Neural-Enhanced Dynamic Range Compression Inversion: A Hybrid Approach for Restoring Audio Dynamic

Colloque Gretsi 2025, Strasbourg

Haoran Sun, Dominique Fourer, Hichem Maaref

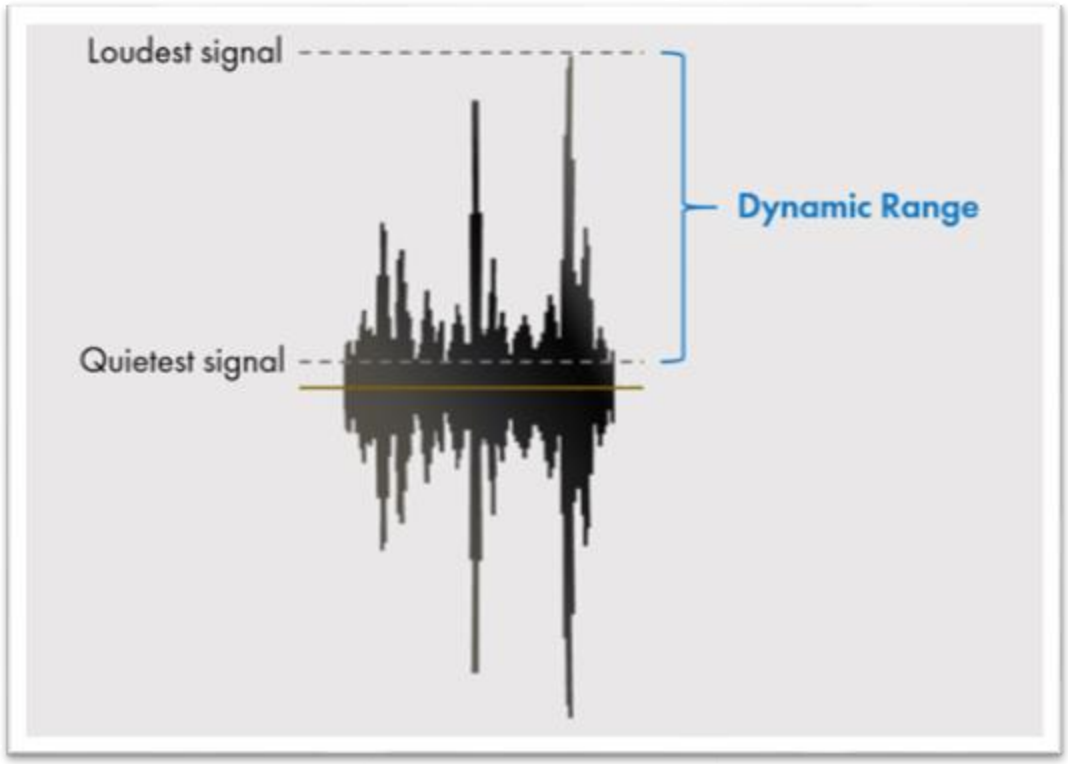
IBISC, Université d'Évry/Paris Saclay

August 29, 2025

Introduction

Dynamic Range Compression (DRC)

DRC is an audio signal processing operation that reduces the volume of loud sounds or amplifies quiet sounds, thus reducing or compressing an audio signal's dynamic range.

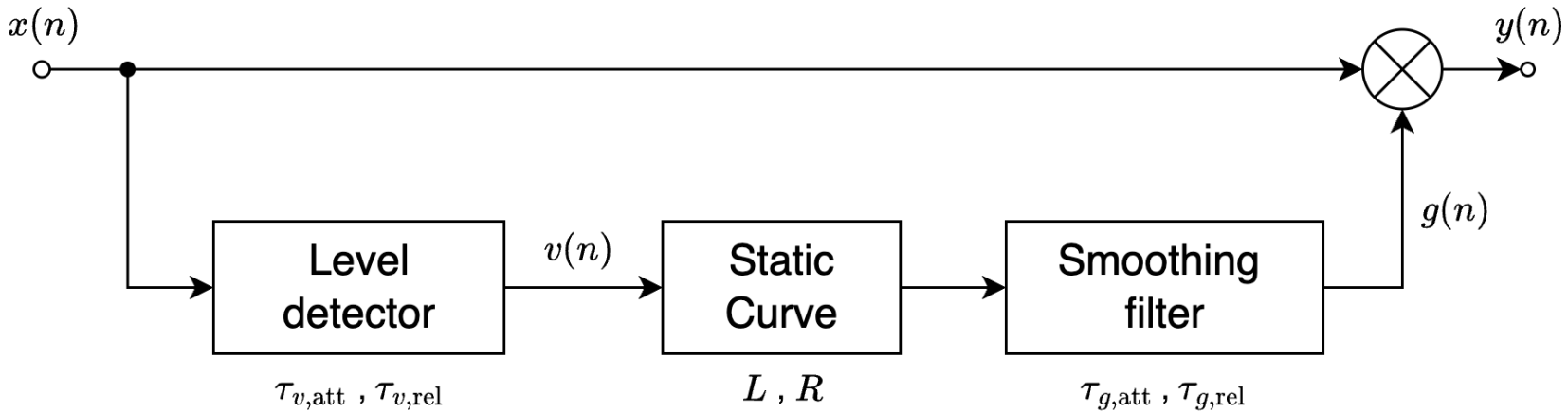


Original

Compressed

Model-based DRC

The DRC model is a commonly used one [1], which employs a profile consisting of 7 parameters:



