NEURAL-DRIVEN MULTI-BAND PROCESSING FOR AUTOMATIC EQUALIZATION AND STYLE TRANSFER

Parakrant Sarkar and Permagnus Lindborg

SoundLab, School of Creative Media
City University of Hong Kong, Hong Kong SAR, China
parakrant.sarkar@my.cityu.edu.hk and pm.lindborg@cityu.edu.hk

ABSTRACT

We present a Neural-Driven Multi-Band Processor (NDMP), a differentiable audio processing framework that augments a static six-band Parametric Equalizer (PEQ) with per-band dynamic range compression. We optimize this processor using neural inference for two tasks: Automatic Equalization (AutoEQ), which estimates tonal and dynamic corrections without a reference, and Production Style Transfer (NDMP-ST), which adapts the processing of an input signal to match the tonal and dynamic characteristics of a reference. We train NDMP using a self-supervised strategy, where the model learns to recover a clean signal from inputs degraded with randomly sampled NDMP parameters and gain adjustments. This setup eliminates the need for paired input–target data and enables end-to-end training with audio-domain loss functions. In the inference, AutoEQ enhances previously unseen inputs in a blind setting, while NDMP-ST performs style transfer by predicting task-specific processing parameters. We evaluate our approach on the MUSDB18 dataset using both objective metrics (e.g., SI-SDR, PESQ, STFT loss) and a listening test. Our results show that NDMP consistently outperforms traditional PEQ and a PEQ+DRC (single-band) baseline, offering a robust neural framework for audio enhancement that combines learned spectral and dynamic control.

1. INTRODUCTION

With the rise of short-form content, podcasting, and online music production, there is increasing demand for tools that deliver studio-quality audio with minimal manual effort. Traditional EQ methods often require expert tuning and are not easily adaptable to diverse content. This motivates learning-based approaches that provide automatic, context-aware enhancement while lowering the barrier to high-quality sound. Equalization (EQ) [1] is a critical component of audio processing that allows for the precise adjustment of tones to improve the clarity and balance of a recording. Often used by audio mixing engineers for audio or music production, it manipulates the acoustic characteristics of individual recordings or complete compositions. EQ does it by enhancing or reducing specific frequency bands in that recording by modifying the EQ parameters. It also ensures that multiple recordings from various instrumental sources like guitar, drums, piano, etc or vocals from singers do not overlap in the same frequency space. In addition to attaining tonal balance, EQ plays a vital role in accentuating the distinct timbral aspects of various instrumental sources or vocals, which results in the overall clarity and definition of the final produced mix.

The Parametric Equalizer (PEQ) [2] is one of the most often used equalization techniques in audio production. It gives users a great deal of control and versatility, allowing them to modify the EQ parameters like gain, frequency, and Q-factor (bandwidth) across multiple frequency bands. It comprises various filters like shelving, peaking, high or low pass filters [3], etc, each handling individual frequency bands to adjust the tonal balance of the recordings precisely. It has become an integral part of studio and live music scenarios due to its minimum latency and real-time adaptability [4]. Traditional PEQs, on the other hand, work with the EQ parameters that don't change throughout the music or audio. This signal's inherent static nature becomes a big problem in dynamic audio situations where the signal's spectral and temporal properties constantly change. For example, an audio signal with many transients, like drums or percussion instruments, needs to be adjusted rapidly to keep clarity and balance. Similarly, dynamically varying content that changes quickly, like vocals with fluctuating intensity or instruments [5] with a wide tonal range, demands real-time adaptability to preserve their natural quality. These problems are not addressed by static EQs, which often lead to sonic shifts, lost information, and an inability to respond correctly to quick changes in sound.

To address these limitations, audio engineers often use a combination of PEQ and Dynamic Range Compression (DRC), a technique designed to control the amplitude envelope and improve balance in dynamically varying signals. DRC modifies gain based on the signal's amplitude and is particularly useful for transient-rich material when static EQs alone are insufficient. Despite their effectiveness, PEQ and DRC remain predominantly manual tools, requiring expertise from audio engineers or relying on rigid, rule-based presets that lack flexibility for automated processes. Moreover, existing implementations are rarely designed for seamless integration into machine learning pipelines. The potential for predicting and optimizing these control parameters in such pipelines remains largely underexplored. In this work, we introduce the *Neural-Driven Multi-Band Processor* (NDMP), a unified framework that combines the static tonal shaping of PEQ with per-band dynamic range compression, both controlled by a neural network. As illustrated in Figure 1, the system processes raw audio by applying static PEQ followed by compressor settings across six frequency bands. NDMP parameters are predicted once per input segment by a neural network trained end-to-end to match reference tonal and dynamic characteristics. While the processing itself is not dynamically adaptive in real time, the neural network enables learned content-aware control, simulating dynamic behavior across segments.

