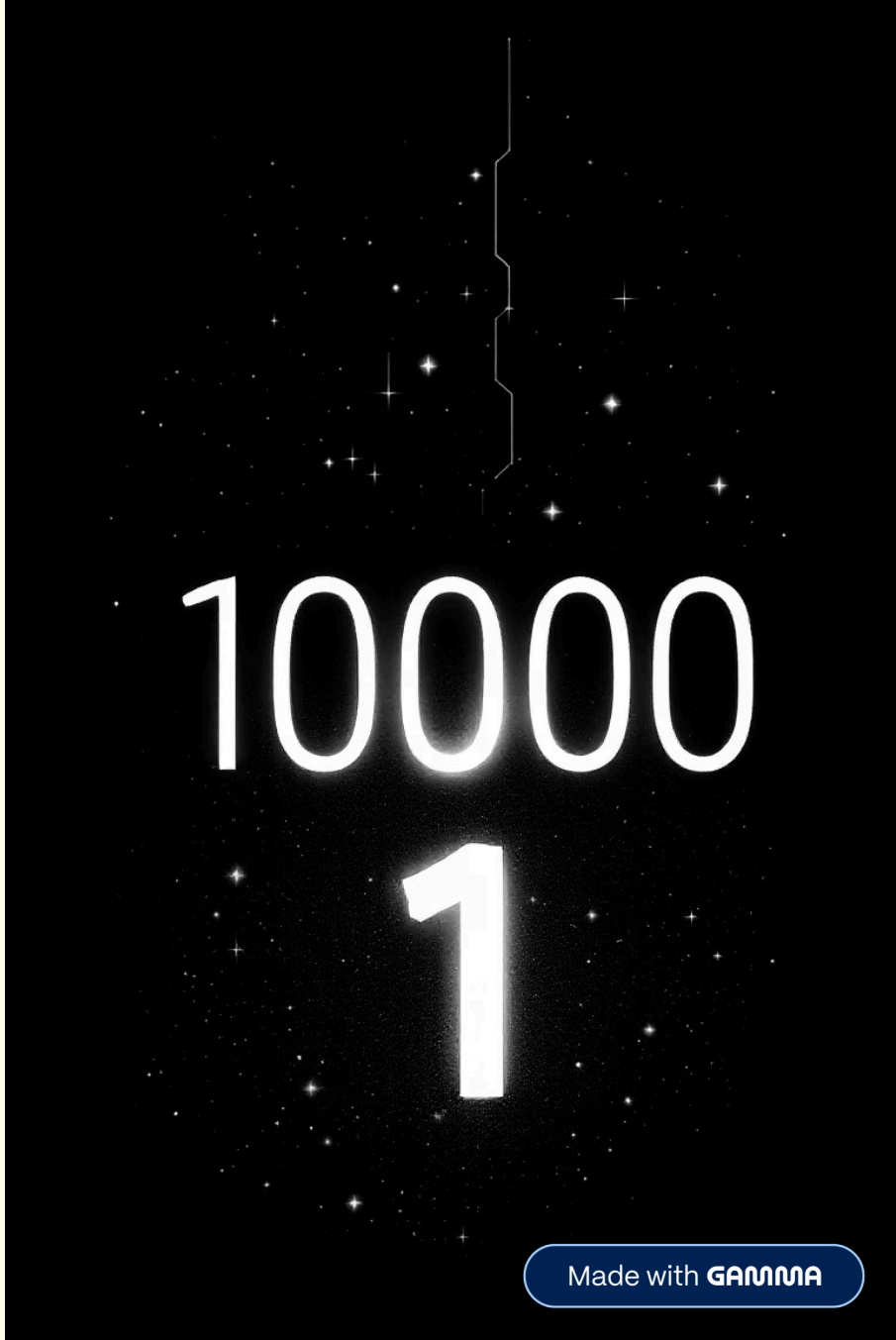
10,000 to 1: The Semantic Leap in Speech Coding

Ultra-Low Bitrate Semantic Speech Coding



Made with GAMMA

We Are Slaves to the Waveform.

