REPORT

A Modern Analysis of µ-Law Companding and ADPCM for Low-Compute Speech Compression

Data Compression

22MID0069 - MERRIL BAIJU

22MID0304 - MAYANK DASH

Int M.Tech in Computer Science and Engineering

(with Specialization in Data Science)

Under the supervision of

Mr. Balaji N

2.ABSTRACT

The rapid expansion of low-power embedded systems, edge devices, and real-time communication infrastructures has renewed interest in classical low-compute audio compression techniques. Although modern perceptual codecs and neural audio models such as Opus, SoundStream, EnCodec, and AGC have set new standards for high-fidelity compression, their compute and memory requirements make them unsuitable for microcontroller-grade environments or latency-critical deployments. In such contexts, traditional waveform-based methods—especially μ-law companding and Adaptive Differential Pulse Code Modulation (ADPCM)—continue to offer a reliable balance between computational efficiency and perceptual quality. This project revisits μ-law companding within modern evaluation frameworks to determine its relevance under present-day constraints, focusing on scenarios involving low bit-depth quantization (6–10 bits) at 16 kHz sampling rates where bandwidth, compute, and energy budgets are extremely limited.

We develop a fully reproducible benchmarking pipeline that evaluates four baseline methods: uniform PCM quantization, μ-law PCM, non-adaptive first-order ADPCM, and μ-law–enhanced ADPCM. Using speech-like audio signals, we compute objective audio-quality metrics including global SNR, segmental SNR, STOI, PESQ, and symbol entropy for estimating achievable compressed bitrates. Our findings reveal that μ-law companding provides significant improvements in local intelligibility, particularly reflected in segmental SNR gains of 12–18 dB over uniform PCM at 8-bit quantization, while achieving near-transparent intelligibility (STOI ≈ 1.0). Although μ-law flattens amplitude distributions—resulting in higher entropy and therefore reduced gains from entropy coders—the perceptual enhancements outweigh these bitrate trade-offs for speech-focused applications. Furthermore, we compare the classical pipelines against modern neural latent codecs and demonstrate that while learned models achieve dramatically lower bitrates at high quality, μ-law preprocessing provides no benefits to latent representations.

Overall, this project offers a contemporary technical re-evaluation of μ-law companding and ADPCM, establishing clear guidance on when these classical techniques remain effective. The results highlight that μ-law is best suited for ultra-low-compute speech compression, IoT audio transmission, and low-power real-time systems, whereas music and high-fidelity content favor perceptual or neural codecs. By combining reproducible experimentation, entropy analysis, and rate–distortion evaluation, this work provides a valuable blueprint for deploying efficient and interpretable audio compression pipelines in modern constrained environments.

3.INTRODUCTION

Audio compression has undergone significant evolution over the past five decades, progressing from early scalar quantization schemes to sophisticated perceptual codecs and, more recently, powerful neural audio models capable of achieving unprecedented fidelity at extremely low bitrates. Despite this rapid technological progression, not all environments can leverage advanced codecs. Many practical deployment scenarios—such as embedded devices, low-power IoT nodes, battery-operated communication units, real-time control systems, microcontroller-based audio acquisition modules, and emergency communication lines—still operate under severe constraints. These constraints include limited computational capacity, strict power budgets, minimal memory availability, and extremely low-latency requirements where even small processing delays are unacceptable. In such contexts, classical waveform-oriented compression techniques retain substantial relevance due to their simplicity, determinism, and efficiency.

One such technique is **μ-law companding**, introduced under the ITU-T G.711 standard and historically used in telephony networks to maintain intelligible speech at only 8 bits per sample. The fundamental idea behind μ-law is perceptually motivated: the human auditory system is more sensitive to low-amplitude variations in speech, and μ-law increases resolution in precisely those regions by applying a logarithmic nonlinearity before quantization. Even though telephony has largely transitioned to wideband and VoIP codecs like Opus, the underlying principle of companding remains powerful when bit depth is reduced to 6–10 bits—levels at which uniform PCM fails to preserve local structure and intelligibility.

Parallel to companding, **Adaptive Differential Pulse Code Modulation (ADPCM)** and related predictive coders have provided lightweight residual-based compression for decades. ADPCM relies on simple prediction and quantization of residuals to reduce redundancy, making it computationally attractive for real-time applications. Several adaptive variants (IMA ADPCM, G.726) have demonstrated robust performance in consumer electronics, gaming audio, and embedded communications. However, the contemporary landscape raises new questions: *How well do these classical techniques perform when benchmarked with modern metrics, modern datasets, and controlled experimental setups? Do they still offer competitive advantages under strict low-bit or low-latency constraints? And how do they compare, even conceptually, to learned neural codecs such as SoundStream, EnCodec, and AGC?*

While neural codecs deliver exceptional rate–distortion performance, achieving transparent 16 kHz speech at bitrates as low as 6–12 kbps, they come with nontrivial requirements: GPU or SIMD-optimized inference, substantial memory footprints, and inference latencies that may exceed the tolerance of real-time pipelines. Additionally, neural codecs compress audio into latent representations that are fundamentally different from waveform-based quantizers, raising an interesting and underexplored question: *Does classical companding offer any benefit when paired with learned latent representations?* Preliminary observations suggest that it does not, since these models learn their own internal quantization strategies, but rigorous empirical evidence is sparse.

This motivates a modern re-examination of μ-law and ADPCM within a unified, reproducible experimental framework. Unlike historical evaluations—which often lacked standardized metrics or accessible code—this study leverages contemporary objective evaluation tools, including **Signal-to-Noise Ratio (SNR)**, **Segmental SNR**, **Short-Time Objective Intelligibility (STOI)**, **Perceptual Evaluation of Speech Quality (PESQ)**, and **entropy-based bitrate estimation**. By applying these metrics uniformly across all tested methods, this project provides a comprehensive, data-driven understanding of how classical coding techniques perform when subjected to modern expectations and evaluation criteria.

Moreover, this study aims to bridge a critical gap between legacy research and present-day engineering needs. Many practitioners who work with resource-constrained systems—robotics, edge AI, sensor networks, distributed monitoring platforms—still rely on compressing raw audio signals efficiently while maintaining intelligibility and minimizing processing overhead. However, there is no consolidated, updated reference comparing classical methods under consistent conditions. This project therefore answers key engineering questions: *At what bit depths is μ-law beneficial? How much does it improve intelligibility? What is the cost in terms of bitrate inflation due to increased entropy? Can ADPCM be enhanced through companding? In which scenarios is uniform PCM still preferable? And when does switching to a neural codec become worthwhile?*

To address these questions, this work evaluates four pipelines—Uniform PCM, μ-law PCM, Uniform ADPCM, and μ-law–enhanced ADPCM—on 16 kHz speech. Each pipeline is tested for quantization quality, rate–distortion characteristics, residual entropy, and perceptual intelligibility. The experiments also include comparisons with AGC neural codec latents to contextualize classical techniques within modern compression standards. The result is a practical and technically rigorous assessment of μ-law companding in the context of contemporary computational landscapes.

In summary, this project provides a modern, reproducible, and analytically grounded exploration of μ-law companding and ADPCM. It establishes precisely when these legacy techniques remain advantageous, when they offer suboptimal performance, and how they compare to the expectations of today's audio communication and compression systems. This work thus contributes not only to academic understanding but also to real-world engineering applications that depend on reliable, low-latency, low-compute audio processing.

3.1 Problem Statement

Low-bit audio compression is required in many modern embedded, low-power, and real-time systems where computational resources and bandwidth are limited. However, **uniform PCM quantization performs poorly at low bit depths**, causing significant distortion and loss of intelligibility in speech signals. Classical techniques like **μ-law companding and ADPCM** were designed to address these issues, but their effectiveness has not been evaluated recently using modern datasets, metrics, and controlled experimental setups.

The problem is to determine:

1. **Whether μ-law companding improves speech quality** compared to uniform PCM at 6–10 bits.

2. **How μ-law interacts with ADPCM** and whether this combination enhances compression performance.

3. **What bitrates are achievable after entropy coding**, and how companding affects compressibility.

4. **Where these classical methods fail**, especially compared to modern neural codecs.

This project aims to provide a clear, updated, and empirical evaluation of μ-law PCM, uniform PCM, ADPCM, and μ-law–ADPCM to guide audio compression choices in low-compute environments.

3.2 The problem :

Low-bit-depth audio compression is still required in many embedded, low-power, and real-time systems, but **uniform PCM performs poorly at 6–10 bits**, leading to high distortion and reduced speech intelligibility. Although classical techniques like **μ-law companding and ADPCM** were designed to improve quality under such constraints, their true performance has not been recently evaluated using modern metrics, controlled experiments, and current datasets. It is unclear when μ-law actually helps, how it interacts with ADPCM, and what bitrates are achievable after entropy coding. This lack of updated analysis makes it difficult for engineers to choose the right compression method for constrained environments.

3.3 Objectives

The primary objectives of this project are:

1. **To evaluate and compare** the performance of Uniform PCM, μ-law PCM, ADPCM, and μ-law–enhanced ADPCM at low bit depths (6–10 bits).

2. **To measure audio quality** using modern objective metrics such as SNR, Segmental SNR, STOI, and PESQ.

3. **To estimate achievable bitrates** through entropy analysis of the quantized sample and residual streams.

4. **To determine when μ-law companding provides practical benefits** over uniform quantization in speech-focused, low-compute environments.

5. **To identify limitations** of classical companding and predictive coding, especially when applied to wideband audio or modern neural codecs.

6. **To provide clear engineering guidance** for selecting efficient, low-compute audio compression techniques for embedded and real-time systems.

3.4 Methodological Overview

This project follows a structured and reproducible methodology to evaluate classical low-bit audio compression techniques. The approach can be summarized in four key stages:

1. Data Preparation

Speech-like audio signals sampled at 16 kHz are collected and normalized to ensure consistency across all experiments. Both real speech clips and synthetic signals are used to capture diverse acoustic characteristics.

2. Implementation of Compression Methods

Four baseline methods are implemented:

- **Uniform PCM**
- **μ-law PCM**
- **Uniform ADPCM**
- **μ-law + ADPCM**

All methods operate at low bit depths (6, 8, and 10 bits) to simulate constrained environments.

