Audio Engineering Society

Convention Express Paper 385

Presented at the AES 159th Convention
2025 October 23–25, Long Beach, CA, USA

This Express Paper was selected on the basis of a submitted synopsis that has been peer-reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This Express Paper has been reproduced from the author's advance manuscript without editing, corrections or consideration by the Review Board. The AES takes no responsibility for the contents. This paper is available in the AES E-Library (http://www.aes.org/e-lib) all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Optimized Loudspeaker Panning for Adaptive Sound-Field Correction and Non-stationary Listening Areas

Yuancheng Luo

Audio Technology, Amazon Inc.

Correspondence should be addressed to Yuancheng Luo (luoyuancheng@gmail.com)

ABSTRACT

Surround sound systems commonly distribute loudspeakers along standardized layouts for multichannel audio reproduction. However in less controlled environments, practical layouts vary in loudspeaker quantity, placement, and listening locations / areas. Deviations from standard layouts introduce sound-field errors that degrade acoustic timbre, imaging, and clarity of audio content reproduction. This work introduces both Bayesian loudspeaker normalization and content panning optimization methods for sound-field correction. Conjugate prior distributions over loudspeaker-listener directions update estimated layouts for non-stationary listening locations; digital filters adapt loudspeaker acoustic responses to a common reference target at the estimated listening area without acoustic measurements. Frequency-domain panning coefficients are then optimized via sensitivity / efficiency objectives subject to spatial, electrical, and acoustic domain constraints; normalized and panned loudspeakers form virtual loudspeakers in standardized layouts for accurate multichannel reproduction. Experiments investigate robustness of Bayesian adaptation, and panning optimizations in practical applications.

1 Introduction

Surround sound systems for multichannel audio reproduction have risen in popularity in home theater setups that accommodate proper loudspeaker selection, layout, acoustic room treatment, and calibration established by the international telecommunication union (ITU) standards [1]. Conversely, the same accommodations present a barrier to entry for extemporary arrangements where loudspeakers differ in quality and placement, and operate in changeable listening locations / areas, and reverberant environments. Deviating from the standards degrade accurate reproduction of multichannel audio content as intended by the content authors. Therefore, methods from sound-field control and reconstruction correct for the effects of irregular loudspeaker placements and room reverberation in the listening area via acoustic measurement system inversion [2, 3, 4], and modal / planewave decomposition [5, 6, 7]; such methods however are inapplicable when acoustic measurements remain unavailable.

In the absence of acoustic measurements, other sensing modalities can infer the loudspeaker layout and listening area location. Inertial measurement unit [8, 9] and bluetooth low energy [10, 11] indoor tracking can estimate changes in loudspeaker position and orientation. Ultrasound [12], camera, and video can track in-room listener and loudspeaker positions within fields-of-view. Such meta-data yields a 2D layout of the estimated loudspeaker placements, listening location, and a front direction. We therefore reproduce multichannel content at the listener's area by incorporating Bayesian uncertainty of the estimated layout inputs with loudspeaker distance and orientation normalization [13, 14] to the listener, and then reformulate conventional amplitude panning methods [15, 16, 17] in terms of constrained optimization along joint spatial [18, 19], electrical [20], and acoustical [21] domains. The paper is organized as follows:

Section 2 introduces our normalization method for aligning loudspeaker acoustic transfer functions in an arbitrary layout to a common axial-reference target at the listener location; acoustic delay and attenuation compensate for varying loudspeaker-listener distances whereas minimum-phase and all-pass factorizations [22] normalize for loudspeaker orientations relative to listener locations. We integrate estimates of the loudspeaker-listener normalization directions via Bayesian posterior updates of a novel circular distribution conjugate prior, and provide a sample calibration for a

sequence of normalization angles. **Section 3** presents our novel normalized loudspeaker panning optimization, which solves for frequency-dependent magnitude-gains that satisfy spatial vector-bases, electrical headroom, and acoustic power constraints; we augment the former vector-base amplitude panning with slack (VBAPS) to accommodate constraints in electric and acoustic domains. Next, we derive a panning sensitivity / efficiency objective from the augmented form that measures panned-source discreteness, and give equivalent primary and null-space formulations in fewer variables. Planewave acoustic covariances model anechoic to diffuse-field assumptions for variable sized listening areas. Optimal solutions are found via second-order cone program [23]. **Section 4** applies our model to several practical applications of loudspeaker correction under varying constraints. For high loudness targets, we find optimal gains across loudspeakers for overdriven content that maximize source discreteness. For anechoic to diffuse-field environments, we show that our panning optimization solutions converge from discrete panning to Rayleigh quotient maximizers [24]. For circular-panning over varying loudspeaker layouts, we evaluate panning sensitivity across azimuth steering-angles and recommend preferred layouts for different number of loudspeakers. **Section 5** discusses results and future work.

2 Loudspeaker Normalization

Let $S(\nu,\theta)$ be the loudspeaker's electrical-acoustical transfer function at frequency ν measured at 1 meter distance along azimuth θ (radians) in the horizontal plane, with the acoustic path-delay removed. Under far-field assumptions, the loudspeaker frequency response attenuates by the inverse-distance and undergoes pure-delay. It is useful to express the far-field transfer function along a listener-centric coordinate frame, which centers the origin at the listener's location and aligns the $+x$ axis with the listener's facing direction. The acoustic transfer function $H_n(\kappa,\boldsymbol{r})$ at coordinate $\boldsymbol{r}\in\mathbb{R}^{2\times1}$ for the n^{th} loudspeaker located at coordinate $\boldsymbol{u}_n\in\mathbb{R}^{2\times1}$ with the orientation unit-vector $\boldsymbol{o}_n\in\mathbb{R}^{2\times1}$ follows

$$\begin{aligned} H_n(\nu,\boldsymbol{r}) &= S(\nu,\theta_n(\boldsymbol{r}))\,\frac{e^{-j\kappa\|\boldsymbol{s}_n(\boldsymbol{r})\|}}{\|\boldsymbol{s}_n(\boldsymbol{r})\|}, \quad \kappa=\frac{2\pi\nu}{c}, \\ \theta_n(\boldsymbol{r}) &= \cos^{-1}\left(\frac{\boldsymbol{o}_n^T\boldsymbol{s}_n(\boldsymbol{r})}{\|\boldsymbol{s}_n(\boldsymbol{r})\|}\right), \quad \boldsymbol{s}_n(\boldsymbol{r})=\boldsymbol{r}-\boldsymbol{u}_n, \end{aligned} \tag{1}$$

where κ is the angular wavenumber, c is the speed of sound in meters/second, $\boldsymbol{s}_n(\boldsymbol{r})$ is the evaluation direction relative to the loudspeaker's location, and $\theta_n(\boldsymbol{r})$ is the evaluation angle relative to the loudspeaker's orientation. We can normalize the loudspeaker's transfer function to approximate the original loudspeaker's response $S(\nu,\theta)$ within a listening window at the listener's location $\boldsymbol{r}=\boldsymbol{0}$.

Consider the following decomposition of the loudspeaker transfer function $S(\nu,\theta)=S_E(\nu)S_A(\nu,\theta)$ into acoustical and electrical domain transfer functions $S_A(\nu,\theta)$ and $S_E(\nu)$ respectively. A filter with frequency response $G_n(\nu)$ that normalizes (1) to the loudspeaker's on-axis response $H_n(\nu,\boldsymbol{0})G_n(\nu)=S(\nu,0)$ is given by

$$G_n(\nu)=Q_n(\nu)\,\|\boldsymbol{u}_n\|\,e^{j\kappa\|\boldsymbol{u}_n\|}, \quad Q_n(\nu)=\frac{S_A(\nu,0)}{S_A\left(\nu,\bar{\boldsymbol{\theta}}_n\right)}, \tag{2}$$

where $\bar{\boldsymbol{\theta}}_n=\theta_n(\boldsymbol{0})$ is the normalization angle between the loudspeaker's orientation and the listener. The electrical domain term $S_E(\nu)$ cancels within the quotient $Q_n(\nu)$ in (2), thereby negating prior signal processing in loudspeaker playback. $Q_n(\nu)$ is therefore the acoustic relative-transfer-function between loudspeaker's axial and listener-direction acoustic responses. Moreover, if $S_A(\nu,\theta)$ share a common acoustic delay and the remainder is minimum-phase for bounded θ that define a listening window, then $Q_n(\nu)$ must also be minimum-phase. Thus, the normalized transfer function $G_n(\nu)$ compensates for both loudspeakers' orientation and distance relative to the listener as shown in Fig. 1.