For decades, audio engineers have been trapped by the conventional approach to sound. Our innovation shatters these limitations.

The Trap: Preserving Voltage

Traditional coding, from PCM to MP3, attempts to replicate the exact voltage fluctuations of a sound wave. This analog-first mindset binds us to an inefficient past.

The Cost: Nyquist-Shannon Bound

We are constrained by fundamental physics. To maintain fidelity, we must capture vast amounts of data, regardless of its linguistic value.

The Result: 64,000 Bits to Say 'Hello'

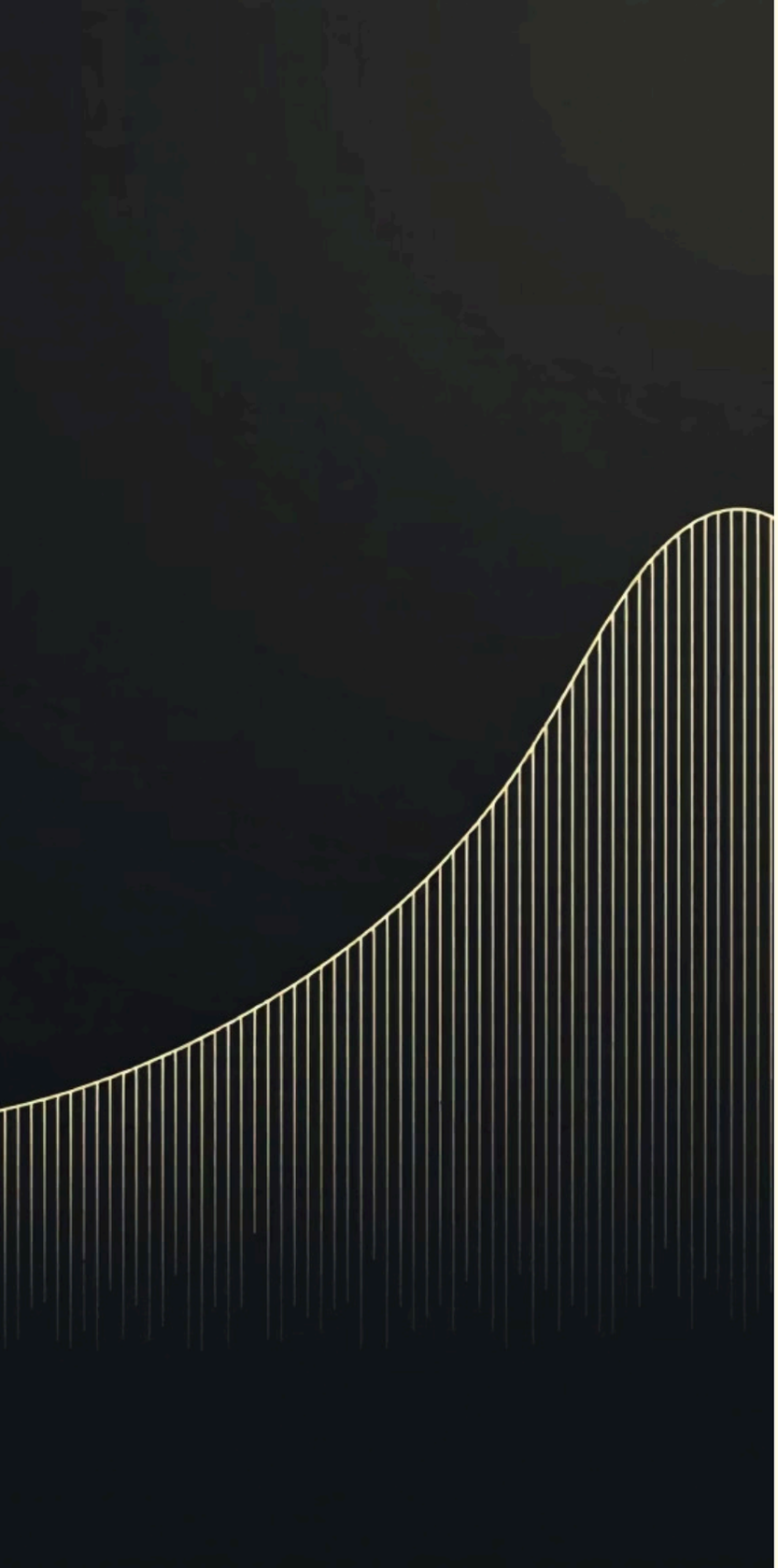
An enormous data footprint just to convey a simple greeting, highlighting the inefficiencies of waveform-centric approaches.