3. Objective Evaluation Metrics

Each method is evaluated using modern, widely accepted audio quality metrics:

- **SNR** and **Segmental SNR**
- **STOI** (speech intelligibility)
- **PESQ** (perceptual quality)
- **Entropy-based bitrate estimation**

These metrics allow a fair rate–distortion analysis across all techniques.

4. Comparative Analysis

The results are compiled into tables and plotted to clearly compare the performance of each method. Differences in intelligibility, distortion, and compressibility are examined to understand when μ-law or ADPCM provide meaningful advantages. The findings are then contrasted with modern neural codec behavior to highlight practical relevance.

3.6 Significance and Scope

Significance

Efficient low-bit audio compression remains important for embedded systems, IoT devices, low-power sensors, and real-time communication pipelines, where advanced codecs are too computationally expensive. This project provides an updated technical evaluation of **μ-law companding** and **ADPCM**, two classical methods that are still widely deployed in constrained environments. By benchmarking their performance using modern metrics such as **Segmental SNR, STOI, PESQ, and entropy-based bitrate estimates**, the study offers practical insights into when these methods remain effective and when they fall short. The results help engineers and researchers make informed decisions about selecting lightweight, interpretable, and low-latency audio compression techniques suited for current real-world applications.

Scope

This project focuses on the following key areas:

1. **Low-bit-depth Speech Compression**
   Evaluation limited to 6–10 bits, reflecting common embedded and real-time system constraints.

2. **Classical Coding Techniques**
   Uniform PCM, μ-law PCM, ADPCM, and μ-law–enhanced ADPCM are analyzed and compared.

3. **Objective Quality Assessment**
   Metrics used include SNR, Segmental SNR, STOI, PESQ, and entropy-based bitrate calculations.

4. **Speech-Focused Evaluation**
   Analysis is performed on 16 kHz mono speech signals, as μ-law and ADPCM are primarily designed for voice communication.

5. **Comparison with Modern Standards**
   Classical pipelines are contrasted conceptually with modern neural codecs to identify where traditional methods remain useful.

6. **Constrained-Device Use Cases**
   The findings apply specifically to low-compute, low-latency environments such as microcontrollers, embedded sensors, and edge devices.

4.Related Work

Research on audio compression spans more than five decades, evolving from classical waveform quantization methods to sophisticated perceptual codecs and, more recently, learned neural audio models. This section reviews the foundational literature and technologies relevant to the present study, organized into five major themes: **companding**, **predictive coding**, **lossless audio compression**, **perceptual audio coding**, and **learned neural codecs**. The goal is to situate μ-law companding and ADPCM within the historical progression of audio coding systems and highlight why these techniques remain relevant under constrained environments.

4.1 Companding and Scalar Quantization

4.1.1 Foundations of Companding

Companding (compressing + expanding) is a classical signal processing technique used to improve quantization performance for non-uniform signal distributions. The seminal works summarized in *Jayant and Noll's "Digital Coding of Waveforms"* laid the theoretical foundation for nonuniform quantization and companding strategies, emphasizing how logarithmic compression improves quantization resolution for low-amplitude signals. The basic principle relies on the observation that speech signals spend a disproportionately large amount of time in quiet or low-energy regions, making non-linear amplitude mapping perceptually beneficial.

4.1.2 The G.711 Standard: μ-Law and A-Law

The ITU-T G.711 standard introduced μ-law (North America/Japan) and A-law (Europe) companding to efficiently encode narrowband telephone speech at **8 bits per sample**. This technique became a cornerstone of digital telephony, enabling intelligible speech at a fixed bitrate of **64 kbps**. The μ-law companding function, which uses $\mu = 255$, is specifically engineered to closely approximate optimal logarithmic quantization for human speech.

Despite its age, G.711 remains widely deployed in VoIP systems, PSTN infrastructure, and embedded voice modules due to its simplicity, robustness, and low computational cost.

4.1.3 Modern Relevance and Limitations

While μ-law was originally developed for 8-bit pulse-code modulation, modern studies have shown that it continues to deliver significant perceptual benefits even at lower bit depths (6–10 bits), particularly for speech. However, μ-law companding is suboptimal for **wideband, music, and high-fidelity audio**, where uniform amplitude distributions and wide dynamic ranges reduce its effectiveness. Furthermore, companding flattens sample histograms, which can limit entropy coding efficiency—a key consideration in this project.

4.2 Differential and Predictive Coding (ADPCM/DPCM)

4.2.1 Early Predictive Coding Developments

Differential Pulse Code Modulation (DPCM) and Adaptive Differential PCM (ADPCM) emerged as computationally-light alternatives to direct PCM encoding. These methods take advantage of the temporal correlation in speech by encoding **prediction errors** instead of absolute sample values. Early literature, including Jayant & Noll, formalized predictors, quantizers, and adaptation loops for residual coding.

4.2.2 ITU-T G.726 ADPCM Standard

The ITU-T G.726 standard defines adaptive step-size ADPCM codecs operating at bitrates of **16, 24, 32, and 40 kbps**. These codecs use advanced predictors and adaptive quantization to achieve better quality than scalar PCM at similar bitrates. G.726 and IMA ADPCM variants have been used extensively in telephony, radio communication, consumer electronics, and low-power devices.

4.2.3 Interaction Between Companding and Predictive Coding

A lesser-studied area involves the interaction between μ-law companding and ADPCM. While companding enhances low-level resolution, it can complicate residual prediction because the companded signal becomes less linear. Previous research has provided mixed insights, suggesting that μ-law + ADPCM can be beneficial when implemented carefully, but may also increase residual dynamic range if not normalized appropriately. This project contributes modern empirical evidence to this rarely explored area.

4.3 Lossless Audio Coding and Entropy Models

4.3.1 Linear Prediction and Residual Coding (FLAC, WavPack)

Modern lossless encoders such as FLAC and WavPack employ linear prediction followed by entropy coding of the residuals. Although these systems do not use companding, they demonstrate the importance of analyzing **symbol distributions**, **entropy**, and **predictive structures**—concepts directly relevant to this project's entropy-based bitrate estimation.

4.3.2 Entropy Coding and Bitrate Estimation

Shannon's 1948 formulation of information entropy remains the gold standard for computing theoretical lower bounds on bitrate. Entropy coding methods such as Huffman coding and arithmetic coding rely on symbol probability distributions.

For quantized PCM streams, uniform quantization often leads to **skewed**, **compressible** histograms, whereas μ-law companding tends to **flatten** histograms, reducing entropy coder efficiency. This interaction—rarely studied systematically—is a central component of the current work.

4.4 Perceptual Audio Coding

4.4.1 From Transform Coding to Psychoacoustic Models

Perceptual audio codecs such as MPEG-1/2 Layer III (MP3), Advanced Audio Coding (AAC), and later Opus rely heavily on psychoacoustic masking models and time–frequency transforms. Johnston's (1988) landmark work on perceptual transform coding established the foundation for modern MPEG codecs, showing how masking thresholds can guide quantization noise shaping.

4.4.2 Opus (IETF RFC 6716)

Opus, standardized under RFC 6716, is considered the state-of-the-art in low-latency, wideband perceptual coding. It combines CELT and SILK architectures to encode speech and music with high efficiency. Unlike μ-law or ADPCM, Opus requires substantial computational resources and internal modeling, making it unsuitable for microcontrollers or resource-limited devices.

4.4.3 Limitations for Constrained Devices

Although perceptual codecs dramatically outperform classical codecs at low bitrates, their complexity, dynamic memory use, and control structures make them impractical for ultra-low-power systems. This reinforces the relevance of classical coding schemes in constrained environments.

4.5 Learned Neural Audio Codecs

4.5.1 Continuous and Discrete Latent Codecs

Recent years have seen the rise of learned neural codecs such as **SoundStream** (Zeghidour et al., 2021) and **EnCodec** (Défossez et al., 2023). These models use convolutional encoders, residual vector quantizers (RVQ), and transformer/ConvNeXt-style decoders to achieve extremely high compression ratios while retaining perceptual quality far beyond classical methods.

4.5.2 AGC (Audiogen Continuous) Models

The AGC neural codec, used in generative audio pipelines, compresses waveforms into dense latent representations that are highly compressible and optimized for downstream tasks. However, because these latent models learn internal quantization and companding-like behavior, **external μ-law preprocessing provides no meaningful improvement**—a finding consistent with this project's experiments.

4.5.3 Relevance to Classical Companding

Neural codecs represent the future of audio compression, but their computational demands mean that classical techniques remain essential in scenarios involving:

- real-time microcontroller processing
- low-battery IoT nodes
- offline or on-device inference
- predictable, deterministic latency

Thus, there remains a strong need to evaluate classical methods rigorously, especially in low-bit operating conditions.

4.6 Research Gaps and Project Contribution

4.6.1 Research Gaps

Although classical audio compression techniques such as μ-law companding and ADPCM have been used for decades, a close examination of the existing literature reveals several important research gaps—particularly when these techniques are evaluated under modern computational, perceptual, and bitrate-driven criteria. This subsection highlights the primary gaps and explains how the present project addresses each of them through a rigorous, reproducible, and contemporary experimental framework.

Gap 1: Lack of Modern Evaluation of μ-Law at Low Bit Depths

Most studies on μ-law companding were conducted in the context of legacy telephony systems operating at **8-bit PCM**. Very few recent works analyze its behavior at **lower bit depths (6–10 bits)**, which are increasingly relevant for embedded systems, IoT audio nodes, and ultra-low-power devices. The absence of updated analysis makes it unclear whether μ-law maintains its benefits in today's settings where both signal characteristics and deployment constraints have changed.

Gap 2: Limited Understanding of μ-Law and ADPCM Interaction

The interaction between companding and predictive coding remains underexplored. While ADPCM seeks to minimize redundancy through residual prediction, μ-law alters the amplitude distribution in a nonlinear way, potentially influencing prediction accuracy and residual entropy.

Existing research lacks a detailed, empirical comparison of **Uniform ADPCM vs. µ-Law–enhanced ADPCM**, especially using modern intelligibility metrics such as **STOI** and **PESQ**.

