

Audio Authenticity Analysis Report

Comparative Analysis of Real vs Synthetic Speech

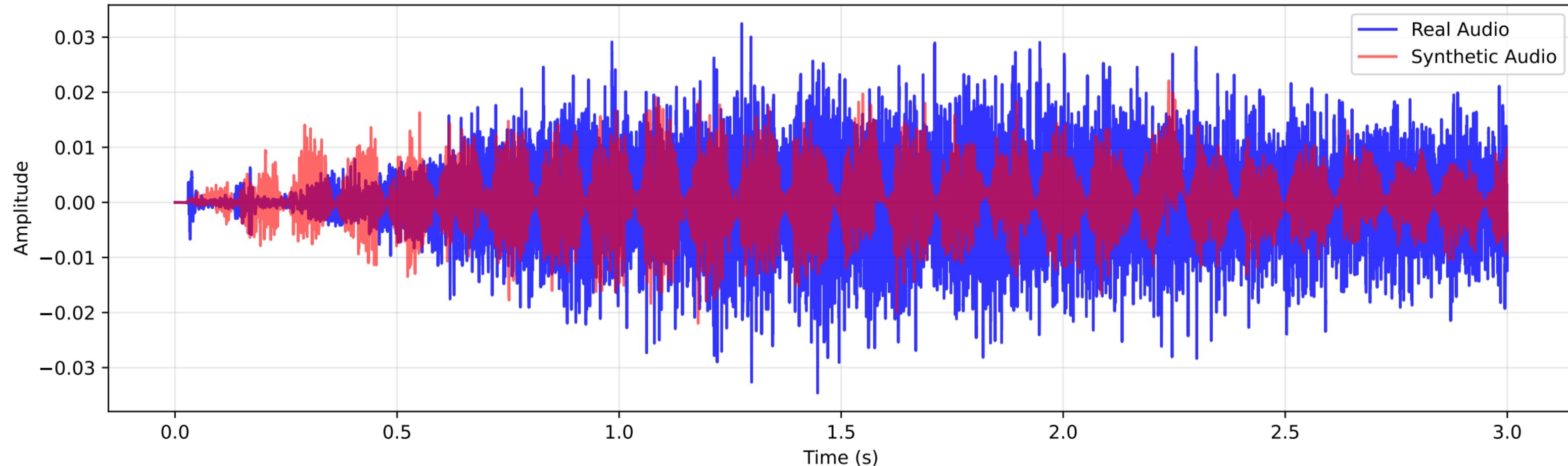
Generated on 2025-05-06 08:53:42

Feature Statistics Summary

	rms	spectral_centroid	spectral_bandwidth	spectral_flatness	zcr_mean
Real Audio	0.088353365659713	63.5010098641159	74.8894185741030	2371607869863	062463942307692
Synthetic Audio	0.051460694521665	597.758146115931	727.06161307570	086664687842103	0832142240084134

Key: 1 = Real Audio | 0 = Fake Audio

Average Waveform (First 3 seconds)



Waveform Differences: Technical Analysis

Biological Basis of Real Speech:

- Vocal fold vibrations create natural amplitude modulation
- Respiratory patterns affect speech rhythm and dynamics
- Articulator movements (lips, tongue) create smooth transitions