Made with GAMMA

The Speech-to-Text-to-Speech Pipeline

This innovative architecture transforms a complex signal processing challenge into an elegant Natural Language Processing problem, redefining audio transmission.

Made with GAMMA



Garbage In, Garbage Out: Cleaning the Canvas.

Before meaning can be extracted, the signal must be pristine. Our meticulous preprocessing lays the foundation for unparalleled accuracy.

Filter Type: Butterworth High-Pass (Order 4)

A precision filter designed to target and eliminate unwanted low-frequency noise without compromising speech clarity.

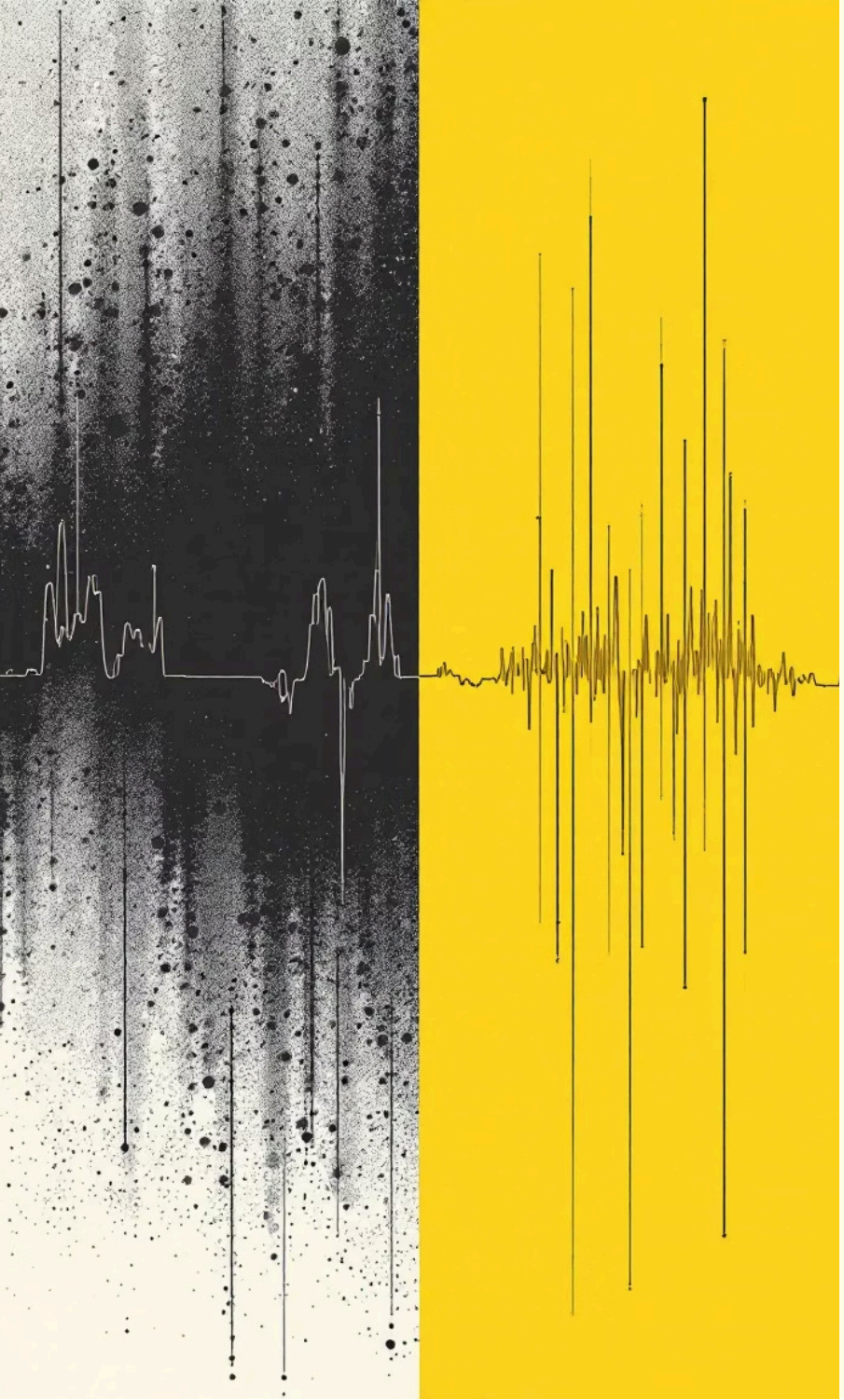
Cutoff: 80 Hz (Removing "Rumble")

