

CONSTRAINT OPTIMIZED MULTICHANNEL MIXER-LIMITER DESIGN

1st Yuancheng Luo, 2nd Dmitriy Yamkovoy, 3rd Guillermo Garcia
Amazon.com, {mikeluo, dpy, guigarg}@amazon.com

ABSTRACT

Multichannel audio mixer and limiter designs are conventionally decoupled for content reproduction over loudspeaker arrays due to high computational complexity and run-time costs. We propose a coupled mixer-limiter-envelope design formulated as an efficient linear-constrained quadratic program that minimizes a distortion objective over multichannel gain variables subject to sample mixture constraints. Novel methods for asymmetric constant overlap-add window optimization, objective function approximation, variable and constraint reduction are presented. Experiments demonstrate distortion reduction of the coupled design, and computational trade-offs required for efficient real-time processing.

Index Terms— Limiter, window optimization, constraint reduction

1. INTRODUCTION

Multichannel audio content reproduction over loudspeaker arrays has grown in popularity in recent years with the proliferation of low-cost sound-bar and smart-speaker consumer electronics. Such audio-reproduction systems are typically resource-constrained compared to professional-grade loudspeakers in terms of available transducer-level digital-electrical headroom, and acoustic output. A conventional method digitally compresses the audio content's dynamic-range throughout channel-mixing stages to stay under digital full-scale levels, maximize loudness, and satisfy studio standards [1–4]. A channel matrix-mixer [5, 6] or matrix-decoder [7–9] therefore allocates headroom between input channels and output transducers; dynamic-range controllers (DRCs) [10–12] such as peak-limiters are thereby placed downstream for transducer and amplifier protection.

Several deficiencies of the conventional method are known: A multichannel matrix-mixer can conservatively pre-allocate headroom for each input channel such that maximum channel mixture levels minimally activate downstream DRCs. Pre-allocation however distorts the audio mixture in the absence of run-time monitoring; a channel's upper dynamic range may be unnecessarily lowered when other channels are sparse, and the choice of mixing gains can alter the spectral and channel balance of the original content. Terminal DRCs operating on mixtures of channels per transducer can limit at different times, intermittently distorting both channel balance and the loudspeaker array's directivity.

We address both pre-allocation and terminal-limiter problems by coupling time-varying channel-gain reduction with per-sample constraints of the channel mixtures via a sequence of quadratic programming (QP) problems. Section 2 presents our QP mixer-limiter design, and relates the QP's feasibility with a novel constant overlap-add (COLA) [13] constrained gain envelope construction. Section 3 presents the channel-mixture's distortion objective and derives the QP objective from the former's optimal Taylor series approximation. Section 4 extends the QP formulation to joint multi-band [14] multi-content mixers, and introduces novel variable and constraint size reduction methods for efficient computation. Section 5 shows experimental results for distortion reduction and computational performance.

2. QUADRATIC PROGRAM MIXER-LIMITER

Let $\mathbf{S} \in \mathbb{R}^{F \times N}$ be a matrix of N input channel column-vectors of F samples in an audio frame. Each of the N input channels are mixed with independent variable gains in vector $\mathbf{x} = [x_1, \dots, x_N]^T$ to produce a single output channel mixture limited in dynamic range and satisfying

the equivalent modulus linear constraints given by

$$-\tau \leq \sum_{n=1}^N S_{mn} x_n \leq \tau, \quad \tau \geq 0, \quad 1 \leq m \leq F, \quad (1)$$

where τ is a user-specified non-negative threshold for any mixture of samples within a frame. The variable gains are subject to non-negative bounds under unity and are equivalent to box-constraints given by

$$0 \leq l_n \leq x_n \leq u_n \leq 1, \quad 1 \leq n \leq N, \quad (2)$$

and therefore only apply gain reduction to each input channel. The gain variables \mathbf{x} , constrained to the feasible space that satisfy the linear constraints, are then found by minimization over a quadratic polynomial objective function. The latter's standard form is given by

$$f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{Q} \mathbf{x} + \mathbf{c}^T \mathbf{x} + d, \quad (3)$$

where symmetric matrix $\mathbf{Q} \in \mathbb{R}^{N \times N}$, vector $\mathbf{c} \in \mathbb{R}^{N \times 1}$, and constant d parameterize our distortion objective in section 3. Minimizing (3) subject to constraints (1), (2) in vectorized form is given by

$$\mathbf{x}_* = \arg \min_{\mathbf{x}} f(\mathbf{x}), \quad \text{s.t.} \quad -\tau \leq \mathbf{S} \mathbf{x} \leq \tau, \quad \mathbf{l} \leq \mathbf{x} \leq \mathbf{u} \leq \mathbf{1}, \quad (4)$$

where $\tau = \tau \mathbf{1} = \tau \sum_{i=1}^F \mathbf{e}_i \in \mathbb{R}^{F \times 1}$ is the vector of constant threshold τ over frame-samples, \mathbf{e}_i is the standard basis, and $\mathbf{l}, \mathbf{u} \in \mathbb{R}^{N \times 1}$ are the vectors of lower and upper bound respectively. We now construct the limiter's gain envelope across frame-wise solutions to (4).