- L: the threshold expressed in dB
- R: the compression ratio (R > 1 : compressor, R ∈]0, 1[: expander, R = 0 : noise gate, etc.)
- $\tau_{v,\mathrm{att}}$, $\tau_{v,\mathrm{rel}}$: the attack and release time used to smooth the detection envelope
- $\tau_{g,\mathrm{att}}$, $\tau_{g,\mathrm{rel}}$: the attack and release time used to smooth the gain function
- p: the compressor detector type (1: peak, 2: RMS)

➢ We set p=2 in this work

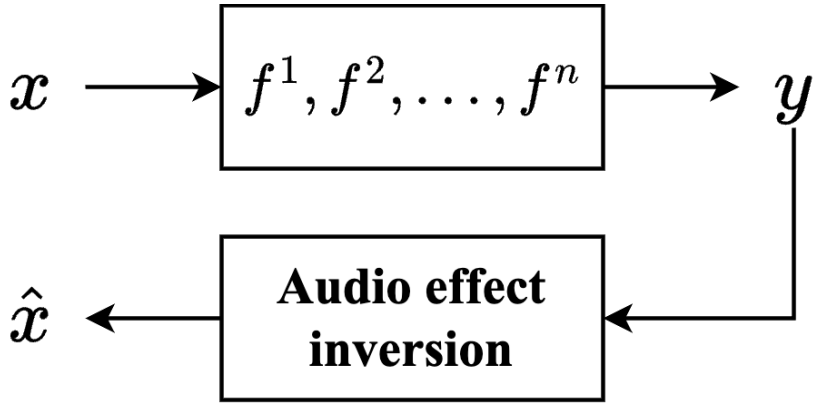
$q_\theta = \{L,\ R,\ \tau_{v,\mathrm{att}},\ \tau_{v,\mathrm{rel}},\ \tau_{g,\mathrm{att}},\ \tau_{g,\mathrm{rel}},\ p\}$ is called a DRC profile

[1] Gorlow, Stanislaw, and Joshua D. Reiss. "Model-based inversion of dynamic range compression." Proc. IEEE Transactions on Audio, Speech, and Language Processing 21.7 (2013): 1434-1444.

Audio Restoration

Problem: DRC can be destructive (distortion, saturation, clipping, etc.)

→ Solution: We invert the DRC by estimating the original signal $\hat{x}$ from a processed signal y, which can be regarded as an audio restoration problem

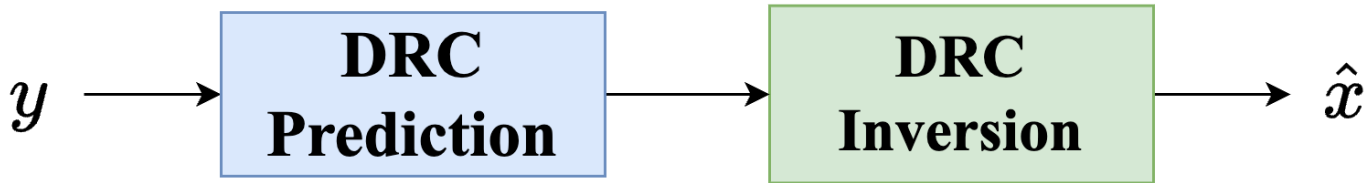


Motivation:

- Restore original signal properties
- Enhance audio clarity and quality
- Essential for an accurate representation and enhancement of the audio content

Objective

Problem: DRC inversion is challenging, with only a limited number of methods available.