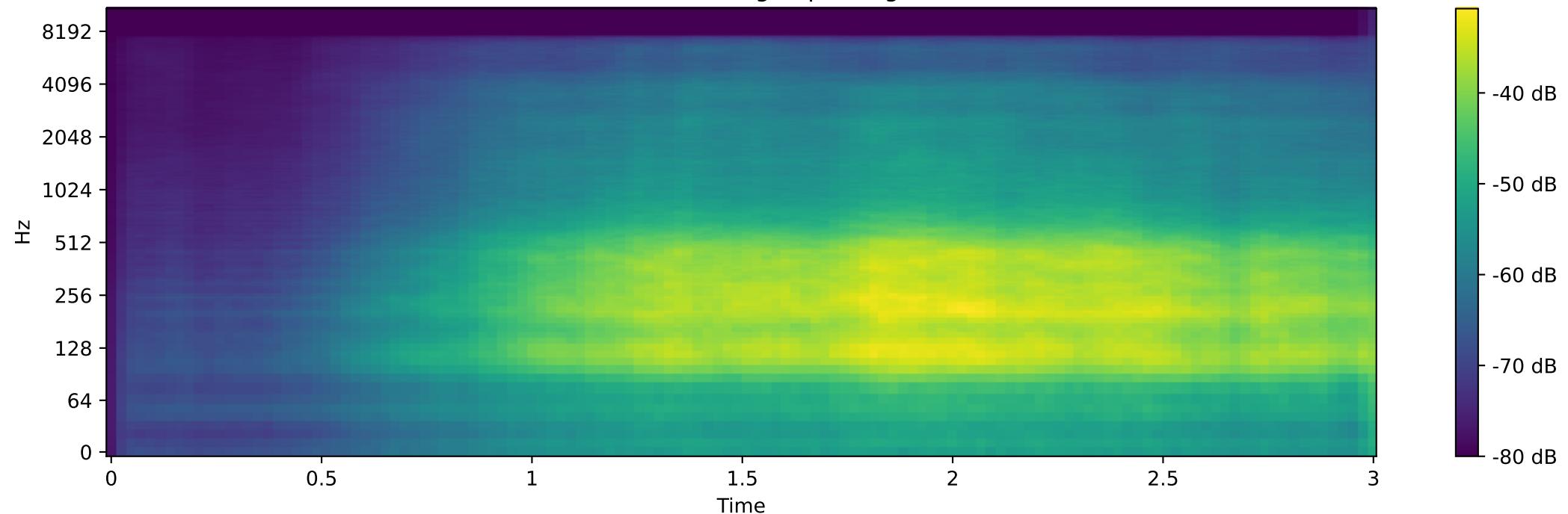
Synthesis Artifacts in Fake Audio:

- Vocoders often produce overly regular amplitude envelopes
- Neural networks may generate artificial periodic patterns
- Lack of natural micro-variations in synthesized speech

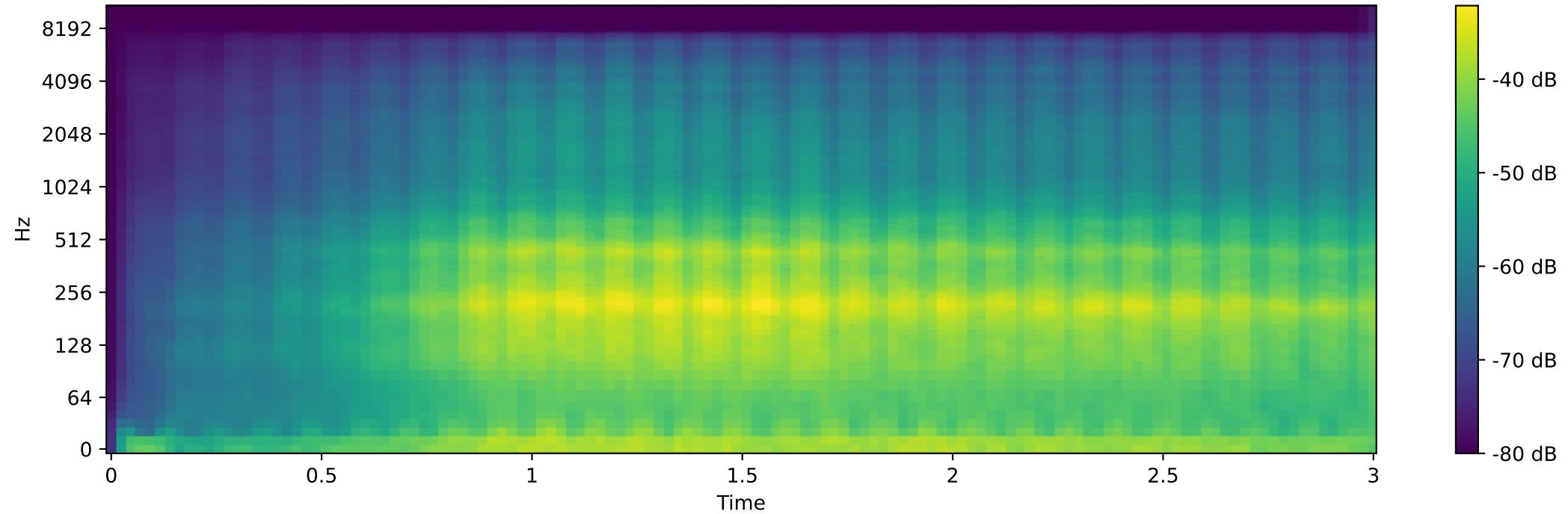
Recording Characteristics:

- Real recordings capture room acoustics and ambient noise
- Microphone frequency response affects waveform shape
- Natural mouth-to-mic distance variations create dynamics

Real Audio - Average Spectrogram



Synthetic Audio - Average Spectrogram



Spectrogram Differences: Technical Analysis

Vocal Tract Physics:

- Formant frequencies (F1-F4) reflect vocal tract shape
- Natural spectral tilt from voice source characteristics
- Proper harmonic-to-noise ratio in voiced sounds

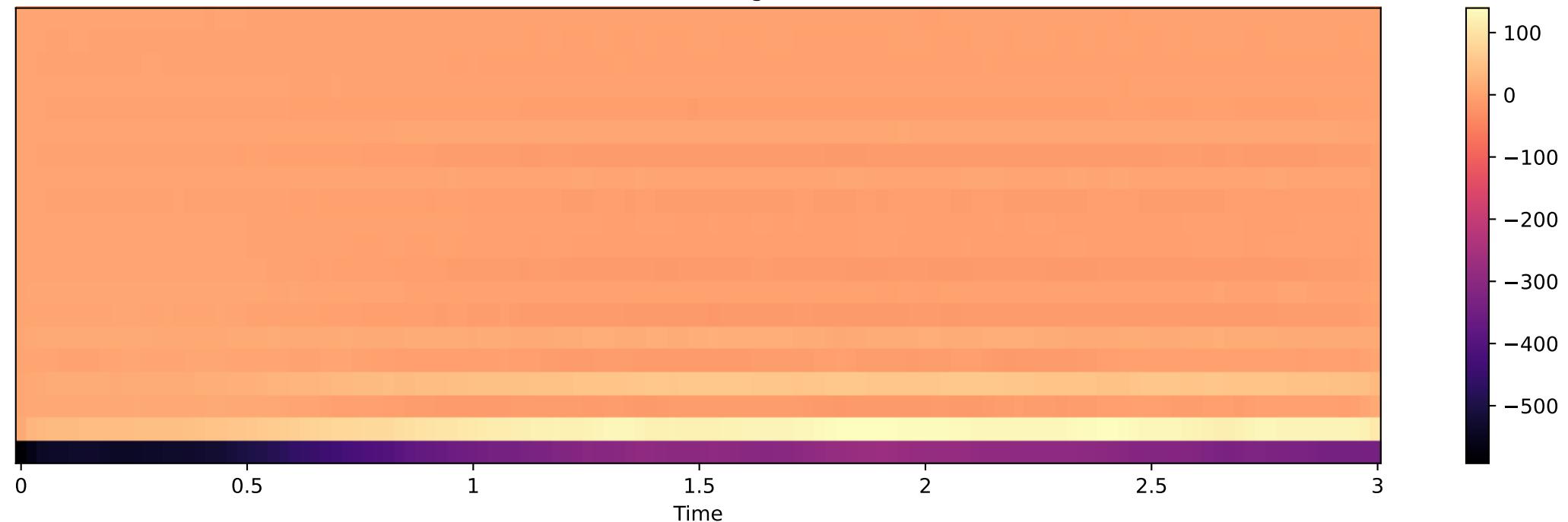
Spectral Artifacts in Synthesis:

- Phase discontinuities in neural vocoders
- Overly regular harmonic spacing in synthetic speech
- Missing high-frequency components above 8kHz