We surgically remove ambient "rumble"—HVAC noise, mic handling—that adds no linguistic value and can confuse ASR engines.

Method: Zero-Phase Filtfilt (Effective Order 8)

Employing a forward-and-backward filtering technique, we eliminate phase distortion and double the effective filter order, ensuring a truly clean signal.

Made with GAMMA



We Don't Reconstruct. We Resynthesize.

Our method moves beyond mere replication, generating an idealized version of the original message.

1

Original: Noisy & Irregular

The original signal is often laden with acoustic "dirt" and unpredictable harmonic patterns.

2

Synthesized: Clean & Idealized

Our output preserves the essence of the speech, removing extraneous noise and presenting a pure message.

3

Result: Meaning Remains Intact

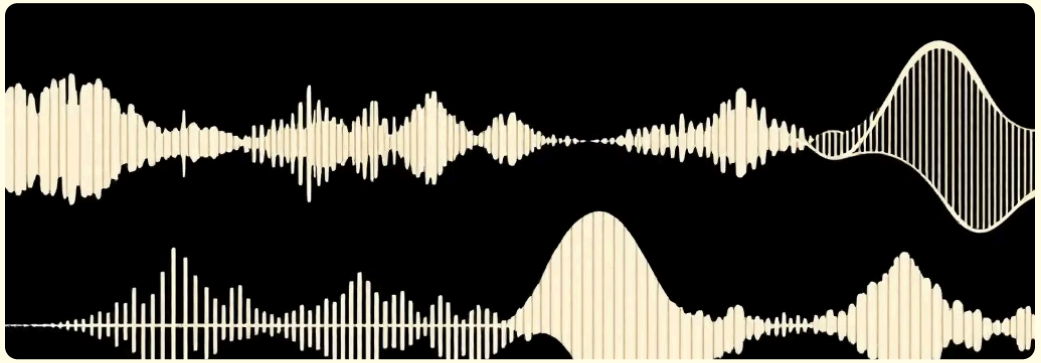
The message is not just compressed; it's purified, retaining its core while shedding acoustic baggage.