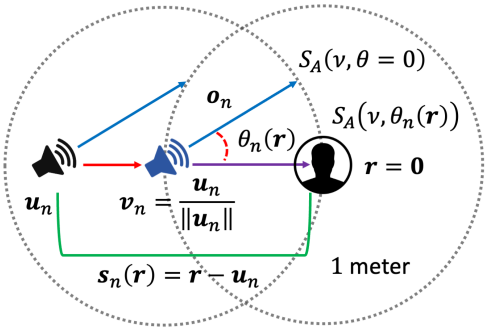


Fig. 1: Acoustic transfer function $G_n(\nu)$ in (2) normalizes the direct acoustic path between the listener and loudspeaker at $\boldsymbol{u}_n$ to be its on-axis response $S_A(\nu,0)$ at the normalized coordinate $\boldsymbol{v}_n$.

In practice, we can find the rational function approximation [25, 26] to $Q_n(\nu)$, expressed in terms of minimum-phase $\mathbb{M}_n(\nu)$ and all-pass $\mathbb{A}_n(\nu)$ transfer functions given by

$$Q_n(\nu)\approx\mathbb{M}_n(\nu)\mathbb{A}_n(\nu), \quad \mathbb{A}_n(\nu)=\bar{\mathbb{A}}_n(\nu)\ddot{\mathbb{A}}_n(\nu), \tag{3}$$

where $\bar{\mathbb{A}}_n(\nu)$ and $\ddot{\mathbb{A}}_n(\nu)$ are all-pass transfer functions belonging to stable and unstable components respectively. The unstable all-pass $\ddot{\mathbb{A}}_n(\nu)$ contains the reciprocal poles and zeros of the Padé approximant outside the complex unit-circle, and is ideally empty or low-order for θ in the listening window. We can realize a causal-stable filter-response $G_n(\nu)$ for an all-passed loudspeaker transfer function in (2) as follows:

$$\begin{aligned} H_n(\nu,\boldsymbol{0})G_n(\nu) &= S(\nu,0)\frac{e^{-j\kappa d}}{\ddot{\mathbb{A}}_{lcm}(\nu)} \quad \Rightarrow \\ G_n(\nu) &= \mathbb{M}_n(\nu)\bar{\mathbb{A}}_n(\nu)\frac{\ddot{\mathbb{A}}_n(\nu)}{\ddot{\mathbb{A}}_{lcm}(\nu)}\,\|\boldsymbol{u}_n\|\,e^{j\kappa\|\boldsymbol{u}_n-d\|}, \end{aligned} \tag{4}$$

where $d = \max_{1\le n\le N} \|\boldsymbol{u}_n\|$ is the furthest loudspeaker distance, and $\ddot{\mathbb{A}}_{lcm}(\nu)$ is the transfer function of the set of least common multiple (LCM) reciprocal poles and zeros across the unstable all-passes $\{\ddot{\mathbb{A}}_1(\nu),\dots,\ddot{\mathbb{A}}_N(\nu)\}$. In the z-domain, we can therefore express the all-pass and LCM transfer functions as follows:

$$\begin{aligned}\ddot{\mathbb{A}}_n(z) &= \prod_{p\in P_n}\left(\frac{1-p^*z}{1-pz^{-1}}\right)^{k_{pn}}, \quad P_n = \{p_{1n},\dots,p_{M_n n}\},\\ \ddot{\mathbb{A}}_{lcm}(z) &= \prod_{p\in P}\left(\frac{1-p^*z}{1-pz^{-1}}\right)^{\max\limits_{1\le n\le N} k_{pn}}, \quad P = \cup_{n=1}^{N} P_n,\end{aligned} \tag{5}$$

where p^* is the conjugate transpose, and P_n is the set of unique poles and k_{pn} is the multiplicity of pole p for the n^{th} loudspeaker. By taking the maximum multiplicity for each unique and unstable pole across all $\ddot{\mathbb{A}}_n(z)$, and dividing by the subsequent LCM $\ddot{\mathbb{A}}_{lcm}(z)$, the unstable poles in $\ddot{\mathbb{A}}_n(z)$ cancel and the remaining all-pass adds minimal additional group-delay in $G_n(\nu)$. The filtered loudspeakers' direct paths are thus matched with a common all-passed on-axis response. Lastly, we gain the loudspeaker filter $G_n(\nu)$ to match the expected acoustic power at a common distance D, such as the median of all loudspeakers-to-listener distances, via the following room acoustic attenuation model:

Let us consider the inverse-distance law $\rho_{DP}(r) = \bar{\rho} r^{-2}$ for the attenuation of the direct acoustic path response's nominal power $\bar{\rho}$ at distance r from a loudspeaker. In a room environment, let $\rho_{IP}(r)$ be the total power of indirect acoustic paths at distance r. We can model the ratio of the direct-to-indirect acoustic path's power at r and total power as follows:

$$\begin{aligned}\frac{\rho_{DP}(r)}{\rho_{IP}(r)} &= \left(\frac{d_c}{r}\right)^{2\beta}, \quad \beta = 10^{\frac{\gamma\,\mathrm{dB/dd}}{10}}, \quad \text{Attenuation rate}\\ \rho(r) &= \rho_{DP}(r) + \rho_{IP}(r) = \bar{\rho} r^{-2}\left(1+\left(\frac{d_c}{r}\right)^{2\beta}\right),\end{aligned} \tag{6}$$

where d_c is the so-called *critical distance* (meters) where the direct and indirect acoustic powers are equivalent, and β a decay-rate parameterized by γ decibels (dB) per double-distance (dd); typical $\gamma \in \{0,-3\}$ and $0.5 \le d_c \le 1.5$ span idealized concert-hall to small-room spaces [27]. Normalizing the power at distance r to D therefore follows

$$F(r, D, d_c) = \sqrt{\frac{\rho(D)}{\rho(r)}} = \frac{r}{D}\sqrt{\frac{d_c^{2\beta} + D^{2\beta}}{d_c^{2\beta} + r^{2\beta}}}, \tag{7}$$

whereby substituting $\|\boldsymbol{u}_n\|$ with $F(\|\boldsymbol{u}_n\|, D, d_c)$ in (4) compensates for loudspeaker distances to the listener in a room.

Model Uncertainty for Non-stationary Targets: In instances where the listener's location changes over time or require online estimation, we normalize the loudspeaker via the mean listener distance $\frac{1}{T}\int_0^T \|\boldsymbol{u}_n(t)\|\,dt$, and treat the normalization angle $\bar{\boldsymbol{\theta}}_n$ relative to the loudspeaker orientation $\boldsymbol{o}_n$ in (2) as a random variable. The target transfer function $G_n(\nu)$ and quotient term $Q_n(\nu)$ are re-defined to minimize the expected squared-differences between the anechoic responses $S_A(\nu,\theta)$ sampled over axial-centered and loudspeaker-listener centered circular probability distribution functions (PDFs) $f_0(\theta)$ and $f_n(\theta)$, $\forall 1 \le n \le N$ respectively; circular PDFs satisfy $f(\theta) = f(\theta + 2\pi k)$, $\forall k \in \mathbb{Z}$. We present two acoustic averages:

$$\begin{aligned}\bar{S}_A(\nu, f(\theta)) &= \mathbb{E}\left[S_A(\nu,\theta)\right] = \int S_A(\nu,\theta) f(\theta) d\theta,\\ \hat{S}_A(\nu, f(\theta)) &= \mathbb{E}\left[|S_A(\nu,\theta)|^2\right] = \int |S_A(\nu,\theta)|^2 f(\theta) d\theta,\end{aligned} \tag{8}$$

where $\bar{S}_A(\nu, f(\theta))$ and $\hat{S}_A(\nu, f(\theta))$ are spatial windowed averages of the acoustic response and power respectively; axial window response average $\bar{S}_A(\nu, f_0(\theta))$ and power average $\hat{S}_A(\nu, f_0(\theta))$ sample from the $f_0(\theta)$ distribution. The modified quotient term $Q_n(\nu)$ in (2) is replaced with the weighted least-squares minimizer of $\arg\min_X \int \left|S_A(\nu,\theta)X - \bar{S}_A(\nu, f_0(\theta))\right|^2 f_n(\theta) d\theta$ given by

$$\bar{Q}_n(\nu) = \bar{S}_A(\nu, f_0(\theta)) \frac{\bar{S}_A^*(\nu, f_n(\theta))}{\hat{S}_A(\nu, f_n(\theta))}, \tag{9}$$

where $\bar{S}_A^*(\nu, f_n(\theta))$ is the conjugate transpose, and $\bar{Q}_n(\nu)$ accounts for both amplitude and phase differences in the averaged responses. The analogous quotient for the spatial windowed acoustic power average follows

$$\hat{Q}_n(\nu) = \sqrt{\frac{\hat{S}_A(\nu, f_0(\theta))}{\hat{S}_A(\nu, f_n(\theta))}}, \tag{10}$$

where $\hat{Q}_n(\nu)$ has zero-phase and therefore compensates for only the amplitude. Both quotients can be efficiently evaluated if $f_0(\theta)$, $f_n(\theta)$ are both uni-modal and smooth over azimuth, have expansions along a common orthogonal basis with $S_A(\nu,\theta)$, and follow the contours of a listening window.

