forum acusticum 2026
8-12 September · Graz, Austria

ANNOYANCE-BASED BEAMFORMING USING PERCEPTUAL COVARIANCE WEIGHTING

Konstantin Fontaine[1], Boaz Rafaely[2]

[1] *Department of Engineering Acoustics, Technische Universität Berlin, Berlin, Germany*
[2] *School of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer-Sheva, Israel*

Abstract

Adaptive beamformers such as the minimum variance distortionless response (MVDR) design commonly suppress interfering sources in proportion to their acoustic power, whereas the annoyance of the residual also depends on its spectral and temporal character. At low spherical harmonics (SH) orders, where only a few channels are available, the limited spatial selectivity may therefore be allocated to loud but tolerable sources while quieter, highly annoying ones remain audible. To allocate the selectivity by annoyance instead, a psychoacoustic annoyance model is embedded in the beamformer design. Directional signals obtained by plane-wave decomposition are evaluated for annoyance and equalized by loudness, and the resulting map defines the directional variances of a synthesized spatial covariance that retains the closed-form MVDR solution. The proposed annoyance-weighted design is evaluated at first SH order in twenty simulated free-field scenes of loud contextual and quieter, highly annoying sources, alongside a per-bin MVDR baseline and a power-weighted counterpart. The MVDR baseline achieves the strongest power reduction, yet its residual is significantly more annoying than a level-equivalent reference, since its selectivity is allocated to the loud contextual sources. The annoyance-weighted design reverses this allocation, attenuating the annoying sources strongly and the contextual sources only marginally, while the overall level remains nearly unchanged. Power reduction and annoyance-selective suppression thus emerge as distinct objectives at low spatial order, and annoyance-derived covariance weighting offers a closed-form means of directing the limited selectivity toward the most annoying sources.

Keywords: *beamforming, spherical harmonics, psychoacoustic annoyance, MVDR*

***Corresponding author**: k.fontaine@tu-berlin.de.*

1. Introduction

Spatial filtering for microphone arrays commonly aims to suppress interfering sound while preserving a target direction. A standard adaptive formulation is the minimum variance distortionless response (MVDR) beamformer in the spherical harmonics (SH) domain [1], which minimizes the output variance subject to a distortionless look direction. Since its criterion is signal energy alone, directional components are penalized in proportion to their power. Perceived annoyance, however, also depends on spectral and temporal properties, including sharpness, roughness, fluctuation strength, and tonality, which psychoacoustic annoyance (PA) models combine with loudness into a single annoyance prediction [2], [3]. Such models are established in product sound quality and soundscape assessment, yet they are commonly formulated for isolated or single-channel sounds and do not explicitly account for binaural listening, where spatial separation can reduce masking and alter the loudness contribution of concurrent sources [4], [5]. In complex acoustic scenes, power-based suppression may thus attenuate loud sources of relatively tolerable character while leaving quieter but more annoying interferers largely intact.

This mismatch becomes particularly relevant at low SH orders, where the $M = (N+1)^2$ channels of an order-N representation limit the filter degrees of freedom (DOF) to the microphone count of compact arrays [1], [6]. A first-order representation, for example, provides only four channels. Once the distortionless constraint is imposed, the remaining DOF are assigned by the power-minimization criterion to the dominant sources, irrespective of their annoyance character. Moreover, beamforming collapses the spatial scene into a single-channel residual, in which the altered source balance may change the masking and exposure of annoyance-relevant components. Two questions therefore arise, namely whether minimum-variance filtering is suboptimal for annoyance suppression under such DOF constraints, and whether annoy-

ance information can shift the source ranking from acoustic power toward psychoacoustic annoyance.

Previous work has used perceptual information around spatial processing mainly for post-hoc psychoacoustic assessment [7], [8], [9], for steering active noise control [10], or for class-selective hear-through [11], rather than within the beamformer optimization itself. The present study instead evaluates the residual annoyance and source-wise suppression of a conventional MVDR beamformer and embeds an annoyance-derived map into the covariance weighting of a closed-form beamformer design. The result is compared with the MVDR baseline and a power-weighted counterpart under matched first-order conditions.