Constrained Limiter-Envelope Design: Attack, hold, and release-time parameters are commonly used to restrict the velocity and shape of a limiter's gain envelope, constraining the latter to be smooth and reducing any audible distortion when multiplied by the input signals [15]; attack refers to the early portion of the gain envelope of increasing gain reduction, hold is the middle portion of constant gain reduction, and release is the late portion of decreasing gain reduction. We can define an envelope function with attack, hold, release dynamics, and supports over the solutions to (4) across overlapped audio frames with look-ahead. Given an audio stream $Y(m, n)$ of N input channels for sample index m and channel index n , let $\mathbf{S}^{\{k\}} \in \mathbb{R}^{\bar{F} \times N}$ be the k^{th} audio frame of size F augmented with L look-ahead samples per channel, where $\bar{F} = F + L$, be defined as follows:

$$\mathbf{S}_{mn}^{\{k\}} = \begin{cases} Y(m + (k-1)F, n), & 1 \leq m \leq F + L \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

For the k^{th} augmented frame, define the QP minimizer $\mathbf{x}_*^{\{k\}}$ s.t. the sample constraint set $\xi^{\{k\}}$ in the frame and look-ahead as follows:

$$\begin{aligned} \mathbf{x}_*^{\{k\}} &= \arg \min_{\mathbf{x}} f(\mathbf{x}) \quad \text{s.t. linear constraints } \xi^{\{k\}}, \\ \xi^{\{k\}} &= \begin{cases} -\tau \leq \mathbf{S}^{\{k\}} \mathbf{x} \leq \tau, & 2\bar{F} \text{ Mixture limits (1)} \\ \mathbf{0} \leq \mathbf{x} \leq \mathbf{u} \leq \mathbf{1}, & 2N \text{ Variable bounds (2)} \end{cases} \end{aligned} \quad (6)$$

where $\mathbf{S}^{\{k\}}$ replaces \mathbf{S} in (4), $\tau = \tau \mathbf{1} \in \mathbb{R}^{\bar{F} \times 1}$, and the variable lower bounds are set to $\mathbf{0}$ to ensure that a feasible solution exists. The enveloped output mixture $y(t)$ at time t is therefore given by the sum of gain enveloped inputs and the envelope function $v: \mathbb{R} \rightarrow \mathbb{R}^{0+}$:

$$y(t) = \sum_{n=1}^N Y(t, n) v_n(t), \quad v_n(t) = \sum_{k=0}^{\infty} W_n(t - kF) x_{*n}^{\{k+1\}}, \quad (7)$$

where $v_n(t)$ are weighted mixtures of solutions across frames. We show the bounds $-\tau \leq y(t) \leq \tau$ and $0 \leq v_n(t) \leq 1$ are satisfied via the design of a weighting function $W_n(t)$.

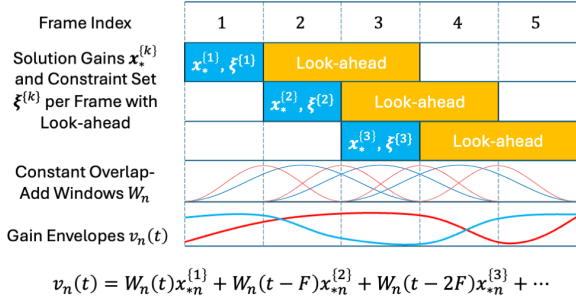


Fig. 1: Blue and red envelopes $v_n(t)$ are COLA weighted mixtures of solutions in (7), belonging to different input channels, and computed over augmented frames $\mathcal{S}^{\{k\}}$ in (5) with different look-ahead times.

Observe in Fig. 1 that the convex set of solutions from consecutive frames $\alpha x_n^{\{k\}} + (1-\alpha)x_n^{\{k+1\}}$, $0 \leq \alpha \leq 1$ is non-empty for the half-space intersection of $\xi^{\{k\}} \cap \xi^{\{k+1\}}$ common mixture-limit constraints from overlapped look-ahead and frame samples of $\mathcal{S}^{\{k\}}, \mathcal{S}^{\{k+1\}}$ respectively. The set of solutions $v_n(t)$ in (7), evaluated across multiple consecutive frames, is convex and satisfies the mixture-limited constraints of (1) at time t if $W_n(t)$ are window functions with bounded non-negative supports in the frame and look-ahead interval, zero-valued elsewhere, and have the COLA [13] property given by

$$\left\{ \begin{array}{ll} 0 \leq W_n(t) \leq 1, & 1 \leq t \leq F+L \\ W_n(t) = 0, & \text{otherwise} \end{array} \right\}, \quad \text{Bounded supports,} \quad (8)$$

$$\sum_{k=-\infty}^{\infty} W_n(t-kF) = 1, \quad \forall t \in \mathbb{R}, \quad \text{COLA property.}$$

For bounded time t , the gain envelopes $\mathbf{v}(t) = [v_1(t), \dots, v_N(t)]^T$ span Barycentric weighted $x_n^{\{k\}}$ of consecutive sequences of frame index k satisfying $1 + (k-1)F \leq t \leq (k-1)F + F + L$. We therefore design window function $W_n(t)$ with bounded COLA and dynamics constraints for smoothly transitioning between supports or frames as to minimize discontinuities in the enveloped output mixture $y(t)$.