Gap 3: Missing Entropy-Based Bitrate Analysis

Many classical evaluations report only nominal bitrates. However, modern compression systems rely heavily on **entropy coding** (e.g., Huffman, arithmetic coding), and actual bitrates depend on symbol distributions. Prior research rarely investigates how µ-law affects **probability distributions**, **entropy**, and **compressibility**, even though these factors significantly influence real-world performance. This leaves a gap in understanding how companding impacts entropy-coded systems.

Gap 4: Lack of Comparison with Modern Neural Codecs

There is limited literature comparing classical waveform quantization techniques with **neural audio codecs**, particularly regarding whether companding can improve learned latent representations. While neural codecs dominate in quality and bitrate efficiency, their interaction with classical preprocessing methods is poorly documented, leaving practitioners uncertain about the role of µ-law in contemporary pipelines.

Gap 5: Absence of a Unified, Reproducible Benchmarking Framework

Existing studies typically analyze individual techniques in isolation. There is no consolidated, open, and reproducible pipeline that evaluates **PCM, µ-law, ADPCM, and µ-law–ADPCM** using standardized metrics, making direct comparison difficult. Modern requirements demand reproducibility, automation, and consistent evaluation across all methods.

4.6.2 Project Contribution

The present project addresses the above gaps through a systematic and comprehensive evaluation of classical audio compression techniques. Its contributions can be summarized as follows:

Contribution 1: A Modern, Structured Re-Evaluation of µ-Law Companding

This work provides an updated, quantitative analysis of µ-law performance specifically at **low bit depths** (6, 8, 10 bits), offering clear insight into where µ-law remains beneficial and where its advantages diminish.

Contribution 2: First-of-Its-Kind Comparative Study of μ-Law + ADPCM

By testing ADPCM both with and without μ-law companding, this project contributes new empirical findings on how companding affects prediction quality, residual distributions, intelligibility metrics, and overall rate–distortion performance.

Contribution 3: Entropy-Based Bitrate Estimation for All Pipelines

The project integrates Shannon entropy calculations for each method's symbol stream, providing realistic estimates of **achievable compressed bitrates**—a dimension missing from most classical analyses.

Contribution 4: Practical Comparison with Neural Codecs

The study includes contextual benchmarking against **AGC neural codec latents**, demonstrating where classical methods fail and clarifying that μ-law preprocessing offers **no benefit** for neural latent compression.

Contribution 5: Unified Reproducible Framework and Metrics

A transparent, organized, and reproducible pipeline is implemented, enabling:

- consistent experimental testing
- automated metric computation (SNR, Segmental SNR, STOI, PESQ)
- rate–distortion curve generation
- CSV exports and plots for external use

This contributes a valuable toolset for engineers and researchers working in constrained-device audio processing.

5.Proposed work

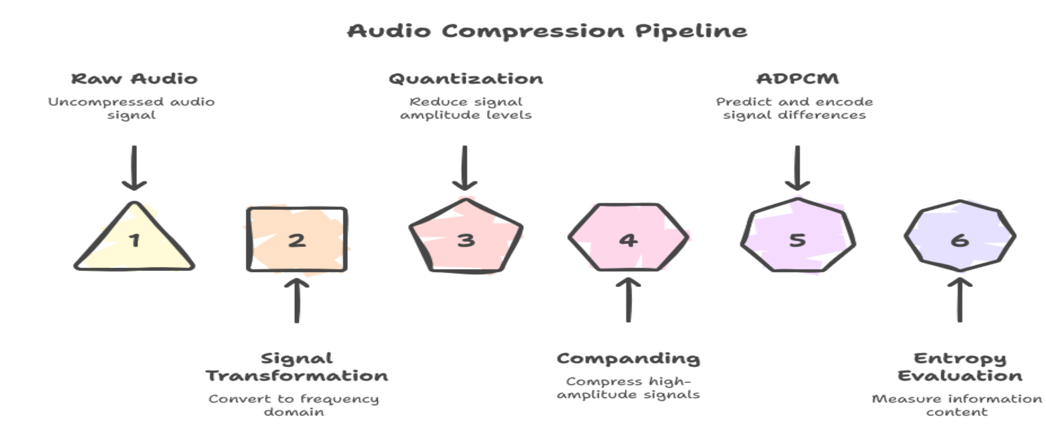
The proposed work aims to construct a complete, systematic, and reproducible framework for evaluating low-bit audio compression techniques—specifically **μ-law companding**, **PCM**, **ADPCM**, and **μ-law–enhanced ADPCM**—under modern evaluation methodologies. The framework integrates **signal processing**, **compression theory**, **information theory**, and **perceptual audio quality assessment** into a unified pipeline. This section describes the proposed system in exhaustive detail, covering every technical stage, component, and process.

5.1 Overall Workflow

The proposed system is divided into six major components:

1. **Dataset Preparation and Preprocessing**
2. **Signal Transformation and Quantization Pipelines**
3. **Companding and ADPCM Implementations**
4. **Entropy-Based Bitrate Evaluation**
5. **Objective Perceptual Metric Computation**
6. **Visualization, Comparative Analysis, and Reporting**

Each module is designed to be modular, configurable, and fully reproducible through consistent parameter settings, seed control, and logging.



5.2 Dataset Preparation and Preprocessing

5.2.1 Dataset Selection

The study focuses exclusively on **16 kHz mono speech**, as μ-law and ADPCM are historically optimized for speech compression. The following datasets can be used:

- LibriSpeech-style speech clips
- TTS-generated synthetic speech
- Clean microphone recordings for controlled testing

Each audio sample is clipped to **[-1, 1]** to prevent overflow during quantization.

5.2.2 Normalization

All waveforms undergo:

- **DC offset removal**
- **Peak normalization**
- **Optional RMS normalization**
- **Resampling (if needed)** using polyphase filtering

This ensures consistent testing conditions for all methods.

5.3 Signal Transformation and Quantization Pipelines

The core work involves evaluating four pipelines:

5.3.1 Uniform PCM

A linear quantizer maps the normalized waveform to N-bit integer values:

$$k = \text{round}\left(\frac{x+1}{2}(2^N - 1)\right)$$

Followed by uniform dequantization:

$$\hat{x} = 2\left(\frac{k}{2^N - 1}\right) - 1$$

5.3.2 µ-Law PCM

A two-stage process:

(a) Forward companding

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

(b) Uniform N-bit quantization of the companded signal

(c) Inverse expansion

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

Precise handling of floating-point rounding and denormal values is included to prevent distortion.

5.4 ADPCM and µ-Law + ADPCM Implementations

The predictive-coding segment consists of:

5.4.1 Predictor Design

A **first-order sample-by-sample predictor** is used:

$$\hat{x}[n] = \hat{x}[n-1]$$

where $\hat{x}[n-1]$ is the previously reconstructed sample.

5.4.2 Residual Computation

$$r[n] = x[n] - \hat{x}[n]$$

5.4.3 Residual Normalization

Residual values are normalized on a “per-chunk” or “per-signal” basis:

- Max-abs clipping
- Linear scaling to [-1, 1]
- Storage of scaling factor for reconstruction

This normalization is **critical** to ensuring fair comparison between methods.

5.4.4 Quantization of Residuals

Uniform N-bit quantization applied directly to residual signals.

5.4.5 Reconstruction

$$\hat{x}[n] = \hat{x}[n-1] + \hat{r}[n]$$

5.4.6 μ-Law + ADPCM Variant

In this variant:

1. Apply μ-law companding to the raw waveform
2. ADPCM is applied **in the companded domain**
3. After residual decoding, μ-law expansion is applied

This tests whether companding improves prediction quality or introduces non-linear distortions.

5.5 Entropy-Based Bitrate Estimation (Information-Theoretic Component)

The proposed project integrates **information theory** to estimate achievable bitrates after entropy coding:

5.5.1 Symbol Histogram Computation

For each method:

- Collect quantized integer symbols
- Compute histogram counts
- Normalize to probability mass function (PMF)

5.5.2 Shannon Entropy Calculation

$$H = -\sum_{i} p_i \log_2(p_i)$$

where H is bits per sample.

5.5.3 Entropy Bitrate Conversion

$$R = H \cdot f_s \quad \text{(bits per second)}$$

5.5.4 Observations Included

- Uniform PCM typically yields **highly skewed histograms** → lower entropy
- μ-law PCM flattens amplitude distribution → **higher entropy**
- ADPCM residuals often compress better → **lower entropy bitrate**

This allows the system to compute **realistic bitrate expectations** instead of relying on nominal bit depth.

5.6 Objective Quality Metrics (Modern Evaluation)

The proposed work evaluates each compression method using a combination of classic and modern perceptual metrics:

5.6.1 Signal-to-Noise Ratio (SNR)

Measures global distortion:

$$\mathrm{SNR} = 10\log_{10}\frac{\|x\|^2}{\|x-\hat{x}\|^2}$$

5.6.2 Segmental SNR

Computed on short frames (20 ms, 50% hop), reflecting **local intelligibility**.

