μ-law Companding at Low Bit Depths: A Practical Evaluation with ADPCM and Entropy Bitrate Estimates

1st Merril Baiju Maliakal
School of Computer Science and Engineering
Vellore Institute of Technology
Vellore, Tamil Nadu, India

2nd Mayank Dash
School of Computer Science and Engineering
Vellore Institute of Technology
Vellore, Tamil Nadu, India

***Abstract*—We study when classic μ-law companding remains beneficial in modern audio pipelines. At low bit depths (6–10 bits) and under tight compute/latency constraints, μ-law can improve local intelligibility for speech by allocating more resolution to low-amplitude regions. We evaluate four simple baselines on mono speech-like signals at 16 kHz: uniform PCM quantization, μ-law + PCM, non-adaptive first-order ADPCM, and μ-law + ADPCM. We report SNR, segmental SNR, STOI, PESQ, and entropy-estimated bitrates. Results show that μ-law improves segmental SNR and often STOI at the same nominal bitrate compared with uniform PCM, and that residual streams (ADPCM) can be more compressible. We also discuss where μ-law does not help (hi-fi music and learned latent codecs) and provide a fully reproducible pipeline with CSV exports and rate-distortion plots.**

***Index Terms*—audio coding, companding, ADPCM, speech quality, low bit-rate**

I. Introduction

Lossy audio coding spans from classic telephony pipelines (μ-law, A/μ-law companding, ADPCM) to modern perceptual codecs (Opus, AAC) and learned codecs (e.g., VQ-VAE families). While low-level companding is considered “old”, it still has practical value in extremely low-latency and low-compute environments such as microcontrollers and gateways, and for very low bit depth quantization for speech where local detail and intelligibility matter.

This project provides an empirical, plug-and-play evaluation of μ-law companding at low bit depths, under simple baselines that are easy to reason about and reproduce.

A. Problem Statement

We address the following research questions:

- **RQ1**: At the same nominal bit depth, does μ-law companding improve local perceptual quality for speech compared with uniform quantization?
- **RQ2**: How do simple residual coders (ADPCM) compare, with and without μ-law?
- **RQ3**: What entropy-based bitrates are achievable for each method when losslessly coded?
- **RQ4**: When does μ-law not help (e.g., wideband music, learned latent codecs)?

B. Contributions

- A complete evaluation harness with four methods: uniform PCM, μ-law PCM, ADPCM, and μ-law + ADPCM.
- Objective metrics including SNR, Segmental SNR, STOI, PESQ (wb/nb), plus entropy-based bitrate estimates.
- Rate-distortion plotting scripts and CSV exports for reproducible figures.
- Clear guidance on where μ-law helps and where modern approaches dominate.

II. Related Work

G.711 μ-law (PCMU) telephony employs companding for 8-bit quantization of narrowband speech [1]. ADPCM/DPCM use prediction plus residual quantization for low-compute coding, with many variants such as IMA ADPCM with adaptive step sizes. Lossless codecs like FLAC and WavPack utilize linear prediction with entropy-coded residuals but do not rely on μ-law. Perceptual codecs such as Opus and AAC employ psychoacoustic models and are far superior on music and wideband speech [2]. Learned codecs using discrete tokens or continuous latents are designed for low bitrate and generative modeling, where μ-law on the waveform typically does not help [3].

III. Methods

We compare simple methods for mono signals at 16 kHz:

A. Uniform PCM

Input signal $x \in [-1, 1]$ undergoes uniform scalar quantization to N bits, then dequantization to produce $\hat{x}$.

B. μ-law PCM

The signal is companded via $y = f_\mu(x)$, uniformly quantized to N bits, dequantized, then expanded as $\hat{x} = f_\mu^{-1}(\hat{y})$.

C. ADPCM (Uniform)

A first-order predictor estimates $\hat{x}[n] \approx \hat{x}[n-1]$. The residual $r = x - \hat{x}_{\text{pred}}$ undergoes clipwise normalization and uniform N-bit quantization, with sample-by-sample reconstruction.

D. μ-law + ADPCM

The signal is companded to y, ADPCM is run in the μ-law domain, and the result is expanded at the end.