2. Background

Consider a sound field modeled as a plane-wave amplitude density $a(\Omega, t)$ over directions $\Omega = (\theta, \phi)$, where the polar angle θ is measured downward from the zenith at $\theta = 0°$, and the azimuth ϕ counterclockwise from the frontal direction at $\phi = 0°$, so that $\phi = 90°$ points to the left. Its SH representation, truncated at order N, is given by the coefficient vector [1]

$$\mathbf{a}_{nm}(t) = \int_{\mathbb{S}^2} a(\Omega, t) \left[Y_n^m(\Omega)\right]^* \mathrm{d}\Omega \in \mathbb{C}^M, \quad (1)$$

with $M = (N+1)^2$, where Y_n^m denotes the complex spherical harmonic of order n and degree m, and $(\cdot)^*$ denotes complex conjugation. The subscript nm marks quantities in this domain. A single plane wave arriving from a direction Ω_0 contributes $\mathbf{a}_{nm}(t) = \mathbf{v}_{nm}(\Omega_0)\, s(t)$ through the steering vector

$$\mathbf{v}_{nm}(\Omega) = \left[\, Y_n^m(\Omega) \,\right]^*_{n\le N,\, |m|\le n} \in \mathbb{C}^M. \quad (2)$$

A beamformer maps the field onto a scalar output $y(t) = \mathbf{w}_{nm}^H \mathbf{a}_{nm}(t)$ through a weight vector $\mathbf{w}_{nm} \in \mathbb{C}^M$, and its spatial response is described by the beampattern $b(\Omega) = \mathbf{w}_{nm}^H \mathbf{v}_{nm}(\Omega)$, which by the frequency independence of (2) is valid at every frequency within the order-limited range.

For a field of J plane waves arriving from $\Omega_1, \ldots, \Omega_J$, (2) gives $\mathbf{a}_{nm} = \mathbf{V}\mathbf{s}$ with the steering matrix $\mathbf{V} = [\mathbf{v}_{nm}(\Omega_1) \cdots \mathbf{v}_{nm}(\Omega_J)]$ and the source vector $\mathbf{s} = [s_1 \cdots s_J]^T$, so that the spatial covariance of the field factorizes as

$$\mathbf{S}_{aa} = \mathrm{E}\{\mathbf{a}_{nm}\mathbf{a}_{nm}^H\} = \mathbf{V}\,\mathbf{S}_{ss}\,\mathbf{V}^H, \quad (3)$$

where $\mathbf{S}_{ss} = \mathrm{E}\{\mathbf{s}\mathbf{s}^H\}$ is the source cross-spectral matrix.

Adaptive MVDR processing commonly operates on the short-time Fourier transform (STFT) of (1), with coefficients $\mathbf{a}_{nm}(k, \tau)$ at bin k and frame $\tau = 1 \ldots L$. The per-bin spatial covariance is estimated as the sample average over the frames,

$$\hat{\mathbf{S}}_{aa}(k) = \frac{1}{L} \sum_{\tau=1}^{L} \mathbf{a}_{nm}(k, \tau)\, \mathbf{a}_{nm}^H(k, \tau), \quad (4)$$

and is stabilized by a trace-relative diagonal loading $\lambda\, \mathrm{tr}\, \hat{\mathbf{S}}_{aa}(k)\, \mathbf{I}/M$. Throughout the beamformer design, $\mathbf{a}_{nm}$ denotes the interferer portion of the field, $\mathbf{a}_{nm} \equiv \mathbf{a}_{nm}^{\mathrm{int}}$, so that (4) is a direct sample estimate of (3), free of target contamination, and therefore serves as the oracle reference.

Based on this estimate, the MVDR beamformer [6] can be formulated and solved as

$$\min_{\mathbf{w}_{nm}(k)} \mathbf{w}_{nm}^H(k)\, \hat{\mathbf{S}}_{aa}(k)\, \mathbf{w}_{nm}(k) \quad \text{s.t.} \quad \mathbf{w}_{nm}^H(k)\, \mathbf{v}_l = 1,$$

$$\mathbf{w}_{nm}(k) = \frac{\hat{\mathbf{S}}_{aa}^{-1}(k)\, \mathbf{v}_l}{\mathbf{v}_l^H \hat{\mathbf{S}}_{aa}^{-1}(k)\, \mathbf{v}_l}, \quad (5)$$

where $\mathbf{v}_l \equiv \mathbf{v}_{nm}(\Omega_l)$ denotes the steering vector of the look direction. The weights are applied per bin as $y(k, \tau) = \mathbf{w}_{nm}^H(k)\, \mathbf{a}_{nm}(k, \tau)$, and the output signal is recovered as $y(t) = \mathrm{iSTFT}\{y(k, \tau)\}$.