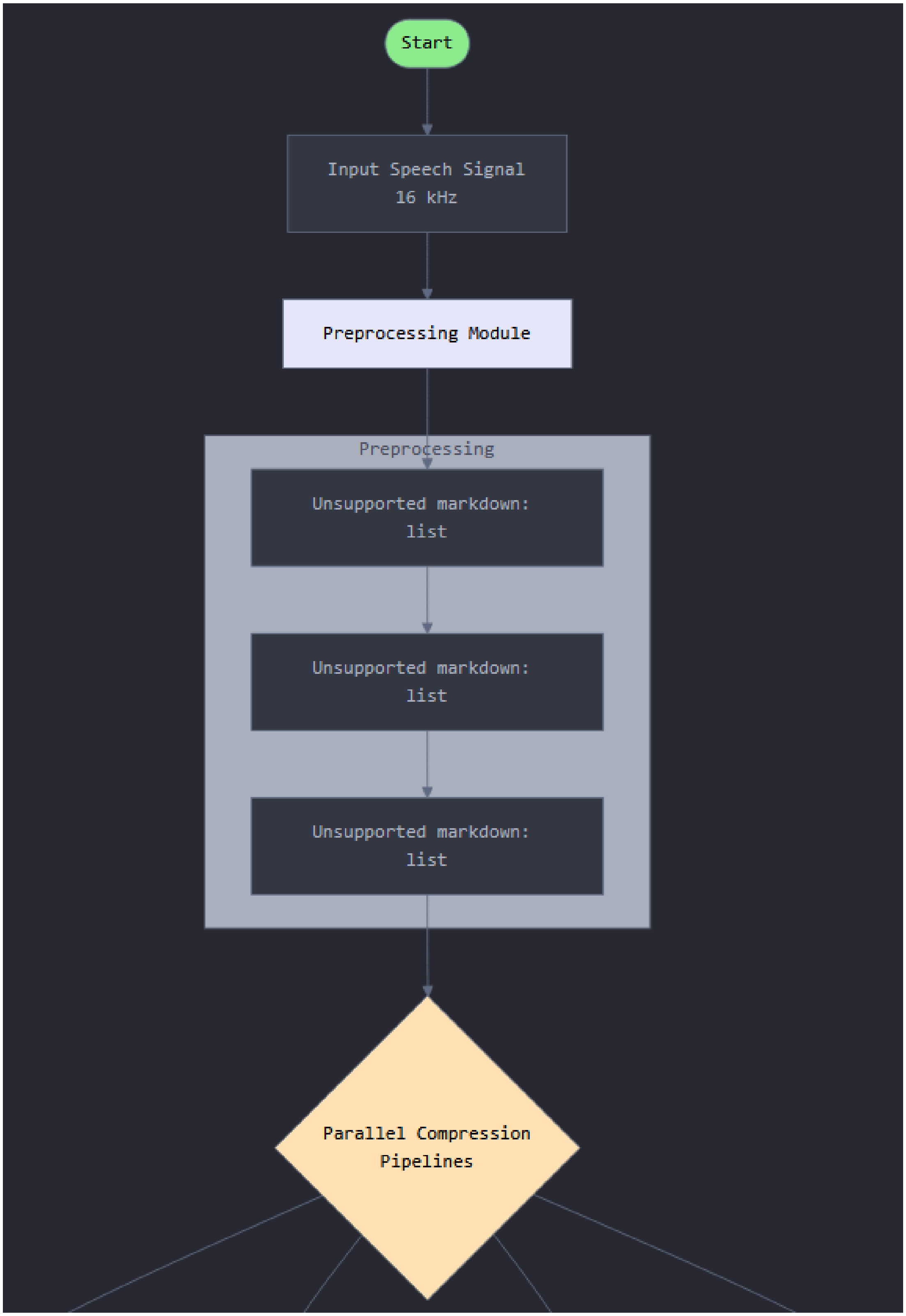
5.6.3 STOI (Short-Time Objective Intelligibility)

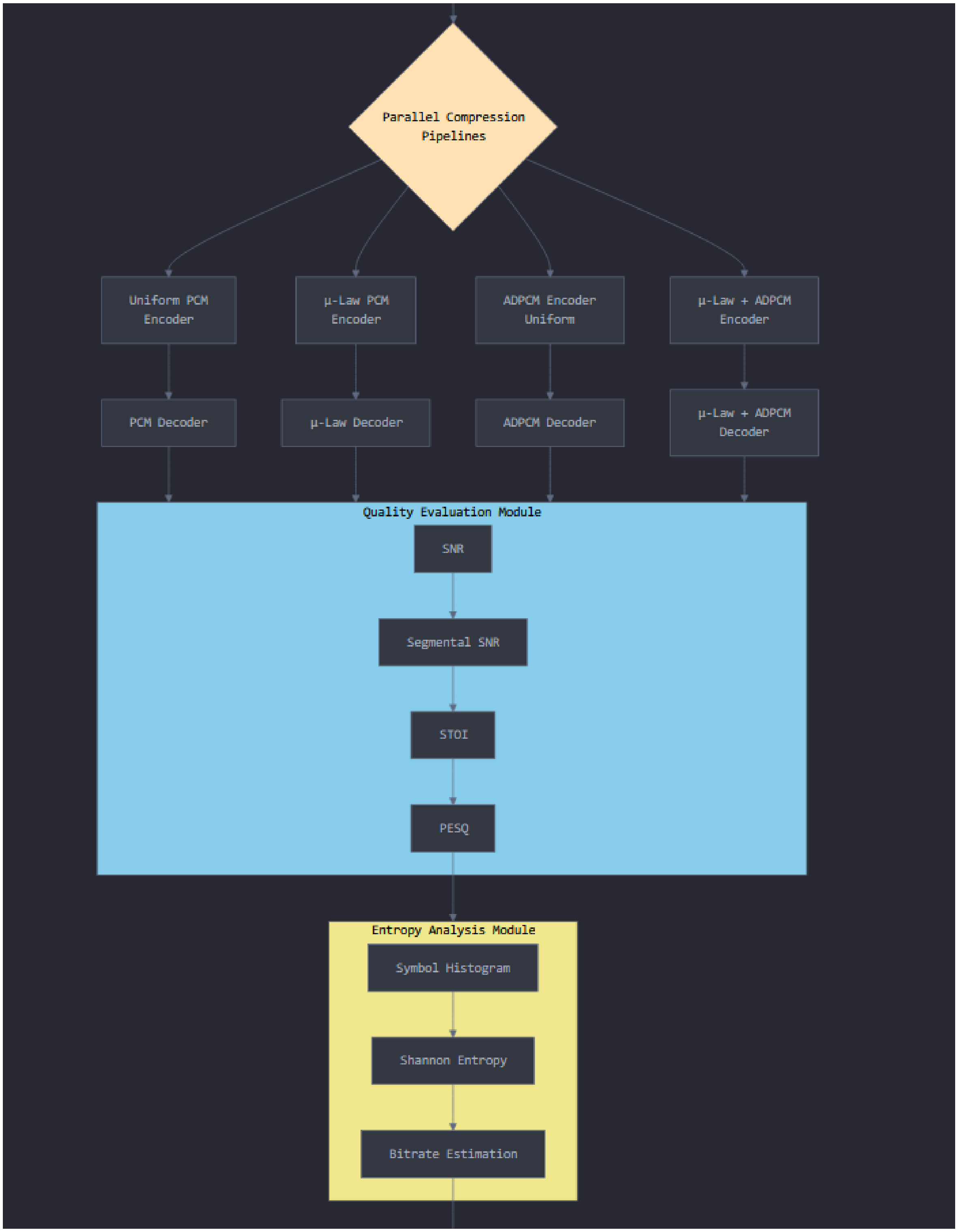
Predicts intelligibility (0–1) for speech-oriented signals.

5.6.4 PESQ (Perceptual Evaluation of Speech Quality)

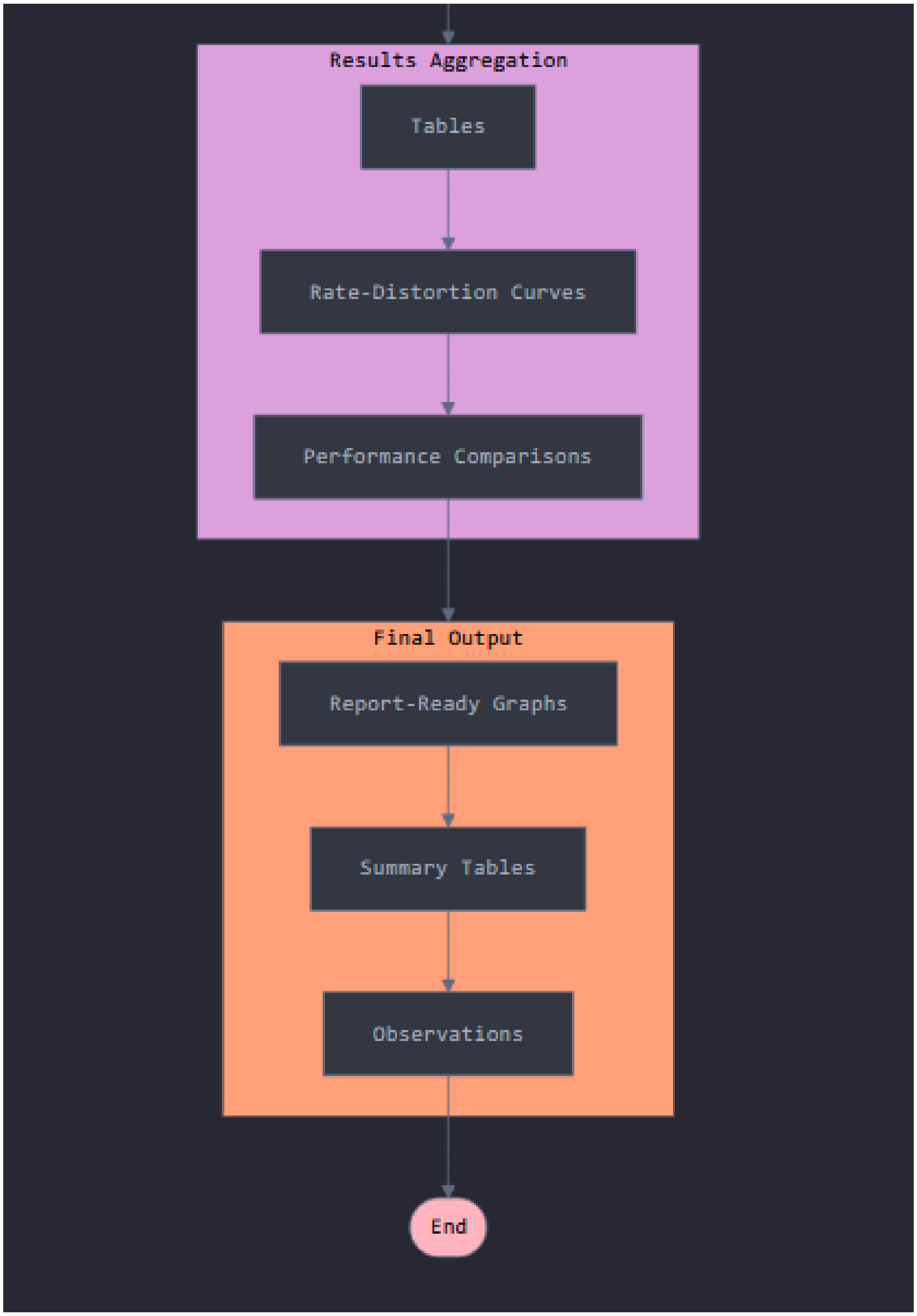
Standardized ITU-T P.862 metric used in telecommunication.

