M-LAW COMPANDING, ADPCM BASELINES, AND NEURAL AUDIO COMPRESSION METRICS

Comparative Analysis of Low-Bit-Rate Audio Coding Techniques

By 22MID0069 Merril Baiju
22MID0304 Mayank Dash

INTRODUCTION

Low-bit-rate audio coding is important in applications where there are strict constraints on compute, memory, and latency.

(e.g., telephony, embedded MCUs, IoT gateways).

There are two categories of codecs:

- **Classical waveform quantization:** PCM, μ-law companding, ADPCM
- **Modern perceptual / neural codecs:** Opus, AAC, SoundStream, EnCodec, AGC

Although neural codecs have excellent quality at low bitrates, Classically obtained techniques remain useful in environments where:

- **Only a few CPU cycles are available**
- **No floating-point hardware exists**
- **Ultra-low latency is required**

This project explores whether μ-law companding still provides benefits. compared to:

- **Uniform PCM**
- **Simple ADPCM residual coding**
- **Neural Latent Codecs (AGC)**

PROBLEM STATEMENT

There is no modern, reproducible comparison between:

- **Uniform PCM**
- **μ-law companding (G.711)**
- **Simple ADPCM variants**
- **Neural audio codecs: AGC, EnCodec, SoundStream**

Existing studies have primarily focused on high-fidelity perceptual codecs, not ultra-simple low-bit-depth waveform quantizers.

Practitioners working with low-power / embedded devices lack guidance on:

- Which method provides the best quality at 6–10 bits?
- How bitrate, entropy and quality change across methods

Need a framework that addresses: When does μ-law help and when does it fail? How does it compare to neural codecs?

OBJECTIVES

Objectives

- Evaluate μ-law companding at low bit depths (6–10 bits)
- Compare against:
  - Uniform PCM
  - Non-adaptive ADPCM
  - Neural audio codec (AGC)
- Measure quality vs bitrate using objective evaluation metrics

CONTRIBUTION

Key Contributions

- Built a unified, reproducible experimental framework
  - Quantization → Reconstruction → Metrics → RD plots
- Generated objective metrics:
  - Segmental SNR, STOI, PESQ, entropy-based bitrate
- Provided a measurement utility for AGC neural codec latents
- Identified where μ-law performs better, and where neural codecs dominate

METHODOLOGY

Methodology Overview

- Audio input: 16 kHz mono speech, normalized in the range [-1, 1]
- Three waveform quantization paths evaluated:
  - Uniform PCM
  - μ-law companding + PCM
  - ADPCM (residual coding) with & without μ-law

μ-law Companding (G.711)

- Compresses amplitude range logarithmically
- Gives more resolution to small amplitude speech signals