Dynamics Constrained COLA Window Design: We can approximate a smooth COLA window $W(t)$ with characteristic attack, hold, and release dynamics via constrained optimization over the uniformly-sampled and integer-spaced window samples $\omega(t) = W(t)$, $t \in \mathbb{Z}$. Let the characteristic dynamics of $W(t)$ be defined by its first-derivative's intervals w.r.t. attack-release onsets as follows:

$$\begin{array}{lll} \text{Attack} & dW/dt \geq 0 & 1 \leq t < T_A \quad \text{Attack-onset } T_A \\ \text{Hold} & dW/dt = 0 & T_A \leq t < T_R \quad \text{Release-onset } T_R \\ \text{Release} & dW/dt \leq 0 & T_R \leq t \leq M \quad \text{Window-size } M. \end{array} \quad (9)$$

The velocity, acceleration, and smoothness of $W(t)$ can be approximated via the following first-order forward, central, and squared central finite-differences of $\omega(t)$ respectively:

$$\begin{aligned} dW/dt &\approx \omega(t+1) - \omega(t) = \omega(t) * \nu_v(t), \\ d^2W/dt^2 &\approx \omega(t+1) - 2\omega(t) + \omega(t-1) = \omega(t) * \nu_a(t), \\ (d^2W/dt^2)^2 &\approx \omega(t) * \nu_R(t) * \omega(t), \quad \nu_R(t) = \nu_a(t) * \nu_a(t), \end{aligned} \quad (10)$$

where $*$ is the discrete convolution operation, $\nu_v(t)$, $\nu_a(t)$, are the velocity, acceleration kernels respectively given by

$$\nu_v(t) = \begin{cases} +1, & t = 1 \\ -1, & t = 0 \\ 0, & \text{otherwise} \end{cases}, \quad \nu_a(t) = \begin{cases} +1, & t = \pm 1 \\ -2, & t = 0 \\ 0, & \text{otherwise} \end{cases}, \quad (11)$$

and $\nu_R(t)$ is the squared-acceleration smoothness kernel. We maximize

the overall smoothness of $W(t)$ with characteristic dynamics of (9) by minimizing the total squared-acceleration of (10) subject to causal COLA of (8) and finite-difference velocity constraints.

Let $\omega = [\omega(1), \dots, \omega(M)]^T \in \mathbb{R}^{M \times 1}$ be the unknown window samples, and the vectorized QP minimization be given by

$$\begin{aligned} \omega_* &= \arg \min_{\omega} \sum_{t=1}^M (\omega(t) * \nu_a(t))^2 = \omega^T \mathbf{T}_0^M (\nu_R) \omega, \quad \text{s.t.} \\ \mathbf{T}_0^F (\nu_c) \omega &= \mathbf{1}, \quad \nu_c(t) = \begin{cases} 1, & t \equiv 0 \pmod{F} \\ 0, & \text{otherwise} \end{cases}, \quad \omega \geq 0, \\ \mathbf{T}_0^{TA} (\nu_v) \omega &\geq 0, \quad \mathbf{T}_{TA}^{TR} (\nu_v) \omega = 0, \quad \mathbf{T}_{TR}^M (\nu_v) \omega \leq 0, \end{aligned} \quad (12)$$

where the listed constraints in (12) satisfy bounded and causal COLA in (8), and attack-hold-release dynamics in (9); a feasible solution can always be found for hold-sizes $T_R - T_A \leq \lfloor \frac{M}{F} \rfloor F$ as COLA rectangle windows can be shifted in time [16]. The operator $\mathbf{T}_a^b(\nu) \in \mathbb{R}^{(b-a) \times M}$ in (12) defines a truncated Toeplitz matrix [17] with constant diagonal entries of the kernel function $\nu(t)$ given by

$$T_{ij} = \begin{cases} \nu(j-i-a), & j-i \leq b \\ 0, & \text{otherwise} \end{cases}, \quad (13)$$

where $\nu(t)$ is shifted by a and upper-bounded by b in time. The objective (12) is convex as the symmetric Toeplitz kernel matrix $\mathbf{T}_0^M(\nu_R) = (\mathbf{T}_0^M(\nu_a))^2 + \mathbf{e}_1 \mathbf{e}_1^T + \mathbf{e}_M \mathbf{e}_M^T$ can be decomposed into the sum of positive semi-definite (PSD) products of the symmetric Toeplitz matrix of $\nu_a(t)$ in (10) with itself and non-negative outer products. Furthermore, $\mathbf{T}_0^M(\nu_R)$ is consistent with the 0-value Dirichlet boundary conditions of (8), causing window tails to taper in the solutions to (12), as shown in Fig. 2 ($M = 1024$, $F = 256$); several regularities for attack-release onset tuples (T_A, T_R) are noted.

Symmetric onsets times, $T_A = M - T_R$ (solid-purple, dashed-red), yield symmetric windows with shapes following convolved rectangle-cosine constructions in [16]. Late attack-onsets $T_A \geq M/2$ (dashed-purple, dotted-green) present identical asymmetric windows. Early attack-onsets $T_A < F$ (solid-blue) exhibit piece-wise flat regions. The remaining cases (red, yellow) have a single flat section in the hold-interval (T_A, T_R) . Lastly, solutions preserve their shapes for constant M/F , and suggest a continuous and closed-form expression.

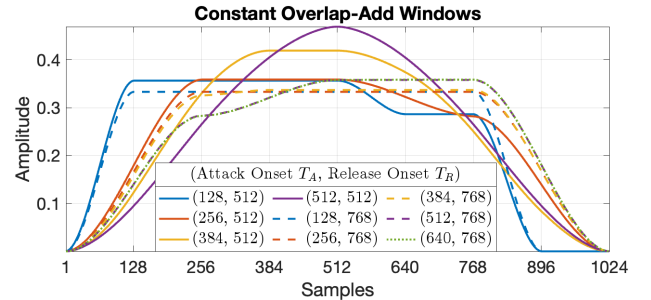


Fig. 2: Solutions (12) vary across attack-onset T_A and release-onset T_R times, and are scale-invariant w.r.t the window/frame size ratio M/F .

