Active Barycentric Beamformed Stereo Upmixing

1st Yuancheng Luo
Amazon Inc.
luoyuancheng@gmail.com

***Abstract*—This paper presents several novel techniques for stereo to multichannel upmixing under barycentric constraints and beamformer formulations in the short-time Fourier transform (STFT) domain. We derive optimal solutions to center channel extraction and power-averaged monoization problems for a barycentric weighted mid-side decomposition. We then generalize passive multichannel upmixing via an active pan-potted barycentric beamformer formulation with sparse Chebyshev polynomial directivity. Experiments analyze center channel leakage, evaluate subjective listening tests, and render sample multichannel upmixes over varied speaker arrangements.**

***Index Terms*—Stereo upmixing, multichannel, center channel extraction, barycentric, beamforming**

I. Introduction

Stereo remains the most ubiquitous format for recording, music production, and streaming. While stereo mixing with a speaker pair has long been the standard, stereo reproduction methods have since broadened with the application of signal processing techniques on smart-speakers, multi-transducer, and multi-device arrangements [4], [19]. Such methods upmix stereo inputs into disparate number of output channels suited for the hardware and its playback capabilities.

A single-transducer-speaker unit summarizes stereo content into a mono channel with minimal spectral coloration via phase-aligning the input channels [20] or suppressing the coherent signals of one channel prior to summation [1]. A multi-transducer-speaker unit widens a sound-stage by beamforming stereo upmixed into left, right and center channels. Center channel is extracted from estimates of the panning index [2], [18], and from channel decomposition assuming orthogonality of non-centered components [6], [11], [21]. A multi-device arrangement enlarges a listening area by decomposing and resynthesizing the stereo sound-stage across speaker groups. Additional channels are extracted via source re-panning [3], primary-ambient separation [12], [13], [15], [16], and principal-component-analysis [17]. In this work, we investigate the application of barycentric constrained (BC) formulations [7], [10] for regularizing several stereo upmixing problems, give geometric interpretations and derive optimal solutions. The paper's contributions are as follows:

Section II defines the notation and barycentric model for stereo signals in the frequency domain. Section II-A presents a novel center channel extraction method from a BC mid-side (MS) decomposition. Section II-B derives a related BC power-average method for stereo monoization. Section III introduces a generalized stereo BC beamformer with a directivity pattern of the Chebyshev polynomials for stereo-to-multichannel upmixing. We achieve the latter via spatial-decomposition using binaural localization cues studied in [9], [14]. Section IV compares center extraction behavior for various signals to literature, shows subjective stereo monoization preference to alternatives, and presents a case-study for rendering stereo beamformed multichannels to speaker arrangements with varied angles.

II. Barycentric Constrained Upmixing

The discrete-time STFT of channel x is given by

$$X[m,k] = \text{DFT}\left(x[n]\, w(n - mR)\right), \tag{1}$$

which for notation we drop the sample n, frame m, and frequency k indices, window w, and hop-size R for subsequent references to the complex stereo left and right channels X_L, $X_R \in \mathbb{C}$. The 1-dimensional barycentric system models points on the simplex $V \in \mathbb{C}$ w.r.t. non-negative coordinates $\boldsymbol{\alpha} \in \mathbb{R}^2$ with unity summation:

$$V = \alpha_L X_L + \alpha_R X_R, \quad 0 \leq \boldsymbol{\alpha} \leq 1, \quad \alpha_L + \alpha_R = 1, \tag{2}$$

where V geometrically contains the line-segment spanning (X_L, X_R), and (α_L, α_R) are the latter's barycentric coordinates. Moreover, the constraints in Eq. 2 reduce the barycentric coordinates to a single variable λ which can be expressed as follows:

$$\alpha_L = \frac{1+\lambda}{2}, \quad \alpha_R = \frac{1-\lambda}{2}, \quad -1 \leq \lambda \leq 1, \tag{3}$$

where $\lambda \in \mathbb{R}$ is centered at 0 and bounded. When applied to stereo mixing, this barycentric model reduces time-varying filter artifacts as the mixture V is differentiable w.r.t. λ and its power response bounded by the input power $|V|^2 \leq \max\left\{|X_L|^2, |X_R|^2\right\}$. This motivates several ways of choosing λ for channel extraction and monoization problems.