Let us consider the circular distribution $f(\theta)$ defined by the squared-exponential of the chordal distance $d(\theta)$ on a unit-disk, which along with $S_A(\nu)$ has a series-expansion over the Legendre polynomials [28], and normalized over the domain of all azimuth angles $-\pi \le \theta \le \pi$:

$$f(\theta) = \frac{e^{\frac{-d^2(\theta-\mu)}{2\ell^2}}}{2\pi e^{-\ell^{-2}} J_0(j\ell^{-2})}, \quad d(\theta) = 2\sin\left(\frac{\theta \bmod 2\pi}{2}\right), \tag{11}$$

where J_0 is the Bessel function of the first kind, μ is the mean azimuth, and ℓ is the dispersion. The function is symmetric w.r.t. its maximum $f(\mu)$ and minimum $f(\mu \pm \pi)$, infinitely differentiable in all azimuths, and its percentiles computable via series expansion in appendix (26). Large dispersion ℓ gives a uniform distribution as $\lim_{\ell\to\infty} f(\theta) = (2\pi)^{-1}$; small dispersion gives the dirac distribution as $\lim_{\ell\to 0} f(\theta-\mu) = \delta$. We can bound the dispersion via design parameters char-

acterizing a listening window's peak such as the full-width at half-maximum (FWHM) measure:

$$\frac{f(\mu)}{2} = f\left(\mu \pm \frac{\text{FWHM}}{2}\right), \quad 0 \le \text{FWMH} \le 2\pi,$$
$$\ell = \frac{2\sin\left(\frac{\text{FWMH}}{4}\right)}{\sqrt{2\ln(2)}} \;\Rightarrow\; 0 \le \ell \le \sqrt{2/\ln(2)}, \tag{12}$$

which defines the angular width where $f(\theta)$ spans half its maximum amplitude as shown in Fig. 2. At the upper-limit FWHM 360°, $f(\theta)$ contains $\{60.9, 33.2, 22.5\}$ % of its mass within the frontal intervals $|\theta - \mu| \le \{90, 45, 30\}^\circ$ respectively. For tighter FWHM $\le 90.22^\circ$ bounds, $f(\theta)$ contains the 95% confidence interval in the half-space $|\theta - \mu| \le 90^\circ$ of its mean azimuth μ. For the axial-centered PDF in (8), we set the window's FWHM to 60° where $f_0(\theta) = f(\theta \,|\, \mu = 0, \ell = 0.4396)$. We now proceed with online adaptation of the normalization angles $\bar{\boldsymbol{\theta}}_n$ over time.

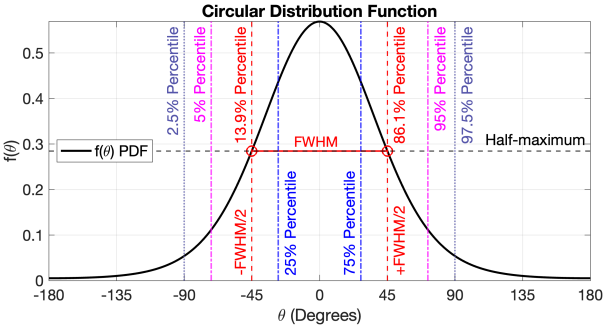


Fig. 2: Circular distribution prior (FWHM 90.2°) contains $\{50, 90, 95\}$ % of normalization angles within $|\theta| \le \{27.4, 72, 90\}^\circ$ of the mean angle.

Suppose we have measured a normalization angle $\bar{\theta}$ belonging to the n^{th} loudspeaker with known measurement dispersion $\bar{\ell}$ such that the likelihood function $f(\bar{\theta} \,|\, \mu = \bar{\boldsymbol{\theta}}_n, \ell = \bar{\ell})$ follows the squared-exponential chordal function in (11). Let the unknown normalization angle $\bar{\boldsymbol{\theta}}_n$ of the n^{th} loudspeaker have a squared-exponential chordal prior-distribution $f(\bar{\boldsymbol{\theta}}_n | \mu = \mu_n, \ell = \ell_n)$ with initial hyperparameters $\mu_n = 0$ mean azimuth and $\ell = \sqrt{2/\ln(2)}$ maximum dispersion. The posterior normalization angle therefore has a conjugate distribution with hyperparameters following appendix (27). Over multiple time-steps t, the likelihood, prior, and posterior functions across measured angles $\bar{\boldsymbol{\theta}}_n^{\{t\}}$ with dispersion $\bar{\ell}_n^{\{t\}}$ are given by

$$L\left(\bar{\boldsymbol{\theta}}_n \,|\, \bar{\boldsymbol{\theta}}_n^{\{t\}}\right) = f\left(\bar{\boldsymbol{\theta}}_n^{\{t\}} \,|\, \mu = \bar{\boldsymbol{\theta}}_n, \ell = \bar{\ell}_n^{\{t\}}\right), \quad \text{Likelihood}$$
$$P(\bar{\boldsymbol{\theta}}_n) = f\left(\bar{\boldsymbol{\theta}}_n \,|\, \mu = \mu_n^{\{t-1\}}, \ell = \ell_n^{\{t-1\}}\right), \quad \text{Prior}$$
$$P\left(\bar{\boldsymbol{\theta}}_n \,|\, \bar{\boldsymbol{\theta}}_n^{\{t\}}\right) \propto L\left(\bar{\boldsymbol{\theta}}_n \,|\, \bar{\boldsymbol{\theta}}_n^{\{t\}}\right) P(\bar{\boldsymbol{\theta}}_n), \quad \text{Posterior} \tag{13}$$

where the reported normalization angle $\bar{\boldsymbol{\theta}}_n^{\{t\}}$ is a point-estimate taken within a measurement session, and the dispersion $\bar{\ell}_n^{\{t\}}$ is proportional to the point-estimate's confidence interval. Both quantities can vary over time as the listener's location may change between sessions (e.g. different seating), and measured under different noise conditions. The initial hyperparameters for mean $\mu_n^{\{0\}} = 0$ and dispersion $\ell_n^{\{0\}} = 0.6515$ (FWHM 90.22°) are informative as loudspeakers generally orient towards the intended listening area. The posterior estimate of $\bar{\boldsymbol{\theta}}_n$ follows Bayes' theorem, where the current mean $\mu_n^{\{t\}}$ and dispersion $\ell_n^{\{t\}}$ hyperparameters are updated from the measurement terms $\bar{\boldsymbol{\theta}}_n^{\{t\}}, \bar{\ell}_n^{\{t\}}$ in the likelihood function and the previous hyperparameters $\mu_n^{\{t-1\}}, \ell_n^{\{t-1\}}$ via appendix (28). Lastly, the normalization filter's quotient terms (9), (10) are updated for PDF $f_n(\theta) = f(\theta \,|\, \mu = \mu_n^{\{t\}}, \ell = \ell_n^{\{t\}})$, and the filters $G_n(\nu)$ are re-computed. Let us step-through the following example:

Consider the sample loudspeaker responses and sequence of estimated normalization angles in Fig. 3 where the listener is 90° offset the loudspeaker axis in azimuth. At $t = 0$ prior to any measurements, the normalization angle assumes a circular distribution centered on the loudspeaker axis $\mu = 0$ with wide dispersion FWHM 90.22°. The first estimate $\bar{\boldsymbol{\theta}}^{\{1\}} = -60^\circ$ is inaccurate with high dispersion FWHM 90° as shown in the $t = 1$ likelihood. Although the posterior shifts its mean halfway between the prior's mean and estimated angle, the dispersion remains high, which gives a similar acoustic windowed power average and correction quotient to that of the prior. The second estimate $\bar{\boldsymbol{\theta}}^{\{2\}} = 75^\circ$ is more accurate with lower dispersion FWHM 45°. The resulting posterior shifts much closer towards the estimate at much reduced dispersion, which distinguishes its windowed power average and correction quotient from the prior. The final and most accurate estimate $\bar{\boldsymbol{\theta}}^{\{3\}} = 90^\circ$ with lowest dispersion FWHM 30° yields a sharp posterior near the true normalization angle, which induces comb-filter patterns in the correction quotient due to lobbing in the loudspeaker's anechoic response in azimuth. Therefore in practice, we avoid equalizing to direct acoustic-paths by enforcing a lower-bound dispersion FWHM 45° for circular distributions $f_n(\theta)$ when computing the correction quotients $\hat{Q}_n(\nu)$.