3. CHANNEL-MIXTURE DISTORTION OBJECTIVE

We define channel-mixture distortion as the weighted summation of channel gain reductions in decibels (dB), which can be expressed as the logarithmic product $g(\mathbf{x})$ of exponentiated gain variables $x_n^{w_n}$:

$$\sum_{n=1}^N w_n 20 \log_{10} x_n = 20 \log_{10} g(\mathbf{x}), \quad g(\mathbf{x}) = \prod_{n=1}^N x_n^{w_n}, \quad (14)$$

where $w_n \in \mathbb{R}_{>0}$ is the positive attenuation rate parameterizing dB reduction of channel $Y(t, n)$ in (7) per dB reduction to x_n . Larger w_n penalizes smaller x_n and minimize distortion to select channels (e.g.

center). The maximizer of the summation in (14) is equal to that of the product of gains as the logarithm is both monotonic and non-positive. For QP objective (3), the quadratic approximation of $g(\mathbf{x})$ follows the Taylor series expansion $h(\mathbf{x})$ at an expansion center $\mathbf{a} \in \mathbb{R}^{N \times 1}$:

$$h(\mathbf{x}) = g(\mathbf{a}) + (\mathbf{x} - \mathbf{a})^T \nabla g(\mathbf{a}) + \frac{1}{2} (\mathbf{x} - \mathbf{a})^T \mathbf{H}(\mathbf{a}) (\mathbf{x} - \mathbf{a}), \quad (15)$$

where the gradient $\nabla g(\mathbf{x}) \in \mathbb{R}^{N \times 1}$ is the vector of partial derivatives

$$\nabla g(\mathbf{x}) = \left[\frac{\partial g}{\partial x_1}, \dots, \frac{\partial g}{\partial x_N} \right]^T, \quad \frac{\partial g}{\partial x_i} = w_i x_i^{w_i-1} \prod_{n=1, n \neq i}^N x_n^{w_n}, \quad (16)$$

and the Hessian $\mathbf{H}(\mathbf{x}) \in \mathbb{R}^{N \times 1}$ is the matrix of second-order partial derivatives $H_{ij} = \frac{\partial^2 g}{\partial x_i \partial x_j}$ with entries given by

$$H_{ij} = \begin{cases} w_i(w_i - 1)x_i^{w_i-2} \prod_{n=1, n \neq i}^N x_n^{w_n}, & i = j \\ w_i w_j x_i^{w_i-1} x_j^{w_j-1} \prod_{n=1, n \neq i, j}^N x_n^{w_n}, & i \neq j \end{cases}. \quad (17)$$

The expansion center at unity $\mathbf{a} = \mathbf{1}$ in (15) incurs no distortion to (14), and has the following gradient (16) and Hessian (17):

$$g(\mathbf{1}) = 1, \quad \nabla g(\mathbf{1}) = \mathbf{w}, \quad \mathbf{H}(\mathbf{1}) = \mathbf{w}\mathbf{w}^T - \text{diag}(\mathbf{w}), \quad (18)$$

where $\mathbf{w} = [w_1, \dots, w_N]^T \in \mathbb{R}^{N \times 1}$ is the vector of attenuation rates, and $\text{diag}(\mathbf{w}) = \sum_{n=1}^N w_n \mathbf{e}_n$ the diagonal operator. Rewriting (15), (18) in the standard form (3) for QP minimization gives

$$f(\mathbf{x}) = 1 - h(\mathbf{x}), \quad \mathbf{Q} = -\mathbf{H}(\mathbf{1}) = \text{diag}(\mathbf{w}) - \mathbf{w}\mathbf{w}^T, \quad (19)$$

$$\mathbf{c} = -(\mathbf{w} + \mathbf{Q}\mathbf{1}) = (\mathbf{w}^T \mathbf{1} - 2)\mathbf{w}, \quad d = \frac{1}{2} \mathbf{1}^T \mathbf{Q} \mathbf{1} + \mathbf{w}^T \mathbf{1},$$

where it is desirable to further constrain the critical point \mathbf{x}_* of $f(\mathbf{x})$ to $\mathbf{x}_* \geq \mathbf{1}$ such that the minimizer of (4) lies on a constraint that is near unity. Setting the gradient $\nabla f(\mathbf{x}) = \mathbf{Q}(\mathbf{x}_* - \mathbf{1}) - \mathbf{w}$ to zero, and computing the Sherman-Morrison [18] matrix-inverse $\mathbf{Q}^{-1} = \text{diag}(\mathbf{w})^{-1} + \frac{\mathbf{1}\mathbf{1}^T}{1 - \mathbf{1}^T \mathbf{w}}$, yields the critical point given by

$$\mathbf{x}_* = \mathbf{Q}^{-1} \mathbf{w} + \mathbf{1} = (2 + q) \mathbf{1}, \quad q = \mathbf{1}^T \mathbf{w} (1 - \mathbf{1}^T \mathbf{w})^{-1}, \quad (20)$$

where \mathbf{x}_* is a scaled vector of unity. Therefore, constraining \mathbf{x}_* to lie outside of unity bounds the sum of rates w_n below unity:

$$\mathbf{x}_* \geq \mathbf{1} \Leftrightarrow q \geq -1 \Leftrightarrow \sum_{n=1}^N w_n \leq 1, \quad w_n > 0. \quad (21)$$

