AUDIO COMPRESSION USING M-LAW COMPANDING + AUTOENCODER

By 22MID0069 Merril Baiju
22MID0304 Mayank Dash

Why Audio Compression?

- **Economic costs:** High storage and transmission costs.
- **Legacy codecs (e.g., MP3, G.711):** Use fixed signal transforms and hand-crafted features.
- **State-of-the-art methods:** Use deep learning (autoencoders) for automatically learning compact, efficient representations that maintain perceptual quality.
- **Effect:** Efficient audio compression is essential for streaming, telephony, and real-time communication in the modern digital world.

BASE PAPER

SoundStream: An End-to-End Neural Audio Codec
https://arxiv.org/pdf/2107.03312

KEY RESULT: OUTPERFORMS TRADITIONAL CODECS LIKE OPUS AT LOWER BITRATES.

SOUNDSTREAM: AN END-TO-END NEURAL AUDIO CODEC (GOOGLE RESEARCH, 2021)

- SoundStream is a neural audio codec that compresses general audio, music, and speech at very low bitrates.
- Utilizes a fully convolutional encoder–decoder with a residual vector quantizer, learned end-to-end.
- Training employs reconstruction + adversarial losses for high-quality output.
- Works at variable bitrates (3–18 kbps) with minimal quality loss.
- Operates in real-time on phones (low latency).
- Is better than Opus (12 kbps) at 3 kbps and comparable to EVS (9.6 kbps).
- Can also do joint compression and noise reduction without additional delay.

ALGORITHMS IN THE BASE PAPER

1. Autoencoder

Encoder: learns a lower-dimensional latent representation of the input audio.

Decoder: reads out the waveform from this compressed representation.

2. Vector Quantization (VQ)

Maps latent vectors to a finite range of discrete codes.

Facilitates efficient transmission and storage at low bitrates.