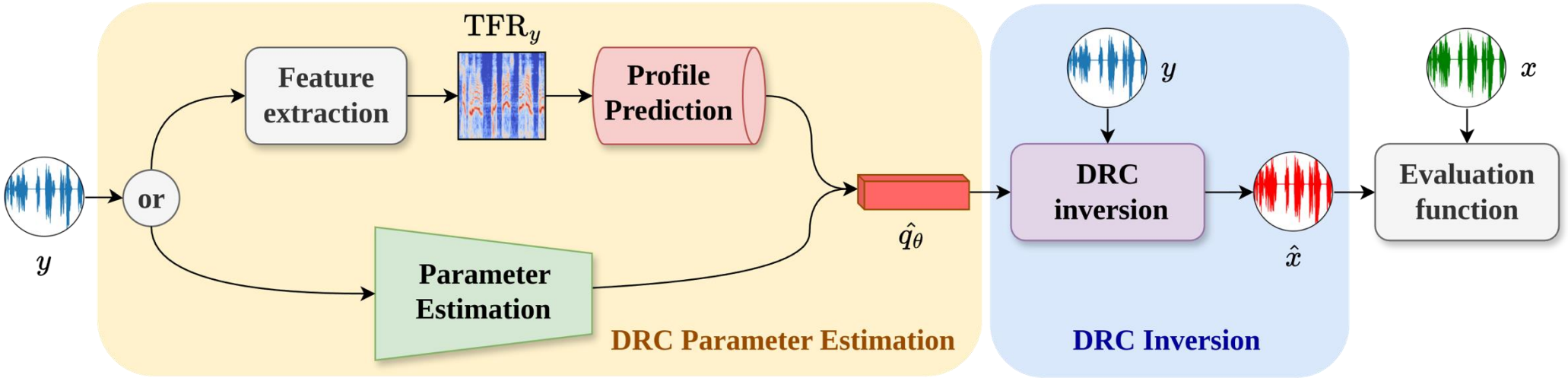
Objective: By focusing on DRC, we propose a model that aims to

1. predict which DRC was applied, based on the observation of the compressed signal y
2. invert DRC and estimate the original signal $\hat{x}$ [1]

[1] Gorlow, Stanislaw, and Joshua D. Reiss. "Model-based inversion of dynamic range compression." Proc. IEEE Transactions on Audio, Speech, and Language Processing 21.7 (2013): 1434-1444.

Methodology

Overall Proposed Method



- x : the ground truth signal
- y : the input compressed signal
- TFR_y : the time-frequency representation of y, used for DRC profile prediction
- θ : the DRC profile index
- $\hat{q}_\theta$: the estimated DRC profile, containing 7 parameters
- $\hat{x}$: the restored signal

Model-based DRC: Forward

Problem Formulation:

$$y[n] = x[n] \cdot g_x[n] \tag{1}$$

1. Compute the detection envelope $v[n]$, using x, p, $\tau_{v,att}$ and $\tau_{v,rel}$:

$$v[n] = (\beta|x[n]|^p + (1-\beta)|v[n-1]|^p)^{1/p} \tag{2}$$

   With β the envelope smoothing factor

2. Compute the compression factor $f[n]$ using v and R as:

$$f[n] = \begin{cases} \left(\frac{l}{v[n]}\right)^{1-\frac{1}{R}} & \text{if } v[n] > l, \text{ with } l = 10^{L/20} \\ 1 & \text{otherwise} \end{cases} \tag{3}$$

3. Smooth the compression factor to obtain the gain as:

$$g[n] = \gamma f[n] + (1-\gamma) g[n-1] \tag{4}$$

   With γ the gain smoothing factor

4. Apply the compression effect to obtain y as Eq.(1)

Algorithm 1 The compressor

function COMP($x_n, \boldsymbol{\theta}, f_s$)
- $\tilde{x}_n \leftarrow 0$
- $g_n \leftarrow 1$
- **for** $n \leftarrow 1, N$ **do**
  - **if** $|x_n|^p > \tilde{x}_n$ **then**
    - $\beta \leftarrow 1 - \exp[-2.2/(f_s \cdot \tau_{v,\text{att}})]$
  - **else**
    - $\beta \leftarrow 1 - \exp[-2.2/(f_s \cdot \tau_{v,\text{rel}})]$
  - **end if**
  - $\tilde{x}_n \leftarrow \beta|x_n|^p + \bar{\beta}\tilde{x}_n$
  - $v_n \leftarrow \sqrt[p]{\tilde{x}_n}$
  - **if** $v_n > l$ **then**
    - $f_n \leftarrow \kappa v_n^{-S}$
  - **else**
    - $f_n \leftarrow 1$
  - **end if**
  - **if** $f_n < g_n$ **then**
    - $\gamma \leftarrow 1 - \exp[-2.2/(f_s \cdot \tau_{g,\text{att}})]$
  - **else**
    - $\gamma \leftarrow 1 - \exp[-2.2/(f_s \cdot \tau_{g,\text{rel}})]$
  - **end if**
  - $g_n \leftarrow \gamma f_n + \bar{\gamma} g_n$
  - $y_n \leftarrow g_n x_n$
- **end for**
- **return** y_n

end function

Model-based DRC: Inversion

- Goal: to estimate the original signal x from compressed signal y and the estimated DRC parameters $\hat{q}_\theta$

$$y[n] = g_\theta[n] \cdot x[n] \rightarrow x[n] = \frac{y[n]}{g_{\hat{\theta}}[n]} \tag{5}$$

- Idea: to estimate the envelope value $\mathbf{v}[n]$ instead of $x[n]$, where $v[n] = \sqrt[p]{\tilde{x}[n]}$
- Solution: the root of the characteristic function is the envelope value $v[n]$:

$$\xi_p(v) = (\gamma\kappa v[n]^{-S} + \bar{\gamma} g[n-1])^p (v[n]^p - \bar{\beta} v[n-1]^p) - \beta |y[n]|^p \tag{6}$$

where $\mathrm{K} = \mathrm{l}^S$ with l the linear threshold, $S = 1 - \frac{1}{R}$

For the inversion, We assume q_θ is known!