3 Loudspeaker Panning Optimization

Let $R_n(\nu, \boldsymbol{r}) = H_n(\nu, \boldsymbol{r}) G_n(\nu)$ be the acoustic response at frequency ν and coordinate $\boldsymbol{r}$ of the n^{th} normalized loudspeaker in (4), and the overall response of the normalized loudspeaker array follows

$$Y(\nu, \boldsymbol{r}) = \sum_{n=1}^{N} R_n(\nu, \boldsymbol{r}) X_n(\nu), \tag{14}$$

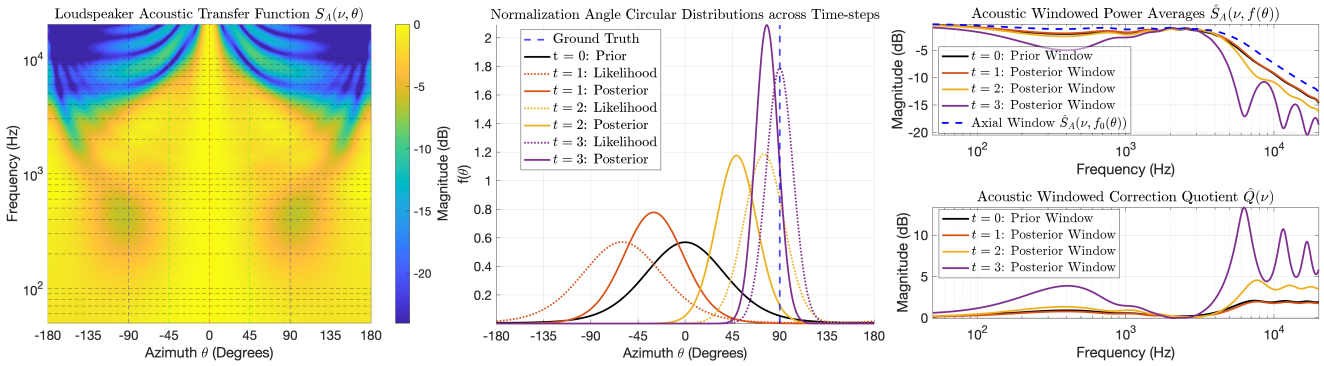


Fig. 3: We equalize a sample loudspeaker with acoustic responses over the horizontal plane (left) between Bayesian estimates of the normalization angle $\bar{\boldsymbol{\theta}}$ in (13) (center) and the axial windowed power average. The acoustic power averages (right) over the posterior circular distribution windows $f(\theta \,|\, \mu = \mu^{\{t\}}, \ell = \ell^{\{t\}})$ update across time-steps to yield a sequence of quotient correction targets in (10).

where $X_n(\nu)$ is the transfer function of the array's weights belonging to the n^{th} loudspeaker. For normalized loudspeaker panning, we constrain $X_n(\nu)$ to have a common phase-component (e.g. delay or all-pass) across loudspeakers and solve for the unknown magnitude components $x_n(\nu) = |X_n(\nu)|$, which are subject to frequency-dependent spatial-electrical-acoustic domain constraints. The magnitude components at frequency ν are therefore expressed as a vector of panning gains $\boldsymbol{x} = [x_1, \dots x_N]^T \in \mathbb{R}^{N\times 1}$, whereby we omit the frequency ν specification for simplifying notation. Further simplifications following the loudspeaker normalization are possible when specifying domain-specific constraints. Loudspeaker coordinates reduce to their unit-directions in the spatial domain given by

$$\boldsymbol{V} = [\boldsymbol{v}_1, \dots, \boldsymbol{v}_N] \in \mathbb{R}^{2xN}, \quad \boldsymbol{v}_n = \frac{\boldsymbol{u}_n}{\|\boldsymbol{u}_n\|}. \tag{15}$$

The normalization filter's electrical gain $|G_n(\nu)|$ bounds the electrical headroom in the electrical domain. The normalized loudspeaker acoustic responses in (4) are matched at the listener's location in the acoustical domain.

Spatial Panning Constraints: The vector-base amplitude panning with slack (VBAPS) constraint is given by

$$\boldsymbol{V}\boldsymbol{x} = \lambda \boldsymbol{s}, \quad \boldsymbol{x} \geq \boldsymbol{0}, \quad \lambda \geq 0, \tag{16}$$

where the panning gains $\boldsymbol{x}$ are non-negative as to preserve the relative-phase between loudspeaker pairs, and constrain the weighted average of the loudspeaker directions $\boldsymbol{V}$ to coincide with the target steering unit-direction $\boldsymbol{s} \in \mathbb{R}^{2\times 1}$ upto non-negative scale given by the slack-variable λ. The latter is an augmented variable for both scaling the target unit-direction $\boldsymbol{s}$ to lie in equality with the panning direction $\boldsymbol{V}\boldsymbol{x}$ as shown in Fig. 4, and to accommodate constraints placed on $\boldsymbol{x}$ from other domains. The feasible steering and panning directions, and panning gains are therefore constrained as follows:

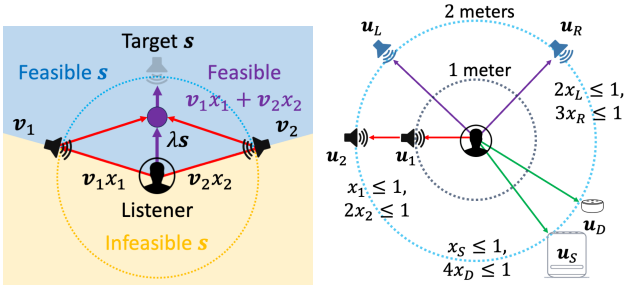


Fig. 4: VBAPS (left) constrains the feasible steering direction $\boldsymbol{s}$ to lie between the minor-arc of the loudspeaker pair coordinates $\boldsymbol{x}_L, \boldsymbol{x}_R$. Sample voltage constraints (right) are proportional to differences in loudspeaker-to-listener distance, orientation, and selection.

Consider a set of N loudspeakers and panning gains satisfying (16). The set of feasible steering unit-directions $\boldsymbol{s}$ must lie in the union of minor-arcs between all pairwise loudspeaker unit-directions. Conversely, steering directions are infeasible along the major-arc of a single loudspeaker-pair $N = 2$ as shown in Fig. 4. For $N > 2$ loudspeakers, the feasible $\boldsymbol{s}$ are all of $\mathbb{R}^2$ iff there exist a set of three loudspeakers where the negative direction of each loudspeaker lies between the minor-arc of the other two loudspeaker directions. The panning direction $\boldsymbol{V}\boldsymbol{x}$ is therefore constrained to be in the set of λ-scaled feasible unit-directions $\boldsymbol{s}$. We now introduce several evaluation metrics or objectives w.r.t. λ.

Let us define *panning sensitivity* by the acoustic-path distance ratio of the panning direction and the summation of component panning gained loudspeaker directions given by

$$\mathbb{S}(\boldsymbol{V}, \boldsymbol{x}, \boldsymbol{s}) = \frac{\|\boldsymbol{V}\boldsymbol{x}\|}{\sum_{n=1}^{N} \|\boldsymbol{v}_n x_n\|} = \frac{\|\lambda \boldsymbol{s}\|}{\sum_{n=1}^{N} x_n} = \frac{\lambda}{\boldsymbol{x}^T \boldsymbol{1}}, \tag{17}$$

which has bounds $0 < \mathbb{S}(\boldsymbol{V},\boldsymbol{x},\boldsymbol{s}) \leq 1$. Sensitivity is maximal iff non-zero panning gains belong to loudspeakers with directions coincident to the steering direction, large if panning gains disproportionately allocate to loudspeakers with directions closer to the steering direction, and minimal when panning gains allocate to loudspeakers with directions that sum to zero. Panning sensitivity therefore gives a similarity measure between panned and discrete sound-sources in the direction of $\boldsymbol{s}$. This contrasts with cross-domain measures of *panning efficiency*, which evaluates the power ratios between panning direction and electric or acoustic gain as follows:

$$\mathbb{F}(\boldsymbol{K},\boldsymbol{V},\boldsymbol{x}) = \frac{\boldsymbol{x}^T\boldsymbol{V}^T\boldsymbol{V}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{K}\boldsymbol{x}} = \lambda^2 \left\| \boldsymbol{K}^{\frac{1}{2}}\boldsymbol{x} \right\|^{-2}, \tag{18}$$

where $\boldsymbol{K} \in \mathbb{C}^{N\times N}$ is a domain-dependent covariance matrix (identity for electrical, model dependent for acoustical). For the electrical domain where $\boldsymbol{K} = \boldsymbol{I}$, the maximum efficiency is N for loudspeakers with directions coincident to the steering direction and uniform panning gains $\boldsymbol{x} = N^{-1}\boldsymbol{1}$. For the acoustic domain, the maximum efficiency is the largest generalized eigenvalues between $\boldsymbol{V}^T\boldsymbol{V}$ and $\boldsymbol{K}$. Thus, higher panning efficiency is realized via more uniformly distributed panning gains across loudspeakers, whereas high panning sensitivity follows sparsely distributed panning gains.