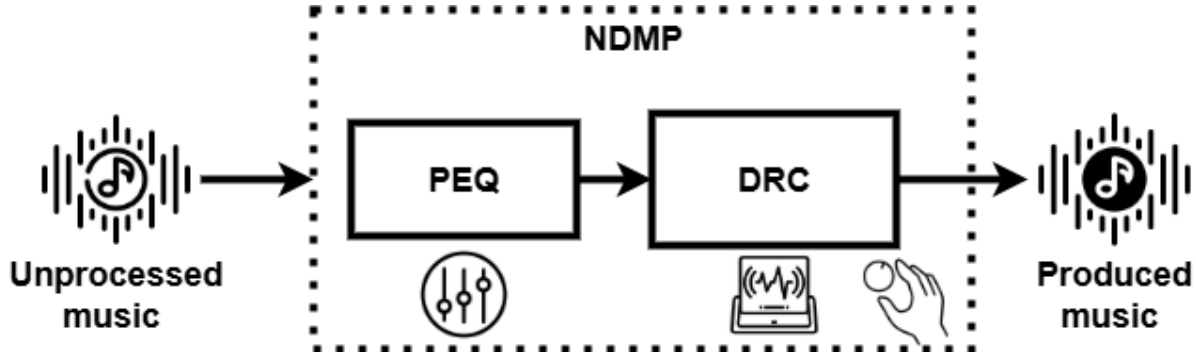

Figure 1: Our proposed Neural-Driven Multi-Band Processor (NDMP) combining a six-band Parametric Equalizer (PEQ) with per-band Dynamic Range Compressor (DRC) for enhanced tonal and dynamic shaping.

We propose an end-to-end, fully differentiable neural network framework that enables NDMP to predict PEQ and DRC parameters for content-aware audio processing. This framework is evaluated on two key audio processing tasks: automatic equalization (AutoEQ), where NDMP predicts and applies tonal adjustments without a reference signal, and production style transfer (NDMP-ST), where NDMP transforms neutral audio to match a production-style reference using segment-level parameter prediction.

The remainder of this paper is organized as follows: Section 2 reviews related work in EQ and style transfer. Section 3 presents our methodology, detailing our proposed NDMP strategy. Sections 4 and 5 describe the model architecture and training setup, while Section 6 discusses the results. Finally, Section 7 summarises contributions and future works.

2. RELATED WORK

In this section, we briefly describe the previous literature in view of automatic equalization for EQ and style transfer on various audio production tasks.In automatic equalization, the focus was on developing EQ matching systems. EQ matching [6] is a process where we automatically adjust the EQ parameters to match the spectral qualities of a reference or target audio signal. This matching can be used based on individual stems like instruments, vocals or on the multi-track mixture for intelligent music production systems, as shown in [7, 1]. In [8], a graphical equalizer was used with fixed-frequency filters to provide precise control of specific frequency bands, inferring only the gain values. [3] introduced efficient filter construction using second-order peaking and shelving filters, highlighting near-log-magnitude self-similarity properties. Neural extensions to this formulation were proposed in [9, 10, 11], where networks predict gain parameters using parameter-domain loss functions. These methods are efficient and robust due to the convex nature of gain optimization and the strong correlation between EQ parameters and magnitude response. [12] proposed a CNN-based end-to-end model that learns content-aware transformations to approximate equalization targets without explicitly computing transfer functions. In [4], differentiable biquad filters were introduced to design a neural parametric equalizer. This approach was inspired by the Differentiable Digital Signal Processing (DDSP) framework [13], using spectral losses to guide perceptually meaningful parameter prediction. In the context of style transfer, [14] introduced *DeepAFX-ST*, combining a parametric equalizer (PEQ) and a dynamic range compressor (DRC) into a differentiable framework. The system learns control parameters for tonal and dynamic effects, achieving content-aware transformations by directly optimizing in the audio domain. While this work pioneered end-to-end differentiable modelling of tonal and dynamic effects, its focus was on reference-style matching with effect conditioning.

Recent works [15, 16, 17] have also explored style transfer using self-supervised or unsupervised models that predict audio processor parameters for transforming raw input into a stylized output. These include EQ, DRC, distortion, and reverb, often implemented with differentiable modules to capture subtle interactions between effects.

While prior work has advanced EQ matching and style transfer with neural and differentiable methods, few approaches have explored predicting both EQ and compression parameters jointly without style labels or reference conditioning. Our proposed NDMP framework fills this gap. NDMP learns to predict static PEQ and per-band DRC parameters directly from audio segments in a self-supervised fashion. Although NDMP does not dynamically adapt parameters during runtime, its learned predictions emulate dynamic control over tone and loudness across varying musical content.

We evaluate NDMP on two key tasks: automatic equalization (AutoEQ), where the model enhances unseen inputs without a reference; and production style transfer (NDMP-ST), where NDMP transforms input audio to adopt a production-style profile via tailored parameter prediction.

3. METHODOLOGY

This section introduces our proposed Neural-Driven Multi-Band Processor (NDMP), a differentiable audio effect pipeline composed of Parametric Equalization (PEQ) and per-band Dynamic Range Compression (DRC). We first describe the underlying PEQ and DRC effects and their differentiability and then present how NDMP is formulated and used for automatic equalization and style transfer tasks.