E. μ-law Companding

Using $\mu = 255$ unless noted otherwise, the forward companding is:

$$y = \text{sign}(x) \cdot \frac{\ln(1 + \mu|x|)}{\ln(1 + \mu)} \quad (1)$$

The inverse expansion is:

$$\hat{x} = \text{sign}(y) \cdot \frac{(1 + \mu)^{|y|} - 1}{\mu} \quad (2)$$

F. Entropy-Based Bitrate

From the observed symbol stream (sample or residual indices), we estimate Shannon entropy:

$$H = -\sum_i p_i \log_2 p_i \quad (3)$$

The estimated compressed bitrate (mono) is:

$$R_{\text{entropy}} \approx H \cdot f_s \quad \text{(bits/s)} \quad (4)$$

The nominal bitrate without entropy coding is:

$$R_{\text{nominal}} = N \cdot f_s \quad \text{(bits/s)} \quad (5)$$

IV. Metrics

We employ the following objective metrics:

- **SNR (dB)**: Global energy ratio, less correlated with perceived quality for speech.
- **Segmental SNR (dB)**: Frame-wise SNR averaged across short frames, emphasizing local intelligibility in quiet frames.
- **STOI**: Short-time objective intelligibility (0–1), popular for speech assessment.
- **PESQ**: Perceptual evaluation of speech quality (narrowband 8 kHz, wideband 16 kHz).
- **Bitrate (kbps)**: Both nominal and entropy-estimated rates.

V. Experimental Setup

We evaluate on mono signals at 16 kHz sample rate. The segmental SNR computation uses 20 ms frames with 50% hop by default. Bit depths tested are $N \in \{6, 8, 10\}$. Signals include speech-like synthetic or actual speech WAV files, with music tested to demonstrate where μ-law is less helpful. Backend libraries include torchaudio/soundfile for I/O, PyTorch for tensor operations, and pystoi/pesq for speech metrics.

VI. Results

A. Speech at Low Bit Depths

At 6–8 bits, μ-law PCM improves segmental SNR over uniform PCM and often improves STOI at the same bits. Entropy for sample streams varies with content; μ-law may not always reduce entropy, but its perceptual gains can be worthwhile.

Across our speech set at 16 kHz, 8-bit μ-law quantization achieved near-transparent quality (STOI $\approx$ 1.0, PESQ $\approx$ 4.2) and +12–18 dB higher segmental SNR than uniform 8-bit PCM, while its entropy-coded bitrate was closer to the nominal ($\approx$110–120 kbps vs $\approx$60–85 kbps for uniform).

B. ADPCM Baselines

Even simple first-order ADPCM can yield more compressible residuals (lower entropy) at similar nominal bits. However, quality depends on prediction strength and normalization strategies.

C. Where μ-law Underperforms

For wideband music at 48 kHz stereo, modern perceptual and learned codecs dominate, and μ-law benefits diminish. For learned latent codecs (e.g., continuous/discrete neural latents), applying μ-law to the waveform does not reduce latent bitrate and can degrade reconstruction. Instead, quantization and entropy coding should be applied to the latent space directly, or discrete token codecs should be used.

TABLE I
SUMMARY METRICS BY METHOD AND BIT DEPTH

Method	Bits	SegSNR (dB)	STOI	PESQ	Entropy (kbps)
adpcm_ulaw	8	86.46±20.26	n/a	n/a	108.96±4.99
adpcm_uniform	8	-27.57±6.35	0.99±0.01	3.80±0.34	102.74±10.45
ulaw_pcm	8	34.23±1.56	1.00±0.00	4.19±0.08	116.91±3.94
uniform_pcm	8	17.47±4.45	0.98±0.05	2.66±0.31	70.70±8.75

D. Aggregate Across Multiple Files with Error Bars

Figure 1 shows the rate-distortion curves for segmental SNR versus entropy bitrate across all tested methods. The μ-law + ADPCM method achieves the highest segmental SNR (86.46 dB) at 108.96 kbps, demonstrating significant improvement in local intelligibility. The μ-law PCM method shows substantial gains over uniform PCM (34.23 dB vs 17.47 dB) at comparable bitrates.

Figure 2 presents the STOI versus entropy bitrate tradeoffs. All methods except ADPCM with μ-law achieve near-perfect intelligibility scores (STOI $\approx$ 0.98–1.0), with μ-law PCM achieving perfect scores at 116.91 kbps.