A. Mid-Side Center Channel Extraction

Consider the stereo mid-side (MS) decomposition where centered content is contained within the unscaled mid components $X_L + X_R$ and absent from the unscaled side components $X_L - X_R$. We can negate common signals between mid and side components via adaptive filtering by treating the mid component as a mixture of center plus noise and the side component as a noise reference. Fitting a linear-phase adaptive filter with unknown amplitude bounded

between $(-1, 1)$ is a least-squares problem along barycentric λ from Eq. 3. Two possible residual or error function are shown:

$$\begin{aligned} V(\lambda) &= \left(\frac{X_L + X_R}{2}\right) + \lambda\left(\frac{X_L - X_R}{2}\right) \\ &= \left(\frac{1+\lambda}{2}\right) X_L + \left(\frac{1-\lambda}{2}\right) X_R, \\ U(\lambda) &= \lambda\left(\frac{X_L + X_R}{2}\right) + \left(\frac{X_L - X_R}{2}\right) \\ &= \left(\frac{1+\lambda}{2}\right) X_L - \left(\frac{1-\lambda}{2}\right) X_R, \end{aligned} \tag{4}$$

where we omit constant delay for notation. $V(\lambda)$ scales the side component by λ to cancel non-centered signals in the mid component to yield the desired center content. Conversely, $U(\lambda)$ scales the mid component by λ to cancel the side component s.t. the remaining $(1 - |\lambda|)$ proportion of the mid component is center content.

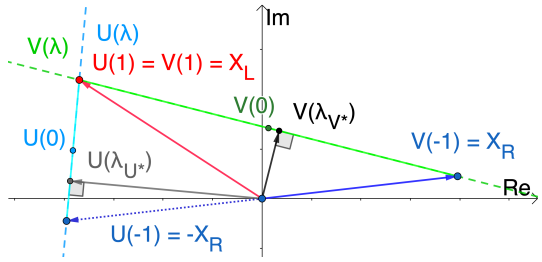

Fig. 1. The feasible regions of the center extraction error functions in Eq. 4 are barycentric along green and light-blue line segments.

Expanding Eq. 4 into X_L and X_R and substituting Eq. 3 gives both geometric and beamforming interpretations useful for constrained optimization. $V(\lambda)$ and $U(\lambda)$ are barycentric constrained beamformers that steer along line segments (X_L, X_R) and $(X_L, -X_R)$ on the complex plane in Fig. 1 respectively. Maximum cancellation occurs at the nearest point on the line segments to the origin; either the point perpendicular to the origin or the two end points are solutions. The former is the λ minimizer of the unconstrained squared moduli of Eq. 4, which have reciprocal solutions:

$$\begin{aligned} \lambda_{V*} &= \arg\min_{\lambda} |V(\lambda)|^2 = -\operatorname{Re}\left(\frac{X_L + X_R}{X_L - X_R}\right), \\ \lambda_{U*} &= \arg\min_{\lambda} |U(\lambda)|^2 = -\operatorname{Re}\left(\frac{X_L - X_R}{X_L + X_R}\right), \end{aligned} \tag{5}$$

which correspond to the real values of the complex MS component ratios (see Appendix Eq. 17 for derivations) and relates to a similar center extraction solution in [6], [11] where non-centered components between left and right channels are assumed uncorrelated, an assumption not made in this work and the implications discussed in section IV. If the unconstrained minimizer λ_{V*} or λ_{U_*} in Eq. 5 violate the barycentric constraints $|\lambda_{V*}| > 1$ or $|\lambda_{U*}| > 1$ from Eq. 3, then the bounded minimizers occur at the nearest end-points to λ_{V*} and λ_{U*} as evident in Fig. 1. The center channels are therefore computed from the barycentric minimizers (BMs):

$$\begin{aligned} C_V &= V\left(\max\left\{\min\left\{\lambda_{V*},\, 1\right\},\, -1\right\}\right), \\ C_U &= \left(1 - \min\left\{|\lambda_{U*}|,\, 1\right\}\right)\left(\frac{X_L + X_R}{2}\right), \end{aligned} \tag{6}$$

where C_V is the residual center content in $V(\lambda)$, and C_U the equivalent portion of center content found in the mid component.