→ How to get the DRC parameters ?

Algorithm 3 The iterative search for the zero-crossing

```
function CHARFZERO(v_n, ε)
    v_i ← v_n
    repeat
        Δ_i ← |ζ_p(v_i)|
        v_i ← v_i − Δ_i · ζ_p(v_i)/[ζ_p(v_i + Δ_i) − ζ_p(v_i)]
        if |ζ_p(v_i)| > Δ_i then
            return v_n
        end if
        v_n ← v_i
    until |ζ_p(v_i)| < ε
    return v_i
end function
```

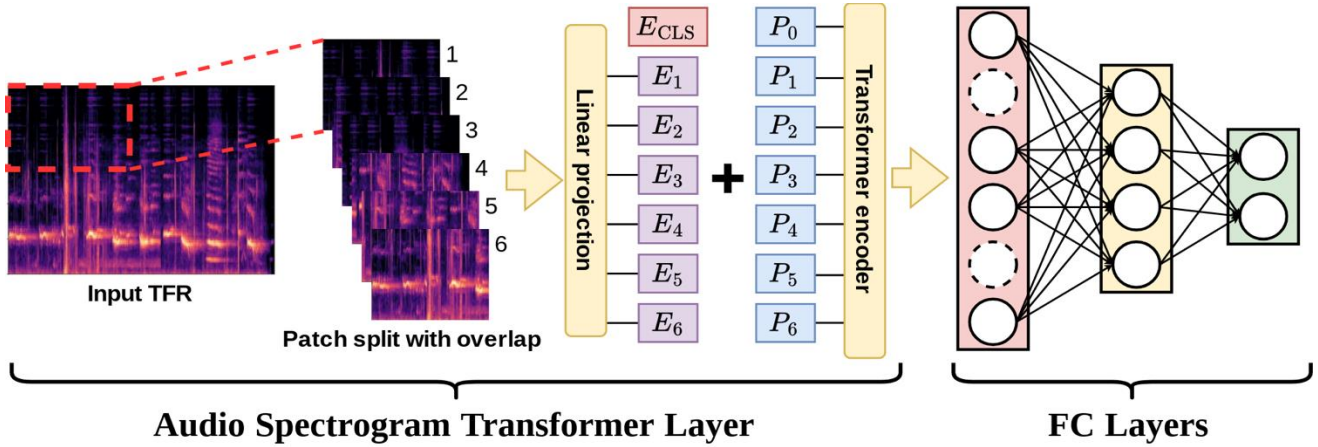
Algorithm 2 The decompressor

```
function DECOMP(y_n, θ, ε, f_s)
    x̃_n ← 0
    g_n ← 1
    for n ← 1, N do
        if |y_n| > ∛x̃_n · g_n then
            β ← 1 − exp[−2.2/(f_s · τ_v,att)]
        else
            β ← 1 − exp[−2.2/(f_s · τ_v,rel)]
        end if
        if |y_n| > ∛[(κ/g_n)^(p/S) − β̄x̃_n]/β · g_n then
            γ ← 1 − exp[−2.2/(f_s · τ_g,att)]
        else
            γ ← 1 − exp[−2.2/(f_s · τ_g,rel)]
        end if
        if |y_n| > ∛(l^p − β̄x̃_n)/β · (γ + γ̄g_n) then
            v_n ← ∛β[|y_n|/(γ + γ̄g_n)]^p + β̄x̃_n
            v_o ← CHARFZERO(v_n, ε)
            |x_n| ← ∛(v_o^p − β̄x̃_n)/β
            x̃_n ← v_o^p
            g_n ← |y_n|/|x_n|
        else
            g_n ← γ + γ̄g_n
            |x_n| ← |y_n|/g_n
            x̃_n ← β|x_n|^p + β̄x̃_n
        end if
        x_n ← sgn(y_n) · |x_n|
    end for
    return x_n
end function
```

Solution I: DRC Profile Classification

We propose to use Audio Spectrogram Transformer (AST) [2] to predict the DRC profile $\hat{q}_\theta$:

- Input: Time-Frequency Representation TFR_y , such as STFT, MFCC, etc.
- Modification: additional MLP at the end, number of layers depends on the experiment
- Output: estimated DRC profile index $\hat{\theta}$, of which the parameters are known



Advantage:

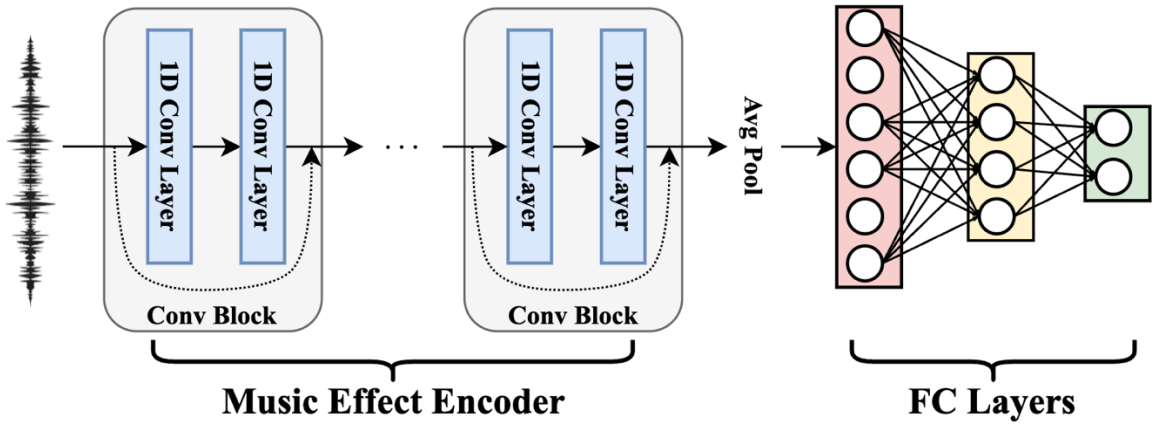
1. The self-attention mechanism captures the audio spectrogram features
2. The compressor creates distinct patterns in the spectrogram, and these patterns match what the AST model was designed to detect
3. The fully connected layers improve the classification performance of the model

[2] Gong, Yuan, Yu-An Chung, and James Glass. "Ast: Audio spectrogram transformer." arXiv preprint arXiv:2104.01778 (2021).

Solution II: DRC Parameter Estimation

Music Effect Encoder (MEE) model used in [3]:

- Input: audio waveform of the compressed signal y
- Modification: additional MLP at the end [4]
- Output: 6 estimated DRC parameters $\hat{q}_\theta$ (we keep the detector type p=2 in this work)



Advantage:

1. Similar experiments in [4] with good results, especially for DRC parameters estimation
2. The fully connected layers improve the classification performance of the model

[3] Koo, Junghyun, Seungryeol Paik, and Kyogu Lee. "End-to-end music remastering system using self-supervised and adversarial training." Proc. IEEE ICASSP 2022, 4608-4612.

[4] Peladeau, Côme, and Geoffroy Peeters. "Blind estimation of audio effects using an auto-encoder approach and differentiable digital signal processing." Proc. IEEE ICASSP 2024,856-860 .

Experimental Setting

Datasets Generation

We source audio from four different datasets, detailed in the Table below:

Name	Data	Sampling Frequency (Hz)	Ref
MedleyDB	Raw	44100	[5]
MUSDB18-HQ	Raw	44100	[6]
DAFX	Mastered	32000	[7]
LibriSpeech	Raw	16000	[8]

Table 1: Datasets information.

We randomly extract 1.6 h of audio from each dataset, then :

- segment each audio file into 5-second-long chunks,
- remove the chunks for which the overall root mean square (RMS) of the amplitude is below -30 dB.

The result is our ground truth original signal x.

[5] Bittner, Rachel M., et al. "Medleydb: A multitrack dataset for annotation-intensive MIR research." Proc. ISMIR. Vol. 14. 2014.
[6] Rafii, Zafar, et al. "MUSDB18-HQ-an uncompressed version of MUSDB18." 2019.
[7] Tardieu, Damien, et al. "Production effect: audio features for recording techniques description and decade prediction." Proc. DAFx 2011.
[8] Panayotov, Vassil, et al. "Librispeech: an asr corpus based on public domain audio books." Proc. IEEE ICASSP 2015.

Datasets Generation

We generate 30 DRC presets according to the range given in [9]:

Parameter	Description	Lower End	Upper End
L (dBFS)	Threshold	-60	-20
R (dB_{in}:dB_{out})	Ratio	2	15
τ_v^{att} (ms)	Envelope attack	5	130
τ_v^{rel} (ms)	Envelope release		
τ_g^{att} (ms)	Gain attack	10	500
τ_g^{rel} (ms)	Gain release	25	2000
p (1 or 2)	Detector type	2	2

Table 2: The range of each parameter for the 30 DRC profiles.