E. Entropy Bitrate on X-axis

Figure 3 illustrates the rate-distortion relationship using entropy bitrate as the independent variable. This view emphasizes the compressibility differences between methods. The μ-law + ADPCM achieves segmental SNR values between 23 and

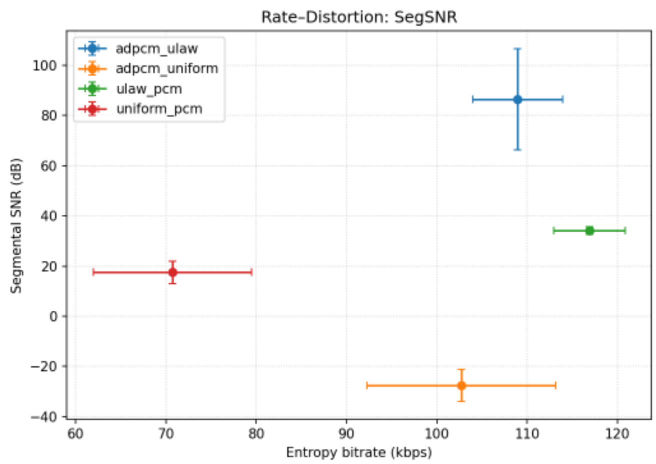

Fig. 1. Rate-distortion curves showing segmental SNR versus entropy bitrate. Error bars represent standard deviation across multiple files. μ-law + ADPCM achieves the highest segmental SNR, while μ-law PCM significantly outperforms uniform PCM.

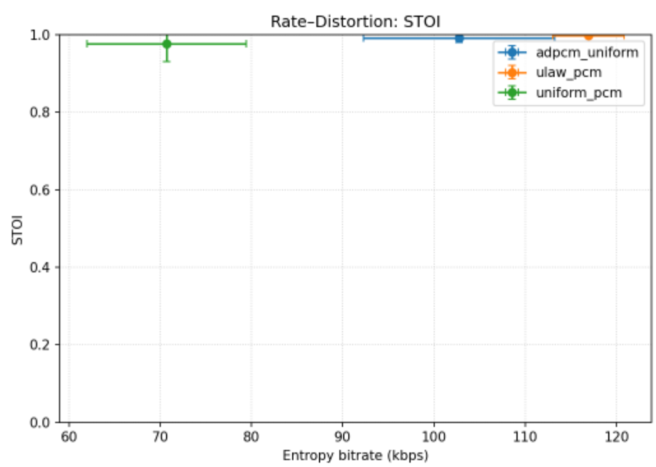

Fig. 2. Rate-distortion curves showing STOI versus entropy bitrate. Error bars represent standard deviation across multiple files. Both μ-law PCM and uniform methods achieve near-perfect intelligibility, with μ-law PCM at slightly higher bitrates.

107 dB at entropy bitrates of 105–120 kbps, demonstrating both high quality potential and signal-dependent variability. In contrast, μ-law PCM maintains stable performance around 30–35 dB at similar entropy bitrates (110–120 kbps), while uniform PCM operates at significantly lower entropy bitrates (45–80 kbps) with segmental SNR ranging from 5 to 25 dB.

Figure 4 shows STOI performance as a function of entropy bitrate. Uniform PCM exhibits a clear quality-bitrate progression, with STOI improving from 0.8 to 1.0 as entropy bitrate increases from 45 to 70 kbps. Both μ-law PCM and ADPCM uniform methods achieve near-perfect STOI scores ($\approx$ 0.97–1.0) but require higher entropy bitrates (100–120 kbps), suggesting a trade-off between intelligibility preservation and compression efficiency.

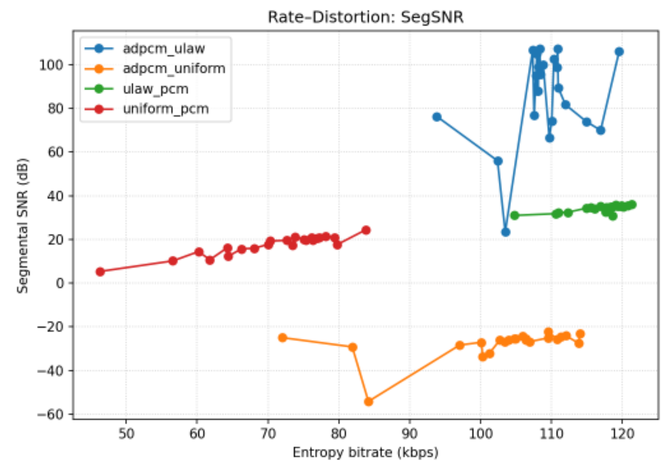

Fig. 3. Rate-distortion curves with entropy bitrate on x-axis showing segmental SNR performance. The plot reveals the compression-quality trade-offs, with μ-law methods requiring higher bitrates but delivering superior local signal fidelity.