3. Proposed Method

3.1. Directional annoyance mapping

Directional signals are extracted on a Lebedev quadrature grid $\{\Omega_q\}_{q=1}^Q$ of Q sampling points, by reusing the beamformer of (5). For a diffuse field the covariance is proportional to the identity, $\hat{\mathbf{S}}_{aa} \propto \mathbf{I}$, for which (5) reduces to the maximum-directivity beamformer $\mathbf{w}_{nm} = \mathbf{v}_{nm}(\Omega_q)/\|\mathbf{v}_{nm}(\Omega_q)\|^2$, steered to each grid point,

$$d_q(t) = \frac{\mathbf{v}_{nm}^H(\Omega_q)}{\|\mathbf{v}_{nm}(\Omega_q)\|^2}\, \mathbf{a}_{nm}(t), \quad (6)$$

where $\|\mathbf{v}_{nm}(\Omega)\|^2 = M/4\pi$ follows from the SH addition theorem [1]. Each $d_q(t)$ therefore estimates the pressure signal arriving from Ω_q at the spatial resolution of that order. The power of each directional signal is then computed as $P_q = \frac{1}{T}\sum_t |d_q(t)|^2$, and its annoyance

$$A_q = \mathrm{PA}\{d_q(t)\} \quad (7)$$

is obtained with the psychoacoustic annoyance model [3],

$$\mathrm{PA} = N_5 \left(1 + \sqrt{w_S^2 + w_{FR}^2 + w_T^2}\right), \quad (8)$$

with

$$\begin{aligned} w_S &= \tfrac{1}{4}\,(S - 1.75)\, \log_{10}(N_5 + 10), \\ w_{FR} &= \frac{2.18}{N_5^{0.4}}\,(0.4\,F + 0.6\,R), \\ w_T &= \frac{6.41}{N_5^{0.52}}\,K, \end{aligned} \quad (9)$$

where N_5 denotes the percentile loudness, S the sharpness, R the roughness, F the fluctuation strength, and K the tonality of the analyzed signal.

Since PA is dominated by N_5 in (8), the raw map $\{A_q\}_{q=1}^Q$ mainly follows the power map $\{P_q\}_{q=1}^Q$ when sources differ considerably in level, as Fig. 1 illustrates for an exemplary scene. Because the present method

aims to steer selectivity by annoyance character rather than by level, the map is equalized by its own loudness,

$$W_q \;=\; \frac{A_q}{N_{5,q}^{\beta}}, \tag{10}$$

where $N_{5,q}$ is the percentile loudness of $d_q(t)$ and β regulates the contribution of level, with $\beta = 0$ recovering the raw map. With $\beta = 1$, substituting (8) into (10) gives the annoyance character map $W_q = 1+\sqrt{w_S^2 + w_{FR}^2 + w_T^2}\,\big|_q$, which ranks directions by their annoyance per unit loudness. The equalization therefore leaves the sensory quantities themselves unaffected and only regulates the contribution of loudness to the predicted annoyance.

3.2. Covariance weighting and beamformer design

The map-driven designs replace the measured covariance of (4) by a model of (3). For mutually uncorrelated sources $\mathbf{S}_{ss}$ is diagonal, and describing the field by its directional distribution rather than by a discrete set of sources turns (3) into

$$\mathbf{S}_{aa} \;\approx\; \mathbf{V}_Q \,\mathrm{diag}(m_1, \ldots, m_Q)\, \mathbf{V}_Q^H, \tag{11}$$

where $\mathbf{V}_Q = [\mathbf{v}_{nm}(\Omega_1) \cdots \mathbf{v}_{nm}(\Omega_Q)]$ collects the steering vectors of the quadrature grid and $m_q \geq 0$ is the variance assigned to direction Ω_q. Expanded over the grid, (11) yields the synthesized spatial covariance