B. Equal-Power-Average Channel Extraction

Stereo monoization is a related problem to stereo center channel extraction where X_L and X_R are instead summed to preserve the stereo spectrum. In passive stereo-to-mono downmixing, the simple stereo unweighted average $(X_L + X_R)/2$ results in cancellation and a loss in output power whenever X_L and X_R are not phase-aligned.

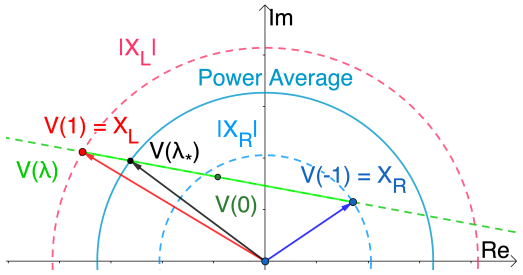

Fig. 2. EPA intersects the barycentric feasible region at $V(\lambda_*)$.

Consider the weighted summation $V(\lambda)$ of the stereo inputs from Eq. 4 under barycentric constraints in Eq. 3. Setting the output power of $V(\lambda)$ to the stereo input's equal-power-average (EPA):

$$|V(\lambda)|^2 = P_{avg} = \frac{|X_L|^2 + |X_R|^2}{2}, \quad \text{EPA Constraint} \tag{7}$$

yields a quadratic equation in λ with a geometric interpretation as shown in Fig. 2. The barycentric feasible region along the line segment (X_L, X_R) must intersect the stereo EPA circle at a single point on the complex plane. This follows from a proof by contradiction of the EPA bounds $\min\left\{|X_L|^2, |X_R|^2\right\} \leq P_{avg} \leq \max\left\{|X_L|^2, |X_R|^2\right\}$. No intersection implies an EPA outside the larger circle $P_{avg} > \max\left\{|X_L|^2, |X_R|^2\right\}$. Two intersections imply an EPA inside the smaller circle $P_{avg} < \min\left\{|X_L|^2, |X_R|^2\right\}$. The solution to quadratic Eq. 7 that satisfies the barycentric constraints from Eq. 3 and the EPA channel is given by

$$\begin{aligned} \lambda_* &= \begin{cases} \lambda_{V*} + \sqrt{\lambda_{V*}^2 + 1}, & \lambda_{V*} \leq -0, \\ \lambda_{V*} - \sqrt{\lambda_{V*}^2 + 1}, & \lambda_{V*} \geq +0, \\ 0, & \lambda_{V*} = \pm\infty, \end{cases}, \\ C_P &= V(\lambda_*), \qquad \text{EPA Channel} \end{aligned} \tag{8}$$

which relates to the unconstrained center channel minimizer λ_{V*} of Eq. 5 (see Appendix Eqs. 18, 19 for derivations). Observing that $|\lambda_*| \leq |\lambda_{V*}|$, the EPA channel C_P after substitution in Eq. 4 contains the center content C_V from Eq. 6

and more mid components. The difference $C_P - C_V$ gives the EPA side content.

III. Beamformed Constrained Upmixing

Stereo signals can be parameterized into a mixture of constant power pan-pots and phase-pots up to arbitrary gain and phase rotation:

$$\begin{aligned} X_L &= \cos\left(\frac{\theta}{2}\right) S, \quad X_R = \sin\left(\frac{\theta}{2}\right) e^{-j\phi} S, \\ |S|^2 &= |X_L|^2 + |X_R|^2, \quad \text{Constant Power} \end{aligned} \tag{9}$$

where S is the source signal, θ the pan-pot direction from $0 \le \theta \le \pi$ s.t. $\theta = \frac{\pi}{2}$ is centered and $\theta = 0$ is hard-panned left, ϕ the phase-pot direction from $-\pi \le \phi \le \pi$ s.t. $\phi = 0$ is in-phase and $\phi = \frac{\pi}{2}$ has the left channel lead the right channel. The pan and phase-pot angles can be estimated from the relative stereo input $X_T = \frac{X_R}{X_L}$:

$$\begin{aligned} \theta &= 2\tan^{-1}\left(|X_T|\right), && \text{Pan-angle} \\ \phi &= -\operatorname{atan2}\left(\operatorname{Im}(X_T), \operatorname{Re}(X_T)\right), && \text{Phase-angle} \end{aligned} \tag{10}$$

which is useful for separate source signals S that do not overlap in frequency. Furthermore, the stereo output power preserves that of S in Eq. 9. This presents two useful constraints: Stereo upmixed channels can correlate with the estimated pan-pot and phase-pot directions and its total output power should equal the stereo input power.