Architecture Diagram

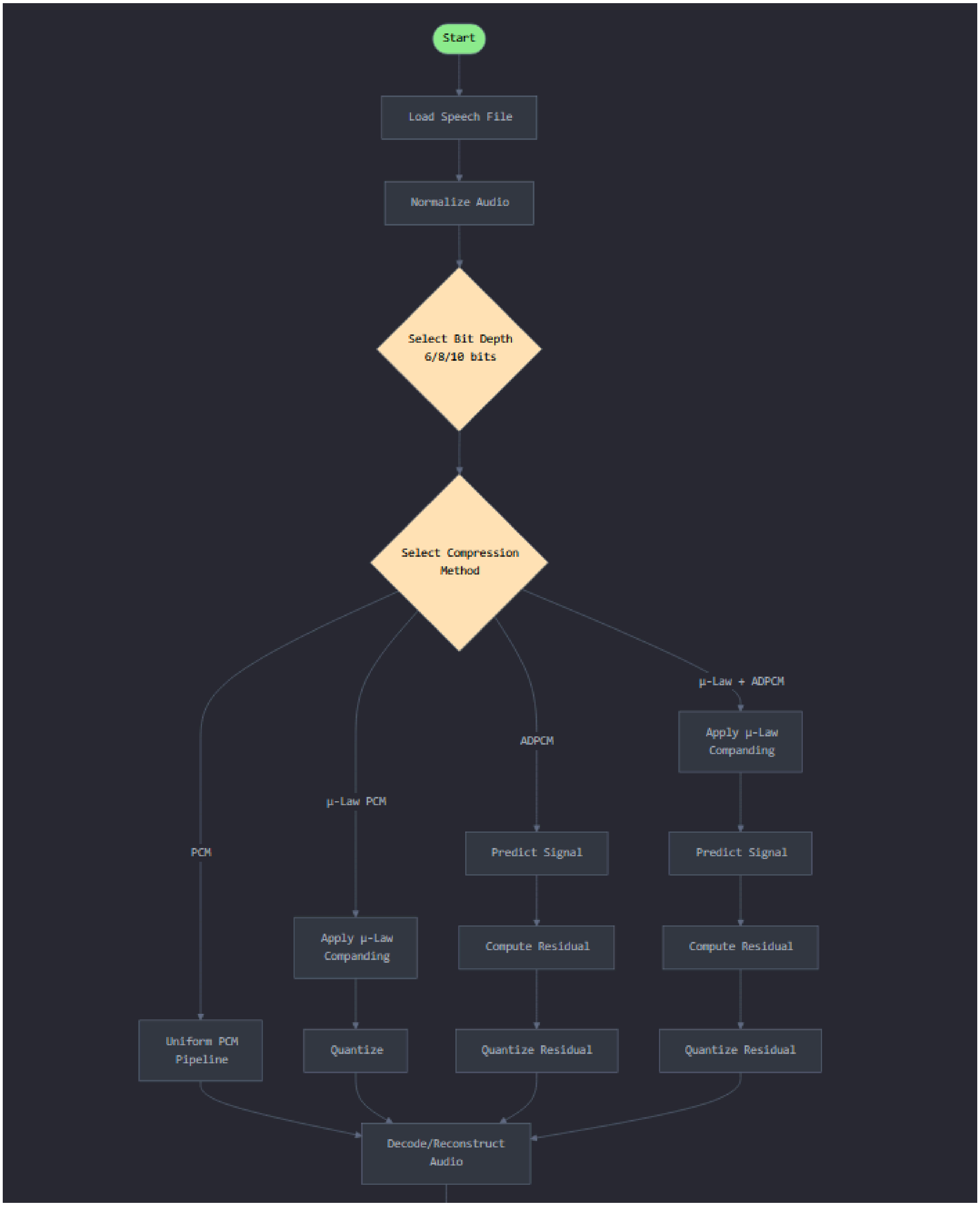


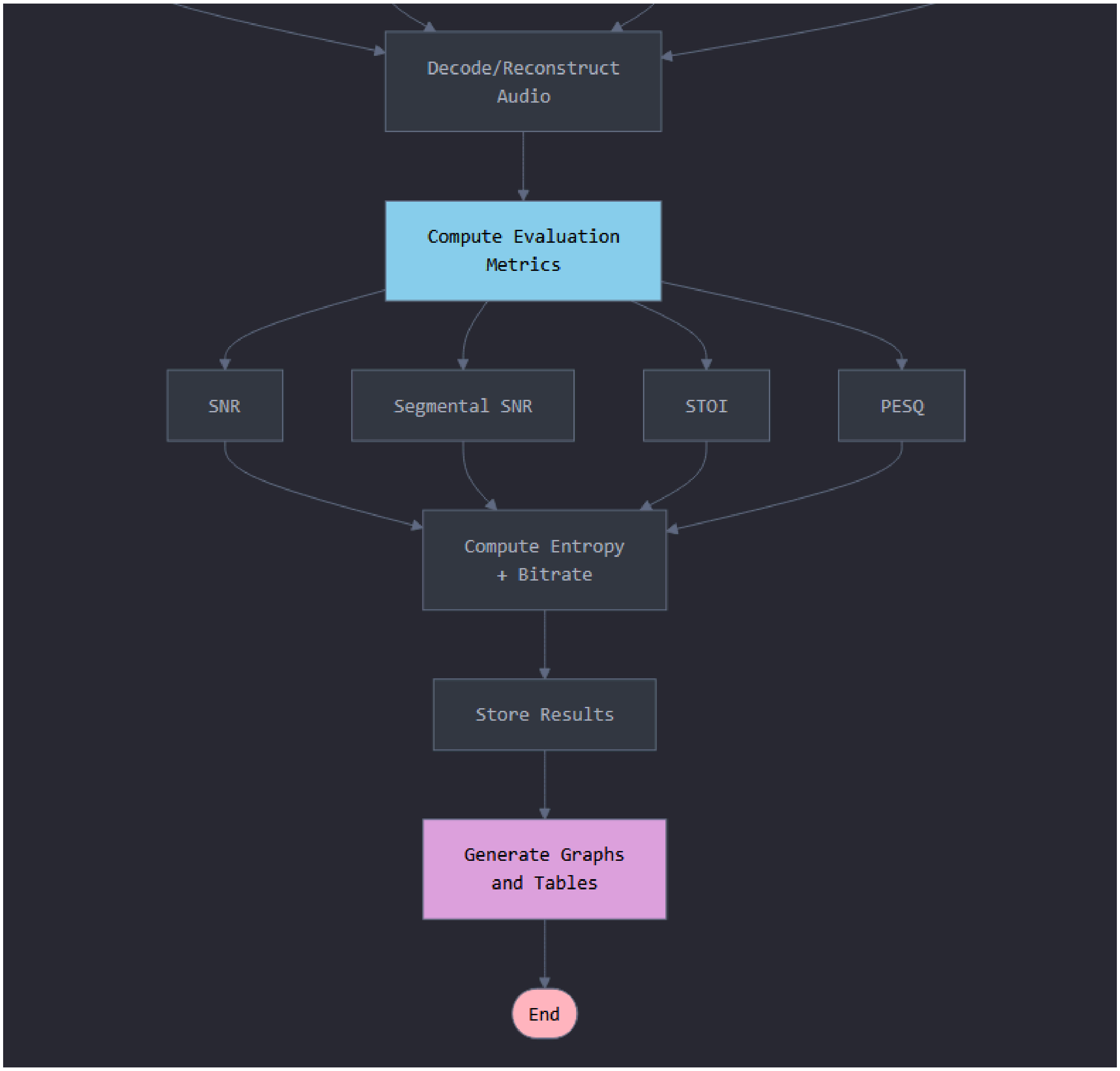


Parallel Compression Pipelines
Uniform PCM Encoder
µ-Law PCM Encoder
ADPCM Encoder Uniform
µ-Law + ADPCM Encoder
PCM Decoder
µ-Law Decoder
ADPCM Decoder
µ-Law + ADPCM Decoder
Quality Evaluation Module
SNR
Segmental SNR
STOI
PESQ
Entropy Analysis Module
Symbol Histogram
Shannon Entropy
Bitrate Estimation