3.1. Parametric Equalization (PEQ)

Parametric Equalizers (PEQs) are commonly implemented as cascaded biquad filters [4, 5], also known as second-order IIR filters [2]. Each filter is parameterized by:

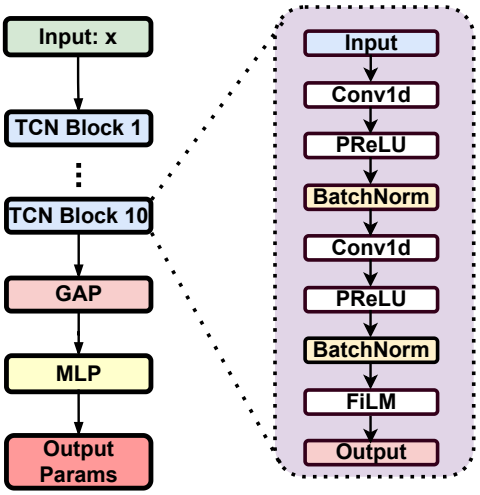

Figure 2: AutoEQ Model Architecture

- **Center/Cutoff Frequency** (f_c): the central or cutoff frequency of the filter.
- **Gain** (g): the amplitude adjustment for the band (in dB).
- **Q-Factor** (Q): the bandwidth control relative to the center frequency.

The filter's transfer function is:

$$H(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2}} \tag{1}$$

where $\{b_i\}$ and $\{a_i\}$ are coefficients determined via bilinear transform to simulate analog response [18].

We use a six-band PEQ configuration from [14], comprising one low-shelf, one high-shelf, and four peaking filters. These are implemented using the differentiable audio DSP toolkit dasp-pytorch.[1] Each filter's coefficients $\{b_i\}, \{a_i\}$ are computed from the normalized parameters f_c, g, and Q using standard digital filter design formulas derived via bilinear transformation [18]. In practice, the neural network predicts normalized values, which are then mapped to physical filter parameters and converted into biquad coefficients as part of the differentiable processing chain.

3.2. Dynamic Range Compression (DRC)

It is a fundamental audio processing technique used to control the dynamic range of a signal by attenuating its amplitudes when they exceed a specified threshold. DRC operates using a gain computer and a ballistics filter, which smooths abrupt changes in gain over time. We use a differentiable compressor design adapted from [19] using a single-pole IIR smoothing filter. We follow a similar strategy in which we simplify the design by approximating the attack and release time constants with a unified constant (α) shown in equation 2

$$y_L[n] = \alpha y_L[n-1] + (1-\alpha) x_L[n] \tag{2}$$

where $x_L[n] = x_G[n] - y_G[n]$ represents gain reduction and α is a shared time constant.

The key compression parameters include are as follows:

- **Threshold** (Th): The dB level above which compression starts.
- **Ratio** (R): It specifies the amount of compression applied to signals exceeding the Th. For example, a ratio of 4:1 indicates that for every 4 dB, the input signal exceeds the threshold and the output level increases by 1 dB.
- **Attack/Release Times** (τ_a, τ_r): It shows how fast the compressor engages/disengages. Short attack times are suitable for transient signals (e.g., drum hits), while longer times preserve the natural dynamics, whereas short release times return the signal to its original dynamics quickly, while longer times ensure smoother transitions.
- **Knee**: It determines the smoothness of the compression onset. A hard knee applies compression abruptly once the threshold is exceeded, while a soft knee results in a gradual transition.
- **Makeup Gain** (G_m): A gain added after compression to restore perceived loudness.

These high-level control parameters: threshold (Th), ratio (R), attack/release times (τ_a, τ_r), knee, and makeup gain are internally converted to smoothing coefficients and gain curves used in the differentiable implementation of Equation 2, following the approach of [14].

3.3. Differentiable Audio Effects

In our work, we propose a neural differentiable DEQ. This framework combines parametric equalization and dynamic range compression into a unified, fully differentiable system. By integrating these audio effects as differentiable operators, our approach supports gradient-based optimization for seamless training within neural networks.