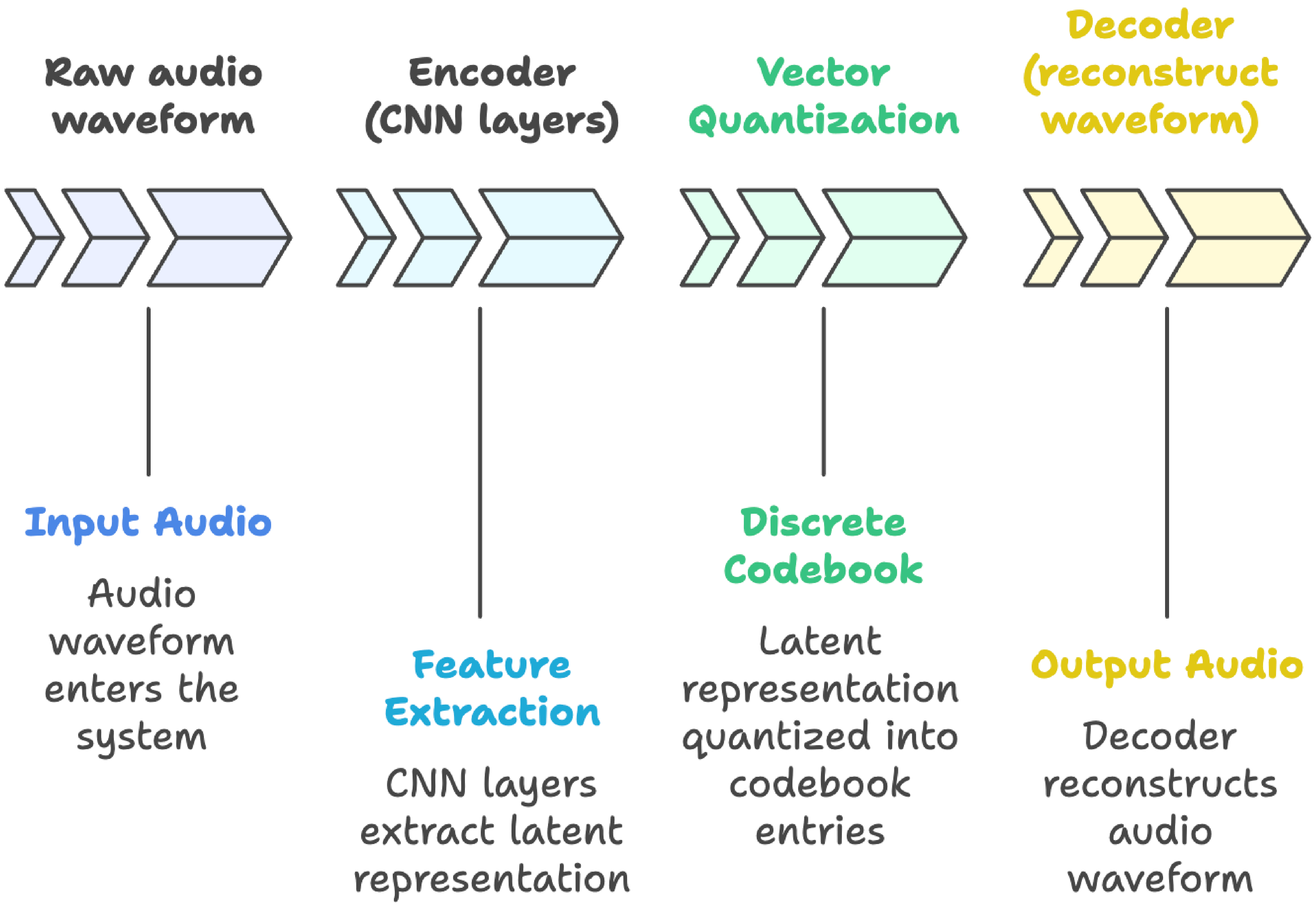
3. Training Objectives

Reconstruction Loss (MSE): maintains signal fidelity.

Perceptual Loss: maintains human listening quality.

Adversarial Loss (GAN-based): encourages natural-sounding audio.

Base System Workflow — SoundStream



WHAT ARE WE DOING DIFFERENTLY

Problem in the Base System

- Raw audio has a large dynamic range, with large differences between loud and quiet sounds.
- The encoder has to model both extremes at the same time, and that makes the compression less efficient.

Our Proposed Solution

