A psycho-acoustic loss function based on a psycho-acoustic model

Gerald Schuller, Muhammad Imran

Asilomar Conference on SSC, 2024

Technische Universität Ilmenau

Introduction

- Perceptual audio quality measurement is important in many applications
- Examples are speech enhancement, music synthesis, etc.
- Reference: [Zwicker and Fastl, 1999], [Thiede et al., 2000]

Motivation

- Traditional loss functions, like mse, are inadequate for perceptual quality
- Psycho-acoustic model offers a better approach
- Reference: [Zwicker and Fastl, 1999]

Traditional Loss Functions

- Log Spectral Difference and Multi-Scale Spectral Loss
- Limitation in representing human auditory perception
- Reference: [Rabiner and Juang, 1993], [Engel et al., 2020]

Introduction to Psycho-Acoustic Models

- Psycho-acoustic models, as from audio coding, mimic human auditory masking
- Application of masking to audio quality measurement
- Reference: [Zwicker and Fastl, 1999]

Psycho-Acoustic Masking

- Concept of auditory masking: louder sounds mask quieter ones
- Used in MDCT domain for perceptual comparison of audio
- Reference: [Zwicker and Fastl, 1999], MPEG-1 Layer III model

The Psycho-Acoustic Loss Function

- Operates in the MDCT domain, comparing spectral differences above masking threshold
- Focus on perceptually significant aspects of audio
- Reference: [Kim, 2020, Schuller, 2024]
- Our loss function uses the model of our Python audio coder in [Schuller, 2023]

Divergence vs. Distance Metric

- Asymmetric behavior: our loss function is a divergence from original
- More aligned with human perceptual evaluation of audio
- Reference: [Vincent et al., 2006]

Implementation in PyTorch

- PyTorch implementation of the loss function
- Modular, easy integration into deep learning models
- Reference: [Schuller, 2024], [Steinmetz, 2020]

Applications: Speech Enhancement

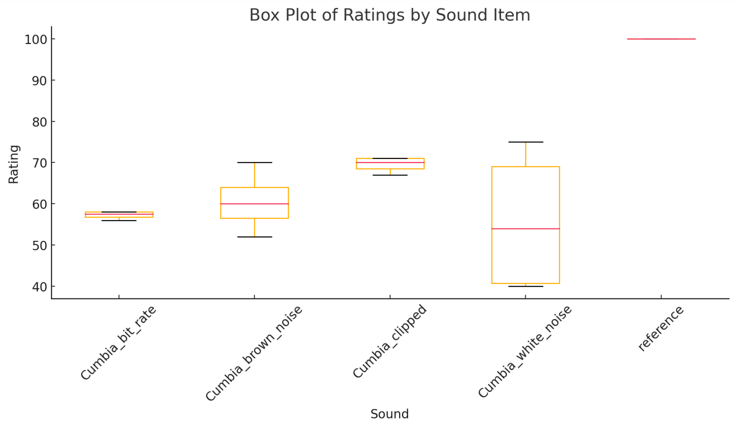
- Use in speech enhancement systems, see how it compares with log-spectral loss
- Perceptual improvements should be confirmed by listening tests
- Reference: [Strauss et al., 2023]

Comparison of our Loss Function with a Listening Test

- a **Listening test** with Multi Stimulus with hidden Reference and Anchor (**MUSHRA**)
- Rating from 0 to 100, 100 being the best.
- "cumbia" music audio item
- Tested distortions:
  -Low bitrate coding,
  -added brownian noise,
  -clipping,
  -added white noise
- Mushra Test with 4 participants

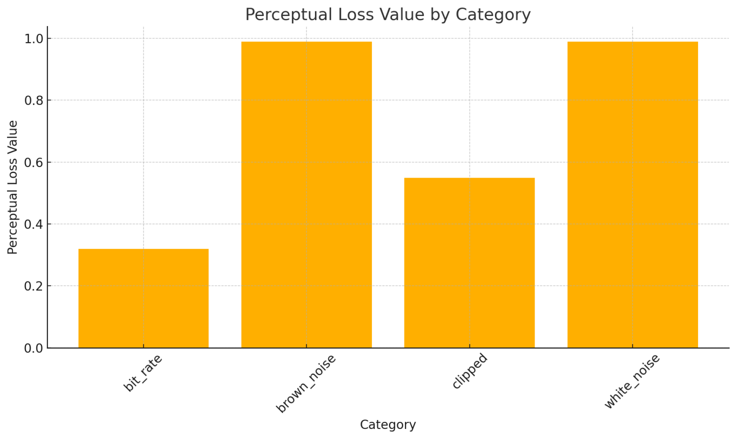
MUSHRA box plot

Higher is better:



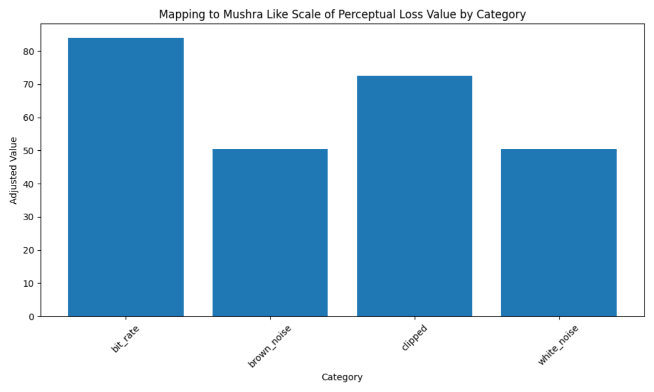
Corresponding Output of our Perc. Loss Function

Here lower ist better!:



Mapped Output of our Perc. Loss Function

- To map the loss function output to a similar scale as the MUSHRA test, the function $50 \cdot (2 - x)$ was applied, where x is the loss function output.
- This also means that now **higher is better** again



Comparing MUSHRA and our Loss Function

- MUSHRA is regarded as ground truth
- For now it is only preliminary because of only few audio items and listeners
- Our loss function seems to predict the noisy items as better than the MUSHRA test.
- Hence more work is needed.

Case Study: Python & PyTorch Implementation

- Practical examples and code from GitHub repositories
- A simple implementation of our psycho-acoustic loss function cane be found in this Colab notebook: `https://colab.research.google.com/github/TUIlmenauAMS/PsychoacousticLoss/blob/main/onlyPsyacLoss.ipynb`
- which is from [Schuller, 2024]

Discussion & Future Work

- Future applications in music generation, real-time systems
- Temporal masking, optimizing for lower latency
- Reference: Conclusion of the paper

Conclusion

- Psycho-acoustic loss improves perceptual quality of audio
- Suitable for machine learning applications in audio
- More data in the paper

Our Github repository:

https://github.com/TUIlmenauAMS/PsychoacousticLoss

Questions?

Engel, J., Hantrakul, L., Gu, C., and Roberts, A. (2020).
Ddsp: Differentiable digital signal processing.
In *International Conference on Learning Representations*.

Kim, M. (2020).
Python model of the mpeg-1 psychoacoustic model.
Available at: `https://github.com/cocosci/pam-nac`.

Rabiner, L. and Juang, B. (1993).
Fundamentals of speech recognition.
PTR Prentice Hall.

Schuller, G. (2023).
Python-Audio-Coder.
`https://github.com/TUIlmenauAMS/Python-Audio-Coder`.
GitHub repository.

Schuller, G. (2024).

Psychoacoustic loss function in pytorch.
Available at: https://github.com/TUIlmenauAMS/PsychoacousticLoss.

Steinmetz, C. (2020).
Auraloss: Audio-focused loss functions in pytorch.
Available at:
https://github.com/csteinmetz1/auraloss.

Strauss, M. et al. (2023).
Sefgan: Harvesting the power of normalizing flows and gans for efficient high-quality speech enhancement.
In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*.

Thiede, T. et al. (2000).
Peaq—the itu standard for objective measurement of perceived audio quality.

Journal of the Audio Engineering Society, 48(1/2):3–29.

Vincent, E., Gribonval, R., and Févotte, C. (2006).
Performance measurement in blind audio source separation.
IEEE Transactions on Audio, Speech, and Language Processing, 14(4):1462–1469.

Zwicker, E. and Fastl, H. (1999).
Psychoacoustics: Facts and models.
Springer.