[1] https://github.com/csteinmetz1/dasp-pytorch

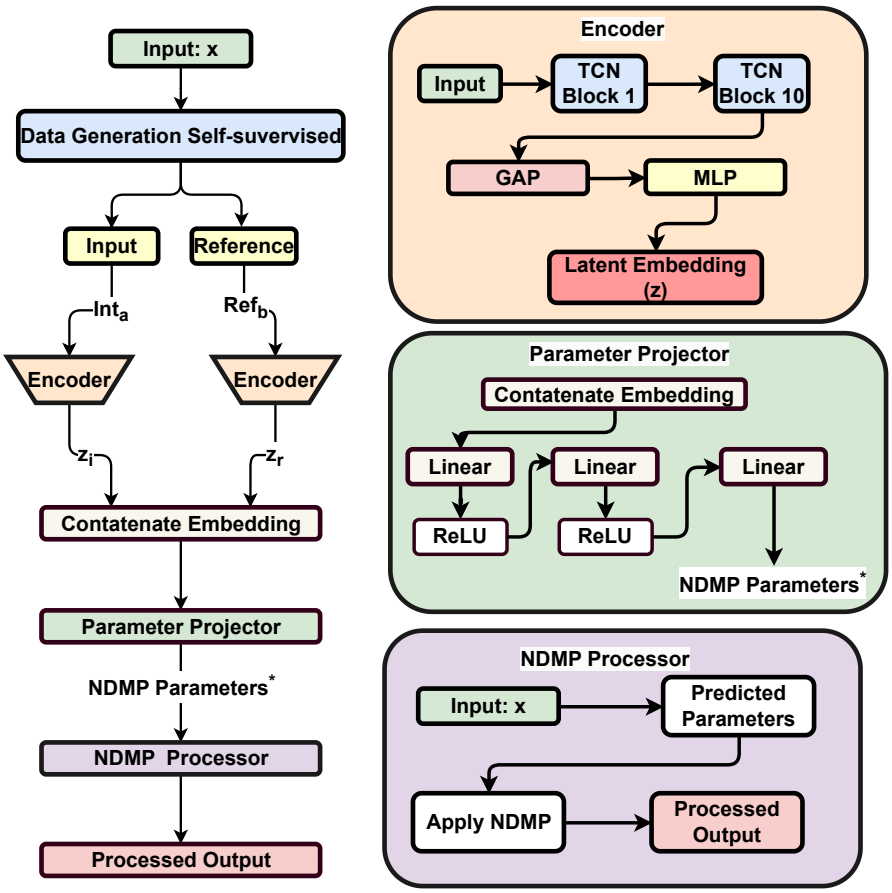

Figure 3: NDMP-ST Model Architecture adapted from [14]

A differentiable audio effect $f(x;\theta)$ supports gradient computation:

$$\nabla_\theta f = \left[\frac{\partial f}{\partial \theta_1}, \ldots, \frac{\partial f}{\partial \theta_n}\right]$$

This enables end-to-end neural network training via backpropagation, optimizing parameters for PEQ (g, f_c, Q) and DRC (Th, R, τ_a, τ_r, G_m). By integrating differentiable PEQ and DRC, NDMP becomes trainable end-to-end using backpropagation.

3.4. NDMP:Neural-Driven Multi-Band Processor

NDMP combines a six-band static parametric equalizer (PEQ) and band-wise dynamic range compressor (DRC) into a single differentiable audio processor. In training, the input audio is degraded using random NDMP parameters and random gain, while the network learns to recover the original audio by predicting the corresponding PEQ and compressor parameters. This enables the model to learn both tonal and dynamic adjustments without requiring paired input-target data.

We have used NDMP in two distinct tasks: (1) **AutoEQ**, which performs blind estimation without access to a reference signal. In the AutoEQ setting, the model predicts NDMP parameters that process the input towards a corrected tonal and dynamic balance, aiming to compensate for degradations or random processing introduced during training; and (2) **NDMP-ST** (Style Transfer), where the model receives both an input and a reference signal and learns to predict NDMP parameters that transform the input audio to match the style (tonal and dynamic characteristics) of the reference. In both cases, NDMP predicts a distinct, fixed set of equalization and compression parameters for each frequency band, and these parameters remain constant throughout each processed audio chunk. The key distinction is that NDMP-ST leverages a reference signal for style matching, while AutoEQ infers processing parameters in a completely reference-free (blind) manner.

Tables 1 and 2 summarize the EQ and compression parameter ranges used in our NDMP formulation.

Table 1: EQ parameters for NDMP used in both AutoEQ and NDMP-ST.

Parameter	Min	Max	Bands
Gain (dB)	-20	20	All bands
Cutoff f_c (Hz)	20	21050	Band-specific
Q-Factor	0.1	6.0	All peaking/shelving

Table 2: Compression parameters for NDMP (used in NDMP-ST).

Parameter	Min	Max
Threshold (dB)	-60	0
Ratio	1.0	10.0
Attack Time (ms)	1.0	100.0
Release Time (ms)	10.0	500.0
Knee (dB)	0.0	12.0
Makeup Gain (dB)	0.0	12.0

4. MODEL ARCHITECTURE

This section outlines the architectures of the proposed NDMP-based models for Automatic Equalization (AutoEQ) and Production Style Transfer (NDMP-ST). Each model is designed to predict NDMP parameters that jointly control equalization and dynamic processing. The architectural details are discussed below.

4.1. AutoEQ Model

Figure 2 illustrates the AutoEQ model, which processes input audio signals $x \in \mathbb{R}^{B\times 1\times N}$, where B is the batch size and N is the number of samples, to predict 42 NDMP control parameters.

Input Layer: The raw waveform input is passed into a stack of 10 Temporal Convolutional Network (TCN) blocks. The input shape is initially $\mathbb{R}^{B\times 1\times N}$.

TCN Blocks: Each TCN block consists of two 1D convolutional layers, followed by PReLU activation and batch normalization.

- The first TCN block maps the input to $\mathbb{R}^{B\times 256\times \frac{N}{2}}$.
- Each subsequent block retains 256 channels and reduces temporal resolution by half due to stride-2 convolutions, resulting in an output of shape $\mathbb{R}^{B\times 256\times \frac{N}{1024}}$ after 10 blocks.

Feature-wise Linear Modulation (FiLM) [20]: Although the architecture supports FiLM-based conditioning, it is disabled in this setup by using a zero vector of shape (1×256) to focus on core equalization and compression learning from the input.

Global Average Pooling (GAP): The temporal dimension is aggregated via global average pooling, resulting in a fixed-length embedding of shape $\mathbb{R}^{B\times 256}$.

Multi-Layer Perceptron (MLP): This embedding is passed through three fully connected layers with dimensions $256 \rightarrow 256 \rightarrow 256 \rightarrow 42$, producing the 42 NDMP parameters.