- Add μ-law companding as a pre-processing step prior to the encoder.
- This non-linear mapping
    - Compresses large amplitudes in order to eliminate redundancy.
    - Maintains finer details in low-amplitude signals.
- More attuned to the human ear's perception, where the ear is more sensitive to soft sounds rather than loud ones.

M-LAW COMPANDING

μ-law Companding

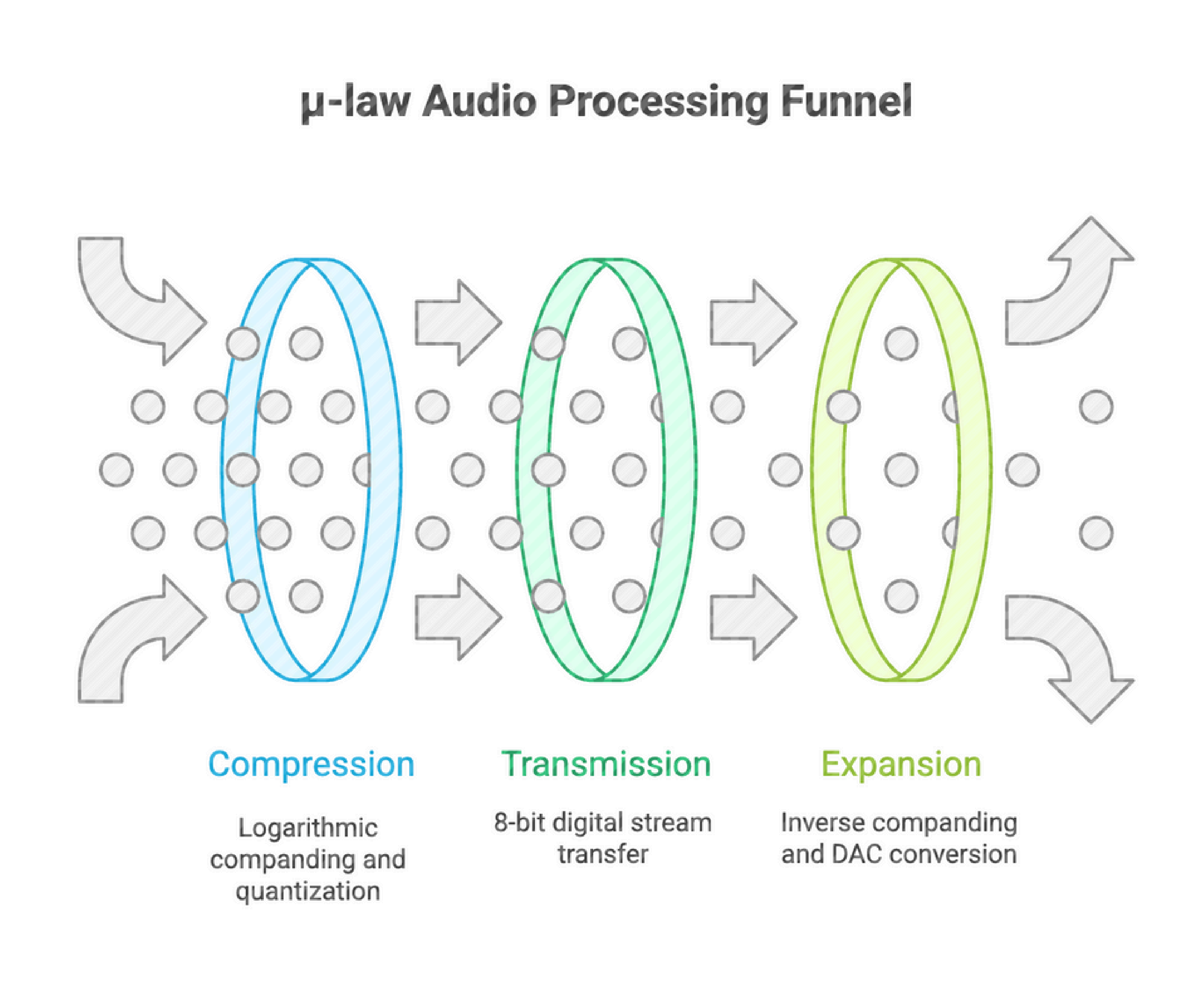
Squeezes audio by providing greater accuracy to soft sounds and less to loud sounds, which corresponds with human hearing.