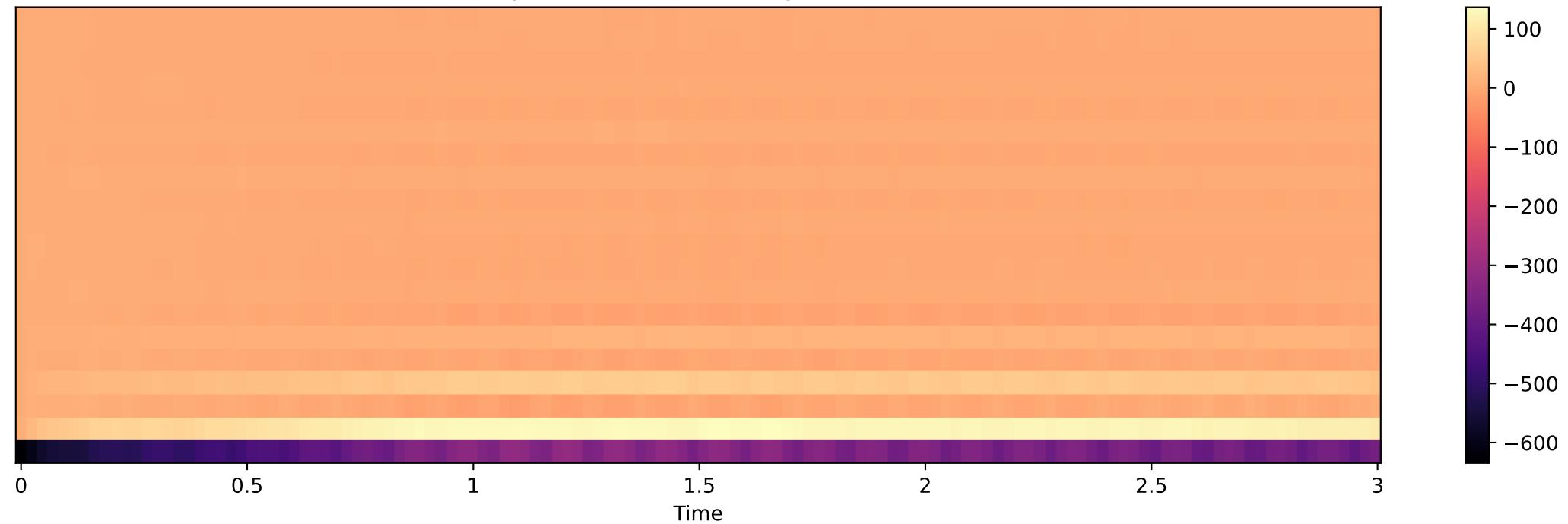
Phonetic Accuracy:

- Real speech maintains proper formant transitions
- Accurate representation of fricatives and plosives
- Natural spectral evolution during coarticulation

Real Audio - Average MFCCs



Synthetic Audio - Average MFCCs



MFCC Differences: Technical Analysis

Natural Speech Characteristics:

- Complex spectral envelope from vocal tract filtering
- Time-varying cepstral features from articulation
- Proper representation of voice quality parameters