4.2. NDMP-ST Model

Figure 3 shows the architecture of the NDMP-ST model, which enables production style transfer by conditioning on both input and reference audio signals.

Data Generation: The input $x \in \mathbb{R}^{B\times 1\times N}$ is paired with a reference signal $r \in \mathbb{R}^{B\times 1\times N}$, created by applying random NDMP parameters and gain. Both are normalized to a target loudness (e.g., -40 LUFS) before encoding.

Encoder: The input and reference signals are processed independently by identical encoders, each consisting of 10 TCN blocks:

- Each encoder outputs a latent representation $z \in \mathbb{R}^{B\times 512}$.
- The TCN blocks retain 256 channels and reduce temporal resolution by half per block.

Embedding Concatenation: The input embedding z_x and reference embedding z_r are concatenated into a joint representation of shape $\mathbb{R}^{B\times 1024}$.

Parameter Projector: This representation is passed through an MLP with dimensions $1024 \rightarrow 256 \rightarrow 256 \rightarrow 42$ to predict the NDMP parameters.

NDMP Processor: The predicted parameters are used to process the input signal x, producing the output $y \in \mathbb{R}^{B\times 1\times N}$ that reflects the reference style.

5. IMPLEMENTATION DETAILS

This section outlines the datasets, preprocessing, training procedures, and evaluation details for the proposed NDMP framework in the context of Automatic Equalization (AutoEQ) and Production Style Transfer (NDMP-ST).

5.1. Dataset

We use MUSDB18 [21], a benchmark dataset for music source separation tasks. Our experiments use the multi-track mixture stems for all tasks. The dataset comprises 150 songs, split into training, validation, and test sets in a $90 : 10 : 50$ configuration. All audio files have a 44.1 kHz sampling rate (f_s) and are converted to mono format. Each audio file is segmented into fixed-length frames for training in both tasks.

The dataset consists of fully produced music tracks that have already been subject to professional equalization, compression, and other audio effects, often with unknown and diverse processing chains. This uncontrolled variability presents challenges for the AutoEQ (blind estimation) task, as it may hinder the model's ability to learn a consistent corrective mapping and can limit the interpretability of results. At the same time, MUSDB18 provides a realistic and diverse testbed for evaluating the robustness and generalization of automatic equalization approaches under real-world conditions, where input material is similarly heterogeneous and seldom available in a truly raw state. In the case of the NDMP-ST (style transfer) task, the objective is to match the processing characteristics of a given reference signal, so the diversity of production styles within MUSDB18 is less problematic.

5.2. AutoEQ Training Details

The AutoEQ model learns to predict 42 control parameters of NDMP from the input audio. The key training details are as follows:

Data Preprocessing: Each audio file is segmented into non-overlapping frames of length $N = 131072$ samples (3 seconds at 44.1 kHz). Low-energy and silent frames are filtered using amplitude and energy thresholds ($1e^{-4}$ and 0.01, respectively).

Data Augmentation: Each input frame is peak-normalized, and randomized NDMP parameters are applied, including PEQ and per-band compression. Additionally, random gain scaling within a ± 24 dB range is applied to simulate diverse production conditions.

Training Setup: The model is trained for 400 epochs with a batch size of 16 using the Adam optimizer. A cosine annealing learning rate schedule is used, starting from 2×10^{-4}.

Loss Function: We use a Multi-Resolution STFT (MRSTFT) loss [22], which captures perceptual and spectral differences between predicted and reference signals. It operates across multiple window lengths: $[128, 256, 512, 1024, 2048, 4096, 8192]$, with hop sizes set to half the window size.

Inference: At test time, inputs are processed using the same segmentation and gain normalization pipeline. The predicted NDMP parameters are applied to the input, and results are evaluated using objective and subjective measures.

5.3. NDMP-ST Training Details

NDMP-ST extends the AutoEQ pipeline by enabling reference-based parameter prediction for production-style transfer.

Data Preprocessing: The audio is segmented into 6-second frames ($N = 262144$). Low-energy and silent frames are excluded using the same thresholds as in AutoEQ. Input and reference audio are normalized to -40 LUFS for consistent loudness.

Self-Supervised Data Generation: Inspired by [14, 23], each training pair consists of a neutral input and a reference created by applying random NDMP parameters. These signals are split into halves, e.g., $x = (x_a, x_b)$ and $r = (r_a, r_b)$, and sub-segments are randomly sampled to serve as inputs and references.

Encoder and Projection: The encoder extracts latent embeddings z_x and z_r from input and reference respectively. These are concatenated and passed through a parameter projector (MLP) to predict 42 NDMP control parameters. The NDMP processor then applies these parameters to the input signal.

Loss Function: The MRSTFT loss is computed between the NDMP-processed output and the reference signal, encouraging both tonal and dynamic similarity.

Training Setup: NDMP-ST is trained for 400 epochs with a batch size of 8. The optimizer is Adam with a learning rate of 3×10^{-4}, decayed via cosine annealing.

Inference: At test time, we apply a neutral EQ preset (mimicking a typical broadcast-style baseline) to the input. The reference retains the original production characteristics. The system evaluates how well the predicted NDMP parameters recreate the reference style from the neutral input.