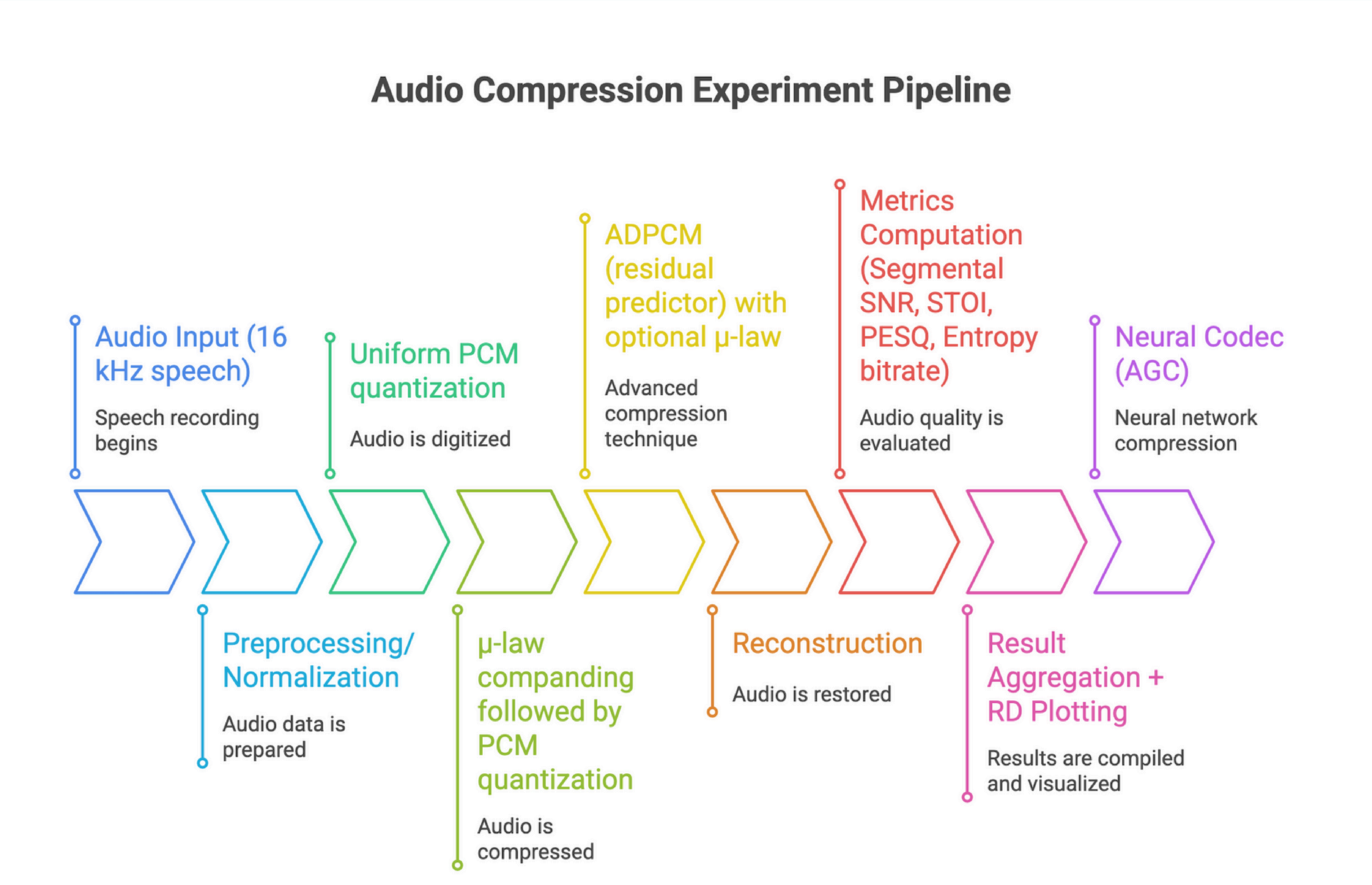
IMPLEMENTATION

- A reproducible pipeline was developed to automate:
  - Audio preprocessing
  - Quantization (PCM / μ-law / ADPCM)
  - Reconstruction
  - Metric calculation & bitrate estimation
  - RD (Rate–Distortion) plot generation

- Reproducibility
  - Single command to run experiments
  - Same audio dataset, sampling rate, and bit-depth across all tests
  - Deterministic runs for fair comparison

- Automation ensures:
  - No manual tuning
  - Results are comparable across all techniques
  - Easy extension to new codecs or bit depths

SYSTEM ARCHITECTURE



ANALYSIS

Method	Quality (SegSNR)	Intelligibility (STOI)	Perceptual Quality (PESQ)	Entropy Bitrate (kbps)
μ-law PCM	High (≈ 34 dB)	1.00 (near perfect)	4.19 (near transparent)	~117 kbps
Uniform PCM	Low (≈ 17 dB)	0.98	2.66	~70 kbps
ADPCM + μ-law	Very high SegSNR but unstable across signal types	—	—	~109 kbps
ADPCM (uniform)	Very low SegSNR on some signals	0.99	3.80	~103 kbps

Key Findings

- μ-law PCM nearly doubles the segmental SNR compared to uniform PCM.
- μ-law maintains speech intelligibility and perceptual clarity even at 8 bits.
- Uniform PCM compresses better (lower entropy bitrate) but sacrifices quality.
- ADPCM requires adaptive step-size to remain stable on all types of audio.

Conclusion: μ-law presents the best trade-off between low complexity and perceptual quality.

NEURAL CODEC CONTEXT (AGC)

Neural Codec AGC Comparison

- The Audiogen Codec (AGC) represents audio in a latent space, not as raw samples.
- Compresses audio significantly more than methods that quantize waveforms.

Observed Compression (from experiments)

- 10-second stereo clip:
  - Raw waveform bitrate: ≈ 293.6 kbps
  - AGC latent bitrate: ≈ 58.7 kbps
  - → ~5× smaller while keeping high perceptual quality

Key Insight

- μ-law / PCM / ADPCM → encode waveform samples directly
- Neural Codecs → Learn and encode high level audio structure

Important Observation

- Applying μ-law before neural encoding does NOT reduce latent size
    - Improvements must come from the latent model; not waveform pre-processing. Neural codecs have better bitrate efficiency and quality compared with classical methods. but require more computation and are not suitable for low-power systems.

- Neural codecs outperform classical methods in bitrate efficiency and quality,but require more compute and are unsuitable for low-power systems.

DISCUSSION

- μ-law companding works best in low-bit, low-compute environments. embedded MCUs, telephony, IoT audio gateways.

- Offers better intelligibility and perceptual quality compared to Uniform PCM at 8 bits.

- While it is more efficient to encode residual error, without adaptivity, ADPCM can be unstable for signals with rapidly changing amplitudes.

- Neural codecs achieve much lower bitrate for higher quality: AGC, EnCodec. But require GPU/accelerated compute and higher latency budgets.

LIMITATIONS

- ADPCM used in experiments is non-adaptive.
  - does not include any step-size tuning, which causes unstable SegSNR results.
- Dataset consists mainly of speech; not music or environmental audio.
- Entropy bitrate is estimated, not fully entropy-coded.
- No subjective listening tests (MOS/MUSHRA/POLQA pending).
- Trade-off summary: μ-law = simplicity + quality Neural = highest quality with the lowest bitrate, yet expensive.

CONCLUSION

- μ-law companding remains highly effective for low-bit-rate speech coding.
- At 8-bit depth, μ-law PCM achieves:
  - Near-transparent intelligibility (STOI ≈ 1.0)
  - Higher perceptual quality (PESQ ≈ 4.2)
  - ~2× better segmental SNR vs. uniform PCM
- Uniform PCM provides lower entropy bitrate but results in audible quality loss.
- ADPCM reduces redundancy through residual coding, but the non-adaptive version is unstable without step-size control.
- Neural codecs (AGC, EnCodec, SoundStream):
  - Achieve much lower bitrates with high fidelity
  - Not suitable for ultra-low-latency or MCU-level compute environments

FUTURE ENHANCEMENTS

- Implement adaptive ADPCM (IMA / G.726)
    - improves stability and prediction accuracy
- Integrate LPC (Linear Predictive Coding)
    - better speech modeling before quantization
- Apply entropy coding (Huffman / Arithmetic)
    - convert estimated entropy bitrate into actual compressed output
- Expand dataset to include:
    - Music
    - Environmental audio
    - Mixed speech + background noise
- Conduct subjective listening tests (MUSHRA / POLQA / MOS)
    - validate perceptual quality beyond objective metrics
- Explore hybrid models
    - μ-law + neural latent compression

THANK YOU