A 14-bit signal can be effectively encoded into 8 bits and then enlarged back without significant loss of quality.

(How it Works)

Compression (Tx): Using a log function → soft portions boosted, loud portions squished → uniformly quantized to 8-bit.

Expansion (Rx): Inverse function returns original dynamic range → effective transmission with understandable speech quality.



M-LAW COMPANDING

- Formula:

$$F(x) = \frac{\operatorname{sgn}(x) * \ln(1+\mu|x|)}{\ln(1+\mu)}$$

What it does:

- Large signals → compressed more.
- Small signals → preserved better.
- Matches human ear sensitivity → better perceptual quality at fewer bits.
- Inverse µ-law expands it back at decoding stage.

OUR PROPOSED SYSTEM (M-LAW + AUTOENCODER)

Step 1: µ-law companding Preprocessing

- Performs a logarithmic mapping to dynamic range compression.
- Quietest sounds amplified, loudest sounds compressed.
- Decreases input entropy → simpler to learn for encoder.

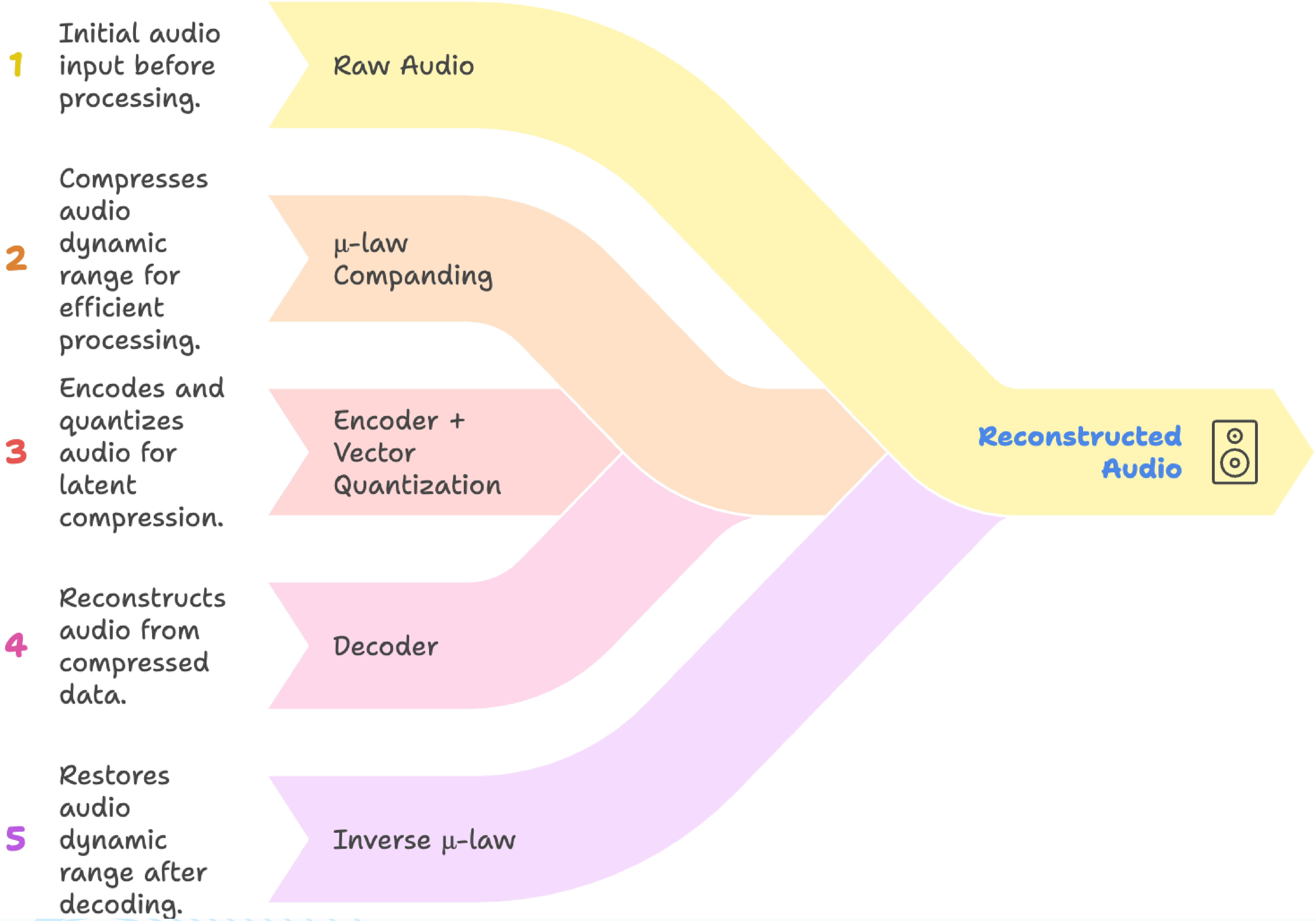
Step 2: Autoencoder + Vector Quantization Compression

- Encoder (CNN layers): captures latent features.
- Vector Quantizer: maps latent features to sparse discrete codes.
- Low bitrate efficient transmission/storage.

Step 3: Reconstruction

- Decoder: reconstructs companded audio waveform from discrete codes.
- Inverse µ-law expansion: returns the original dynamic range of the audio.

Audio Processing Workflow



Key Findings & Benefits

Observations from the Base Paper (SoundStream):

Neural autoencoders can outperform conventional codecs in compression as well as perceptual quality.

Residual vector quantization allows efficient, low-dimensional, discrete representations of audio.

Contributions of Our Proposed System (with μ-law):

Lower input entropy: μ-law companding reduces the complexity of the signal, easing modeling for the autoencoder.

Improved speech quality: Better clarity and perceptual fidelity at constant bitrate.

Towards near-lossless compression: More accurate reconstruction with low bandwidth needs.

Key Findings & Benefits

Observations from the Base Paper (SoundStream):

Neural autoencoders can outperform conventional codecs in compression as well as perceptual quality.

Residual vector quantization allows efficient, low-dimensional, discrete representations of audio.

Contributions of Our Proposed System (with μ-law):

Lower input entropy: μ-law companding reduces the complexity of the signal, easing modeling for the autoencoder.

Improved speech quality: Better clarity and perceptual fidelity at constant bitrate.

Towards near-lossless compression: More accurate reconstruction with low bandwidth needs.

Key Advantages

- Reduces complexity for the autoencoder.
- Improves efficiency of latent compression.
- Expected to deliver better perceptual quality at the same or lower bitrate.
- Moves closer to near-lossless audio compression.

CONCLUSION

Base paper (SoundStream) = solid neural codec baseline

- Our system = combines traditional signal processing (µ-law) with contemporary deep learning (autoencoder).

Expected gains:

- Improved perceptual quality.
- Efficient encoding.
- Future Work: Experiment on bigger datasets, compare metrics (SNR, PESQ, MOS).

THANK YOU