A second desirable property is a PSD $\mathbf{Q} \succeq 0$ or negative semi-definite $\mathbf{H}(\mathbf{1}) \preceq 0$ to ensure that $f(\mathbf{x})$ is convex in (4) and solvable in polynomial time [19]. We show that the summation bound $\sum_{n=1}^N w_n \leq 1$ in (21) is necessary and sufficient for PSD \mathbf{Q} . Let the eigenvalues $\lambda \in \mathbb{R}$ of symmetric \mathbf{Q} be the zeros of secular equation $S(\lambda)$ of a rank-1 update to a diagonal matrix [20]:

$$S(\lambda) = 1 - \sum_{n=1}^N \frac{w_n^2}{w_n - \lambda} = 1 - \sum_{n=1}^N \frac{(w_n^2 - \lambda^2) + \lambda^2}{w_n - \lambda} \quad (22)$$

$$= 1 - \sum_{n=1}^N w_n - \lambda \sum_{n=1}^N \frac{w_n}{w_n - \lambda}. \quad \text{Diff. of Squares}$$

Setting (22) to zero and applying bounds (21) gives the necessary conditions that constrain the eigenvalues as follows:

$$\sum_{n=1}^N w_n = 1 - \lambda \sum_{n=1}^N \frac{w_n}{w_n - \lambda} \leq 1 \Rightarrow \lambda \geq 0, \quad (23)$$

whereby λ must be non-negative via proof by contrapositive as $\lambda < 0 \Rightarrow \sum_{n=1}^N \frac{w_n}{w_n - \lambda} > 0 \Rightarrow 1 - \lambda \sum_{n=1}^N \frac{w_n}{w_n - \lambda} > 1$. For the sufficient condition, let rates $0 < w_1 \leq w_2 \leq \dots \leq w_N$ and eigenvalues $\lambda_1 \leq$

$\lambda_2 \leq \dots \leq \lambda_N$ of \mathbf{Q} be sorted in ascending order. We prove $\lambda_1 \geq 0 \Rightarrow \sum_{n=1}^N w_n \leq 1$ via the contrapositive $\sum_{n=1}^N w_n > 1 \Rightarrow \lambda_1 < 0 < \lambda_2$ and examine $S(\lambda)$ in (22) as follows:

$$S(-\infty) = 1, \quad S(0) = 1 - \sum_{n=1}^N w_n < 0, \quad (24)$$

$$\lim_{\lambda \rightarrow w_n^\pm} S(\lambda) = \pm\infty, \quad \frac{\partial S(\lambda)}{\partial \lambda} = - \sum_{n=1}^N \frac{w_n^2}{(w_n - \lambda)^2} < 0,$$

where the partial derivative $\partial S(\lambda)/\partial \lambda$ is negative. The intermediate value theorem therefore guarantees the existence of the two smallest roots in the intervals $\lambda_1 < 0$ and $w_1 < \lambda_2 < w_2$ respectively. Thus in practical implementations, we normalize w_n to sum to or below unity.

4. VARIABLE AND CONSTRAINT REDUCTION

QP solvers exhibit quadratic to cubic compute costs w.r.t. the number of constraints and variables [21], [22] in (6), which is prohibitive for real-time processing within an audio frame's period in practical applications. Consider the case of multi-band, multi-content channel inputs with multiple output-mixer terminals. Let the channel gains be the column matrix $\mathbf{X} \in \mathbb{R}^{N_B \times N_C}$ of N_B multi-band gains per N_C content channels, and the sample data be the tensor $\mathbf{S} \in \mathbb{R}^{\bar{F} \times N_B \times N_C \times N_M}$ of \bar{F} samples per frame and look-ahead, N_B bands, N_C content channels, summed into N_M mixers. The constraint set is given by

$$-\tau_m \leq \sum_{j=1}^{N_B} \sum_{k=1}^{N_C} S(i, j, k, m) X_{jk} \leq \tau_m, \quad 0 \leq X_{jk} \leq 1, \quad (25)$$

where $1 \leq i \leq \bar{F}$, $1 \leq m \leq N_M$, and the variables $\mathbf{x} = \text{vec}(\mathbf{X}) \in \mathbb{R}^{N \times 1}$ can be vectorized by stacking the columns of matrix \mathbf{X} for consistency with 6. Both the number of variables $N = N_B N_C$, and sample constraints $M = 2\bar{F} N_M$ are multiplicative and large. We give two methods for reducing the number of variables and constraints.

Pre-mixing Variable Reduction: We can consider pre-mixing subsets of input channels as an approximation to the QP in 6 that constrains the variables \mathbf{x} to a linear transformation $\mathbf{x} = \mathbf{P}\mathbf{y}$, $\mathbf{P} \in \mathbb{R}^{N \times N_P}$ of fewer $N_P < N$ variables in vector $\mathbf{y} \in \mathbb{R}^{N_P \times 1}$:

$$\mathbf{y}_* = \arg \min_{\mathbf{y}} f(\mathbf{y}) = \frac{1}{2} \mathbf{y}^T \mathbf{P}^T \mathbf{Q} \mathbf{P} \mathbf{y} + \mathbf{c}^T \mathbf{P} \mathbf{y} + d, \quad (26)$$

$$\text{s.t. } -\tau_m \leq \mathbf{S}_m \mathbf{P} \mathbf{y} \leq \tau_m, \quad 0 \leq \mathbf{P} \mathbf{y} \leq \mathbf{u},$$

whereby $\mathbf{P}^T \mathbf{Q} \mathbf{P} \succeq 0$ preserves convexity, channel matrix $\mathbf{S}_m \in \mathbb{R}^{\bar{F} \times N}$ reshapes $S(:, :, :, m)$ in (25), $\tau_m = \tau_m \mathbf{1} \in \mathbb{R}^{\bar{F} \times 1}$ specifies mixer-dependent thresholds, and $\mathbf{S}_m \mathbf{P}$ downmixes channel samples. The following pre-mixer matrices \mathbf{P} are lossless if $\mathbf{P}\mathbf{1} = \mathbf{1}$, uncouples variables in \mathbf{y} if each row contains exactly one non-zero entry, and induces finite upper-bounds on \mathbf{y} if all entrants are non-negative:

Pre-mixer	Definition \mathbf{P}	N_P	Uncoupled
Single channel	$\mathbf{1}_N$	1	True
Multi-band	$\mathbf{I}_{N_C} \otimes \mathbf{1}_{N_B}$	N_C	True
Multi-content	$\mathbf{1}_{N_C} \otimes \mathbf{I}_{N_B}$	N_B	True
Concatenation	$[\alpha \mathbf{1}_{N_C} \otimes \mathbf{I}_{N_B}, \mathbf{I}_{N_C} \otimes (1 - \alpha) \mathbf{1}_{N_B}]$	$N_B + N_C$	False

where \otimes is the Kronecker product operator. The single, multi-band, and multi-content pre-mixers sums all, banded, and content channels respectively. The concatenation pre-mixer contains multi-band, and multi-content weighted downmixes with lossless $0 < \alpha < 1$, and box-constraints $\mathbf{0} \leq \mathbf{y} \leq [\alpha^{-1} \mathbf{1}_{N_B}; (1 - \alpha)^{-1} \mathbf{1}_{N_C}]$. We omit downmix matrices that are both lossy and coupled (cross-format AC-3 [23]), but can be used to preserve left-right content balance in practice.

Occlusion Culling Constraint Reduction: We can efficiently identify and remove a class of mixture limit constraints in (1) that do not support the QPs feasible space. Let $\mathbf{V}^{\{m\}}$ be the set of extreme vertices in \mathbb{R}^N of the convex hull $\mathbb{H}^{\{m\}}$ defined by the following half-space intersection of the signed $|m|^{\text{th}}$ mixture constraint $s_m(\mathbf{x}) =$

$\text{sgn}(m) \sum_{n=1}^N S_{|m|n} x_n \leq \tau$ and the box-constraints in (6):

$$\mathbb{H}^{\{m\}} = \left\{ \mathbf{x} \in \mathbb{R}^N : s_m(\mathbf{x}) \leq \tau \wedge \mathbf{0} \leq \mathbf{x} \leq \mathbf{u} \right\}, \quad (28)$$

where $\text{sgn}()$ is signum, $\mathbf{V}^{\{m\}} \subset \mathbf{V}^{\{m\}}$ is the subset of vertices that also lie in equality of the m^{th} constraint $s_m(\mathbf{x}) = \tau$, and has cardinality $|\mathbf{V}^{\{m\}}| \leq 2^{N-1}$ bounded by maximum number of edges $2^{N-1}N$ in the N -dimensional hyper-cube. A convex hull $\mathbb{H}^{\{i\}}$ is fully contained in convex hull $\mathbb{H}^{\{j\}}$ iff the i^{th} mixture constraint *occludes* the j^{th} mixture constraint via the following indicator function:

$$I_o(i, j) = \begin{cases} 1, & s_i(\mathbf{x}) > \tau, \quad \forall \mathbf{x} \in \mathbf{V}^{\{j\}} \\ 0, & \text{otherwise} \end{cases}, \quad (29)$$

whereby all vertices in $\mathbf{V}^{\{j\}}$ lie on the negative half-space of the i^{th} constraint. The j^{th} constraint $s_j(\mathbf{x}) \leq \tau$ can therefore be removed from constraint set Ξ in (6) as it does not support the intersection of convex hulls $\hat{\mathbb{H}} = \cap_i \mathbb{H}^{\{i\}}$ that defines the feasible space in Fig. 3. We compare the costs of finding $\hat{\mathbb{H}}$ to solving the QP given M number of constraints and $\hat{M} \leq M$ supports of $\hat{\mathbb{H}}$ in N variables, where $M \gg N$.

Let the cost of finding the support constraints of $\hat{\mathbb{H}}$ via dual space methods [24] be $O(MM^{[N/2]}/(\hat{M} \lfloor N/2 \rfloor!))$, which exceeds that of solving the QP via interior-point methods [21] in $O(M^{3/2}N^2)$ or first-order methods [22] in $O((M+N)^3)$. It is therefore efficient to solve the QP (6) over a small super-set of support constraints, initially determined by pre-processing [25] in $O(MN)$, before applying further constraint reduction in sub-quadratic time. Consider the following set Ξ of mixture constraints not occluded by any other constraints:

$$\Xi = \{s_j(\mathbf{x}) \leq \tau : I_o(i, j) = 0, \forall i \neq j\}, \quad \hat{M} = |\Xi|, \quad (30)$$

which contains the supports of $\hat{\mathbb{H}}$. The costs of computing $\mathbf{V}^{\{j\}}$ and evaluating $I_o(i, j)$ in (29) are $O(2^N N^2)$ and $O(2^N N)$ respectively. We can determine Ξ in $O(2^N(N^2 + N\hat{M} + \log M)M)$ time via sorted constraints defined in the following propositions:

$$I_o(i, j) = 1 \Rightarrow \min_{\mathbf{x} \in \mathbf{V}^{\{i\}}} \|\mathbf{x}\| < \min_{\mathbf{x} \in \mathbf{V}^{\{j\}}} \|\mathbf{x}\|, \quad (31)$$