$$\begin{aligned} \tilde{\mathbf{S}}_{aa} &= \sum_{q=1}^{Q} m_q\, \mathbf{v}_{nm}(\Omega_q)\, \mathbf{v}_{nm}^H(\Omega_q), \\ \mathbf{w}_{nm}^H \tilde{\mathbf{S}}_{aa} \mathbf{w}_{nm} &= \sum_q m_q\, |b(\Omega_q)|^2, \end{aligned} \tag{12}$$

whose quadratic form penalizes beampattern gain in proportion to the map. Built at the filter order, $\tilde{\mathbf{S}}_{aa}$ replaces $\hat{\mathbf{S}}_{aa}(k)$ in (5) and retains the closed-form solution. Since the maps are broadband and (2) is frequency independent, a single broadband weight vector results, which remains less adaptive than the per-bin MVDR formulation of (5).

An annoyance-weighted beamformer (AWB) is formulated using this character map. By the grid symmetry, however, $\sum_q \mathbf{v}_{nm}(\Omega_q)\, \mathbf{v}_{nm}^H(\Omega_q) = (Q/4\pi)\, \mathbf{I}$ at the filter order, so that a constant offset in m_q contributes an identity term to $\tilde{\mathbf{S}}_{aa}$ and acts as diagonal loading in (5). Since W_q has a near-unit floor by (10), this loading would dominate the design and leave it close to the non-adaptive plane-wave solution. The map is therefore used with its floor removed and its contrast expanded,

$$\tilde{W}_q = \Big(W_q - \min_{q'} W_{q'}\Big)^{\kappa}, \qquad m_q = \tilde{W}_q \Big/ \max_{q'} \tilde{W}_{q'}, \tag{13}$$

with κ restoring a power-like dynamic range.

Setting $m_q = P_q$ instead yields the power-weighted beamformer (PWB), whose $\tilde{\mathbf{S}}_{aa}$ estimates the directional variances from the field itself and is thus the closest available reconstruction of (3), serving as the energy-based counterpart. Both map covariances receive the same trace-relative loading, so that PWB and AWB share (12), the solver, and the regularization, and differ only in the map.

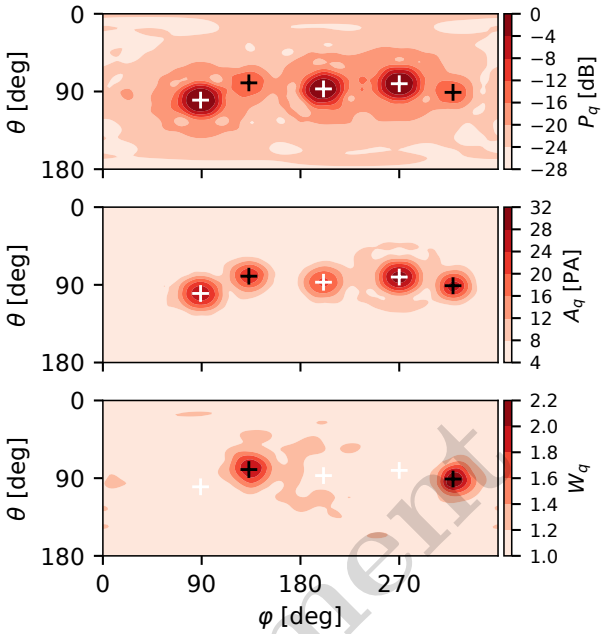

Figure 1: Directional maps for an exemplary scene from the simulation study, showing the distribution of power P_q (top), psychoacoustic annoyance A_q (middle), and loudness-equalized character W_q (bottom). Markers indicate contextual source positions (white) and annoying source positions (black).