Electrical Headroom Constraints: The electrical-power headroom of normalized loudspeakers decreases in proportion to the normalization filter power responses $|G_n(\nu)|^2$. Under non-negative panning constraint, the quadratic electrical-power constraint are linearized as follows:

$$x_n x_n^* \leq |G_n(\nu)|^{-2}, \quad x_n \geq 0, \quad \Rightarrow \quad \boldsymbol{0} \leq \boldsymbol{x} \leq \boldsymbol{\tau}, \tag{19}$$

where $\boldsymbol{\tau} = \left[|G_1(\nu)|^{-1}, \ldots, |G_N(\nu)|^{-1} \right]^T \in \mathbb{R}^{N\times 1}_{\geq 0}$ is a vector containing the digital headroom per loudspeaker that bounds the feasible space of panning gains to the upper box-orthant. We give several examples of voltage headroom consumed by normalization in Fig. 4. Doubling the loudspeaker $\boldsymbol{u}_1$'s distance to the listener to that of $\boldsymbol{u}_2$ halves the voltage headroom. Re-orienting the loudspeaker $\boldsymbol{u}_R$ to face the listener at $\boldsymbol{u}_L$ lowers high-frequency headroom. Equalizing the mid-range loudspeaker at $\boldsymbol{u}_D$ to match the full-range loudspeaker at $\boldsymbol{u}_S$ decreases the low-frequency headroom.

Acoustical Power Constraints: The acoustic covariances between the normalized loudspeaker transfer functions $R_n(\nu,\boldsymbol{r})$ in (14), over coordinates $\boldsymbol{r}$ in the listening area, specify quadratic power constraints in equality to the acoustic power target ρ as follows:

$$\boldsymbol{x}^T\boldsymbol{K}\boldsymbol{x} = \rho, \quad K_{ij} \approx \mathbb{E}_{\boldsymbol{r}\sim g(\boldsymbol{r})}\left[R_i(\nu,\boldsymbol{r}) R_j^*(\nu,\boldsymbol{r}) \right], \tag{20}$$

whereby $\boldsymbol{r}$ is sampled from a disc of radius τ_r with a truncated uniform PDF $g(\boldsymbol{r}) = \frac{1}{\pi\tau_r^2}, \forall \|\boldsymbol{r}\| \leq \tau_r$, and 0 otherwise. For loudspeaker transfer functions in the far-field, spherical-waves can be approximated by plane-waves which give the acoustic covariance matrix $\bar{\boldsymbol{K}}$ with analytic terms $\bar{K}_{ij}$ as derived in appendix (31) as follows:

$$\bar{K}_{ij} = |S(\nu,0)|^2 \begin{cases} \frac{2J_1(D_{ij}\kappa\tau_r)}{D_{ij}\kappa\tau_r}, & D_{ij}\kappa\tau_r > 0 \\ 1, & D_{ij}\kappa\tau_r = 0 \end{cases}, \tag{21}$$

where $D_{ij} = \left\| \boldsymbol{v}_i - \boldsymbol{v}_j \right\|$ is the distance between loudspeaker unit-directions, and $J_1(z)$ is the Bessel function of the first kind. Note that at the listener location $\boldsymbol{r} = \boldsymbol{0}$, the normalized loudspeaker transfer functions are constant in (4). Thus, the acoustic covariance matrix $\bar{\boldsymbol{K}}$ degenerates to the rank-1 matrix $\mathring{\boldsymbol{K}} = |S(\nu,0)|^2 \boldsymbol{1}\boldsymbol{1}^T$ as the evaluation radius decreases to zero in $\lim_{\tau_r\to 0} g(\boldsymbol{r}) = \delta$. We therefore decompose the acoustic covariance as follows:

Let the acoustic covariance matrix in (20) be a mixture of the listener location, and listening area covariances given by

$$\boldsymbol{K} = (1-\alpha)\mathring{\boldsymbol{K}} + \alpha\bar{\boldsymbol{K}}, \quad 0 \leq \alpha \leq 1, \tag{22}$$

where the acoustic covariance for $\alpha = 0$ evaluates only the direct acoustic transfer function from loudspeakers to the listener location. The quadratic constraints in (20) linearize to $\boldsymbol{x}^T\boldsymbol{1} = \sqrt{\rho}\,|S(\nu,0)|^{-1}$ for non-negative $\boldsymbol{x}$; maximizing λ s.t. the linear gain summation constraint maximizes the panning sensitivity. Conversely, the acoustic covariance for $\alpha = 1$ evaluates the acoustic transfer functions over a larger listening area; maximizing λ s.t. the quadratic equality constraint maximizes panning efficiency. Moreover, the loudspeaker acoustic covariances in the listening area at the limits are correlated in low-frequency $\lim_{\kappa\to 0}\bar{\boldsymbol{K}} = \mathring{\boldsymbol{K}}$, and uncorrelated in high-frequency or large evaluation radii $\lim_{\kappa\to\infty}\bar{\boldsymbol{K}} = \lim_{\tau_r\to\infty}\bar{\boldsymbol{K}} = \boldsymbol{I}$. Therefore, the mixture of covariances (22) are proportional to $\boldsymbol{K} \propto (1-\alpha)\boldsymbol{1}\boldsymbol{1}^T + \alpha\boldsymbol{I}$. We now formulate the loudspeaker steering optimization w.r.t. spatial, electrical, and acoustical constraints.

Optimal Panning Sensitivity and Efficiency (OPSE): Maximizing the panning sensitivity λ subject to spatial, acoustical, and electrical constraints is the second-order cone problem [23] given by

$$\begin{aligned} (\lambda_*,\boldsymbol{x}_*) = \arg\max_{\lambda,\boldsymbol{x}} \lambda \qquad \text{s.t.} \quad \lambda \geq 0, \\ \boldsymbol{V}\boldsymbol{x} = \lambda\boldsymbol{s}, \quad \boldsymbol{x}^T\boldsymbol{K}\boldsymbol{x} \leq \rho, \quad \boldsymbol{0} \leq \boldsymbol{x} \leq \boldsymbol{\tau}, \end{aligned} \tag{23}$$

where a feasible solution always exist if the acoustic loudness's equality constraint in (20) is relaxed to be in inequality; acoustic loudness is tight w.r.t. ρ if panning sensitivity (17) or efficiency (18) is also maximized. We can eliminate λ by left-multiplying both sides of the equality constraints in (23) by unit-direction $\boldsymbol{s}^T$ to yield $\lambda = \boldsymbol{s}^T\boldsymbol{V}\boldsymbol{x}$, and the equality constraint matrix $\boldsymbol{A} = (\boldsymbol{I} - \boldsymbol{s}\boldsymbol{s}^T)\boldsymbol{V}$. The equivalent optimization in only $\boldsymbol{x}$ is expressed as follows:

$$\begin{aligned} \boldsymbol{x}_* = \arg\max_{\boldsymbol{x}} \boldsymbol{c}^T\boldsymbol{x} \qquad \text{s.t.} \quad \boldsymbol{c}^T\boldsymbol{x} \geq 0, \\ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}, \quad \boldsymbol{x}^T\boldsymbol{K}\boldsymbol{x} \leq \rho, \quad \boldsymbol{0} \leq \boldsymbol{x} \leq \boldsymbol{\tau}, \end{aligned} \tag{24}$$