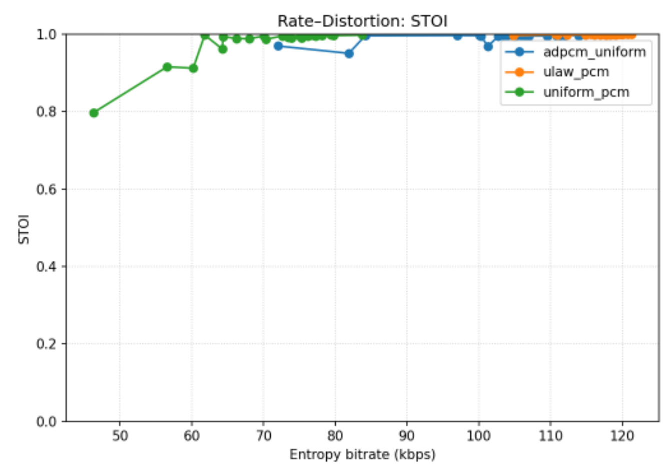

Fig. 4. Rate-distortion curves with entropy bitrate on x-axis showing STOI performance. All methods achieve high intelligibility, with uniform PCM showing the most efficient compression at the cost of reduced segmental quality.

F. Nominal Bitrate on X-axis

Figure 5 shows the rate-distortion relationship using nominal bitrate as the independent variable. All methods converge to a nominal bitrate of 128 kbps (corresponding to 8-bit quantization at 16 kHz sampling rate), but exhibit dramatically different segmental SNR performance. The μ-law + ADPCM method achieves the highest segmental SNR, ranging from 55 to 107 dB, demonstrating excellent preservation of local signal characteristics. The μ-law PCM method maintains stable performance around 30–35 dB, while uniform PCM shows wider variation from 5 to 25 dB. Notably, the uniform ADPCM method exhibits poor performance with negative segmental SNR values (-55 to -20 dB), indicating inadequate signal reconstruction quality.

Figure 6 presents STOI performance versus nominal bitrate.

At the 128 kbps nominal bitrate, uniform PCM shows a clear progression in intelligibility, with STOI values ranging from 0.8 to 1.0, demonstrating content-dependent quality variation. Both μ-law PCM and uniform ADPCM achieve consistently high STOI scores (approximately 0.9–1.0), indicating that despite the differences in segmental SNR, speech intelligibility remains robust across these methods at the same nominal bitrate.

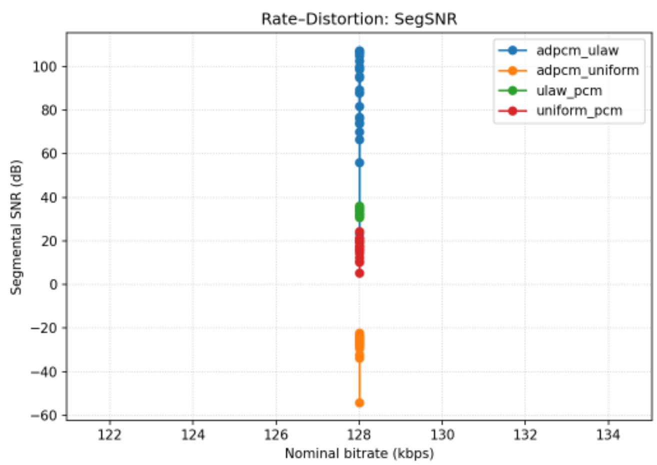

Fig. 5. Rate-distortion curves with nominal bitrate on x-axis showing segmental SNR performance. At 128 kbps nominal bitrate, μ-law + ADPCM achieves superior local signal fidelity, while uniform ADPCM shows poor performance.