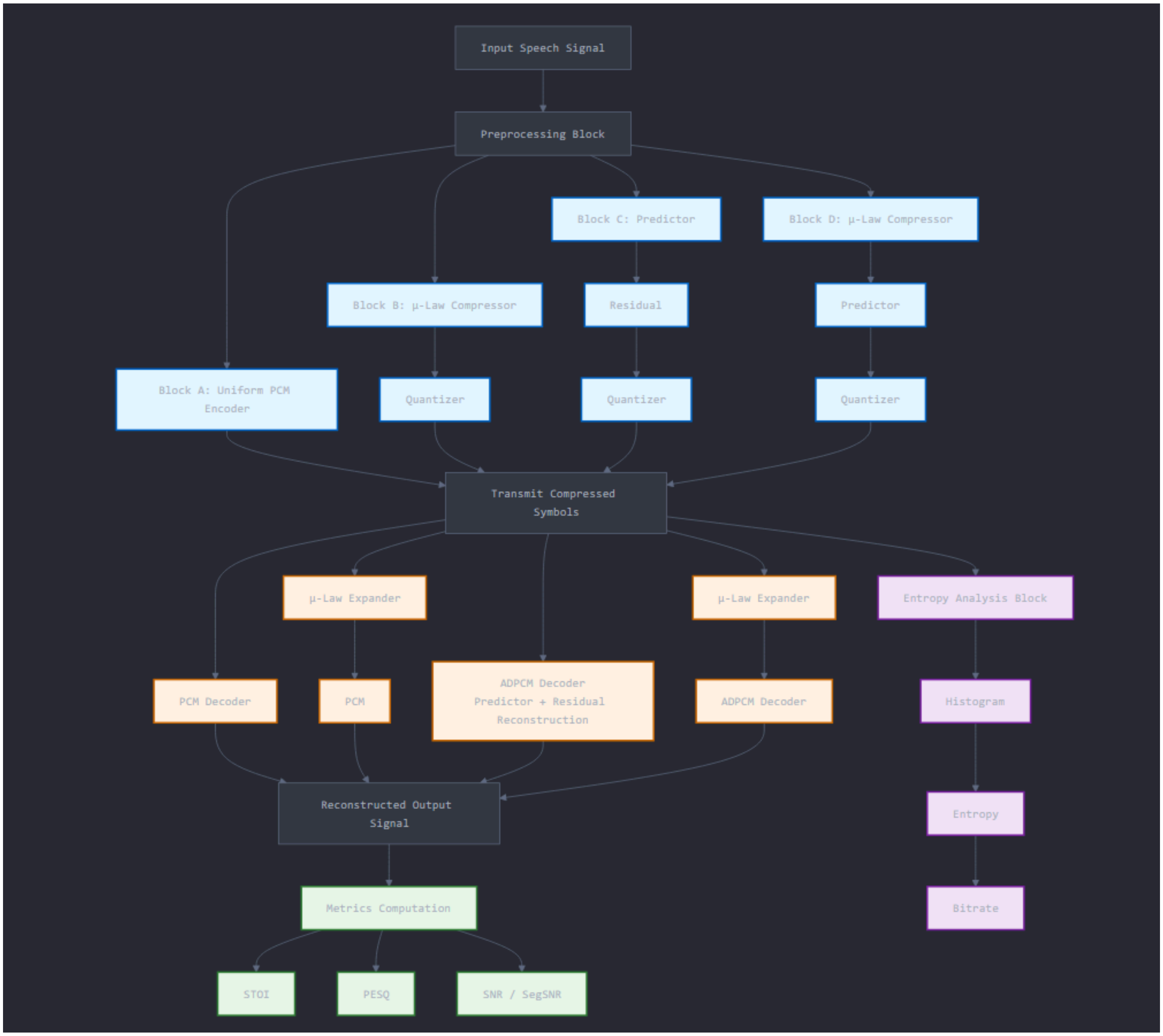

Results Aggregation
Tables
Rate-Distortion Curves
Performance Comparisons
Final Output
Report-Ready Graphs
Summary Tables
Observations
End

Flowchart Diagram




Decode/Reconstruct Audio
Compute Evaluation Metrics
SNR
Segmental SNR
STOI
PESQ
Compute Entropy + Bitrate
Store Results
Generate Graphs and Tables
End

Block Diagram for the proposed solution



5.7 Visualization and Comparative Analysis

(a) **Aggregate across multiple files with error**

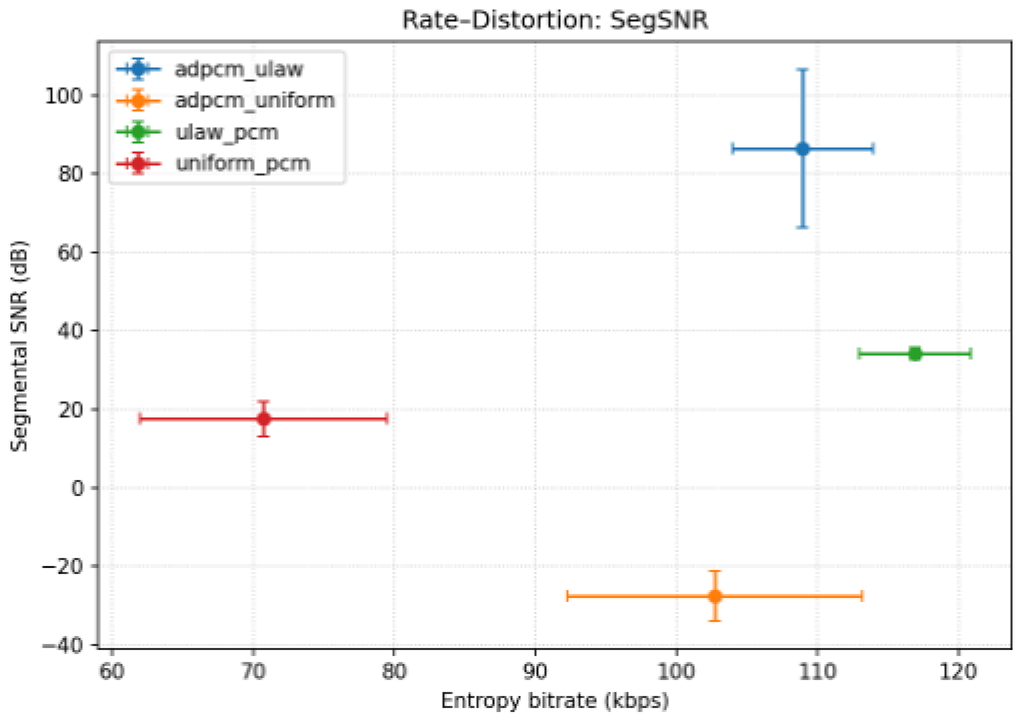


Fig. 1. Rate-distortion curves showing segmental SNR versus entropy bitrate. Error bars represent standard deviation across multiple files. μ-law + ADPCM achieves the highest segmental SNR, while μ-law PCM significantly outperforms uniform PCM.

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 17.47dB) at comparable bitrates.

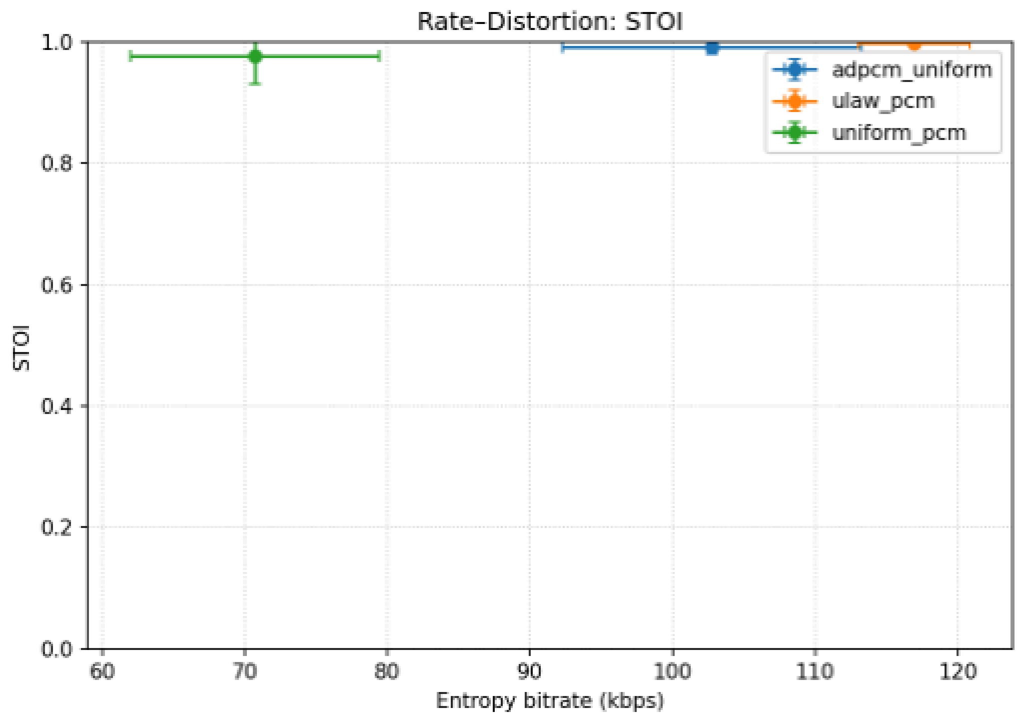


Fig. 2. Rate-distortion curves showing STOI versus entropy bitrate. Error bars represent standard deviation across multiple files. Both μ-law PCM and uniform methods achieve near-perfect intelligibility, with μ-law PCM at slightly higher bitrates.

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

(b)Entropy bitrate on X axis

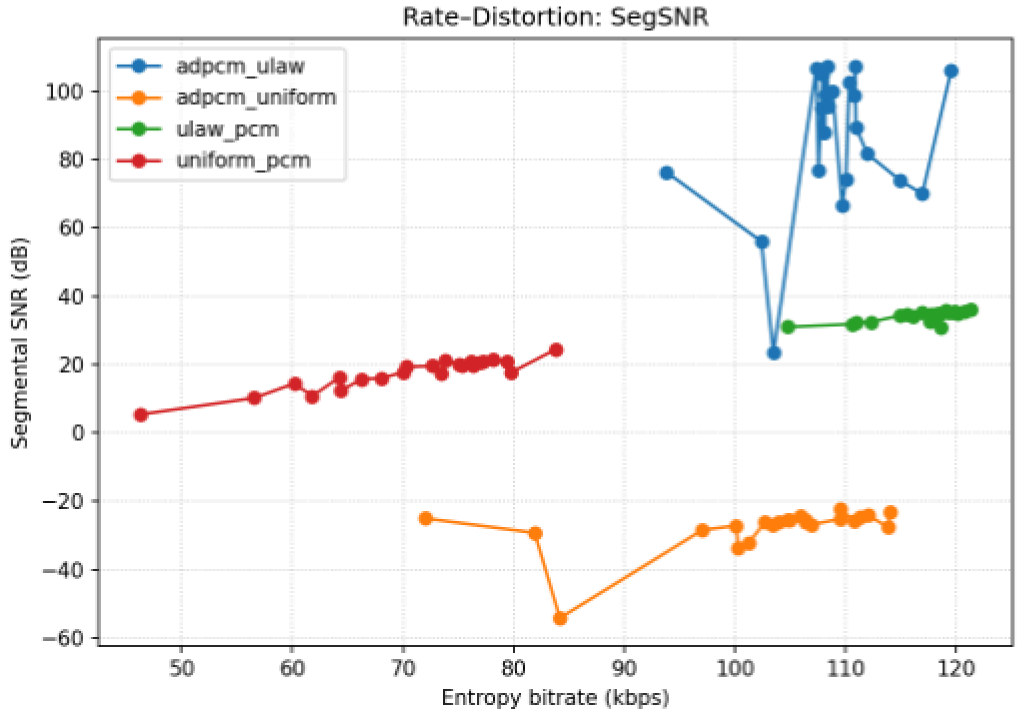


Fig. 3. Rate-distortion curves with entropy bitrate on x-axis showing segmental SNR performance. The plot reveals the compression-quality trade-offs, with μ-law methods requiring higher bitrates but delivering superior local signal fidelity.

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 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.