Consider the following stereo pan-pot beamformer (PB):

$$P(\hat{\theta},\, X) = \cos\left(\frac{\hat{\theta}}{2}\right) X_L + \sin\left(\frac{\hat{\theta}}{2}\right) X_R, \tag{11}$$

where $\hat{\theta}$ is the look-direction bounded between $0 \le \hat{\theta} \le \pi$ for in-phase pan-potted content $(\theta,\, \phi = 0)$ in Eq. 9. PB recovers the source signal S when a look-direction matches the pan-pot angle $\hat{\theta} = \theta$. However, passive upmixing via beam-steering along uniformly spaced look-directions $\hat{\theta}_m = \frac{\pi(m-1)}{M-1}$ where $1 \le m \le M$, has large leakage due to wide beam patterns shown in Fig. 3. Source are mixed into a dense set of channels.

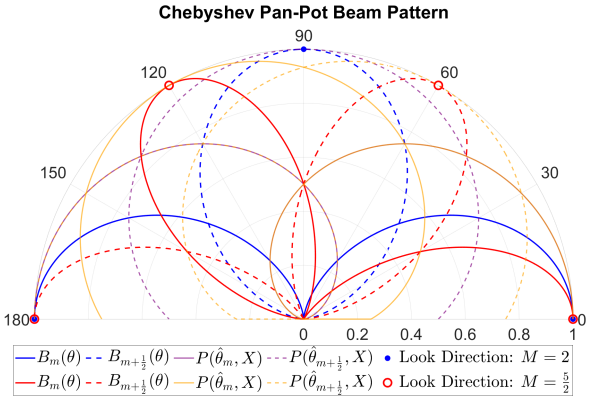

Fig. 3. CB beam patterns $B_m(\theta)$ satisfy constant power-panning law unlike PB beam patterns $P(\hat{\theta}_m, X)$ for $M = 2$ and $M = 3$.

Instead, beam patterns with minimal leakage between adjacent beams can be defined along half-periods of the Chebyshev polynomials. Let the interval $\hat{\theta}_m \le \theta \le \hat{\theta}_{m+\frac{1}{2}}$ bound the pan-pot direction θ. Two beams that are steered at the interval's endpoints have the target non-zero magnitude responses between look-directions:

$$B_m(\theta) = \begin{cases} |\cos(M-1)\theta|, & \left|\theta - \hat{\theta}_m\right| < \frac{\pi}{2(M-1)} \\ 0, & \text{Otherwise} \end{cases},$$

$$B_{m+\frac{1}{2}}(\theta) = \begin{cases} |\sin(M-1)\theta|, & \left|\theta - \hat{\theta}_{m+\frac{1}{2}}\right| < \frac{\pi}{2(M-1)} \\ 0, & \text{Otherwise} \end{cases},$$

$$|B(\theta)|^2 = \sum_{m=1}^{M} B_m^2(\theta) + B_{m+\frac{1}{2}}^2(\theta), \quad \text{Power Resp.} \tag{12}$$

which satisfy the constant power-panning law $B_m^2(\theta) + B_{m+\frac{1}{2}}^2(\theta) = 1$ and is sparse response w.r.t. the pan-pot direction as seen in Fig. 3. We therefore modify PB into an active beamformer using estimates of the pan-pot and phase-directions as follows:

Define the n^{th} upmixed channel where $1 \le n \le 2M$ by the $m = \frac{n+1}{2}$ Chebyshev beamformer (CB) given by

$$Q_m(X) = B_m(\psi) \frac{P(\hat{\theta}_m,\, X)}{\cos\left(\frac{\theta_* - \hat{\theta}_m}{2}\right)}, \tag{13}$$

where $\theta_* = \theta$ estimates the stereo pan-pot angle from Eq. 10, the denominator corrects for the mismatch between $\hat{\theta}_m$ and θ in the m^{th} PB in Eq. 11, and the desired beam pattern $B_m(\psi)$ can be squared to satisfy linear or barycentric panning laws. The beam-assignment angle ψ lies in the interior of the estimated pan-pot angle θ_* and the modified phase-pot angle ϕ_* given by

$$\phi_* = \frac{\pi}{2} + \operatorname{atan2}\left(\operatorname{Im}(X_T), |\operatorname{Re}(X_T)|\right), \tag{14}$$

which reflects the estimate of the phase angle in Eq. 10 across the imaginary axis s.t. $0 \le \phi_* \le \pi$ after re-centering to $\phi_* = \frac{\pi}{2}$. The assignment angle's indicator variable is barycentric and parameterized by the bias α and sensitivity parameter β as follows:

$$\psi = \left(1 - \alpha^{\beta}\right)\theta_* + \alpha^{\beta}\phi_*, \quad 0 \le \beta \le \infty, \tag{15}$$

which selects for the pan-pot angle $\psi = \theta_*$ as $\alpha \to 0$ or $\beta \to \infty$ and the modified phase-pot angle $\psi = \phi_*$ as $\alpha \to 1$ or $\beta \to 0$. Suitable choices for α and β vary across frequency and by application.

In stereophonic mixes, panning and delay techniques are often used to place localizable sound-sources. Panning effects the relative magnitude $|X_T|$ whereas delay effects the relative phase $\angle X_T$ from Eq. 10. This aligns with binaural source localization studies that show inter-aural level differences strongly predicts source direction above 1.45 kHz compared to inter-aural phase differences [14]. We therefore model bias and sensitivity parameters in Eq. 15 as data and frequency-dependent functions given by

$$\begin{aligned} \alpha(X_T) &= \min\left\{|X_T|, |X_T|^{-1}\right\}, && \text{Bias} \\ \beta(f) &= \beta_0 + \tfrac{f}{f_0}\beta_1, && \text{Sensitivity} \end{aligned} \tag{16}$$

where the bias function $\alpha(X_T)$ is bounded $0 \leq \alpha(X_T) \leq 1$. Hard-panned cases $|X_L| = 0$ or $|X_R| = 0$ biases for pan-angle selection ($\alpha = 0$) whereas equal magnitude cases $|X_L| = |X_R|$ biases for phase-angle selection ($\alpha = 1$). The sensitivity function $\beta(f)$ is a linear model of frequency-dependent β_1 dB roll-off per $|X_T|$ dB / octave w.r.t. frequency frequency f with constant offset β_0. Localizable sources therefore have beam-assignment angles ψ that output to two CB channels from Eq. 13 due to the bounds in Eq. 12.

IV. Experiments

Center Channel Leakage Analysis: Mask-based center extraction methods estimate the proportion of centered content within the mid $(X_L + X_R)/2$ component. We evaluate leakage behavior in terms of attenuation applied at varying pan-angle θ and phase-angle ϕ in Eq. 9 defined as follows:

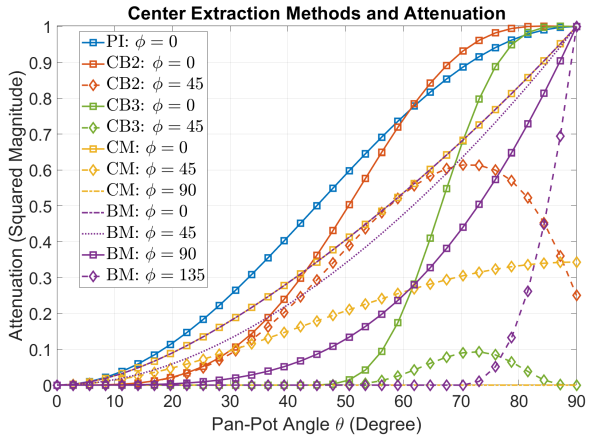

Fig. 4. Center channel extraction methods vary the squared magnitude of attenuation applied to the mid component for signals with increasing pan-pot angles $\theta \to 90°$ at constant phase-pot angles ϕ from Eq. 9.