4. Simulation Study

4.1. Setup

Twenty free-field scenes are simulated, each consisting of three loud contextual sources and two quieter annoying interferers at 2 m distance. The contextual class comprises an ambient background recording, a ventilation recording, and a synthetic crowd murmur, all steady broadband signals covering the full audio bandwidth, presented at 78 dB SPL. The annoying class draws two sources per scene from a pool of four, namely amplitude-modulated white noise at 70 and 4 Hz modulation rate as well as a male and a female speech recording, presented at 65 dB SPL and thus 13 dB below the contextual class. All signals are sampled at 48 kHz with a duration of 5 s, and their levels are specified at 1 m distance from the source. Interferer positions are chosen randomly within $80° \leq \theta \leq 100°$ and $30° \leq \phi \leq 330°$, subject to a minimum angular spacing of 30° from the frontal look direction $\Omega_l = (90°, 0°)$ and from every other source. Table 1 reports the psychoacoustic metrics of the seven sources at a common level of 78 dB SPL,

Table 1: Psychoacoustic metrics of the source signals equalized to a common level of 78 dB SPL and grouped into the contextual class (top) and the highly annoying class (bottom). Reported quantities are percentile loudness N_5, sharpness S, roughness R, fluctuation strength F, tonality K, the resulting annoyance PA, and the character ratio $W = \mathrm{PA}/N_5$.

	N_5	S	R	F	K	PA	W
ambient background	25.4	1.5	0.11	0.38	0.05	29.1	1.14
ventilation	18.0	1.1	0.08	0.04	0.11	20.8	1.16
crowd murmur	21.6	1.4	0.12	0.39	0.07	25.4	1.17
am. white 70 Hz	22.8	3.0	2.86	0.06	0.03	49.8	2.18
am. white 4 Hz	25.2	3.0	0.08	4.60	0.03	56.3	2.23
speech male	26.6	3.4	1.86	3.85	0.13	71.7	2.69
speech female	26.7	2.5	3.62	3.60	0.23	84.3	3.15

at which the classes reach similar loudness while the annoying class far exceeds the contextual class in annoyance, primarily through the modulation metrics.

4.2. Methodology

The scenes are synthesized at order $N = 35$, assuming an ideal higher-order receiver. Three designs are compared on every scene, namely the per-bin MVDR beamformer of (5) as the oracle baseline and the broadband pair PWB and AWB of Section 3.2, which share the construction of (12) and differ only in the map entering it. The STFT of the per-bin MVDR uses a Hann window of 2048 samples with 50 % overlap. The maps of Section 3.1 are computed at order 7 on a Lebedev grid with $Q = 302$ points, whose lobe width keeps map peaks attributable to individual sources at the imposed spacing, whereas the filters operate on the first-order truncation with $M = 4$ channels. The character map uses $\beta = 1$ and the contrast expansion $\kappa = 3$. The trace-relative diagonal loading is set to $\lambda = 10^{-3}$ for the MVDR covariance and to $\lambda = 10^{-1}$ for the map pair. The psychoacoustic metrics are computed with the SQAT MATLAB toolbox [12], whose implementation evaluates them as 5 % exceedance percentiles of their time series and sets w_S to zero for $S \leq 1.75$.

4.3. Evaluation metrics

Since every distortionless design passes the target unaltered, the annoyance of the full output may be dominated by the clean target. All metrics are therefore computed on the residual $y_{\mathrm{res}}(t) = \mathbf{w}_{nm}^H \mathbf{a}_{nm}^{\mathrm{int}}(t)$, obtained by applying the full-field weights to the interferer-only field. With p_{omni} denoting that field at an omnidirectional sensor, the overall power attenuation $\sigma_P = 10 \log_{10} \left(\|p_{\mathrm{omni}}\|^2 / \|y_{\mathrm{res}}\|^2 \right)$ and annoyance attenuation $\sigma_A = 10 \log_{10} \left(\mathrm{PA}\{p_{\mathrm{omni}}\} / \mathrm{PA}\{y_{\mathrm{res}}\} \right)$ quantify the energy and annoyance objectives, respectively.

As annoyance decreases with level, σ_A does not necessarily separate the effect of the spatial filtering from that of the level reduction it achieves. Each residual is therefore compared with the unfiltered omnidirectional scene attenuated to the same residual power,

Table 2: Beamformer comparison across all scenes as medians, with E as the geometric mean. Reported are the level-equivalent annoyance gain E, the overall power and annoyance attenuation σ_P and σ_A, the residual character W_{res}, and the annoying-class share of the residual power ρ_{ann}. The unprocessed mixture anchors the residual metrics.