whereby the vertex in $\mathbf{V}^{\{i\}}$ closest to the origin is closer than that of $\mathbf{V}^{\{j\}}$ if constraint i occludes constraint j following $\mathbb{H}^{\{i\}} \subset \mathbb{H}^{\{j\}}$. The contrapositive of (31) is therefore true and expressed by

$$\min_{\mathbf{x} \in \mathbf{V}^{\{i\}}} \|\mathbf{x}\| \geq \min_{\mathbf{x} \in \mathbf{V}^{\{j\}}} \|\mathbf{x}\| \Rightarrow I_o(i, j) = 0, \quad (32)$$

where if the vertex in $\mathbf{V}^{\{i\}}$ closest to the origin is further than or equidistant to that of $\mathbf{V}^{\{j\}}$, then constraint i does not occlude constraint j . We can therefore sort the M constraints in ascending order by the min-norm vertex $\min_{\mathbf{x} \in \mathbf{V}} \|\mathbf{x}\|$ in $O(2^N M \log M)$ time, add the first constraint to Ξ for the base case, and add subsequent sorted constraints j to Ξ if all constraints i in Ξ do not occlude constraint j ($I_o(i, j) = 0, \forall s_i(\mathbf{x}) \leq \tau \in \Xi$) in $O(2^N N \hat{M} M)$ time.

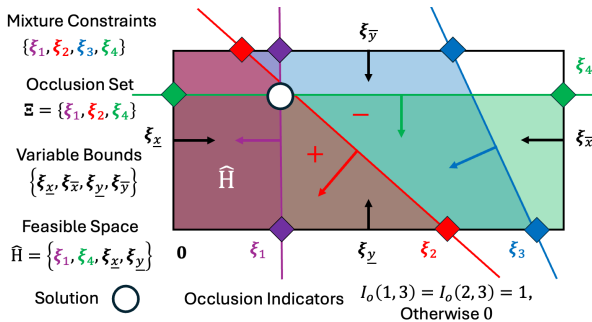


Fig. 3: The sample constraint set's feasible space $\hat{\mathbb{H}}$ is contained in the non-occluded set of constraints Ξ in (30) which excludes ξ_3 .

5. EXPERIMENTS

We evaluate the distortion objective $g(\mathbf{x})$ in (14) in terms of the QP objective $f(\mathbf{x})$ in (6) across pre-mixers in (27) of decreasing number of channels compared to the full mixer-limiter. The multi-band and multi-content channel tensor \mathbf{S} in (25) is populated by amplitude modulated signals $S(t, j, k) = \sin(2\pi a_j t) \sin(2\pi(b_k t + \phi_{jk}))$, where $\mathbf{a} = [101, 443, 1627]$ and $\mathbf{b} = [2, 5, 11]$ are lists of frequencies (Hz) for the carrier band and message content respectively, and $\phi_{jk} = \frac{(k-1)N_B + j}{N_B N_C}$ the latter's phase-offset. We plot the time-evolution of $g(\mathbf{x})$ over a 1-second duration in Fig. 4, and show the distortion improvement gap with the full mixer-limiter decreases for summative $N_B + N_C$ number of channels. The mean and standard deviation of $f(\mathbf{x})$ across pre-mixers validate the trend: Single channel limiter 0.23 ± 0.23 , multi-band and multi-content channel limiters 0.2 ± 0.21 , concatenation ($\alpha = 1/2$) 0.19 ± 0.2 , and full 0.16 ± 0.18 .

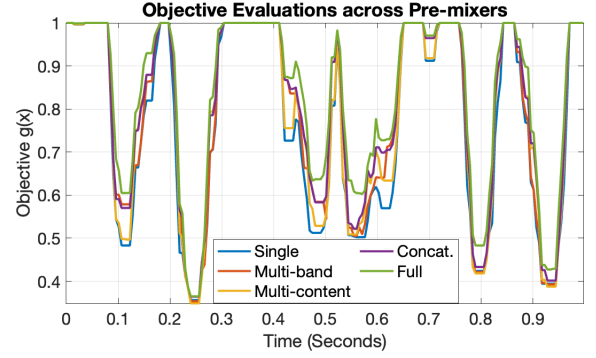


Fig. 4: Distortion objective $g(\mathbf{x})$ in (14) for (27) pre-mixers converges to the full mixer-limiter as the number of channels increases.

We evaluate constraint reduction performance by simulating a multi-band mixer-limiter in (6) of frame-size $F = 256$, look-ahead $L = 768$, mixture threshold $\tau = 1$, and upper bound $\mathbf{u} = 1$. The input signals consist of N full-scale sine tones at $[101, 443, 1627, 4153, 8747, 15733]$ Hz (1-second duration at 48 kHz sample rate). Pre-processing [25] implied-bounds and tightening \mathbf{x} methods reduce the number of mixture-limit constraints (originally 2048). The non-occluded sets Ξ (30) computed with or without pre-processing are equivalent. We compare its cardinality \hat{M} to that of the convex hull's supports in Table 1 across frames for mixer-limiter instances of increasing number of channels, containing the first N tones. The ratio of constraint to convex-support set sizes reduces in range from (8.07, 51.24) (pre-processing) to (1.37, 1.88) (non-occluding set), thereby approaching the cardinality's lower-limit.