Made with GAMMA

The Ultimate Vector Quantization

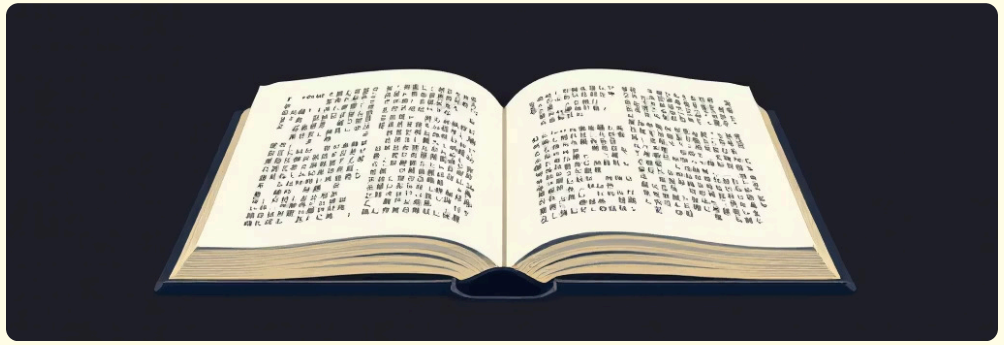
We have reimagined Vector Quantization, moving from acoustic shapes to the profound precision of language itself.

Standard VQ: Acoustic Codebook



Acoustic Codebook

Semantic VQ: Linguistic Codebook



Linguistic Codebook

Input: Infinite Signal Vectors

The vast, analog world of sound offers endless variations, making traditional compression inefficient.