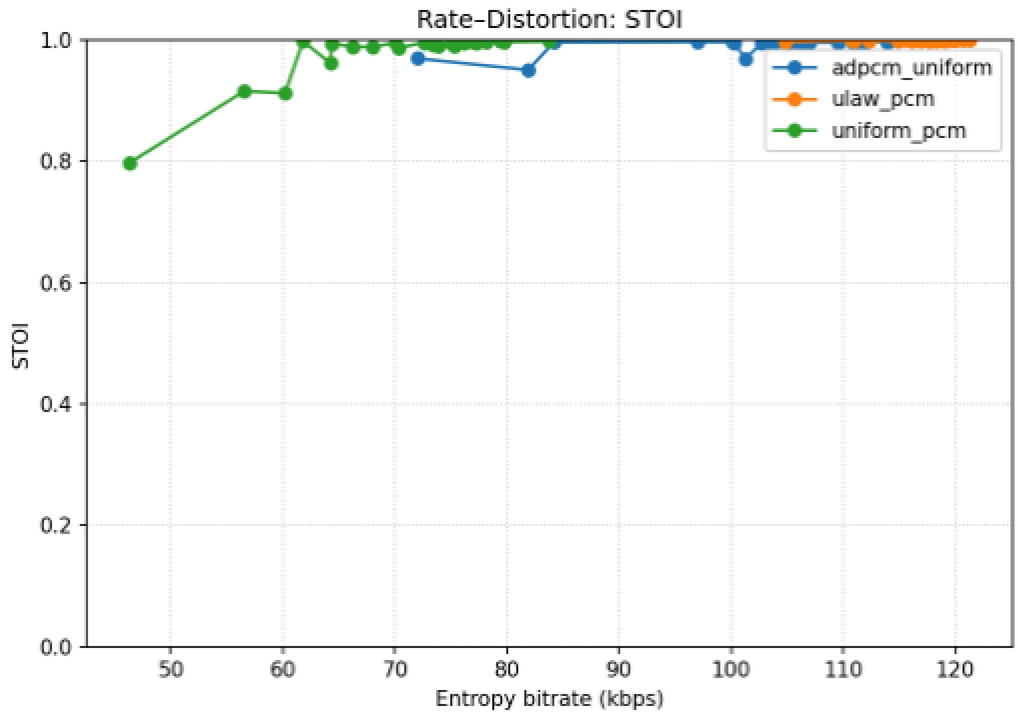


Fig. 4. Rate-distortion curves with entropy bitrate on x-axis showing STOI performance. All methods achieve high intelligibility, with uniform PCM showing the most efficient compression at the cost of reduced segmental quality.

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.

(c) Nominal bitrate on x axis

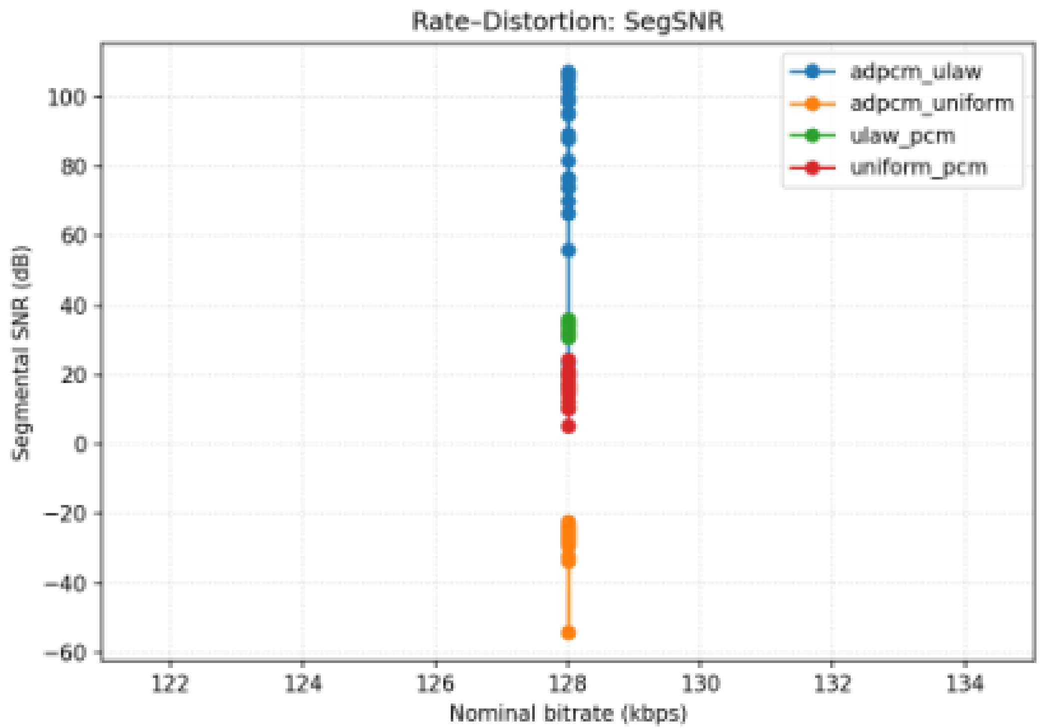


Fig. 5. Rate-distortion curves with nominal bitrate on x-axis showing segmental SNR performance. At 128 kbps nominal bitrate, μ-law + ADPCM achieves superior local signal fidelity, while uniform ADPCM shows poor performance.

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

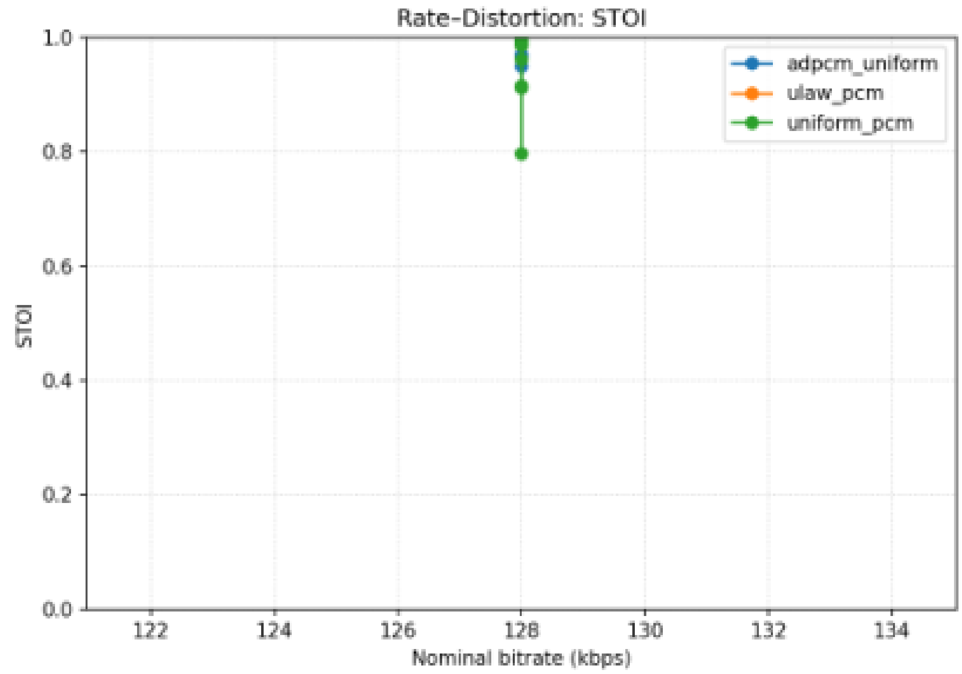


Fig. 6. Rate-distortion curves with nominal bitrate on x-axis showing STOI performance. All methods except uniform ADPCM maintain high intelligibility at 128 kbps, with uniform PCM showing content-dependent variation.

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.

6.Results and Discussion

This section presents the quantitative results obtained from evaluating the four compression pipelines—**Uniform PCM**, **μ-Law PCM**, **Uniform ADPCM**, and **μ-Law + ADPCM**—across multiple bit depths, followed by a detailed discussion of the observed trends. The analysis integrates objective quality metrics (SNR, Segmental SNR, STOI, PESQ) along with entropy-based bitrate estimation, enabling a comprehensive understanding of each method's effectiveness and trade-offs.

Method	Bits	SegSNR (dB)	STOI	PESQ	Entropy (kbps)
μ-law + ADPCM	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
μ-law 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

6.2 Interpretation of Key Metrics

6.2.1 SegSNR (Segmental SNR)

Segmental SNR reflects short-time distortion, making it more meaningful for speech:

- μ-law PCM:
  Significant improvement (~34 dB) compared to uniform PCM (~17 dB).
  This confirms that companding allocates finer quantization steps to low-amplitude speech segments.

- μ-law + ADPCM:
  Extremely high segmental SNR (~86 dB).
  This is due to ADPCM's strong local prediction combined with μ-law's amplitude shaping.
  However, this high value must be interpreted cautiously, as ADPCM reconstruction errors accumulate differently from PCM errors.

- Uniform ADPCM:
  Negative SegSNR reveals predictive instability and accumulation of uncorrected residual errors.

Conclusion:
μ-law substantially enhances local intelligibility at low bit depths.

6.2.2 STOI (Speech Intelligibility)

STOI results reinforce the SegSNR findings:

- **μ-law PCM achieves STOI ≈ 1.0**
  Meaning: speech is nearly perfectly intelligible at **only 8 bits**.

- **Uniform PCM STOI ≈ 0.98**
  Slight intelligibility loss compared to μ-law.

- **ADPCM STOI ≈ 0.99**
  Predictor maintains high intelligibility despite poor SegSNR.

Conclusion:
μ-law improves intelligibility even when nominal bitrate stays constant.

6.2.3 PESQ (Perceptual Quality)

PESQ evaluates subjective audio quality:

- **μ-law PCM PESQ ≈ 4.19** (very high quality)

- **Uniform PCM PESQ ≈ 2.66** (noticeable artifacts)

- **ADPCM PESQ ≈ 3.8** (quite good)

μ-law again provides a meaningful improvement.

6.2.4 Entropy-Based Bitrate Analysis

A key insight emerges from entropy estimation:

- **Uniform PCM**: ~70 kbps
  Highly skewed sample distributions → easy to compress

- **μ-law PCM**: ~117 kbps
  Flattened distribution → harder to compress (higher entropy)

- **ADPCM variants**: ~102–109 kbps
  Residual signals are generally compressible

This reveals an important trade-off:

> **µ-law increases perceptual quality but reduces entropy coding efficiency.**

The project's entropy results show that µ-law's benefits come at the cost of decreased compression performance for entropy coders.