where the objective maximizes the panning gains $\boldsymbol{x}$ in the direction of vector $\boldsymbol{c} = \boldsymbol{V}^T\boldsymbol{s}$, consisting of cosine similarities between the target and loudspeaker unit-directions. Moreover, the equality constraints restrict $\boldsymbol{x}$ to the null space of $\boldsymbol{A}$, which has nullity $N-1$. Thus for real-time applications and small number of loudspeakers ($N \leq 5$), we remove the equality constraints and reduce the number of variables via the linear transformation of the panning gains $\boldsymbol{x} = \bar{\boldsymbol{A}}\boldsymbol{y}$ along an orthonormal basis $\bar{\boldsymbol{A}}^T\bar{\boldsymbol{A}} = \boldsymbol{I}$ of the null space $\bar{\boldsymbol{A}} \in \text{span}(\ker(\boldsymbol{A})) \in \mathbb{R}^{N\times N-1}$. The optimization in the kernel space reduces to linear and quadratic inequality constraints given by

$$\boldsymbol{y}_* = \arg\max_{\boldsymbol{y}} \bar{\boldsymbol{c}}^T\boldsymbol{y} \quad \text{s.t.} \quad \begin{matrix} \bar{\boldsymbol{c}}^T\boldsymbol{y} \geq 0, \\ \boldsymbol{0} \leq \bar{\boldsymbol{A}}\boldsymbol{y} \leq \boldsymbol{\tau}, \end{matrix} \quad \boldsymbol{y}^T\bar{\boldsymbol{K}}\boldsymbol{y} \leq \rho, \quad (25)$$

where $\bar{\boldsymbol{c}} = \bar{\boldsymbol{A}}^T\boldsymbol{c}$, and $\bar{\boldsymbol{K}} = \bar{\boldsymbol{A}}^T\boldsymbol{K}\bar{\boldsymbol{A}}$, and the feasible region is convex. Lastly, the steering direction $\boldsymbol{s}$ can be infeasible where only the trivial solution $\boldsymbol{x} = \boldsymbol{0}$ satisfies the VBAPS equality constraint; dropping the VBAPS constraints $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ and $\boldsymbol{c}^T\boldsymbol{x} \geq 0$ in the primary form (24) relaxes the feasible space to be convex. Therefore, optimal solutions for both the null space (25) and relaxed primary forms can be efficiently found via interior-point methods. Let us now investigate the solutions to (23), (24), (25) under various acoustic power, covariance, and loudspeaker layouts in practical applications.

4 Experiments

Distributed Center Channel: In the 5.0 multichannel standard, the center content channel is fully sent to a center loudspeaker in a 5.0 ITU layout (left = $-30°$, right = $30°$, center = $0°$, surround left = $-110°$, surround right = $110°$), where the maximum acoustic power (unity) is limited to that of a single loudspeaker. Under OPSE, we can specify a larger acoustic power target ρ via the equality constraint $\boldsymbol{x}^T\boldsymbol{K}\boldsymbol{x} = \rho$, spatial panning constraints of a center steering direction $\boldsymbol{s} = [1;0]$, and unity electrical constraints $\boldsymbol{x} \leq \boldsymbol{1}$ WLOG. The optimal panning sensitivity gains for the listener location's acoustic covariance $\boldsymbol{K} = \boldsymbol{1}\boldsymbol{1}^T$ are shown in Fig. 5 for increasing acoustic power ρ targets. For acoustic power targets $0 < \rho \leq 1$, only the center loudspeaker is active $0 < x_C \leq 1$, and panning sensitivity is maximum. For $1 < \rho \leq 9$, the center loudspeaker exhausts its headroom and the left and right loudspeakers equally engage ($0 < x_{L,R} \leq 1, x_C = 1$), resulting in a slight loss in panning sensitivity (0.9 at $\rho = 9$), and increase in both panning/electric and acoustic/electric efficiency. For $9 < \rho \leq 25$, the left and right loudspeakers exhausts their headroom and the surround loudspeakers equally engage ($0 < x_{SL,SR} \leq 1, x_{L,R,C} = 1$), resulting in a sharper loss to panning sensitivity and degradation to panning/electric efficiency as the center steering direction lies in the infeasible sector of the surround loudspeaker pair. Note that for inequality constraints $\boldsymbol{x}^T\boldsymbol{K}\boldsymbol{x} \leq \rho$, the surround panning gains remain in-active as the quadratic constraint is not tight for $\rho > 9$. Panning sensitivity therefore monotonically decreases for larger acoustic power targets.

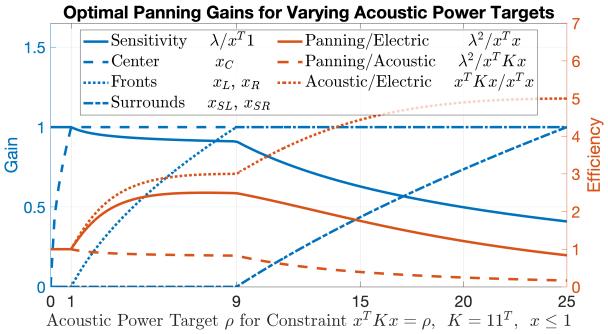


Fig. 5: OPSE center content more uniformly distributes across 5.0 ITU loudspeakers for increasing acoustic power targets ρ, and constant electrical headroom.

Diffuse-field Panning: In reverberant environments, acoustic covariance between well-separated loudspeakers in the listening area decreases due to increasing variations in acoustic reflection path responses. Normalized loudspeakers produce a mixture of correlated sound-fields from their direct acoustic paths, and less correlated diffuse-fields from their reflection paths over a listening area. The acoustic covariance in the listening area is therefore proportional to (22). Let us reconsider the previous case of distributed center channel over a 3.0 ITU layout (left = $-30°$, right = $30°$, center = $0°$). Under OPSE, we constrain the acoustic power to unity $\boldsymbol{x}^T\boldsymbol{K}\boldsymbol{x} = 1$, relax the electrical headroom $\boldsymbol{x} \leq \boldsymbol{10}$, and vary the mixture of acoustic covariances as shown in Fig. 6. For correlated sound-fields $0 \leq \alpha \leq 1 - \boldsymbol{s}^T\boldsymbol{v}_L$, only the center loudspeaker is active as panning sensitivity is maximum. For less correlated sound-fields $1 - \boldsymbol{s}^T\boldsymbol{v}_L < \alpha \leq 1$, the center loudspeaker attenuates relative to the left and right loudspeakers as more uniform-distributed gains yield both higher acoustic/panning and panning/electric efficiency. The gap between acoustic/electric efficiency and its theoretical Rayleigh quotient maximum, given by the largest eigenvalue of $\boldsymbol{K}$, closes at the diffuse-field limit $\alpha = 1$. OPSE therefore converges to the largest eigenvector of $\boldsymbol{K}$ under diffuse-field conditions where source-localization is difficult.

Circular Panning Across Loudspeaker Layouts: For adaptive multichannel reproduction, it is desirable to render content channels over common loudspeaker layouts shown in Fig. 7 for any listener location and front-direction. Under OPSE, we can evaluate the panning sensitivity for all steering directions in azimuth in both anechoic $\boldsymbol{K} = \boldsymbol{1}\boldsymbol{1}^T$ and diffuse-field $\boldsymbol{K} = \boldsymbol{I}$ conditions. Let us constrain the acoustic power to unity $\boldsymbol{x}^T\boldsymbol{K}\boldsymbol{x} = 1$, relax the electrical headroom $\boldsymbol{x} \leq \boldsymbol{10}$, and vary $\boldsymbol{s} = [\cos\theta; \sin\theta]$ for the half-circle $0 \leq \theta \leq \pi$ as the layouts are symmetric w.r.t. $\theta = 0$. For layouts with only frontal loudspeakers such as LRC, and wide LRC, the

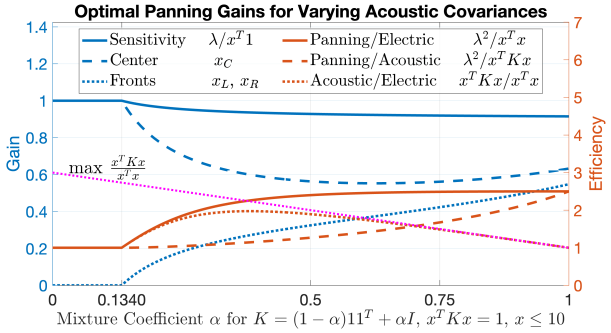


Fig. 6: OPSE center content gains for 3.0 ITU loudspeakers converge to the acoustic/electric Rayleigh quotient maximizer in diffuse-field conditions.

panning sensitivity remains high > 0.85 for feasible steering directions. For infeasible steering directions, the VBAPS constraints are dropped in (24), and the panning sensitivity, taken to be $\boldsymbol{c}^T\boldsymbol{x}/\boldsymbol{x}^T\mathbf{1}$, decrease for larger θ. The solutions are continuous w.r.t. θ for the anechoic covariance but discontinuous for the diffuse-field covariance at the feasibility boundary of θ. For triangular loudspeaker layouts (surround LRC, LRRear) containing the listener, only $2/3$ loudspeakers are active for any given θ. The solutions therefore uniquely satisfy the VBAPS constraints and are equivalent in both anechoic and diffuse-field conditions. LRRear has acceptable panning sensitivity between $|\theta| \leq 30°$, but minimal panning sensitivity near surround steering angles $100 \leq \theta \leq 110$. Surround LRC has low panning sensitivity for the left and right steering angles $\theta = \pm 30°$. For the LRSLSR layout, the panning sensitivity degrades in diffuse-field conditions for frontal angles $|\theta| \leq 60°$, and is minimal in the surround loudspeaker pair's gap $110° \leq \theta \leq 250°$. For the pentagon layout of uniformly spaced loudspeakers, anechoic and diffuse-field conditions have acceptable > 0.8 and borderline > 0.7 panning sensitivity respectively, with the latter also having lower variance. Under OPSE, the pentagon layout is therefore suited for uniform directional circular panning, LRSLSR for non-rear directional panning, and wide LRC for frontal to semi-surround directional panning for content reproduction.