5.4. Baseline Training

To evaluate the contribution of dynamic processing, we establish a baseline using a differentiable Parametric Equalizer (PEQ) without dynamic range compression. The PEQ model employs a six-band equalizer, predicting 18 control parameters (gain, center frequency, and Q-factor for each band). Both PEQ and the full NDMP models are implemented using differentiable audio effect modules from [14] via `dasp-pytorch`. We include a PEQ + single-band DRC baseline, inspired by the DeepAFX-ST approach [14]. In this setup, the model predicts the same six-band PEQ parameters as above, but augments them with a single set of dynamic range compression (DRC) parameters (threshold, ratio, attack, release, etc.) that are applied globally to the entire audio signal using a single-band compressor. This ablation isolates the effect of using per-band (multi-band) versus single-band dynamic range control within our neural framework.

To have a fair comparison, we use the same neural architecture, training schedule, and data preprocessing pipeline for all the models. This ensures that performance differences can be attributed directly to the presence or absence of dynamic range compression. By contrasting PEQ (static equalization), PEQ + DRC (single-band), and NDMP (static equalization plus per-band dynamic range control), we isolate the benefits of band-specific dynamic processing for tasks such as automatic equalization and production style transfer.

Note that in all models, we do not perform explicit band splitting (such as via crossover filters), as is common in traditional multi-band compressors. Instead, each of the six parametric EQ filters and, for NDMP, their associated compressors is applied sequentially to the full-band audio signal. Band-specific effects are realized through the parameterization of each filter and compressor, without extracting separate sub-band signals. We initialize the center frequencies of each EQ band with log-spaced defaults to encourage coverage across the spectrum and reduce redundancy during training.

6. RESULTS AND ANALYSIS

We compared NDMP, PEQ+DRC (Single-Band), and PEQ on our test set using both objective and subjective measures. We selected metrics following prior work in neural audio effects and production modeling [24, 25, 14]. More specifically, we evaluated overall enhancement quality with SI-SDR, perceptual audio quality with PESQ, time-domain fidelity via RMSE, and loudness consistency by measuring LUFS difference. We assessed frequency-domain alignment using a multi-resolution STFT loss and further quantified spectral and dynamic processing through spectral centroid error, spectral bandwidth and flatness differences, harmonic and phase distortion, transient preservation, and crest factor difference.

6.1. Automatic Equalization Task

We observed that NDMP delivered the highest SI-SDR at 14.47 dB, marginally exceeding PEQ (14.38 dB) and PEQ+DRC (14.25 dB). In terms of perceptual quality, NDMP achieved a PESQ score of 4.18 compared to 4.11 for PEQ and 4.03 for PEQ+DRC. The lowest RMSE was produced by PEQ+DRC (0.0479), with NDMP close behind (0.0498); all three models matched loudness equally (–2.63 LUFS). PEQ+DRC achieved the best frequency-domain alignment (STFT loss 1.52), followed by NDMP (1.57) and PEQ (1.61). NDMP also reduced the spectral centroid error to 316 Hz (from PEQ's 327 Hz) and the bandwidth difference to 1.35 (from 1.48), confirming more accurate spectral shaping. The differences in harmonic and phase distortion, transient preservation, and crest factor were minimal, although NDMP retained a slight advantage. These results demonstrate that adding DRC, whether global or per-band, consistently outperforms PEQ, with NDMP providing the most balanced improvements.

6.2. Production Style Transfer Task

In this task, PEQ+DRC achieved the highest SI-SDR (9.94 dB) and PESQ (3.91), with NDMP close behind at 9.40 dB and 2.97 PESQ, both well above PEQ (6.46 dB, 3.73). NDMP produced the smallest LUFS difference (0.20 LUFS), indicating superior loudness matching to the reference. The global dynamics, as measured by crest factor, favoured PEQ+DRC (1.00), while NDMP excelled in spectral bandwidth (48.6 Hz vs. 59.7 Hz for PEQ) and reduced phase distortion (0.64 vs. 0.50). The NDMP continued to trail in transient preservation and spectral centroid error, highlighting opportunities for transient-aware enhancements. Overall, NDMP's

Table 3: Results for automatic equalization and production style transfer using NDMP, PEQ+DRC (Single-Band), and PEQ. Higher SI-SDR and PESQ indicate better performance, while lower RMSE, LUFS Difference, STFT Loss, and spectral metrics suggest improved audio quality and alignment with the reference.

Metric	**Automatic Equalization**			**Production Style Transfer**		
	NDMP	**PEQ+DRC**	**PEQ**	**NDMP**	**PEQ+DRC**	**PEQ**
SI-SDR (dB)	14.47	14.25	14.38	9.40	9.94	6.46
PESQ	4.18	4.03	4.11	2.97	3.91	3.73
RMSE	0.0498	0.0479	0.0511	0.0358	0.0374	0.0471
LUFS Difference	-2.63	-2.63	-2.63	0.20	0.88	1.46
STFT Loss	1.57	1.52	1.61	0.69	0.81	0.77
Spectral Centroid Error (Hz)	316.18	322.68	327.65	481.96	98.78	117.66
Spectral Bandwidth Diff	1.35	1.18	1.48	193.75	48.56	59.72
Spectral Flatness Diff	0.0197	0.0211	0.0199	0.0299	0.0158	0.0153
Harmonic Distortion	0	0	0	3.38e-05	-9.73e-05	-1.11e-04
Phase Distortion	1.02	0.79	0.77	0.66	0.64	0.50
Transient Preservation	0.0140	0.0043	0.0171	1.07	2.62	1.61
Crest Factor Difference	0.41	0.52	0.47	1.86	1.00	1.39

Table 4: Mean Opinion Scores (MOS) for overall audio quality in automatic equalization and production style transfer. Statistical significance: * ($p < 0.05$), ** ($p < 0.01$).