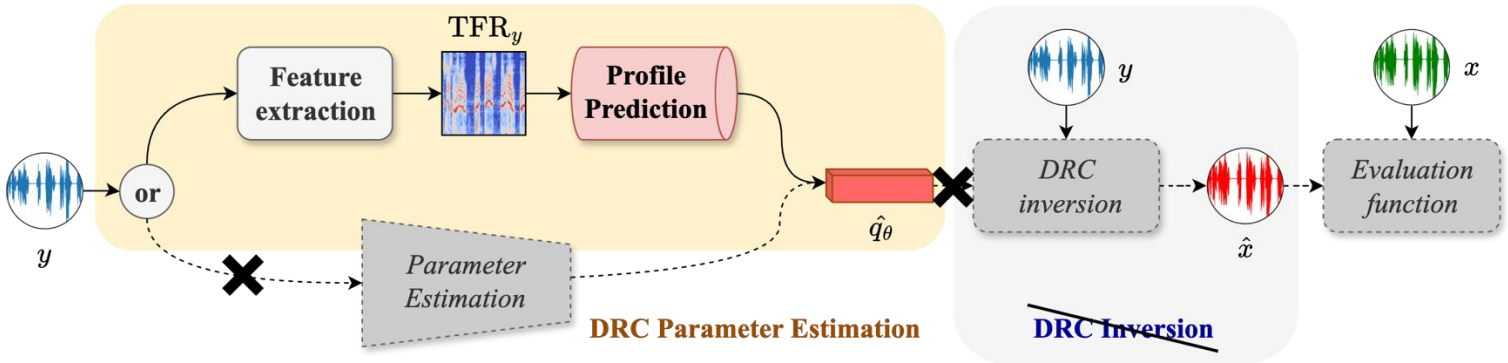
And then, we generate 4 sub-datasets using the 30 DRC profiles.

- As a results, each sub-datasets contains **35,867** chunks of 5 seconds in 31 classes (30 DRC profiles + Neutral signal).

[9] Zölzer, Udo, et al. DAFX-Digital audio effects. John Wiley & Sons, 2002.

Parameters Estimation Results

DRC Profile Classification: Input Feature Selection



Goal: to find the most suitable TFR_y and the corresponding size:

- TFR_y: STFT, MFCC, Mel Spectrogram (MelS.) and CQT
- Time scale resolution of the TFR_y: 2.9, 5.8, 22.6, 23.1;
- Number of frequency bins of the TFR_y: 32, 64, 126, 256, according to [10]

Dataset	Input	Size	Time/epoch(h)	Acc
MedleyDB	MFCC	128 × 431	0.093	0.66
	STFT	64 × 431	**0.037**	**0.82**
	MelS	128 × 431	0.074	0.81
	CQT	128 × 862	0.79	0.55
MUSDB18-HQ	MFCC	128 × 431	0.093	0.65
	STFT	64 × 431	**0.037**	**0.80**
	MelS	128 × 431	0.074	0.78
	CQT	128 × 862	0.79	0.53
DAFX	MFCC	64 × 431	0.091	0.72
	STFT	64 × 431	**0.036**	**0.84**
	MelS	64 × 431	0.072	0.83
	CQT	64 × 862	0.75	0.66
LibriSpeech	MFCC	64 × 431	0.089	0.73
	STFT	64 × 431	**0.035**	**0.85**
	MelS	64 × 431	0.070	0.84
	CQT	64 × 862	0.71	0.67

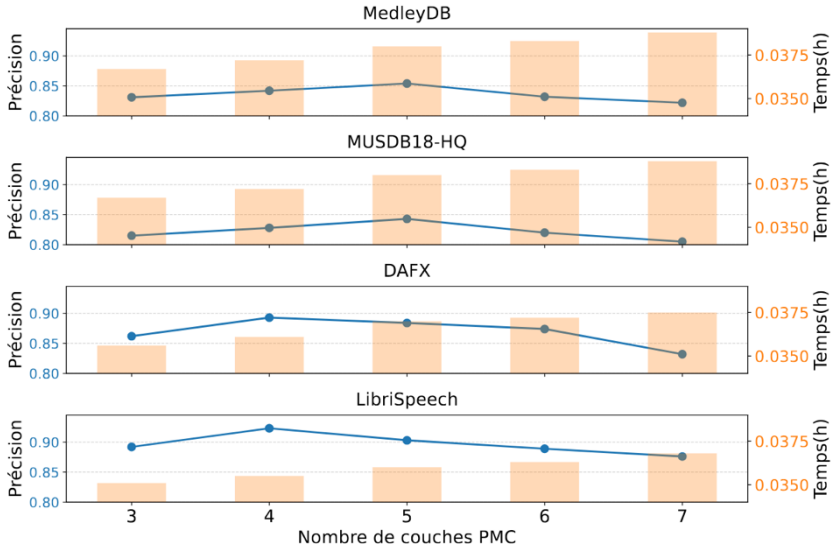
Table 3: Input feature exploration results.

★ The STFT of size 64×431 is always the best choice !

[10] Huzaifah, Muhammad. "Comparison of time-frequency representations for environmental sound classification using convolutional neural networks." arXiv preprint arXiv:1706.07156 (2017).