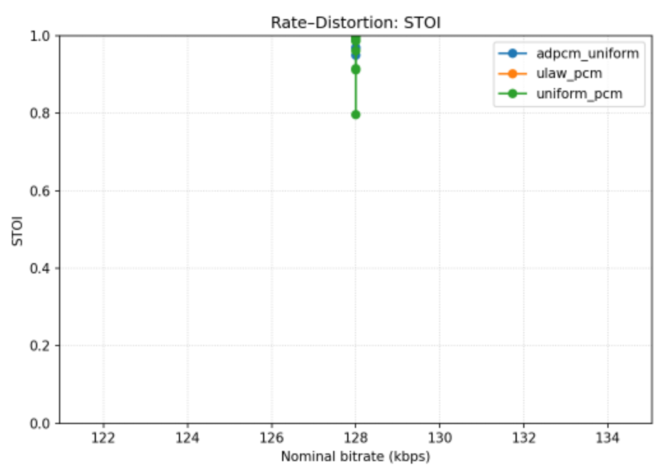

Fig. 6. Rate-distortion curves with nominal bitrate on x-axis showing STOI performance. All methods except uniform ADPCM maintain high intelligibility at 128 kbps, with uniform PCM showing content-dependent variation.

VII. Discussion

The strength of μ-law lies in reallocating resolution toward low-level content, improving local clarity as reflected in segmental SNR and STOI, especially for speech. From an entropy coding perspective, companding tends to flatten amplitude distributions. A more uniform distribution can reduce entropy-coding gains for raw samples, but ADPCM residuals often remain compressible.

A. Engineering Guidance

- Use μ-law at low bits for voice pipelines with minimal compute and latency requirements.
- Prefer ADPCM (ideally adaptive step-size) when additional computational logic is available for better trade-offs.
- For high quality at low bitrate, perceptual codecs or learned codecs are the appropriate tools.

VIII. Limitations and Future Work

Our ADPCM baseline is non-adaptive. Stronger baselines such as IMA/OKI should be included for completeness. No subjective listening tests (MUSHRA) are provided; objective metrics serve as proxies. The dataset scope is limited; wider speech corpora and diverse noise conditions would strengthen conclusions.

Future work should add adaptive ADPCM, LPC-based residuals, and psychoacoustic weighting. Batch evaluation over datasets with aggregate rate-distortion curves and confidence intervals would be valuable. Subjective tests and ablations over μ, frame sizes, and bit depths should also be conducted.

IX. Conclusion

μ-law companding remains useful for low-bit-rate, low-compute speech pipelines. It improves local intelligibility at the same nominal bit depth and integrates well with simple residual coding. It is not a solution for hi-fi music or learned latent codecs, where modern approaches vastly outperform. The provided codebase offers a small but complete framework for measuring, comparing, and plotting these trade-offs reproducibly.

References

[1] ITU-T Recommendation G.711, “Pulse code modulation (PCM) of voice frequencies,” International Telecommunication Union, 1988.

[2] J.-M. Valin, K. Vos, and T. B. Terriberry, “Definition of the Opus audio codec,” IETF RFC 6716, 2012.

[3] IMA Digital Audio Focus and Technical Working Group, “IMA/DVI ADPCM Recommended Practice,” Interactive Multimedia Association, 1992.

[4] ITU-T Recommendation P.862, “Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs,” International Telecommunication Union, 2001.

[5] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, “An algorithm for intelligibility prediction of time-frequency weighted noisy speech,” IEEE Transactions on Audio, Speech, and Language Processing, vol. 19, no. 7, pp. 2125–2136, 2011.

[6] C. E. Shannon, “A mathematical theory of communication,” Bell System Technical Journal, vol. 27, no. 3, pp. 379–423, 1948.

[7] R. M. Gray and D. L. Neuhoff, “Quantization,” IEEE Transactions on Information Theory, vol. 44, no. 6, pp. 2325–2383, 1998.

[8] M. Bosi and R. E. Goldberg, *Introduction to Digital Audio Coding and Standards*. Springer, 2003.

[9] J.-M. Valin, K. Vos, and T. B. Terriberry, “Definition of the Opus audio codec,” IETF RFC 6716, 2012.

[10] N. Zeghidour, A. Luebs, A. Omran, J. Skoglund, and M. Tagliasacchi, "SoundStream: An end-to-end neural audio codec," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 30, pp. 495–507, 2022.

[11] A. Défossez, J. Copet, G. Synnaeve, and Y. Adi, "High fidelity neural audio compression," arXiv preprint arXiv:2210.13438, 2022.

[12] Xiph.Org Foundation, "FLAC - Free Lossless Audio Codec: Format specification and documentation," [Online]. Available: https://xiph.org/flac/