	E	σ_P [dB]	σ_A [dB]	W_{res}	ρ_{ann} [%]
unprocessed	–	0	0	1.13	3.2
MVDR	1.19	14.0	3.1	1.20	22.6
PWB	1.05	7.6	1.9	1.16	9.6
AWB	0.99	0.2	0.2	1.13	0.0

yielding the level-equivalent annoyance gain

$$E = \frac{\mathrm{PA}\{y_{\mathrm{res}}\}}{\mathrm{PA}\{g\, p_{\mathrm{omni}}\}}, \tag{14}$$

where $g = \|y_{\mathrm{res}}\| / \|p_{\mathrm{omni}}\|$ scales the reference to the residual power and $E > 1$ marks a residual more annoying than a plain level reduction of the same depth. The gains are tested as $\log E$ with a two-sided Wilcoxon signed-rank test and reported as geometric means.

Source-wise allocation is measured by applying each set of weights to the isolated source fields, yielding the power attenuation Δ_P and annoyance attenuation Δ_A per source, defined in the same manner and reported as medians of the per-scene class medians. Since loudness roughly doubles per 10 dB, the just-noticeable level difference of 1 dB corresponds to only about 0.3 dB in the annoyance metrics, which is adopted as the threshold for perceptually relevant differences. The residual composition is further described energetically by the annoying-class share of the residual power ρ_{ann} and perceptually by the residual character $W_{\mathrm{res}} = W\{y_{\mathrm{res}}\}$, according to (10).

4.4. Results

Table 2 summarizes the residual metrics for the three designs. As expected, MVDR achieves the strongest energy reduction, with median overall attenuations of 14.0 dB in power and 3.1 dB in annoyance. PWB yields a smaller reduction of 7.6 dB and 1.9 dB, respectively. Relative to the level-equivalent reference, however, the MVDR residual carries significantly more annoyance, with a geometric mean gain of $E = 1.19$ (Wilcoxon $p = 0.006$). Thus, the reduction in total level does not correspond to an equivalent reduction in annoyance. This is reflected in the residual composition, where the annoying class holds 3.2 % of the interferer power in the unprocessed mixture, but 22.6 % in the MVDR residual, while the residual character increases from $W = 1.13$ to $W_{\mathrm{res}} = 1.20$. PWB shows the same tendency at weaker suppression, although its excess over the reference is not significant ($E = 1.05$, $p = 0.17$). AWB, in contrast, yields little total reduction, with median overall power and annoyance attenuation both at 0.2 dB, below the 0.3 dB threshold. Its residual character remains at the mixture value ($W_{\mathrm{res}} = 1.13$),

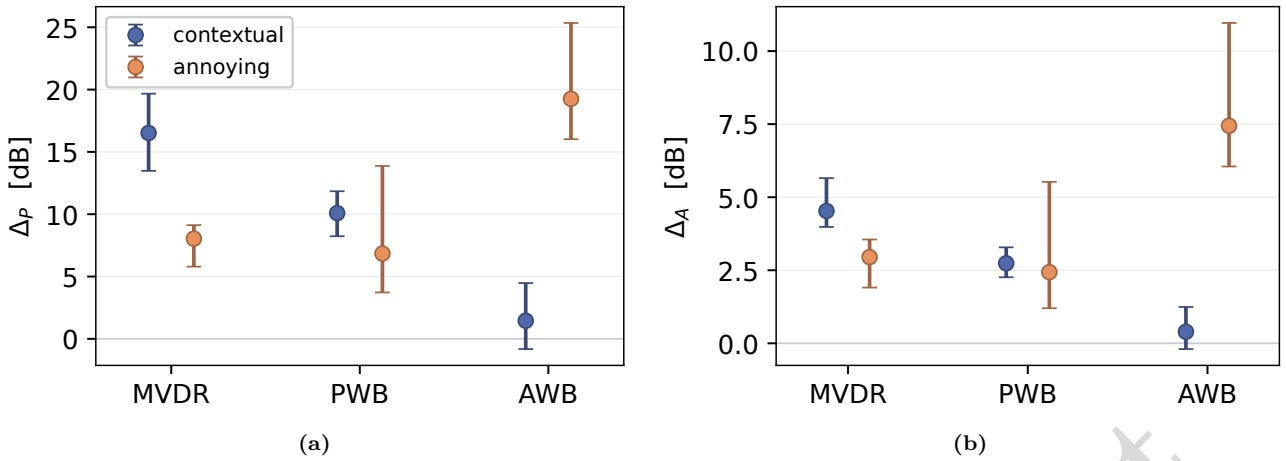

Figure 2: Class-wise source attenuation for MVDR, PWB, and AWB. Panels show the power attenuation Δ_P (a) and annoyance attenuation Δ_A (b) as medians of the per-scene class medians with 95 % percentile confidence intervals.