DRC Profile Classification: FC Layer Selection

- Goal: We try to find the most suitable number of FC layers in MLP
- Results: 5 FC layers is suitable for the MedleyDB and MUSDB18-HQ datasets, 4 for DAFX and LibriSpeech datasets:



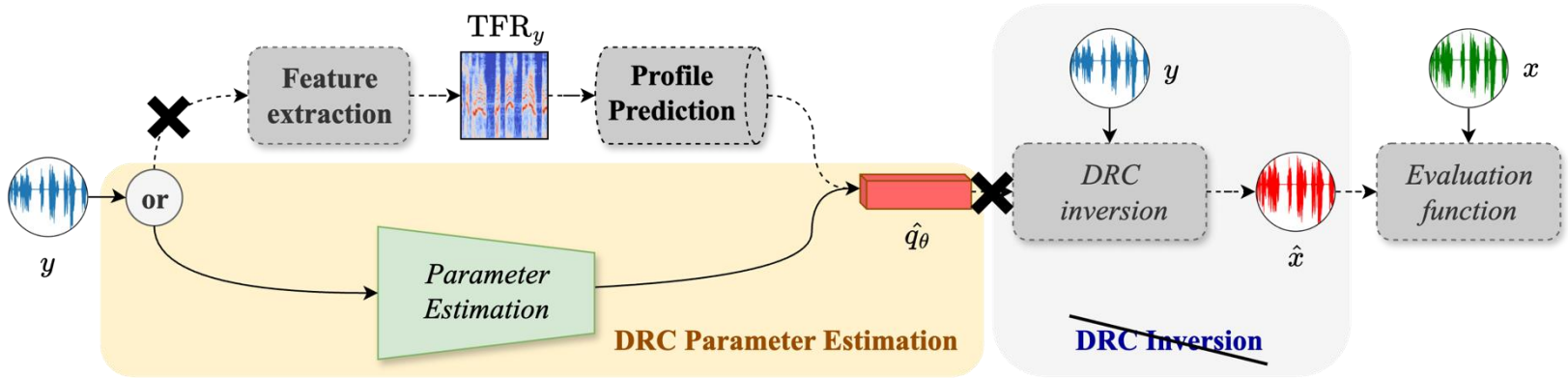
→ The best results achieved for the DRC profiles classification task:

Dataset	MLP	Time/epoch(h)	Accuracy
MedleyDB	5	0.09	0.854
MUSDB18-HQ	5	0.09	0.843
DAFX	4	0.08	0.893
LibriSpeech	4	0.08	0.923

Table 4: The best classification results achieved on each dataset.

DRC Parameter Estimation: Model Comparison

- Goal: estimating the individual 7 parameters instead of the profiles



- We made a comparison between the MEE and Time-Frequency Encoder (TFE) [4] :

MEE model

Dataset	Time/epoch(h)	$\mathcal{L}^{\text{MSE}}_{\hat{q_\theta},q_\theta}$
MedleyDB	0.084	**0.031**
MUSDB18-HQ	0.087	**0.039**
DAFX	0.065	**0.025**
LibriSpeech	0.040	**0.012**

TFE model

Dataset	Time/epoch(h)	$\mathcal{L}^{\text{MSE}}_{\hat{q_\theta},q_\theta}$
MedleyDB	0.032	0.058
MUSDB18-HQ	0.038	0.058
DAFX	0.022	0.054
LibriSpeech	0.015	0.052

- The input feature of the TFE model is STFT of size (64, 431).

[4] Peladeau, Côme, and Geoffroy Peeters. "Blind estimation of audio effects using an auto-encoder approach and differentiable digital signal processing." Proc. IEEE ICASSP 2024.

DRC Inversion Results

DRC Inversion: Baseline Models Comparison

We address the DRC inversion task: $y, \theta \rightarrow \hat{x}$ with the well-trained AST and MEE model.