Metric	**Automatic Equalization**			**Production Style Transfer**		
	NDMP	**PEQ+DRC**	**PEQ**	**NDMP**	**PEQ+DRC**	**PEQ**
Overall MOS (1–5)	3.85**	3.75	3.55	3.65**	3.60*	3.45

per-band dynamic control delivers the greatest flexibility and effectiveness, and the simpler PEQ+DRC baseline captures much of the benefit of integrating learned compression into audio-effect pipelines.

6.3. Listening Test

We conducted a listening test with 20 participants (mean age 26) using a 5-point MOS scale. The Table 4 reports the results: for automatic equalization, MOS were 3.85* (NDMP), 3.75* (PEQ+DRC), and 3.55 (PEQ); for production style transfer, MOS were 3.65** (NDMP), 3.60* (PEQ+DRC), and 3.45 (PEQ). A Shapiro–Wilk test confirmed normality ($p > 0.05$). The paired t-tests showed that in the automatic equalization task, NDMP significantly outperformed both PEQ+DRC and PEQ ($p < 0.05$), and PEQ+DRC also outperformed PEQ ($p < 0.05$). In the style transfer task, NDMP significantly outperformed both baselines ($p < 0.01$), and PEQ+DRC outperformed PEQ ($p < 0.05$). These results confirm that listeners perceptually prefer NDMP's outputs. The audio examples are available here.

6.4. Discussion

We observe that NDMP consistently surpasses both PEQ+DRC (single-Band) and PEQ across objective metrics: SI-SDR, STFT loss, and LUFS consistency, and subjective MOS ratings, demonstrating its superior ability to capture both spectral and temporal nuances. The per-band dynamic control in NDMP yields more precise tonal shaping and reference matching than either global compression or static equalization alone, while the PEQ+DRC baseline confirms that even a single-band learned compressor provides substantial gains, as shown in this work [14]. In our subsequent work [26], we extend this framework by inserting crossover filters to create truly band-split streams and apply per-band PEQ+DRC in a differentiable architecture, further improving control and transparency for the speech post-production. We follow a PEQ before DRC arrangement to ensure that spectral imbalances are corrected prior to applying level-dependent dynamics, which is a common production practice for shaping tonal balance before controlling loudness variation. We also note areas for further improvement. NDMP's lower PESQ scores and transient preservation metrics suggest the need for transient-aware loss functions or dedicated modules to better capture attack and release characteristics. We need to explore more expressive architectures such as attention-based encoders or hybrid time–frequency representations, which may enhance perceptual smoothness and phase coherence.

7. CONCLUSION

We introduced a Neural-Driven Multi-Band Processor (NDMP) for two core audio processing tasks: Automatic Equalization and Production Style Transfer. The proposed framework integrates parametric equalization and dynamic range compression in a fully differentiable architecture, enabling end-to-end learning from unpaired audio examples via a self-supervised training strategy. Our evaluation shows that NDMP outperforms traditional parametric equalization (PEQ) across multiple objective and perceptual metrics, offering improved tonal balance, loudness consistency, and temporal shaping. We also introduced a PEQ+DRC (single-band) baseline inspired by DeepAFX-ST, which captures some benefits of learned compression, but NDMP still consistently yields the most robust performance, highlighting its effectiveness in modeling both spectral and dynamic aspects of audio within a unified, learnable framework.

In future work, we plan to further contextualize NDMP's performance by exploring additional differentiable audio effect architectures and extending evaluation to broader production tasks such

as mixing and mastering. We also aim to investigate attention-based neural architectures and improve the modeling of transients and phase-related behavior. More controlled datasets with known processing parameters could provide deeper insight and improved interpretability for blind equalization scenarios.

8. REFERENCES

[1] Vesa Välimäki and Joshua D. Reiss, "All About AudioEqualization: Solutions and Frontiers," *Applied Sciences*, vol. 6, no. 5, pp. 129, May 2016, Number: 5 Publisher: Multidisciplinary Digital Publishing Institute.

[2] Herwig Behrends, Adrian von dem Knesebeck, Werner Bradinal, Peter Neumann, and Udo Zölzer, "Automatic Equalization Using Parametric IIR Filters," *Journal of the Audio Engineering Society. Audio Engineering Society*, vol. 59, pp. 102–109, Mar. 2011.

[3] J. Abel and D. Berners, "Filter Design Using Second-Order Peaking and Shelving Sections," in *ICMC*, 2004.

[4] Shahan Nercessian, "Neural Parametric Equalizer Matching Using Differentiable Biquads," in *International Conference on Digital Audio Effects (DAFx)*, eDAFx, 2020, DAFx, ISSN: 2413-6689.