5 Discussion

While Bayesian loudspeaker normalization and OPSE formulations are acoustic measurement-free sound-field correction methods, their accuracy may be improved with indirect acoustic measurements. Loudspeaker acoustic covariances at the listening area, critical distances, and distance attenuation rates may be estimated from simplified room acoustic models such as image-sources [29] if the room dimensions are known, and from loudspeaker-to-microphone acoustic transfer functions co-located on supporting smart-loudspeaker devices. In such instances, we may couple loudspeaker normalization with OPSE by substituting the in-situ loudspeaker acoustic covariance estimates in-place of anechoic and planewave mixture. Extension to 3D layouts can be considered for some loudspeaker arrangements, but may be impractical for satisfying VBAPS constraints and ensuring feasible coverage over spherical coordinates.

6 Conclusion

We presented a loudspeaker filtering method that normalizes multiple loudspeakers to a common acoustic target for a non-stationary listening location or area. Loudspeaker normalization angles w.r.t. the listener location were adapted via Bayesian posteriors over circular distribution probability density functions. We then formulated panning gain optimization problems by relaxing VBAP constraints to give a novel panning sensitivity / efficiency objective, and specified electrical, acoustical domain constraints. Augmented, primary, null-space, and relaxed forms of the OPSE problem were derived. Lastly, practical experiments quantified the OPSE solutions in applications for distributing excess content channels, evaluating efficiency in anechoic to diffuse-field conditions, and recommending loudspeaker layouts for circular panning and multichannel content reproduction.

7 Appendix

Circular Distribution: We can evaluate $P(-a \leq \theta \leq a)$ for $f(\theta)$ in (11) via the power-series approximation and cosine-integral [30] given by

$$\int_{-a}^{a} e^{\frac{-d^2(\theta)}{2\ell^2}} d\theta = \int_{-a}^{a} e^{\frac{\cos(\theta)-1}{\ell^2}} d\theta = \sum_{n=0}^{\infty} \frac{\int_{-a}^{a} \cos^n(\theta) d\theta - 1}{\ell^{2n} n!}. \tag{26}$$

The product of circular distributions is a circular distribution as the sum of weighted and phase-shifted cosines in the exponents' terms is a cosine given by

$$\begin{aligned} f_i(\theta) \propto e^{\frac{-d^2(\theta-\mu_i)}{2\ell_i^2}} \propto e^{\frac{\cos(\theta-\mu_i)}{\ell_i^2}} \quad \Rightarrow \\ f_i(\theta) f_j(\theta) \propto e^{\frac{\ell_j^2\cos(\theta-\mu_i)+\ell_i^2\cos(\theta-\mu_j)}{\ell_i^2\ell_j^2}} = e^{\frac{\cos(\theta-\mu)}{\ell^2}}, \end{aligned} \tag{27}$$

where the mean and dispersion of the product are given by

$$\begin{aligned} \mu_{ij} &= \operatorname{atan2}\left(\frac{\sin(\mu_i)}{\ell_i^2} + \frac{\sin(\mu_j)}{\ell_j^2}, \frac{\cos(\mu_i)}{\ell_i^2} + \frac{\cos(\mu_j)}{\ell_j^2}\right), \\ \ell_{ij}^2 &= \ell_i^2\ell_j^2\left(\ell_i^4 + 2\cos(\mu_i-\mu_j)\ell_i^2\ell_j^2 + \ell_j^4\right)^{-\frac{1}{2}}. \end{aligned} \tag{28}$$

The posterior dispersion depends on both prior mean and dispersion unlike that of a normal distribution. Substituting the empirical mean $\mu_i = \bar{\boldsymbol{\theta}}_n^{\{t\}}$, dispersion $\ell_i = \bar{\ell}_n^{\{t\}}$, and the prior's mean $\mu_j = \mu_n^{\{t-1\}}$, dispersion $\ell_j = \ell_n^{\{t-1\}}$ in (13) gives the posterior mean $\mu_{ij} = \mu_n^{\{t\}}$, dispersion $\ell_{ij} = \ell_n^{\{t\}}$.

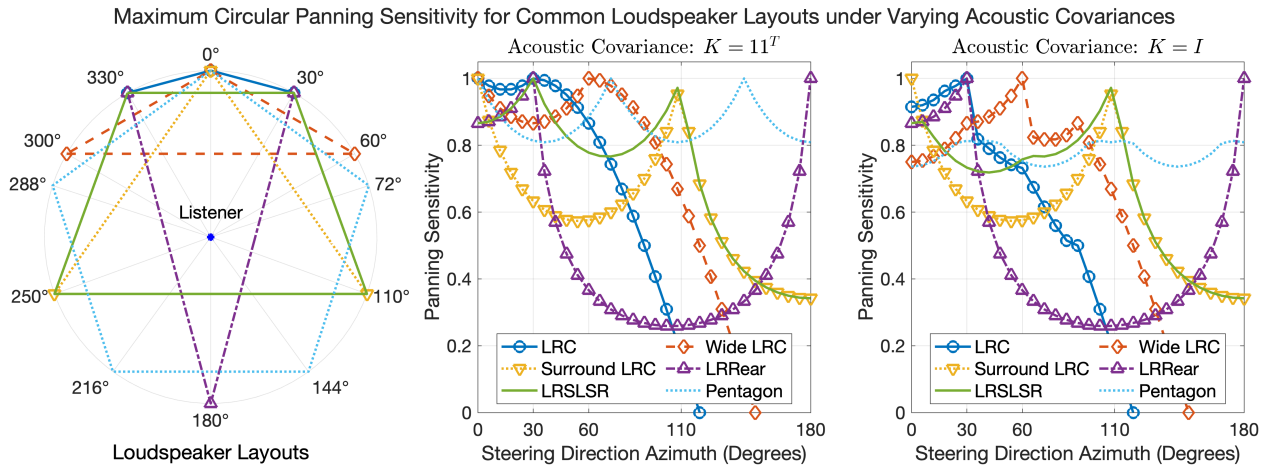


Fig. 7: Maximum panning sensitivity in (24) varies across azimuth steering directions for different loudspeaker layouts, and in anechoic and diffuse-field conditions. All triangle arrangements have low-sensitivity gaps. A minimum of five uniform-spaced loudspeakers (Pentagon) achieves moderate sensitivity in diffuse-field conditions across all of azimuth.

