# Audio Dequantization Using (Co)Sparse (Non)Convex Methods

Pavel Záviška, Pavel Rajmic, Ondřej Mokrý

Brno University of Technology
Brno, Czech Republic

ICASSP2021 TORONTO Canada
June 6-11, 2021
Metro Toronto Convention Centre

2021 IEEE International Conference on Acoustics, Speech and Signal Processing

6–11 June 2021 • Toronto, Ontario, Canada

IEEE Signal Processing Society

BRNO UNIVERSITY OF TECHNOLOGY

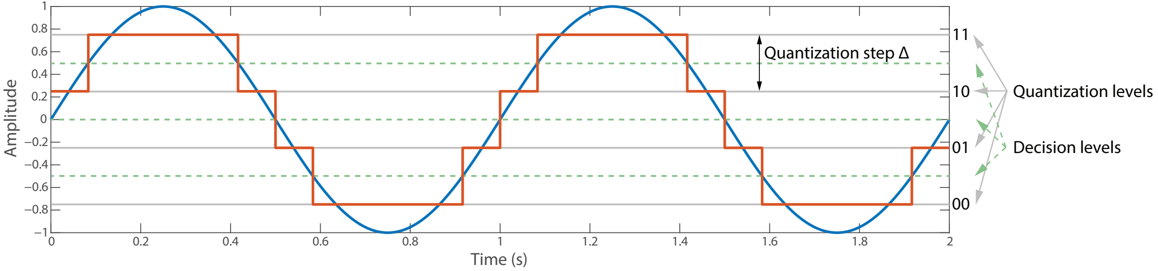
Signal Processing LABORATORY

IEEE

# Quantization

- Nonlinear limitation of signal values.
- Necessary step in signal digitization.
- Number of quantization levels, word length $w$ (bits per sample).
- Mid-riser uniform quantization:

$$(x^q)_n = \operatorname{sgn}^+(x_n)\Delta\left(\left\lfloor \frac{|x_n|}{\Delta} \right\rfloor + \frac{1}{2}\right), \text{ where } \Delta = 2^{-w+1} \text{ is the quantization step.}$$



# Quantization Example

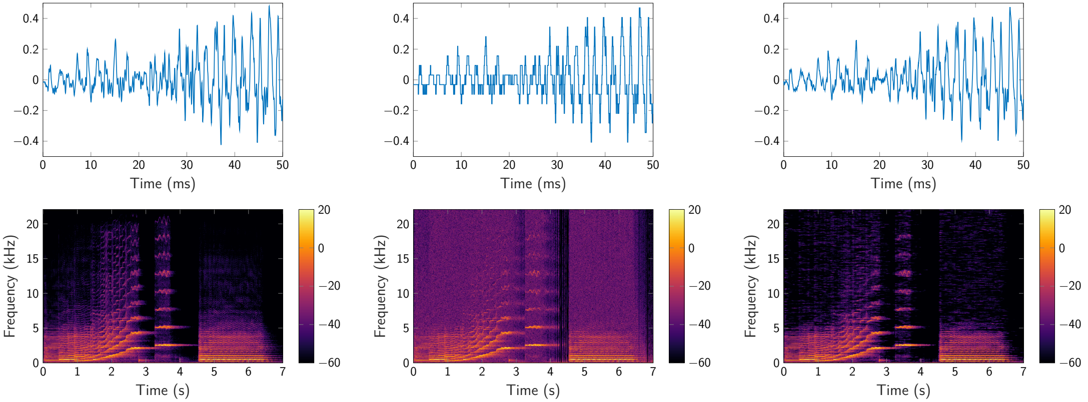
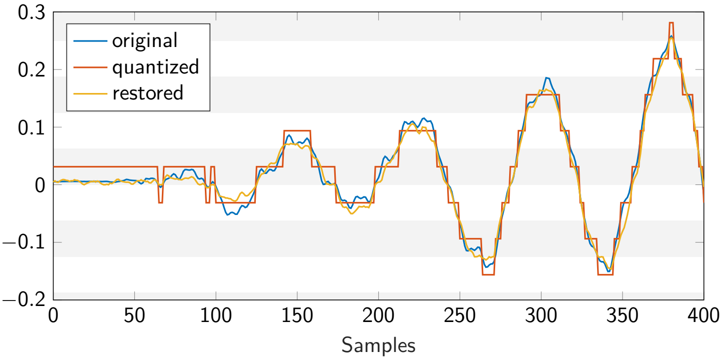



Figure: Waveforms (top) and Spectrograms (bottom) of the original (left), quantized (middle), and restored (right) signals.

