Report on the 2024 Asilomar Conference on Signals, Systems, and Computers
Special Session on
Perceptual and Higher Level Loss and Distance Functions for Audio and Acoustics

Gerald Schuller
Ilmenau University of Technology, Germany

Outline

Asilomar Conference Overview

- The Asilomar Conference on Signals, Systems, and Computers is an annual conference focused on digital signal processing for media and communications.
- In 2024, a special session was dedicated to *Perceptual and Higher Level Loss and Distance Functions for Audio and Acoustics*.
- The session aimed to explore new loss functions and distance metrics that improve optimization in audio and acoustic applications.

Importance of Loss Functions in Audio Processing

- Loss functions are critical in training models, particularly in deep learning
- In audio processing, designing perceptually relevant loss functions can lead to better sound quality and more efficient algorithms.
- Challenges include ensuring smoothness, continuity, and convexity to facilitate optimization, or to focus on desired features.

Key Properties for Optimization

- **Smoothness**: Ensures gradient-based optimization methods perform effectively.
- **Continuity**: Prevents abrupt changes that could hinder convergence.
- **Convexity**: Helps in finding global minima, avoiding local minima traps.
- **Few Local Minima**: Simplifies the optimization landscape.
- **Focus on Desired Features**: This should lead to a desired optimum.

Talk 1: Similarity Metrics for Late Reverberation

Presenter: Gloria Del Santo et al.

Highlights:

- Introduced metrics like *Energy Decay Convergence* and *Averaged Power Convergence*.
- Focused on room impulse response similarity for optimizing artificial reverberation.
- Proposed methods smooth the distance function, aiding optimization.

Talk 2: Time-Frequency Audio Similarity using Optimal Transport

Presenter: Linda Fabiani et al.

Highlights:

- Applied *Optimal Transport* theory to spectrograms for measuring audio similarity.
- Provided a framework that is both mathematically rigorous and perceptually relevant.
- Enhanced the smoothness and continuity of the loss function in audio comparisons.

Tailoring Loss Functions to Specific Properties

- Beyond optimization-friendly properties, loss functions can focus on perceptual aspects.
- Incorporating human auditory models can lead to improvements in tasks like compression and synthesis.
- Emphasis on properties such as timbre, pitch, and intelligibility.

Talk 3: A Psycho-Acoustic Loss Function Based on a Psycho-Acoustic Model

Presenter: Gerald Schuller et al.

Highlights:

- Developed a loss function incorporating masking effects of the human ear.
- Used a psycho-acoustic model from a Python audio coder.
- `https://github.com/TUIlmenauAMS/Python-Audio-Coder`
- Focused on perceptual relevance rather than convexity.
- This means non-convex optimization might be needed.
- Presented listening test comparisons demonstrating effectiveness.
- ▸ Open

Talk 4: Instrumental Timbre Transfer Based on Disentangled Representation of Timbre and Pitch

Presenter: Yin Le et al.

Highlights:

- Proposed a method for timbre transfer using disentangled representations.
- Developed a loss function focusing on timbre characteristics.
- Included sound examples and interactive elements like a chatbot.
- Advanced the field of musical style transfer and synthesis.
- ▸ Open

Talk 5: STOI-Optimized Pruned Recurrent Deep Autoencoders

Presenter: Reemt Hinrichs et al.

Highlights:

- Targeted low-complexity compression for cochlear implants at zero delay.
- Utilized the *Vocoder Short Time Objective Intelligibility* (VSTOI) score.
- Introduced a "pruning aware loss" for smooth network retraining during pruning.
- Improved speech intelligibility for cochlear implant users.

Key Takeaways

- The session highlighted the importance of designing loss functions that are both optimization-friendly and perceptually relevant.
- Integrating domain-specific knowledge, such as psycho-acoustic models, enhances the performance of audio processing systems.
- Future work may explore combining these approaches for even more effective audio and acoustic applications.

Acknowledgments

- Thanks to all the presenters and attendees for their valuable contributions.
- Special thanks to the organizing committee of the Asilomar Conference 2024.

References

- Special session outline: `https://cmsworkshops.com/Asilomar2024/view_session.php?SessionID=1126`
- Presentation materials:
  - Psycho-Acoustic Loss Function: `https://github.com/TUIlmenauAMS/PsychoacousticLoss`
  - ▸ Open
  - Instrumental Timbre Transfer: `https://github.com/TUIlmenauAMS/timbre-transfer`
  - ▸ Open

Contact Information

Gerald Schuller

- Email: gerald.schuller@tu-ilmenau.de
- Institution: Ilmenau University of Technology, Germany