[5] Florian Mockenhaupt, Joscha Simon Rieber, and Shahan Nercessian, "Automatic Equalization for Individual Instrument Tracks Using Convolutional Neural Networks," in *27th International Conference on Digital Audio Effects (DAFx24)*, Guildford, Surrey, UK, Sept. 2024.

[6] François G. Germain, Gautham J. Mysore, and Takako Fujioka, "Equalization matching of speech recordings in real-world environments," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 609–613.

[7] Brecht De Man, Joshua Reiss, and Ryan Stables, "Ten Years of Automatic Mixing," Salford, UK, Sept. 2017.

[8] Juho Liski and V. Välimäki, "The quest for the best graphic equalizer," Edinburgh, UK, Sept. 2017, p. 95102.

[9] Vesa Välimäki and Jussi Rämö, "Neurally Controlled Graphic Equalizer," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 12, pp. 2140–2149, Dec. 2019.

[10] Jussi Rämö and Vesa Välimäki, "Neural third-octave graphic equalizer," in *International Conference on Digital Audio Effects*. 2019, University of Birmingham.

[11] Valeria Bruschi, Vesa Välimäki, Juho Liski, and Stefania Cecchi, "Linear-Phase Octave Graphic Equalizer," *Journal of the Audio Engineering Society*, vol. 70, no. 6, pp. 435–445, 2022, Publisher: Audio Engineering Society.

[12] Marco A. Martínez Ramírez and Joshua D. Reiss, "End-to-end equalization with convolutional neural networks," in *21th International Conference on Digital Audio Effects (DAFx18)*, 2018, ISSN: 2413-6689.

[13] Jesse Engel, Lamtharn (Hanoi) Hantrakul, Chenjie Gu, and Adam Roberts, "DDSP: Differentiable Digital Signal Processing," in *International Conference on Learning Representations (ICML)*, 2020.

[14] Christian J. Steinmetz, Nicholas J. Bryan, and Joshua D. Reiss, "Style Transfer of Audio Effects with Differentiable Signal Processing," *Journal of the Audio Engineering Society*, vol. 70, no. 9, pp. 708–721, Sept. 2022.

[15] Ondrej Cifka, Alexey Ozerov, Umut Simsekli, and Gael Richard, "Self-Supervised VQ-VAE for One-Shot Music Style Transfer," *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 96–100, 2021.

[16] Junghyun Koo, Marco A. Martínez-Ramírez, Wei-Hsiang Liao, Stefan Uhlich, Kyogu Lee, and Yuki Mitsufuji, "Music Mixing Style Transfer: A Contrastive Learning Approach to Disentangle Audio Effects," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023.

[17] Côme Peladeau and Geoffroy Peeters, "Blind estimation of audio effects using an auto-encoder approach and differentiable signal processing," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2024)*, Seoul, Korea, Apr. 2024, IEEE, arXiv:2310.11781 [cs, eess].

[18] Jonathan S. Abel and David P. Berners, "Discrete-Time Shelf Filter Design for Analog Modeling," *Journal of the Audio Engineering Society*, , no. 5939, Oct. 2003.

[19] Joseph Colonel, Joshua D Reiss, and others, "Approximating ballistics in a differentiable dynamic range compressor," in *Audio Engineering Society Convention 153*. 2022, Audio Engineering Society.

[20] Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron Courville, "Film:Visual Reasoning with a General Conditioning Layer," Dec. 2017, arXiv:1709.07871 [cs].

[21] Zafar Rafii, Antoine Liutkus, Fabian-Robert Stöter, Stylianos Ioannis Mimilakis, and Rachel Bittner, "MUSDB18 - a corpus for music separation," 2017.

[22] C. Steinmetz and Joshua D. Reiss, "auraloss: Audio-focused loss functions in PyTorch," in *Digital Music Research Network One-Day Workshop (DMRN*, London, UK, 2020.

[23] Soumya Vanka, Christian Steinmetz, Jean-Baptiste Rolland, Joshua Reiss, and György Fazekas, "Diff-mst: Differentiable mixing style transfer," in *Proc. of the 25th Int. Society for Music Information Retrieval Conf. (ISMIR)*, San Francisco, USA, 2024.

[24] Soumya Sai Vanka, Maryam Safi, Jean-Baptiste Rolland, and György Fazekas, "The Role of Communication and Reference Songs in the Mixing Process: Insights From Professional Mix Engineers," *Journal of the Audio Engineering Society*, vol. 72, no. 1/2, pp. 5–15, Jan. 2024.

[25] Brecht De Man, Brett Leonard, Richard L. King, and Joshua D. Reiss, "An analysis and evaluation of audio features for multitrack music mixtures," in *Proceedings of the 15th International Society for Music Information Retrieval Conference, ISMIR 2014, Taipei, Taiwan, October 27-31*, 2014, pp. 137–142.

[26] Parakrant Sarkar and PerMagnus Lindborg, "Diff-DEQ: Differentiable dynamic equalization for studio-quality speech processing," in *Proceedings of the 33rd European Signal Processing Conference, (EUSIPCO), Palermo, Italy, September 08-12*, 2025.