time room equalization based on complex smoothing: Robustness results. In Proceedings of the Audio Engineering Society Convention 116, Berlin, Germany, 8–11 May 2004.
33. Rakerd, B.; Hartmann, W.M. Localization of sound in rooms. V. Binaural coherence and human sensitivity to interaural time differences in noise. *J. Acoust. Soc. Am.* **2010**, *128*, 3052–3063. [CrossRef] [PubMed]
34. Kripac, L.; Reeves, J. The Show. In *Lenka*; Epic Records: New York, NY, USA, 2008.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.