Table 1: Mixer-limiter Number of Constraints (Mean \pm Std. Dev)

N	Implied-bounds	Tightening	Non-occluded	Convex
2	384.7 ± 52	374.1 ± 50.5	10 ± 4.3	7.3 ± 2.6
3	805.8 ± 103.7	799 ± 102.8	41.8 ± 14.7	25.9 ± 7.5
4	1167 ± 149.4	1164 ± 149	99.1 ± 22.9	58.5 ± 14
5	1442 ± 184.5	1441 ± 184.3	226.3 ± 64	130.1 ± 35.7
6	1636 ± 209.3	1636 ± 209.2	381.5 ± 78.6	202.8 ± 41.8

6. CONCLUSIONS

We presented a coupled multichannel mixer-limiter design for adaptive channel-headroom allocation and loudspeaker protection. A minimum distortion objective was approximated and optimized via a QP formulation over a frame-based processor. Limiter envelopes satisfying the QP constraints were constructed from dynamics-constrained COLA window formulations. Pre-mixing and occlusion-set variable-constraint reduction methods decreased the QP problem size and achieved comparable performance with the full mixer-limiter in experiments.

7. ACKNOWLEDGMENT

This work is funded by Amazon.com incorporated.

8. COMPLIANCE WITH ETHICAL STANDARDS

This is an audio signal-processing study for which no ethical approval was required.

9. REFERENCES

- [1] R EBU-Recommendation, “Loudness normalisation and permitted maximum level of audio signals,” *Eur. Broadcast. Union*, 2011.
- [2] Martin Colloms, *High performance loudspeakers: optimising high fidelity loudspeaker systems*, pp. 13–14, John Wiley & Sons, 2018.
- [3] Floyd Toole, *Sound reproduction: the acoustics and psychoacoustics of loudspeakers and rooms*, pp. 410–432, Routledge, 2017.
- [4] Anthony Chiarella and Matthew Polk, “Home theater handbook,” 2006.
- [5] ITUR Recommendation, “Itu-r recommendation bs. 775-4,” *Multi-channel stereophonic sound system with or without accompanying picture*, 2022.
- [6] Jürgen Herre, Kristofer Kjörling, Jeroen Breebaart, Christof Faller, Sascha Disch, Heiko Purnhagen, Jeroen Koppens, Johannes Hilpert, Jonas Rödén, Werner Oomen, et al., “Mpeg surround-the iso/mpeg standard for efficient and compatible multichannel audio coding,” *Journal of the Audio Engineering Society*, vol. 56, no. 11, pp. 932–955, 2008.
- [7] Roger Dressler, “Dolby surround pro logic decoder principles of operation,” *Dolby White paper*, 2000.
- [8] David Griesinger, “Multichannel matrix surround decoders for two-eared listeners,” in *Audio Engineering Society Convention 101*. Audio Engineering Society, 1996.
- [9] David J Meares and Günther Theile, “Matrixed surround sound in an mpeg digital world,” *Journal of the Audio Engineering Society*, vol. 46, no. 4, pp. 331–335, 1998.
- [10] Guy W McNally, “Dynamic range control of digital audio signals,” *Journal of the Audio Engineering Society*, vol. 32, no. 5, pp. 316–327, 1984.
- [11] Dimitrios Giannoulis, Michael Massberg, and Joshua D Reiss, “Digital dynamic range compressor design—a tutorial and analysis,” *Journal of the Audio Engineering Society*, vol. 60, no. 6, pp. 399–408, 2012.
- [12] James M Kates, “Principles of digital dynamic-range compression,” *Trends in amplification*, vol. 9, no. 2, pp. 45–76, 2005.
- [13] Julius O. Smith, *Spectral Audio Signal Processing*, W3K, 2011.
- [14] Jon C Schmidt and Janet C Rutledge, “Multichannel dynamic range compression for music signals,” in *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*. IEEE, 1996, vol. 2, pp. 1013–1016.
- [15] John M. Eargle, *Compressors, Limiters, and Noise Gates*, pp. 303–312, Springer US, Boston, MA, 1996.
- [16] Christian Borß and Rainer Martin, “On the construction of window functions with constant-overlap-add constraint for arbitrary window shifts,” in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 337–340.
- [17] Robert M Gray et al., “Toeplitz and circulant matrices: A review,” *Foundations and Trends® in Communications and Information Theory*, vol. 2, no. 3, pp. 155–239, 2006.
- [18] Jack Sherman and Winifred J. Morrison, “Adjustment of an inverse matrix corresponding to a change in one element of a given matrix,” *Annals of Mathematical Statistics*, vol. 21, pp. 124–127, 1950.
- [19] M. K. Kozlov, Sergey P. Tarasov, and Leonid Khachiyan, “The polynomial solvability of convex quadratic programming,” *Ussr Computational Mathematics and Mathematical Physics*, vol. 20, pp. 223–228, 1980.
- [20] Gene H. Golub, “Some modified matrix eigenvalue problems,” *SIAM Review*, vol. 15, no. 2, pp. 330, 1973.
- [21] Arkadi Nemirovski, “Interior point polynomial time methods in convex programming,” *Lecture notes*, vol. 42, no. 16, pp. 3215–3224, 2004.
- [22] B. Stellato, G. Banjac, P. Goulart, A. Bemporad, and S. Boyd, “OSQP: an operator splitting solver for quadratic programs,” *Mathematical Programming Computation*, vol. 12, no. 4, pp. 637–672, 2020.
- [23] ATSC Standard, “Digital audio compression (ac-3, e-ac-3),” 2012, pp. 91–96.
- [24] C. Bradford Barber, David P. Dobkin, and Hannu Huhdanpaa, “The quickhull algorithm for convex hulls,” *ACM Transactions on Mathematical Software*, vol. 22, no. 4, pp. 469–483, Dec. 1996.
- [25] Nick Gould and Philippe L Toint, “Preprocessing for quadratic programming,” *Mathematical Programming*, vol. 100, pp. 95–132, 2004.