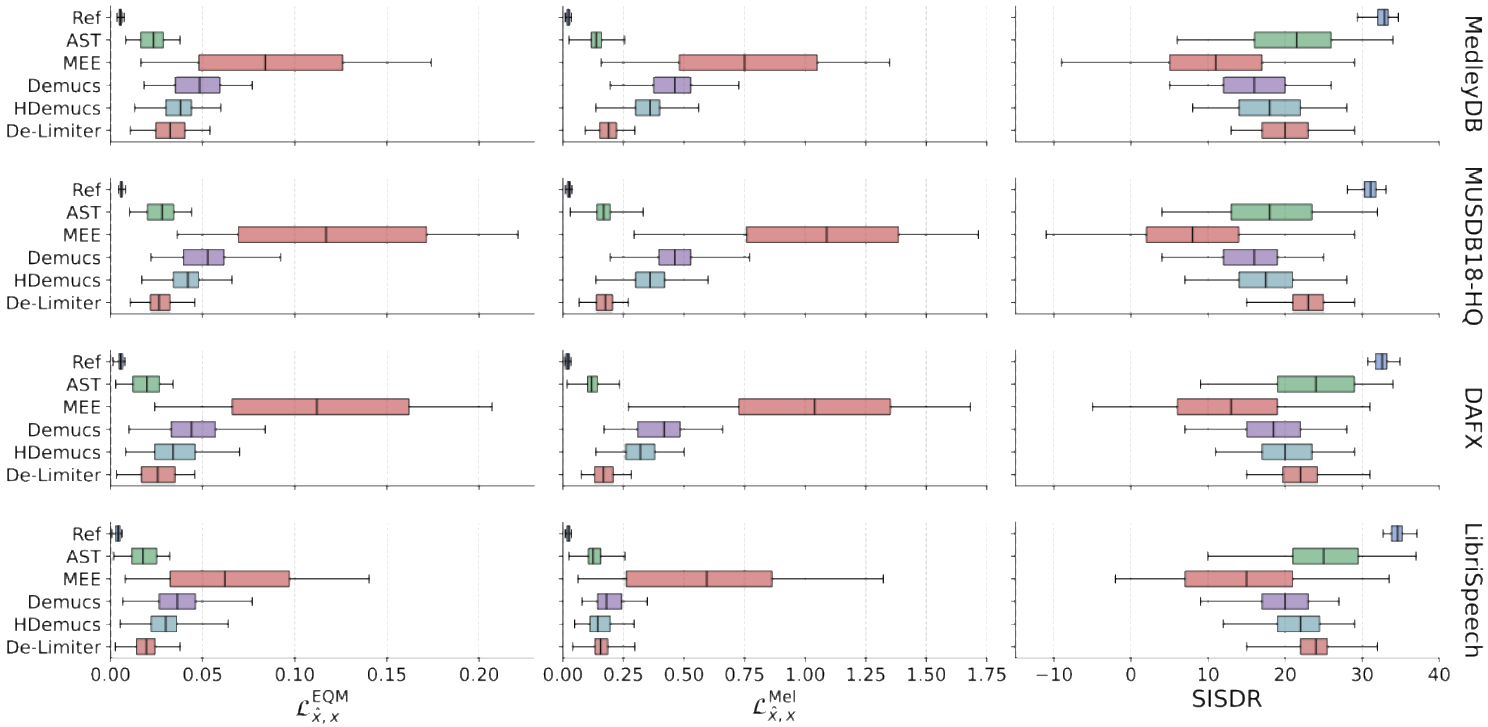
We compare our model with three state-of-the-art models:

- **Demucs**[11]: Used in [12] to remove distortion and clipping applied to guitar tracks for music production and get good results
- **Hybrid Demucs (HDemucs)** [13]: An improved version of the Demucs, audio processing capabilities is improved especially for high-frequency details and complex arrangements. It has shown good results on audio restoration task [14]
- **De-Limiter** [15]: Originally designed to recover original signals from heavily compressed audio (delimiter), making it well-suited for our task.

[11] Défossez, Alexandre, et al. "Music source separation in the waveform domain." arXiv preprint arXiv:1911.13254 (2019).
[12] Imort, Johannes, et al. "Removing distortion effects in music using deep neural networks." arXiv preprint arXiv:2202.01664 (2022).
[13] Rouard, Simon, Francisco Massa, and Alexandre Défossez. “Hybrid transformers for music source separation.” Proc. IEEE ICASSP 2023.
[14] Rice, Matthew, et al. “General purpose audio effect removal.” Proc. IEEE WASPAA 2023.
[15] Jeon, Chang-Bin, and Kyogu Lee. "Music de-limiter networks via sample-wise gain inversion." Proc. IEEE WASPAA.

DRC Inversion: Baseline Models Comparison

- For this task, we use $L^{MSE}_{\hat{x},,x}$, $L^{Mel}_{\hat{x},,x}$ and SISDR as evaluation matrices:



The AST model significantly outperforms the MEE method because:

1. **Task property**: The classification task is less complex than the regression task
2. **Experimental difference**: A correct classification provides the ground truth parameters directly, which are often difficult to obtain accurately in a regression task
3. **Many-to-one mapping challenge**: Different parameter settings can lead to the same compression effect

DRC Inversion: Baseline Models Comparison

Original

Compressed

AST

De-Limiter

Conclusion

In this work, we:

- Propose an end-to-end method for audio DRC inversion
- Use hybrid approach for DRC parameters estimation
  - Classification task for profile prediction
  - Regression task for parameters estimation

Limitation:

- DRC parameters estimation performs poorly
- The Model is limited by the priori knowledge of the DRC parameters

Perspectives:

- Improve the performance of the DRC parameters estimation task
- Contribute on more audio effects, such as dereverberation, equalization inversion, etc.

Thank You for Attention!

Contact: haoran.sun@upsaclay.fr

Github page: https://github.com/SunHaoRanCN/DRC_Inversion

- [1]Math model: Gorlow, Stanislaw, and Joshua D. Reiss. "Model-based inversion of dynamic range compression." Proc. IEEE Transactions on Audio, Speech, and Language Processing 21.7 (2013): 1434-1444.
- [2]AST model: Gong, Yuan, Yu-An Chung, and James Glass. "Ast: Audio spectrogram transformer." arXiv preprint arXiv:2104.01778 (2021).
- [4]MEE model: Peladeau, Côme, and Geoffroy Peeters. "Blind estimation of audio effects using an auto-encoder approach and differentiable digital signal processing." Proc. IEEE ICASSP 2024,856-860 .