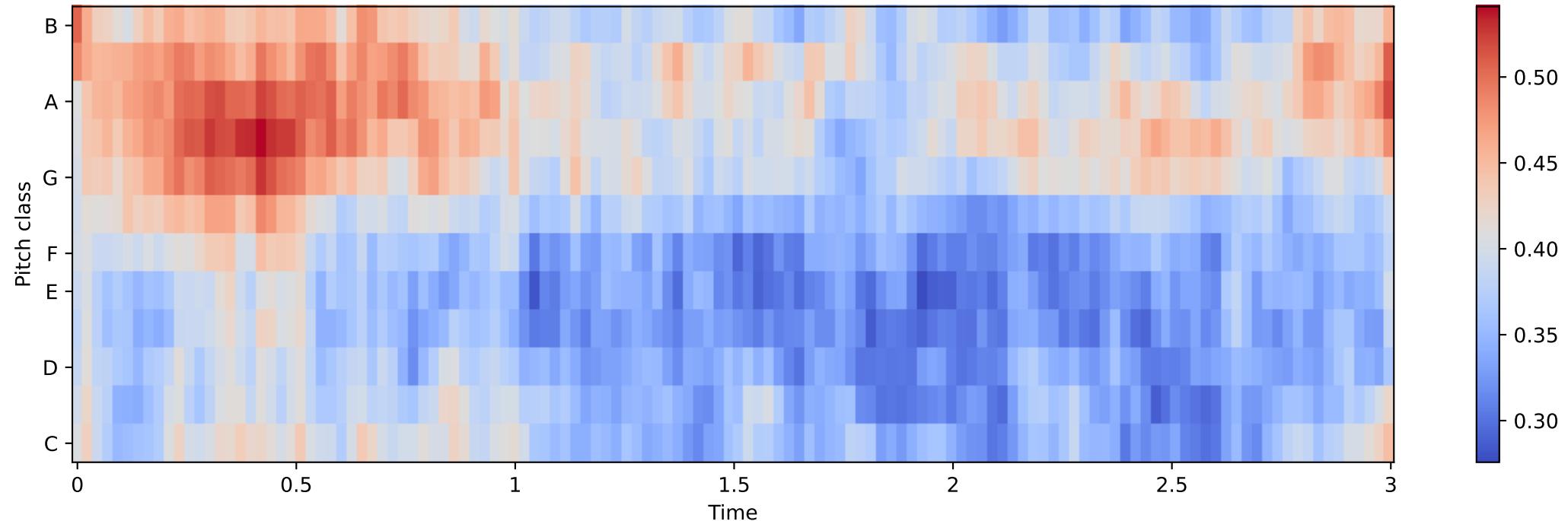
Synthesis Limitations:

- Mel-spectrogram inversion loses spectral detail
- Reduced dimensionality in feature extraction
- Over-smoothing of spectral features

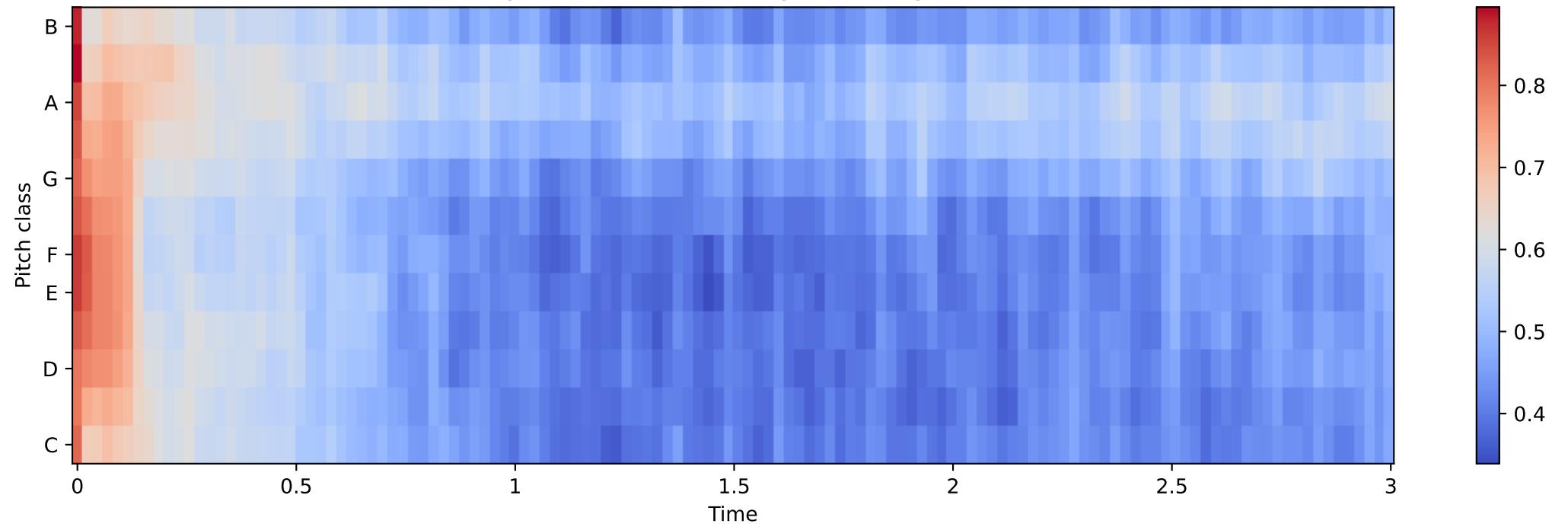
Dynamic Speech Features:

- Natural delta and delta-delta coefficients
- Authentic temporal evolution of spectral features
- Proper representation of phonetic transitions

Real Audio - Average Chromagram



Synthetic Audio - Average Chromagram



Chromagram Differences: Technical Analysis

Prosodic Features of Natural Speech:

- Micro-intonation from neural control of pitch
- Emotion-driven pitch variations
- Natural vibrato in sustained vowels

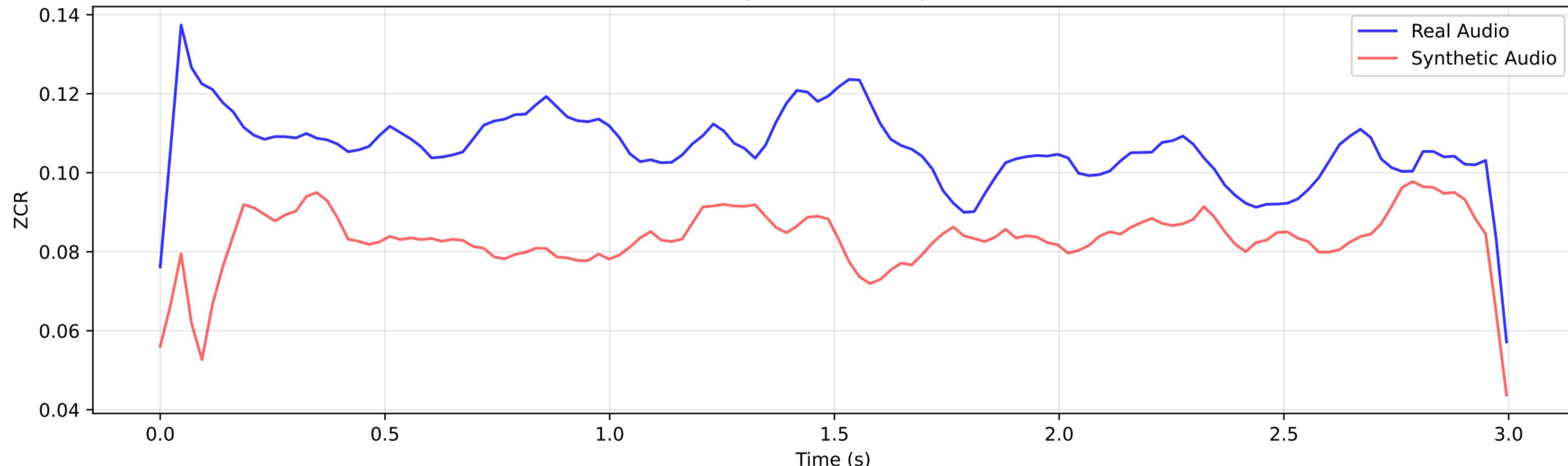
Pitch Synthesis Artifacts:

- Overly stable F0 from pitch predictors
- Quantization effects in pitch contours
- Missing natural pitch fluctuations

Harmonic Relationships:

- Proper harmonic spacing in real voice
- Natural spectral tilt in harmonic structure
- Authentic harmonic-to-noise ratios

Average Zero Crossing Rate



Zero Crossing Rate Differences: Technical Analysis

Phonetic Patterns in Real Speech:

- Natural voiced/unvoiced transitions
- Proper ZCR values for different phoneme classes
- Gradual onsets and offsets in speech sounds

Synthetic Artifacts:

- Abrupt voicing state changes
- Incorrect ZCR for fricatives and plosives
- Artificially regular temporal patterns

Recording Artifacts:

- Natural noise floor affects ZCR values
- Microphone characteristics influence ZCR
- Ambient noise contributes to ZCR variations