Plane-wave Covariance: Let $p(\boldsymbol{r}) = e^{-j\kappa \boldsymbol{v}^T \boldsymbol{r}}$ be the 2D plane-wave equation with incident direction $\boldsymbol{v}$, and the region of integration be the 2D disc of radius R where $\boldsymbol{r} = [x, y]^T$, $-R \le x \le R$, $y = \sqrt{R^2 - x^2}$. We may express the incident angle and evaluation point in polar coordinates as follows:

$$\begin{aligned} \boldsymbol{v} &= \|\boldsymbol{v}\| \left[\cos\theta_v, \sin\theta_v\right]^T, \quad \boldsymbol{r} = r\left[\cos\theta, \sin\theta\right]^T, \\ \boldsymbol{v}^T\boldsymbol{r} &= \|\boldsymbol{v}\| r\left(\cos\theta_v\cos\theta + \sin\theta_v\sin\theta\right) \\ &= \|\boldsymbol{v}\| r\cos\left(\theta - \theta_v\right), \\ -\boldsymbol{v}^T\boldsymbol{r} &= \|\boldsymbol{v}\| r\cos\left(\pi - \theta + \theta_v\right), \quad \text{Cosine reflection} \end{aligned} \tag{29}$$

where $0 \le r \le R$ and $0 \le \theta \le 2\pi$, which removes the dependence on the plane-wave incident angle in subsequent integrals. The first moment is analytic w.r.t. the Bessel function of the first kind $J_n(x)$ and given by

$$\begin{aligned} \mathbb{E}\left[p(\boldsymbol{r})\right] &= \frac{1}{\pi R^2}\int_0^{2\pi}\int_0^R re^{-jk\boldsymbol{v}^T\boldsymbol{r}}\,dr\,d\theta \\ &= \frac{1}{\pi R^2}\int_0^R r\int_0^{2\pi} e^{jk\|\boldsymbol{v}\|r\cos(\pi-\theta+\theta_v)}\,d\theta\,dr \\ &= \frac{2}{R^2}\int_0^R rJ_0(k\|\boldsymbol{v}\|r)\,dr \quad \text{Hansen-Bessel [31]} \\ &= \frac{2J_1(k\|\boldsymbol{v}\|R)}{k\|\boldsymbol{v}\|R}. \quad \text{Bessel integral identity [32]} \end{aligned} \tag{30}$$

Let $p_n(\boldsymbol{r}) = e^{-j\kappa\boldsymbol{v}_n^T\boldsymbol{r}}$ be the plane-wave equation of the n^{th} loudspeaker. The second moment can be expressed as $\boldsymbol{v} = \boldsymbol{v}_m - \boldsymbol{v}_n$, which after substitution in (30) follows

$$\begin{aligned} \mathbb{E}\left[p_m(\boldsymbol{r})p_n^*(\boldsymbol{r})\right] &= \frac{1}{\pi R^2}\int_0^{2\pi}\int_0^R re^{-jk(\boldsymbol{v}_m-\boldsymbol{v}_n)^T\boldsymbol{r}}\,dr\,d\theta \\ &= \begin{cases} \frac{2J_1(k\|\boldsymbol{v}_m-\boldsymbol{v}_n\|R)}{k\|\boldsymbol{v}_m-\boldsymbol{v}_n\|R}, & \|\boldsymbol{v}_m-\boldsymbol{v}_n\| > 0 \\ 1, & \|\boldsymbol{v}_m-\boldsymbol{v}_n\| = 0 \end{cases}. \end{aligned} \tag{31}$$

References

[1] Recommendation, I., "ITU-R Recommendation BS. 775-4," *Multi-channel stereophonic sound system with or without accompanying picture*, 2022.

[2] De Sena, E., Hacıhabiboğlu, H., and Cvetković, Z., "Analysis and design of multichannel systems for perceptual sound field reconstruction," *IEEE Trans. Audio Speech Lang. Process.*, 21(8), pp. 1653–1665, 2013.

[3] López, J. J., González, A., and Fuster, L., "Room compensation in wave field synthesis by means of multichannel inversion," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2005.*, pp. 146–149, IEEE, 2005.

[4] Brännmark, L.-J., Bahne, A., and Ahlén, A., "Compensation of loudspeaker–room responses in a robust MIMO control framework," *IEEE Trans. Audio Speech Lang. Process.*, 21(6), pp. 1201–1216, 2013.

[5] Jin, W. and Kleijn, W. B., "Theory and design of multizone soundfield reproduction using sparse methods," *IEEE/ACM Trans. Audio Speech Lang. Process.*, 23(12), pp. 2343–2355, 2015.

[6] Talagala, D. S., Zhang, W., and Abhayapala, T. D., "Efficient multi-channel adaptive room compensation for spatial soundfield reproduction using a modal decomposition," *IEEE/ACM Trans. Audio Speech Lang. Process.*, 22(10), pp. 1522–1532, 2014.

[7] Kolundzija, M., Faller, C., and Vetterli, M., "Reproducing sound fields using MIMO acoustic channel in-

version," *Journal of the Audio Engineering Society*, 59(10), pp. 721–734, 2011.

[8] Poulose, A., Eyobu, O. S., and Han, D. S., "An indoor position-estimation algorithm using smartphone IMU sensor data," *Ieee Access*, 7, pp. 11165–11177, 2019.

[9] Won, S.-h. P., Melek, W. W., and Golnaraghi, F., "A Kalman/particle filter-based position and orientation estimation method using a position sensor/inertial measurement unit hybrid system," *IEEE Transactions on Industrial Electronics*, 57(5), pp. 1787–1798, 2009.

[10] Daniş, F. S. and Cemgil, A. T., "Model-based localization and tracking using bluetooth low-energy beacons," *Sensors*, 17(11), p. 2484, 2017.

[11] Wang, Y., Yang, X., Zhao, Y., Liu, Y., and Cuthbert, L., "Bluetooth positioning using RSSI and triangulation methods," in *2013 IEEE 10th Consumer Communications and Networking Conference (CCNC)*, pp. 837–842, IEEE, 2013.

[12] Kim, K.-W., Kwon, J., Lee, C.-G., and Han, J., "Accurate indoor location tracking exploiting ultrasonic reflections," *IEEE Sensors Journal*, 16(24), pp. 9075–9088, 2016.

[13] Greenfield, R. and Hawksford, M. J., "Efficient filter design for loudspeaker equalization," *Journal of the Audio Engineering Society*, 39(10), pp. 739–751, 1991.

[14] Mäkivirta, A., Liski, J., and Välimäki, V., "Modeling and delay-equalizing loudspeaker responses," *AES: Journal of the Audio Engineering Society*, 66(11), pp. 922–934, 2018.

[15] Pulkki, V., "Virtual sound source positioning using vector base amplitude panning," *Journal of the audio engineering society*, 45(6), pp. 456–466, 1997.

[16] Pulkki, V. et al., *Spatial sound generation and perception by amplitude panning techniques*, Helsinki University of Technology Ph. D. dissertation, 2001.

[17] 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*, pp. 187–190, IEEE, 1999.

[18] Franck, A., Fazi, F. M., and Hamdan, E., "An optimization approach to control sound source spread with multichannel amplitude panning," in *24th International Congress on Sound and Vibration*, 2017.

[19] Luo, Y., "Active Barycentric Beamformed Stereo Upmixing," in *2023 31st European Signal Processing Conference (EUSIPCO)*, pp. 316–320, 2023.

[20] Sadek, R. and Kyriakakis, C., "A novel multichannel panning method for standard and arbitrary loudspeaker configurations," 2004.

[21] Batke, J.-M. and Keiler, F., "Investigation of Robust Panning Functions for 3-D Loudspeaker Setups," in *Audio Engineering Society Convention 128*, Audio Engineering Society, 2010.

[22] Oppenheim, A. V., *Discrete-time signal processing*, Pearson Education India, 1999.

[23] Alizadeh, F. and Goldfarb, D., "Second-order cone programming," *Mathematical programming*, 95(1), pp. 3–51, 2003.

[24] Horn, R. and Johnson, C., *Matrix Analysis*, Cambridge University Press, 1990, ISBN 9780521386326.

[25] Sanathanan, C. and Koerner, J., "Transfer function synthesis as a ratio of two complex polynomials," *IEEE Trans. Autom. Control*, 8(1), pp. 56–58, 2003.

[26] Lai, X. and Lin, Z., "Minimax design of IIR digital filters using a sequential constrained least-squares method," *IEEE Trans. Signal Process*, 58(7), pp. 3901–3906, 2010.

[27] Toole, F. E., "Loudspeakers and rooms for sound reproduction—A scientific review," *Journal of the Audio Engineering Society*, 54(6), pp. 451–476, 2006.

[28] Luo, Y., "Spherical harmonic covariance and magnitude function encodings for beamformer design," *EURASIP Journal on Audio, Speech, and Music Processing*, 2021, pp. 1–17, 2021.

[29] Lehmann, E. A. and Johansson, A. M., "Prediction of energy decay in room impulse responses simulated with an image-source model," *The Journal of the Acoustical Society of America*, 124(1), pp. 269–277, 2008.

[30] Weisstein, E. W., "Cosine Integral," `https://mathworld.wolfram.com/CosineIntegral.html`, 2025, From MathWorld–A Wolfram Web Resource.

[31] Temme, N. M., *Special Functions: An Introduction to the Classical Functions of Mathematical Physics*, pp. 228–231, Wiley-Interscience, 1996.

[32] Weisstein, E. W., "Bessel Function of the First Kind," `https://mathworld.wolfram.com/BesselFunctionoftheFirstKind.html`, 2023, From MathWorld–A Wolfram Web Resource.