the annoying-class share of the residual power falls below 0.1 %, and the level-equivalent annoyance gain remains close to the unfiltered reference, with no systematic deviation ($E = 0.99$, $p = 0.55$). Within the tested setting, AWB therefore changes the residual composition rather than its overall level.

The per-source analysis in Fig. 2 shows how attenuation is allocated across classes. MVDR attenuates the loud contextual class more strongly than the annoying class, with class-median power attenuations of $\Delta_P = 16.5$ dB and 8.0 dB, respectively. PWB, which uses the power map, preserves this class ordering at smaller attenuation depth ($\Delta_P = 10.1$ dB and 6.8 dB). AWB reverses the ordering and attenuates the annoying class more strongly than the contextual class in both power and annoyance. In power, the class medians are $\Delta_P = 19.3$ dB for the annoying class and 1.5 dB for the contextual class. The annoyance attenuation of the annoying class follows this allocation, with $\Delta_A = 3.0$ dB under MVDR, 2.4 dB under PWB, and 7.4 dB under AWB.

5. Discussion

The MVDR analysis shows how a power-minimizing criterion structures the residual of a low-order spatial filter when source level and annoyance character are not aligned. With exact knowledge of the interference statistics, MVDR provides the strongest reduction in residual power and, in absolute terms, also the lowest residual PA. Comparison with the level-equivalent reference, however, shows that this reduction is not annoyance-equivalent. The residual contains a larger relative contribution from the annoying class because the limited selectivity is mainly assigned to the louder contextual sources. In the single-channel output, this changes the composition of the mixture, which is consistent with reduced masking of the components that carry the higher annoyance character.

The weighted-covariance comparison separates this source allocation from the particular MVDR implementation. PWB preserves the energy-driven ordering and therefore resembles the MVDR allocation at reduced depth. AWB, by contrast, uses the loudness-equalized character map and shifts the attenuation toward the annoying class. The selectivity shift appears both in residual power and in residual PA, indicating that an annoyance-derived map can redirect the scarce spatial selectivity when embedded in the beamformer design.

This redirection also makes the central trade-off explicit. AWB reduces the annoying components most strongly but leaves the louder contextual background largely unattenuated. Its total residual power and absolute residual PA therefore remain high compared with MVDR. The gain is compositional rather than energetic, since the residual retains the broader scene while containing little energy from the sources with high annoyance character. At the present first SH order, power reduction and annoyance-selective suppression therefore appear as distinct objectives rather than jointly optimized outcomes.

The observed reversal depends on how PA enters the design. The raw PA map is dominated by loudness and therefore approaches the power map when sources differ considerably in level. The loudness equalization removes this contribution so that the covariance weighting is determined by the spectral and temporal character terms. The equalization therefore isolates the effect studied here, but uses only the character terms of the annoyance model, whereas the evaluation retains the full model. The results thus provide model-based support that spatial selectivity can be redirected by annoyance character within a closed-form design, while perceptual validation of the residual annoyance remains open.

The interpretation of the findings is constrained by the use of PA as an acoustic model prediction and by the controlled simulation setup. Perceived annoyance in spatial mixtures may also depend on masking, source attitude, and listening context, so a percep-

tual evaluation is required to determine whether the predicted residual differences correspond to listener judgments. A practical implementation would additionally require time-varying covariance and map estimates, robustness to non-ideal array conditions, and target exclusion during processing. The loudness equalization, which isolates annoyance character here, should also be evaluated perceptually, since practical systems may need to relax it so that level is readmitted as a relevant cue, or to gate it so that very quiet sources do not receive excessive weight.

Overall, the study indicates that psychoacoustic information can not only assess a beamformer residual but also define how the limited selectivity of a low-order array is allocated. In the tested scenes, the allocation induced by acoustic power differs from that induced by annoyance character. The proposed covariance weighting offers one way of exploiting this difference, while making the accompanying cost in residual level explicit.