# Dequantization

- Inverse problem to quantization.
- Restore the quantized observation to be as close to the original signal as possible.
- Ill-posed without additional knowledge.
- Assumption of sparsity w.r.t. STFT.

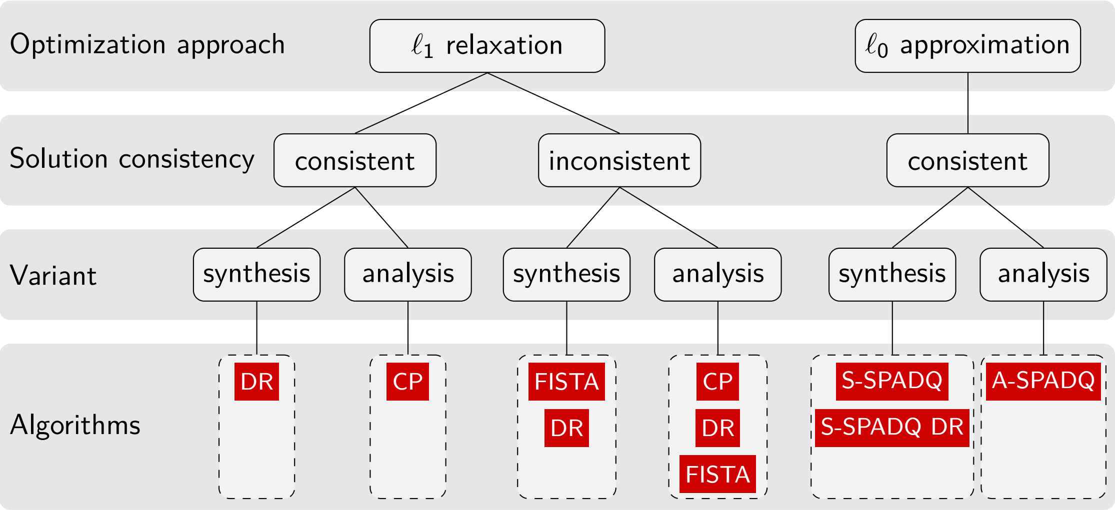


# Motivation

- Standard consumer quality audio is 16 bits per sample (bps).
- Every extra bit improves a signal-to-noise ratio (SNR) by approx. 6 dB.
- Usage of audio dequantization:
  - Enhance standard 16-bit audio,
  - Restore audio in special cases, where less than standard bit depth had to be used (scenario for the paper),
  - Enhance audio generated by Flow-based Neural Vocoder[1].
- The goal is not to compete with current lossy compression standards.
- Can be applied to any types of signals.

[1] Hyun-Wook Yoon, Sang-Hoon Lee, Hyeong-Rae Noh, and Seong-Whan Lee, "Audio dequantization for high fidelity audio generation in flow-based neural vocoder," in Proc. Interspeech 2020, Shanghai, China, Oct. 2020, pp. 3545–3549.

# Overview of algorithms



# Experiments

- Audio database
  - 10 musical audio excerpts,
  - approximate length 7 seconds,
  - sampling rate 44.1 kHz,
  - bit-depth 16 bps.
- Quantized to 7 different levels, $w = 2, 3, \dots, 8$ bps using Mid-riser quantization.
- Signals restored using algorithms based on sparsity.
- Discrete Gabor Transform (DGT/STFT), 8192 samples long Hann window, 75% overlap.
- Evaluation
  - Signal-to-distortion ratio improvement ($\Delta$SDR)
  - PEMO-Q ODG (perceptually motivated metric)
- Implementation available on GitHub:

`https://github.com/zawi01/audio_dequantization`

# ΔSDR Results

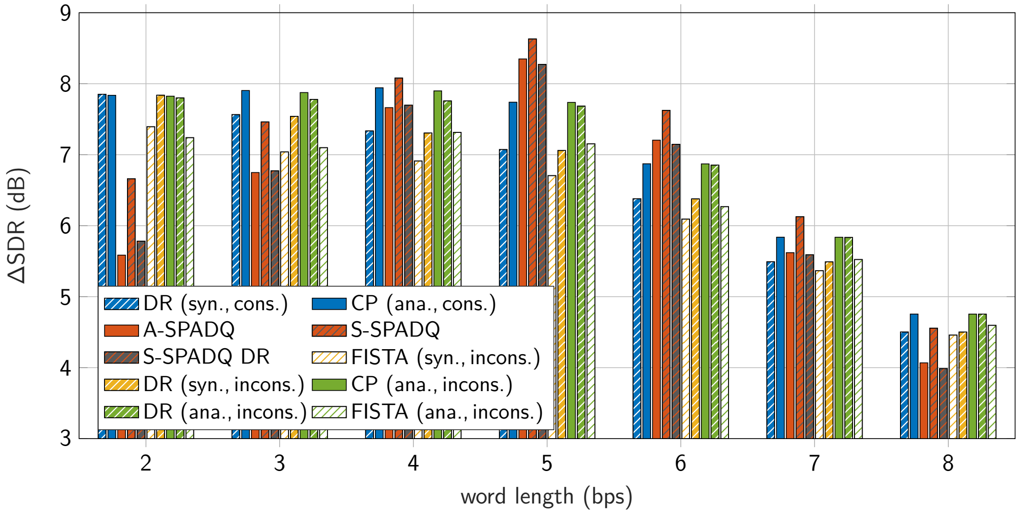



Figure: Average ΔSDR results.

# PEMO-Q Results

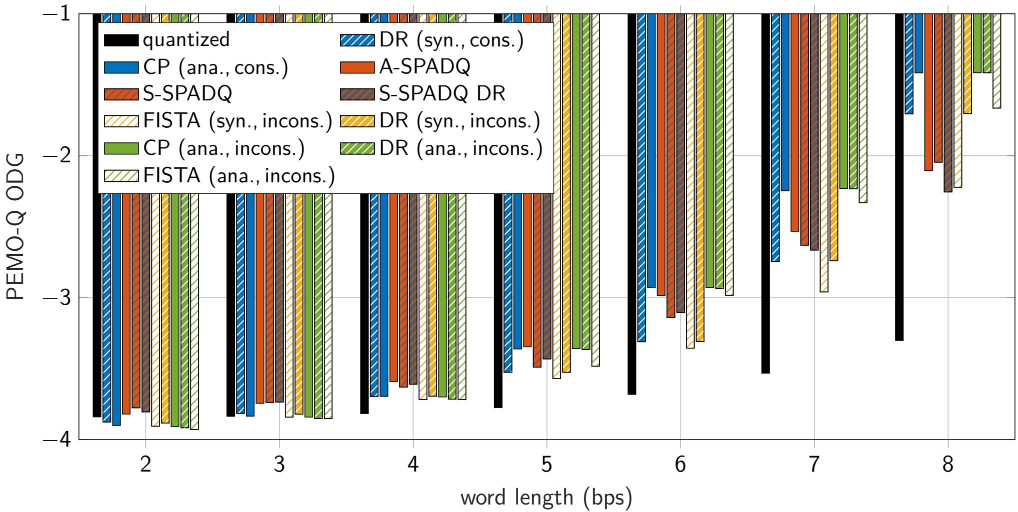



Figure: Average PEMO-Q ODG results.

# Conclusion

- 10 sparsity-based approaches to audio dequantization.
  - Convex $\ell_1$ relaxation, nonconvex $\ell_0$ approximation.
  - Strict or only approximate compliance of the solution consistency
  - Synthesis and analysis model.
- All methods improve the quality of the signal.
- No clear winner of all presented methods.
- Analysis model seems to outperform its synthesis counterpart.

Thank you for your attention!