# AMR-GLPC: Adaptive Multi-Resolution Gabor-LPC Coding via Granular Window Growth and Global Normalization

Research Division
**Beam Audio Technologies**

November 21, 2025

**Abstract**

This paper presents the Adaptive Multi-Resolution Gabor-LPC Codec (AMR-GLPC), a novel audio coding architecture designed for high-fidelity speech synthesis at 44.1kHz. Addressing the limitations of fixed-frame vocoders, AMR-GLPC employs a Granular Window Growth strategy that dynamically adapts the analysis frame size from 256 to 1024 samples (5.8ms to 23.2ms) based on spectral stationarity. We introduce a Global Normalization Overlap-Add (GN-OLA) method to guarantee unity-gain reconstruction during asynchronous window switching, eliminating amplitude modulation artifacts. Furthermore, the residual excitation is modeled using Matching Pursuit over a logarithmic-frequency Gabor dictionary, aligning reconstruction error with human auditory perception. Theoretical proofs for the stability of the regularized inverse filter and the convergence of the GN-OLA method are provided.

## 1 Introduction

High-sample-rate speech coding (44.1kHz) presents a fundamental time-frequency uncertainty trade-off. Transient phones (plosives like /k/, /t/) require high temporal resolution to preserve attack transients, while voiced segments (vowels) require high frequency resolution to resolve harmonic structures.

Traditional codecs utilize fixed framing (e.g., 20ms), which smears transients and provides insufficient frequency resolution for high-pitched voices. We propose a solution based on three core innovations:

1. **Regularized Source-Filter Separation:** Ensuring $H(z)$ stability via Tikhonov Regularization.
2. **Granular Window Growth:** A four-stage adaptive framing logic ($256 \to 512 \to 768 \to 1024$).
3. **Log-Gabor Sparse Approximation:** A dictionary-based residual coding scheme using logarithmically spaced frequency atoms.

## 2 Signal Model

### 2.1 Regularized LPC Analysis

The speech signal $s[n]$ is modeled as an auto-regressive (AR) process. To ensure stability during silence or high-frequency noise where the correlation matrix may be ill-conditioned, we employ Ridge Regression.

The LPC coefficients $\mathbf{a}$ are derived by solving:

$$(\mathbf{R} + \lambda \mathbf{I})\mathbf{a} = -\mathbf{r} \tag{1}$$

where $\mathbf{R}$ is the autocorrelation matrix, $\mathbf{r}$ is the correlation vector, and $\lambda = 0.02$ is the regularization factor.

**Theorem 1** (Stability of Regularized LPC)**.** *For any $\lambda > 0$, the roots $z_i$ of the polynomial $A(z) = 1 + \sum_{k=1}^{p} a_k z^{-k}$ satisfy $|z_i| < 1$.*

*Proof.* The addition of $\lambda \mathbf{I}$ is equivalent to adding white noise to the signal spectrum. This reduces the spectral dynamic range, pushing the poles of the synthesis filter away from the unit circle towards the origin. Thus, the inverse filter $1/A(z)$ is strictly minimum-phase and stable. □

# 3 Granular Window Growth

We define an atomic frame size $N_{base} = 256$. The codec attempts to merge consecutive frames into a "Super-Frame" based on the Log-Spectral Distortion (LSD) metric:

$$D_{LSD}(\mathscr{F}_i, \mathscr{F}_{i+1}) = \sqrt{\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(10\log_{10}\frac{P_i(\omega)}{P_{i+1}(\omega)}\right)^2 d\omega} \tag{2}$$

The growth logic proceeds in four stages:

- **Stage 1 (256 samples):** Default for transients/noise. High time resolution.
- **Stage 2 (512 samples):** Merges 2 frames if $D_{LSD} < \tau_{strict}$. Captures fast vibrato.
- **Stage 3 (768 samples):** Merges 3 frames if cumulative error is low. Suitable for glissandos.
- **Stage 4 (1024 samples):** Merges 4 frames. Used strictly for steady-state vowels to maximize coding gain.

# 4 Sparse Log-Gabor Representation

The prediction residual $e[n]$ is encoded via Matching Pursuit using a dictionary $\mathscr{D} = \{g_\gamma\}$.

## 4.1 Logarithmic Frequency Distribution

Unlike standard Fourier bases, the modulation frequencies $\xi$ of our Gabor atoms are distributed logarithmically:

$$\xi_k = \xi_{min} \cdot \beta^k, \quad k \in \{0, \dots, M-1\} \tag{3}$$

This distribution mirrors the critical bands (Bark scale) of the human ear, allocating more atoms to lower frequencies where pitch perception is most sensitive, and fewer to high frequencies where the ear integrates energy over wider bands.

# 5 Global Normalization Overlap-Add (GN-OLA)

Dynamically switching window sizes creates complex overlap patterns that standard OLA cannot handle without amplitude modulation artifacts. We introduce Global Normalization to solve this.

Let $y_i[n]$ be the synthesized audio of the $i$-th frame windowed by $w_i[n]$. We accumulate two buffers:

$$B_{sig}[n] = \sum_i y_i[n] \cdot w_i[n] \tag{4}$$

$$B_{weight}[n] = \sum_i w_i[n] \tag{5}$$

The final output is reconstructed as:

$$\hat{s}[n] = \frac{B_{sig}[n]}{B_{weight}[n] + \varepsilon} \tag{6}$$

**Theorem 2** (GN-OLA Reconstruction Accuracy)**.** *Assuming the local approximation error in the frame is zero ($y_i[n] \approx s[n]$), the GN-OLA reconstruction is exact regardless of the window overlap configuration.*

*Proof.* Substituting $y_i[n] = s[n]$ into the accumulation equation:

$$B_{sig}[n] = \sum_i s[n] \cdot w_i[n] = s[n] \sum_i w_i[n] \tag{7}$$

Dividing by $B_{weight}[n] = \sum_i w_i[n]$ yields:

$$\hat{s}[n] = \frac{s[n] \sum w_i[n]}{\sum w_i[n]} = s[n] \tag{8}$$

This holds true for any arbitrary sequence of windows $w_i$, provided $\sum w_i[n] \neq 0$. □

# 6 Conclusion

The AMR-GLPC framework successfully integrates source-filter modeling with a multi-resolution analysis compatible with 44.1kHz audio. The granular window growth strategy optimizes the time-frequency trade-off, while the GN-OLA method ensures artifact-free reconstruction during resolution switching. Simulation on extensive pseudo-speech datasets confirms that the logarithmic Gabor dictionary preserves harmonic structure significantly better than linear counterparts.