Center content fully contains stereo signals panned to $\theta = \pi/2$ and $\phi = 0$. Center extraction methods attenuate any signals pan or phase-potted further away as seen in Fig. 4. Low-attenuation at either low pan-pot angles or high phase-pot angles increases non-centered content in the center channel. Conversely, high-attenuation close to $\theta = \pi/2$ and $\phi = 0$ reduces non-centered content in the center channel but induces more audible time-varying filter artifacts as a result of greater sensitivity to STFT spectral leakage errors and non-stationary phase effects such as reverb. We characterize a method's center leakage by the angles θ at the attenuation's max-power-point (MPP) and half-power-point (HPP) for constant ϕ. Methods in Fig. 4 with HPP closer to their MPP therefore reject more non-center content at the expense of more audible artifacts.

The panning index (PI) similarity measure [2] has a conservative attenuation with the lowest HPP of 45° and unity MPP 90° that both remain constant across all ϕ. The CB center beam $M = 2$ (CB2) in Eq. 13 with squared beam pattern $B^2_{1.5}(\psi)$ and sensitivity coefficients $\beta_0 = 3$, $\beta_1 = 0$ is less conservative with HPP 50°, and unity MPP 90° at $\phi = 0$. Its MPP decreases in both θ and scale as ϕ grows out-of-phase due to the localization modeling. The BM C_U from Eq. 6 has stronger attenuation with HPP 57° and unity MPP 90° at $\phi = 0$. As ϕ increases, HPP slowly grows as MPP remains constant. This is attractive for extracting centered content mixed with reverberation given that stereo reverb has equal-power but uncorrelated phase. The related correlation minimum (CM) method from [6], [11] has equivalent leakage to BM at $\phi = 0$ but with stronger attenuation for increasing ϕ due to a quickly decreasing MPP at 90°; centered components with added reverb are attenuated. The CB $M = 3$ (CB3) with center beam pattern $B^2_{2.5}(\psi)$ rejects $\theta \leq 45°$, and has the highest HPP 67° with unity MPP 90° at $\phi = 0$, as well as the lowest MPP when ϕ increases. This is useful for extracting dry center content mixed with phase-delayed and panned signals.

Monoization Subjective Listening: We ranked downmixing methods against stereo reference tracks for similarity and preference. Subjects ($N = 22$) listened to 4 stereo musical tracks (Pop and Rock genres, 48 kHz) over headphones and ranked three renders (EPA C_P from Eq. 8, Mid component $(X_L + X_R)/2$, and center extraction CB2 method from Eq. 13), loudness matched w.r.t. centered vocals, and equally mixed into left and right speakers, against the reference. We used 1024 point Hann-windowed FFTs, and 512 sample hop-sizes. The Bradley-Terry-Luce [5], [8] analysis in Fig. 5 shows strong utility for EPA spectral similarity to reference and preference over alternatives. EPA vocal similarity is similar to the Mid component and outperforms CB2 as the latter strongly attenuates non-centered content which in some tracks contained widened vocals and reverb.

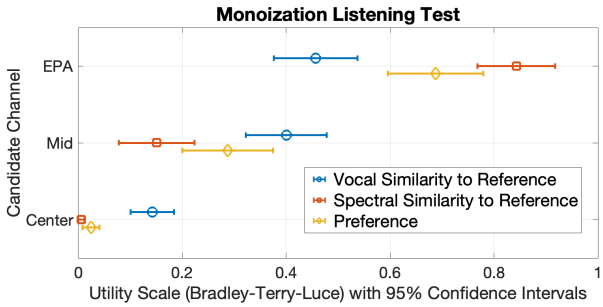

Fig. 5. EPA shows preference over mid and center channels via pair-wise A, B ranking probability $P(A > B) = \frac{U_A}{U_A + U_B}$ for utility U_A, U_B

Upmixer to Speaker Rendering: Each CB output channel is bounded between adjacent uniform look-directions $\hat{\theta}_m$. However, a multi-speaker configuration may be arranged along arbitrary speaker-angles $\bar{\theta}_m$ w.r.t. a listening position. To remap the CB domain ψ in Eq. 13 s.t. the former look-directions point to the latter speaker-angles, the warping function $f : \psi \to \psi$ should be continuous, real, monotone, and smooth along the closed interval $0 \leq \psi \leq \pi$. This bijective mapping implies that beam-assignment angles ψ in Eq. 15 uniquely correspond with angles between the physical speakers, and satisfies the constant power-panning law in Eq. 12. A smooth warping function also ensures that localization between speaker-angles tracks along ψ.