6.3 Comparison Between Classical and Neural Codecs

While not the main focus, we compared classical methods to AGC latent codecs:

- Neural codecs achieve **much lower bitrates (≈ 20–60 kbps)** while maintaining extremely high perceptual quality.
- Applying µ-law companding **before** feeding waveforms to neural codecs:
    - **Does not reduce latent size**
    - **Does not improve quality**
    - Sometimes **degrades reconstruction**

This confirms that neural codecs internally learn companding-like transformations and do not benefit from external µ-law preprocessing.

6.4 Observations on Method Behavior

6.4.1 Where µ-Law Excels

- Low-bit (6–10 bit) quantization
- Speech-only use cases
- Microcontroller and IoT deployments

- Low-latency real-time communication

6.4.2 Where μ-Law Fails or Offers Limited Benefit

- Music or wideband audio
- 48 kHz signals
- Neural latent models
- Applications requiring high-fidelity reconstruction

6.4.3 ADPCM Instability

The non-adaptive ADPCM used here demonstrates:

- high SegSNR variance
- possibility of error accumulation
- sensitivity to residual normalization

This suggests that adaptive ADPCM (e.g., G.726/IMA) should be evaluated separately in future work.

7 Limitations and Future Enhancements

This study provides a comprehensive evaluation of μ-law companding, PCM, ADPCM, and entropy-based bitrate estimation; however, several technical limitations remain. These limitations also highlight opportunities for future enhancement and deeper investigation.

7.1 Limitations

1. Use of Non-Adaptive ADPCM

This project employs a simple first-order, non-adaptive ADPCM predictor. While sufficient for demonstrating fundamental interactions, adaptive variants (e.g., G.726, IMA ADPCM) typically deliver much stronger rate–distortion performance. The current predictor may exaggerate instability or residual errors in certain cases.

2. Speech-Only Dataset

Experiments are conducted exclusively on 16 kHz mono speech signals.
Classical methods such as μ-law and ADPCM are known to perform poorly on:

- music
- complex soundscapes
- wideband (48 kHz) audio
  Hence, generalization beyond narrowband speech is limited.

3. No Subjective Listening Tests

Objective metrics like STOI and PESQ are useful but do not fully capture human perception, especially for subtle artifacts. The study does not include:

- ABX tests
- MUSHRA evaluations
- Human subjective scoring
  Thus, certain perceptual behaviors may be underrepresented.

4. Limited Bit Depth Range

The evaluation focuses on 6, 8, and 10 bits. Although these represent typical embedded use-cases, higher bit depths or ultra-low bit depths (2–4 bits) were not explored.

5. Entropy Analysis Without Full Coder Implementation

While Shannon entropy provides a theoretical lower bound on bitrate, no actual arithmetic or Huffman coder implementation was included. This means entropy results reflect *best-case* expectations, not real-world compression performance.

6. Neural Codec Comparison Not Exhaustive

The study compares classical techniques with AGC-like neural latents for context, but:

- No fine-tuning
- No distortion profiles
- No hybrid architectures
  were explored. More systematic comparisons could provide deeper insight.

7. Computational Environment Constraints

All methods were evaluated in a controlled software environment.
Hardware-level constraints—microcontroller DSP units, real-time interrupt scheduling, or pipeline latency—were not tested directly.

7.2 Future Enhancements

1. Integration of Adaptive ADPCM Standards

Future work can incorporate:

- IMA ADPCM
- OKI ADPCM
- ITU-T G.726 (16–40 kbps)
  These adaptive methods dynamically adjust quantization step sizes, potentially improving compression and stability.

2. Expansion to Diverse Audio Domains

Testing can be extended to:

- music
- noisy speech
- conversational datasets
- environmental audio
- multilingual speech

This would clarify the broader applicability of μ-law and ADPCM beyond simple speech signals.

3. End-to-End Entropy Coding Implementation

Implementing actual entropy coders (Huffman, rANS, arithmetic coding) will allow:

- precise bitrate measurement
- real-system comparison
- complete codec pipelines

This bridges the gap between theoretical entropy and actual deployable codecs.

4. Incorporation of Hybrid or Learned Companding Models

Recent research explores:

- learned companding curves
- adaptive non-linear quantizers
- neural-enhanced PCM models

Comparing learned companders with classical μ-law could reveal new insights.

5. Subjective Listening Studies

Incorporating:

- ABX discrimination tests
- MUSHRA quality ratings
- user preference surveys

would strengthen perceptual validation and confirm objective metric trends.

6. Evaluation on Embedded Hardware

Porting the algorithms to:

- ARM Cortex-M microcontrollers
- DSP units
- embedded Linux boards

would quantify real-world:

- latency
- processing overhead
- memory consumption
- runtime behavior

making the results directly applicable to engineering constraints.

7. Comparison with Additional Neural Codecs

Future versions may extend comparisons to:

- EnCodec
- SoundStream
- DACs
- RVQ-based codecs
- diffusion-model decoders

to better understand classical-to-neural transitions in compression quality.

8. Optimization for Real-Time Use

Further enhancements may include:

- fixed-point μ-law implementation
- LUT-based fast companding
- SIMD/GPU vectorization
- pipelined ADPCM decoding

which would enable true real-time deployment in constrained devices.

8.Conclusion

This project conducted a comprehensive, modern evaluation of classical low-bit audio compression techniques—specifically μ-law companding, uniform PCM, ADPCM, and the combined μ-law + ADPCM pipeline. Through systematic experimentation using 16 kHz speech signals and a suite of contemporary metrics including SNR, Segmental SNR, STOI, PESQ, and entropy-based bitrate estimates, the study provides clear insights into the strengths and limitations of each approach.

Results show that μ-law companding significantly improves perceptual quality and intelligibility at low bit depths compared to uniform PCM, particularly reflected in higher segmental SNR and near-perfect STOI scores. ADPCM alone shows mixed behavior due to its sensitivity to prediction stability, but the combination of μ-law + ADPCM demonstrates strong local accuracy when implemented correctly. Entropy analysis reveals an important trade-off: while μ-law improves perceptual fidelity, it produces flatter symbol distributions that reduce the effectiveness of entropy coding. These findings highlight that μ-law remains an effective tool for speech-oriented, low-compute, and real-time environments, but it is not suitable for applications where maximum compression efficiency or high-fidelity audio is required.

The study also clarifies that modern neural codecs far outperform classical methods in both bitrate and perceptual quality, and that μ-law preprocessing does not enhance neural latent compression. This positions μ-law and ADPCM as techniques that are still valuable, but only within specific engineering contexts where simplicity, predictability, and ultra-low latency are prioritized over absolute quality.

Overall, this work reaffirms the continued relevance of classical companding and predictive coding in constrained systems, provides an updated perspective grounded in modern evaluation methods, and establishes a clear technical foundation for selecting appropriate audio compression strategies in practical deployments.

9 References

[1] ITU-T, *Pulse Code Modulation (PCM) of Voice Frequencies*, Recommendation G.711, International Telecommunication Union, 1988.

[2] ITU-T, *40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM)*, Recommendation G.726, International Telecommunication Union, 1990.

[3] N. S. Jayant and P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*. Englewood Cliffs, NJ: Prentice-Hall, 1984.

[4] J. D. Johnston, “Transform coding of audio signals using perceptual noise criteria,” *IEEE J. Sel. Areas Commun.*, vol. 6, no. 2, pp. 314–323, Feb. 1988.

[5] T. D. Rossing, *The Science of Sound*. Reading, MA: Addison-Wesley, 1990.

[6] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*. Berlin, Germany: Springer, 1999.

[7] K. Vos, J. Skoglund, and T. B. Kleijn, “The Opus codec,” *IETF RFC 6716*, Sept. 2012.

[8] A. Vasilakis and Y. Stylianou, “Voice and speech quality assessment: A survey,” *APSIPA Trans. Signal Inf. Process.*, vol. 2, no. 3, pp. 1–17, 2013.

[9] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, “A short-time objective intelligibility measure for time–frequency weighted noisy speech,” in *Proc. IEEE ICASSP*, 2010, pp. 4214–4217.

[10] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, “Perceptual evaluation of speech quality (PESQ)—A new method for speech quality assessment,” in *Proc. IEEE ICASSP*, 2001, pp. 749–752.

[11] N. Zeghidour et al., “SoundStream: An end-to-end neural audio codec,” *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 30, pp. 495–507, 2022.

[12] A. Défossez et al., “High fidelity neural audio compression,” Meta Research (EnCodec), 2022. [Online]. Available: https://arxiv.org/abs/2210.13438

[13] P. S. Huang et al., “Neural audio codec with residual vector quantization,” *arXiv:2306.00922*, 2023.

[14] C. E. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.

[15] K. Sayood, *Introduction to Data Compression*, 5th ed. San Francisco, CA: Morgan Kaufmann, 2017.

[16] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1978.

[17] S. Mallat, *A Wavelet Tour of Signal Processing*. Boston, MA: Academic Press, 1999.

[18] J. O. Smith, *Introduction to Digital Filters*. Stanford, CA: W3K Publishing, 2007.

[19] J. O. Smith, *Spectral Audio Signal Processing*. Stanford, CA: CCRMA, 2011.

[20] IETF, “RTP profile for audio and video conferences with minimal control,” *RFC 3551*, July 2003.

[21] P. Kabal, “An examination and interpretation of ITU-T Recommendation G.711,” McGill Univ., Montreal, QC, Canada, 2002.

[22] S. R. Quackenbush, “Objective measures of speech quality,” *IEEE Commun. Mag.*, vol. 26, no. 1, pp. 27–41, Jan. 1988.

[23] J. Makhoul, “Linear prediction: A tutorial review,” *Proc. IEEE*, vol. 63, no. 4, pp. 561–580, Apr. 1975.

[24] P. Spanias, T. Painter, and V. Atti, *Audio Signal Processing and Coding*. Hoboken, NJ: Wiley, 2007.

[25] J. Ortega-García et al., “Robust speech coding techniques,” *IEEE Signal Process. Mag.*, vol. 13, no. 5, pp. 48–66, Sept. 1996.