6. Conclusions

An annoyance-weighted beamformer has been presented that embeds a loudness-equalized character map, derived from a psychoacoustic annoyance model, into the spatial covariance of a closed-form SH-domain design. Its residual annoyance and per-source attenuation were evaluated against MVDR and power-weighted beamforming with first-order SH filters in simulated free-field scenes of loud contextual and quieter annoying sources.

Although the MVDR beamformer attains the deepest power and annoyance attenuation, its residual is significantly more annoying than a level-equivalent reference, since the limited selectivity favors the loud contextual sources and exposes the annoying components. The annoyance-weighted design reverses this allocation and attenuates the annoying class most strongly while leaving the contextual background almost unchanged, so that the residual consists almost entirely of contextual sources.

Power reduction and annoyance-selective suppression thus prove to be distinct objectives at low spatial order. Psychoacoustic weighting can redirect the scarce selectivity toward the most annoying sources, at the cost of residual level. Future work should validate the predicted differences perceptually and extend the weighting to time-varying, level-aware processing.

Acknowledgments

This work was funded by the HEAD-Genuit Foundation as part of the MOSAIC project, and the authors thank Dr. André Fiebig for his valuable advice on the topic.

References

[1] B. Rafaely, *Fundamentals of Spherical Array Processing* (Springer Topics in Signal Processing), 2nd ed. Cham: Springer, 2019, vol. 16.

[2] U. Widmann, "Ein Modell der psychoakustischen Lästigkeit von Schallen und seine Anwendung in der Praxis der Lärmbeurteilung," Ph.D. dissertation, Technische Universität München, Munich, Germany, 1992.

[3] G.-Q. Di, X.-W. Chen, K. Song, B. Zhou, and C.-M. Pei, "Improvement of Zwicker's psychoacoustic annoyance model aiming at tonal noises," *Applied Acoustics*, vol. 105, pp. 164–170, 2016.

[4] B. C. J. Moore, B. R. Glasberg, and T. Baer, "A model for the prediction of thresholds, loudness, and partial loudness," *Journal of the Audio Engineering Society*, vol. 45, no. 4, pp. 224–240, 1997.

[5] J. F. Culling and M. Lavandier, "Binaural unmasking and spatial release from masking," in *Binaural Hearing*, ser. Springer Handbook of Auditory Research, R. Y. Litovsky, M. J. Goupell, R. R. Fay, and A. N. Popper, Eds., vol. 73, Cham: Springer, 2021, pp. 209–241.

[6] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. New York: Wiley-Interscience, 2002.

[7] W. Song, W. Ellermeier, and J. Hald, "Using beamforming and binaural synthesis for the psychoacoustical evaluation of target sources in noise," *The Journal of the Acoustical Society of America*, vol. 123, no. 2, pp. 910–924, 2008.

[8] W. Song, W. Ellermeier, and J. Hald, "Psychoacoustic evaluation of multichannel reproduced sounds using binaural synthesis and spherical beamforming," *The Journal of the Acoustical Society of America*, vol. 130, no. 4, pp. 2063–2075, 2011.

[9] M. Alkmim et al., "Drone noise directivity and psychoacoustic evaluation using a hemispherical microphone array," *The Journal of the Acoustical Society of America*, vol. 152, no. 5, pp. 2735–2745, 2022.

[10] P. Zachos, G. Moiragias, and J. Mourjopoulos, "Targeted beamforming active noise control based on disturbance metrics," *Acta Acustica*, vol. 8, p. 39, 2024.

[11] B. Veluri, M. Itani, J. Chan, T. Yoshioka, and S. Gollakota, "Semantic hearing: Programming acoustic scenes with binaural hearables," in *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST)*, San Francisco, CA, USA, 2023.

[12] G. Felix Greco, R. Merino-Martínez, A. Osses, and S. C. Langer, "SQAT: A MATLAB-based toolbox for quantitative sound quality analysis," in *Proceedings of the 52nd International Congress and Exposition on Noise Control Engineering (INTER-NOISE)*, vol. 268, Chiba, Japan, 2023, pp. 7172–7183.