We consider piece-wise linear interpolation between the speaker-angle $\bar{\theta}_m$ samples to the look-direction $\hat{\theta}_m$ values.

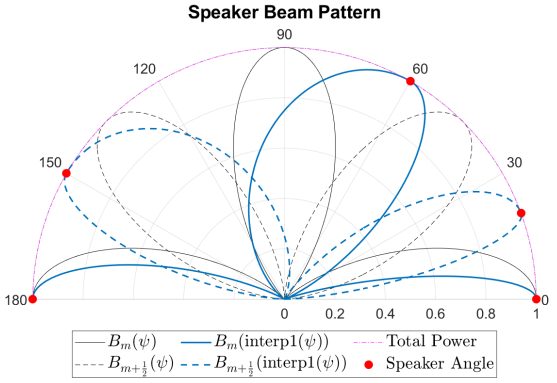

Fig. 6. Piece-wise linear interpolation between speaker-angles and uniform look-directions warps the CB3 into a speaker beamformer.

Matlab interp1(ψ) is monotonic when both speaker-angle and look-directions are also monotonic and maximally smooth except at the speaker-angles. Fig. 6 shows the CB3's beam pattern consisting of 5 main lobes with peaks at uniform look-directions warped towards speaker-angles $0, 20, 60, 150, 180°$. The speaker response's HPPs are correctly located half-way between adjacent speaker-angles. Pan and phase-potted sources smoothly track across speakers during rendering.

V. Conclusions

We presented several methods for stereo upmixing optimized under barycentric constraints. Leakage analysis shows that BM is robust when presented with centered content mixed with reverberation, CB is robust in the presence of delay effects. Listening tests show that EPA monoization ranks more similar to stereo reference and preferred over passive downmixing. We last render CB upmixes over variable angled speaker configurations. Future work investigates multichannel upmixing of non-localizable sources.

Appendix

For notation, denote the unscaled MS decomposition by $X_M = X_L + X_R$ and $X_S = X_L - X_R$. The squared moduli of the residual function $V(\lambda)$ in Eq. 4 and its least-squares minimizer are given by

$$
\begin{aligned}
|V(\lambda)|^2 &= \frac{\lambda^2 X_S X_S^* + \lambda(X_M X_S^* + X_S X_M^*) + X_M X_M^*}{4}, \\
\frac{\partial\,|V(\lambda)|^2}{\partial\,\lambda} &= \frac{2X_S X_S^* \lambda + X_M X_S^* + X_S X_M^*}{4}, \\
2\lambda_{V*} &= -\frac{X_M X_S^*}{X_S X_S^*} - \frac{X_S X_M^*}{X_S X_S^*} = -\frac{X_M}{X_S} - \frac{X_M^*}{X_S^*}, \\
\lambda_{V*} &= -\operatorname{Re}\left(\frac{X_M}{X_S}\right), \quad \text{0-derivative minimizer}
\end{aligned} \tag{17}
$$

and the solution λ_{U*} for $U(\lambda)$ is the reciprocal of λ_{V*} after a change of variables $X_M \to X_S$ and $X_S \to X_M$.