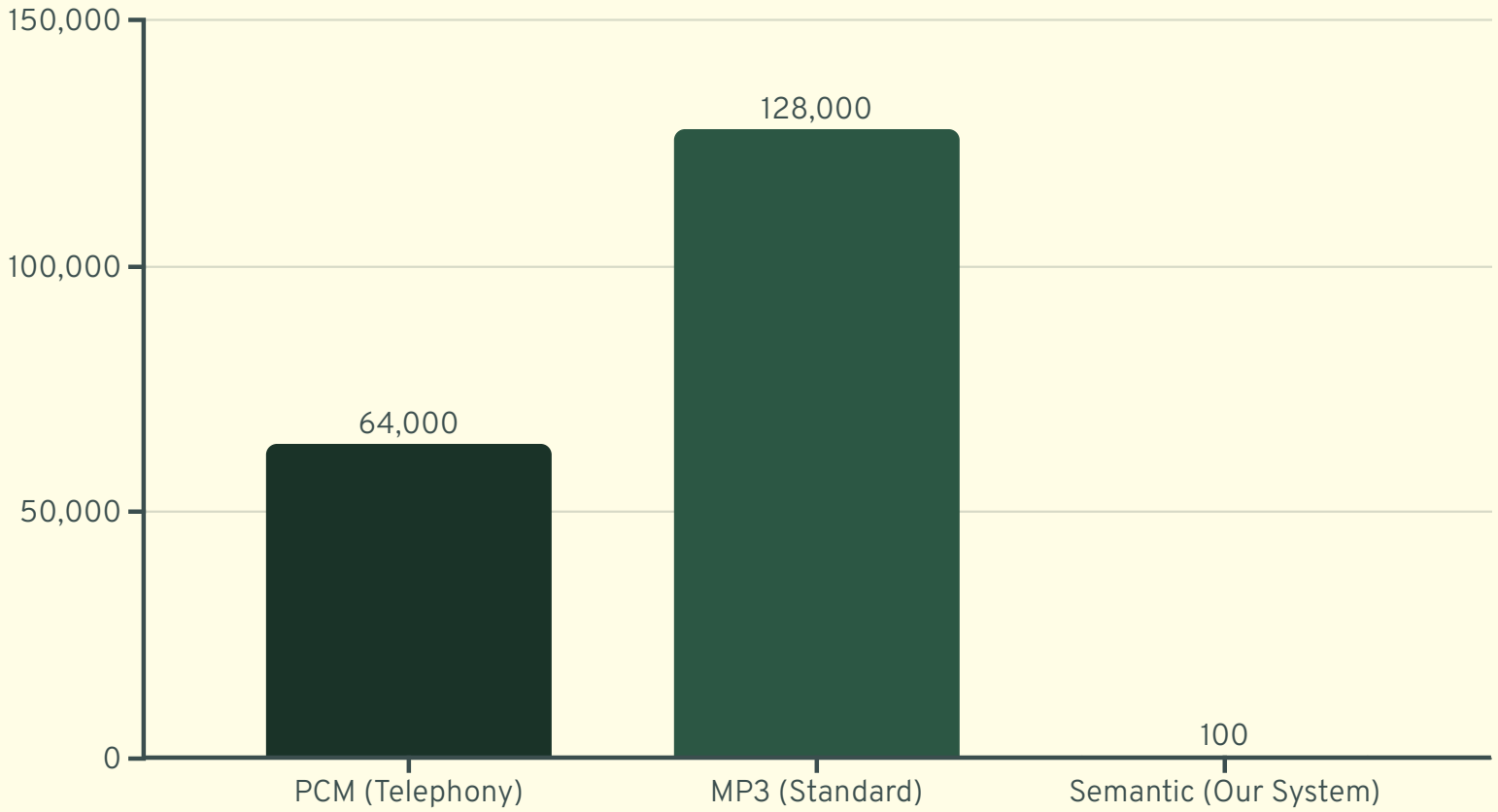
Codebook: Finite Text Tokens

We distill this infinite input into a discrete set of linguistic tokens—the words themselves.

Made with GAMMA

640 Times More Efficient than Telephony

Our breakthrough isn't just an improvement; it's a monumental leap in data efficiency, shattering previous benchmarks.



From a firehose to a mere drip, our system transmits crystal-clear voice at 12.5 bytes per second. That’s a 10,000:1 compression ratio against CD quality and 640:1 against standard telephony.

Made with GAMMA

We Lose the Voice. We Keep the Message.

While the raw waveform is sacrificed, the essence of communication, and even speaker identity, can be elegantly preserved.

Identity Lost ❌

By stripping the waveform, we initially lose the unique acoustic fingerprint of the speaker's voice.

Identity Reclaimed ✅

A compact 128-dimensional speaker embedding vector can be transmitted alongside the text, allowing for voice reconstruction.

Made with GAMMA

12.5 Bytes Per Second.

This isn't just about efficiency; it's about enabling communication in the most challenging environments.

Extreme Low-Bandwidth Operations

Enabling critical voice comms in environments where traditional audio is impossible.

Massive Archival Storage

Revolutionizing how we store and access vast libraries of spoken content with minimal footprint.

The Future is Semantic.

We're not just compressing sound; we're redefining its very essence.

Made with GAMMA

Standing on the Shoulders of Giants.

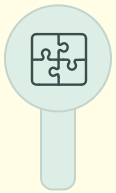
Our project builds upon a rich legacy of foundational theories and enabling technologies, culminating in a paradigm shift for speech communication.

1948: The Foundation

A Mathematical Theory of Communication (Shannon): Defined the bit-level limit, which we now challenge with semantic coding.

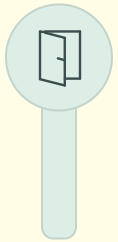
2017: The Architecture

Transformer/Attention Mechanism: Revolutionized sequence modeling, laying groundwork for advanced NLP models like Whisper.



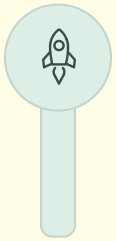
2021: The Pivot

Semantic Communication (Weng & Qin): The first proposal to use STT/TTS for radical bandwidth reduction.



2022: The Enabler

Robust Speech Recognition (OpenAI Whisper): Provided the noise-robust encoder essential for making semantic coding viable.



2025: The Implementation

This Project: Deploying Deep Semantic Communication concepts, leveraging semantic tokens for unprecedented efficiency.

Made with GAMMA