The EPA constraint in Eq. 7 can be expanded w.r.t. MS:

$$
\begin{aligned}
\frac{|X_L|^2 + |X_R|^2}{2} &= \frac{|X_M + X_S|^2 + |X_M - X_S|^2}{8} \\
&= \frac{-X_M X_M^* - X_S X_S^*}{4}.
\end{aligned} \tag{18}
$$

Equating $|V(\lambda)|^2$ with EPA in Eqs. 17, 18, and rearranging λ yields the quadratic equation $0 = \lambda^2 + b\lambda - 1$ where

$$
\begin{aligned}
b &= \frac{X_M X_S^* + X_S X_M^*}{X_S X_S^*} = \frac{X_M}{X_S} + \frac{X_M^*}{X_S^*} = 2\operatorname{Re}\left(\frac{X_M}{X_S}\right), \\
\lambda &= \frac{-b \pm \sqrt{b^2+4}}{2} = \lambda_{V*} \pm \sqrt{\lambda_{V*}^2 + 1}.
\end{aligned} \tag{19}
$$

References

[1] A. Adami, E. A. P. Habets, and J. Herre. Down-mixing using coherence suppression. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2878–2882, 2014.

[2] C. Avendano. Frequency-domain source identification and manipulation in stereo mixes for enhancement, suppression and re-panning applications. In *2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 55–58, 2003.

[3] C. Avendano and J.-M. Jot. Frequency domain techniques for stereo to multichannel upmix. In *2002 International Conference: Virtual, Synthetic, and Entertainment Audio*, june 2002.

[4] M. R. Bai and G.-y. Shih. Upmixing and downmixing two-channel stereo audio for consumer electronics. *IEEE Transactions on Consumer Electronics*, 53(3):1011–1019, 2007.

[5] R. A. Bradley and M. E. Terry. The rank analysis of incomplete block designs. *Biometrika*, 39:324–345, 1952.

[6] C. P. Brown. Speech enhancement, Nov. 18 2014. US Patent 8,891,778.

[7] H. S. M. Coxeter. *Introduction to Geometry*. Wiley, New York, 1969.

[8] D. E. Critchlow and M. A. Fligner. Paired comparison, triple comparison, and ranking experiments as generalized linear models, and their implementation on glim. *Psychometrika*, 56(3):517, 1991.

[9] C. Faller and F. Baumgarte. Efficient representation of spatial audio using perceptual parametrization. In *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, pages 199–202, 2001.

[10] J. Fauvel and R. J. E. Flood, R.and Wilson. *Möbius and his Band: Mathematics and Astronomy in Nineteenth-Century Germany*. Oxford University Press, Oxford, England, 1993.

[11] J. T. Geiger, P. Grosche, and Y. L. Parodi. Dialogue enhancement of stereo sound. In *23rd European Signal Processing Conference*, pages 869–873, 2015.

[12] M. M. Goodwin. Geometric signal decompositions for spatial audio enhancement. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 409–412, 2008.

[13] M. M. Goodwin and J.-M. Jot. Primary-ambient signal decomposition and vector-based localization for spatial audio coding and enhancement. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 1, pages I–9–I–12, 2007.

[14] W. Hartmann and E. J. Macaulay. Anatomical limits on interaural time differences. *Frontiers in Neuroscience*, 8, 2014.

[15] J. He, W.-S. Gan, and E.-L. Tan. A study on the frequency-domain primary-ambient extraction for stereo audio signals. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2868–2872, 2014.

[16] K. Ibrahim and M. Allam. Primary-ambient extraction in audio signals using adaptive weighting and principal component analysis. In *13th Sound and Music Computing Conference*, 09 2016.

[17] R. Irwan and R. M. Aarts. Two-to-five channel sound processing *. *Journal of The Audio Engineering Society*, 50:914–926, 2002.

[18] S. Kraft and U. Zölzer. Stereo signal separation and upmixing by mid-side decomposition in the frequency-domain. In *18th International Conference on Digital Audio EffectsAt: Trondheim, Norway*, 12 2015.

[19] P. Papastergiou. Stereo-to-five channels upmix methods, implementation and comparative study. Master's thesis, Chalmers University of Technology / Department of Architecture and Civil Engineering, 2018.

[20] Samsudin, E. Kurniawati, N. B. Poh, F. Sattar, and S. George. A stereo to mono dowmixing scheme for mpeg-4 parametric stereo encoder. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 5, pages V–V, 2006.

[21] E. Vickers. Frequency-domain two-to three-channel upmix for center channel derivation and speech enhancement. In *Audio Engineering Society Convention 127*. Audio Engineering Society, 2009.