Audio Authenticity Analysis Report

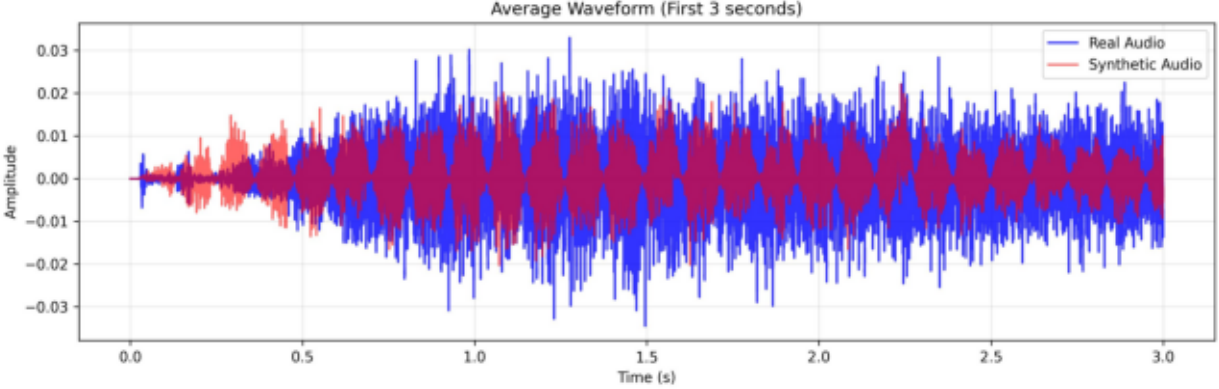
Comparative Analysis of Real vs Synthetic Speech

Generated: 2025-04-27 23:19:50

Feature Statistics Summary

	rms	spectral_centroid	spectral_bandwidth	spectral_flatness	zcr_mean
Real Audio	0.08839119970798	1779.982209070191	590.1422330534	0.032306760549545	0.109591346153846
Synthetic Audio	0.05146083235740	1604.656641487301	583.9897546960	0.020408866209909	0.083662259615384

Key: 1 = Real Audio | 0 = Fake Audio


Average Waveform (First 3 seconds)
Real Audio
Synthetic Audio
Amplitude
0.03
0.02
0.01
0.00
−0.01
−0.02
−0.03
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Time (s)

Waveform Analysis: Real vs Synthetic Speech

Real Audio Characteristics:

- Natural amplitude variations from breathing patterns and vocal cord vibrations

- Smooth transitions between phonemes (speech sounds)

- Gradual attack and decay of speech sounds

Fake Audio Indicators:

- Overly uniform amplitude modulation (too perfect)

- Abrupt transitions between sounds

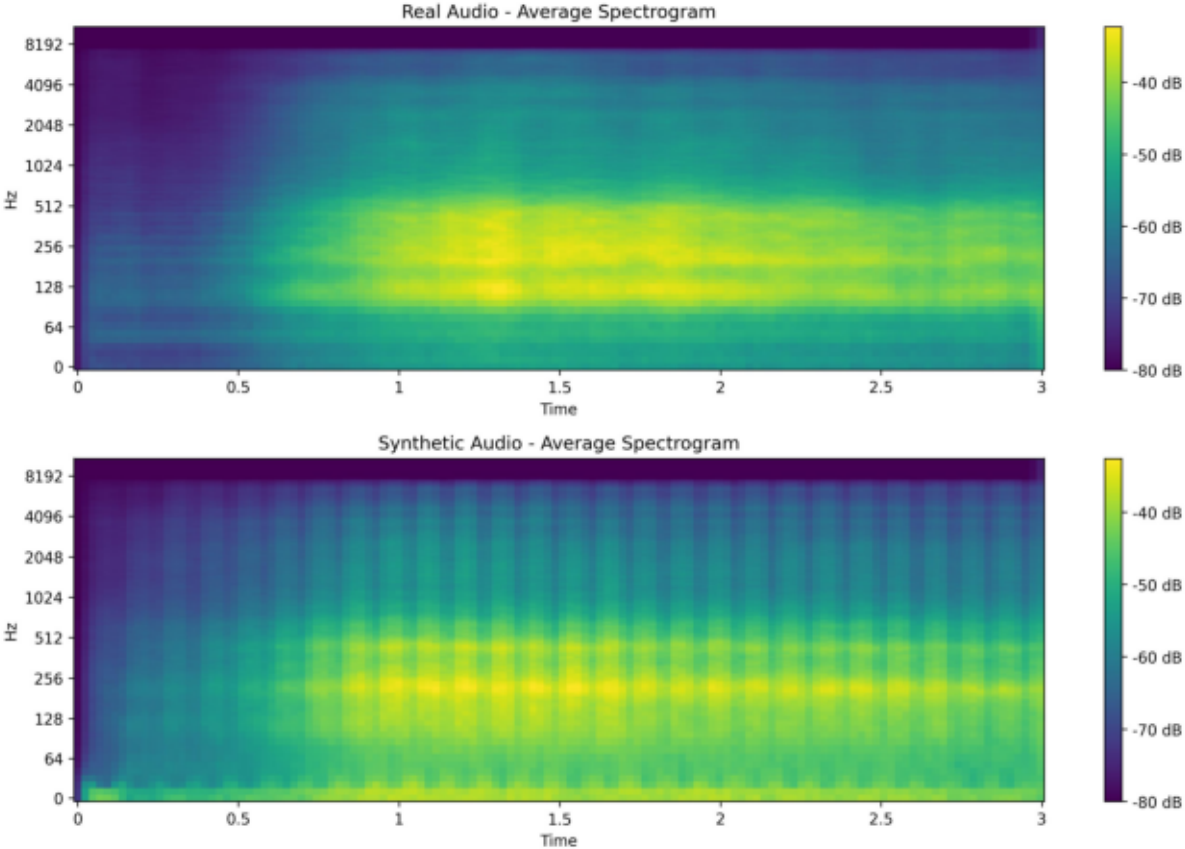
- Mechanical repetition patterns in longer samples

Key Differences to Observe:

1. Amplitude variation patterns (organic vs synthetic)

2. Transient smoothness in sound transitions

3. Presence of periodic artifacts in synthetic audio


Real Audio - Average Spectrogram
8192
4096
2048
1024
512
256
128
64
0
Hz
0
0.5
1
1.5
2
2.5
3
Time
-40 dB
-50 dB
-60 dB
-70 dB
-80 dB
Synthetic Audio - Average Spectrogram
8192
4096
2048
1024
512
256
128
64
0
Hz
0
0.5
1
1.5
2
2.5
3
Time
-40 dB
-50 dB
-60 dB
-70 dB
-80 dB

Spectrogram Analysis: Real vs Synthetic Speech

Real Audio Characteristics:

- Clear formant structure (vowel resonance bands)

- Natural harmonic spacing that follows physics of vocal cords

- Smooth frequency transitions between phonemes

Fake Audio Indicators:

- Blurred or distorted harmonics

- Missing high-frequency components (>8kHz)

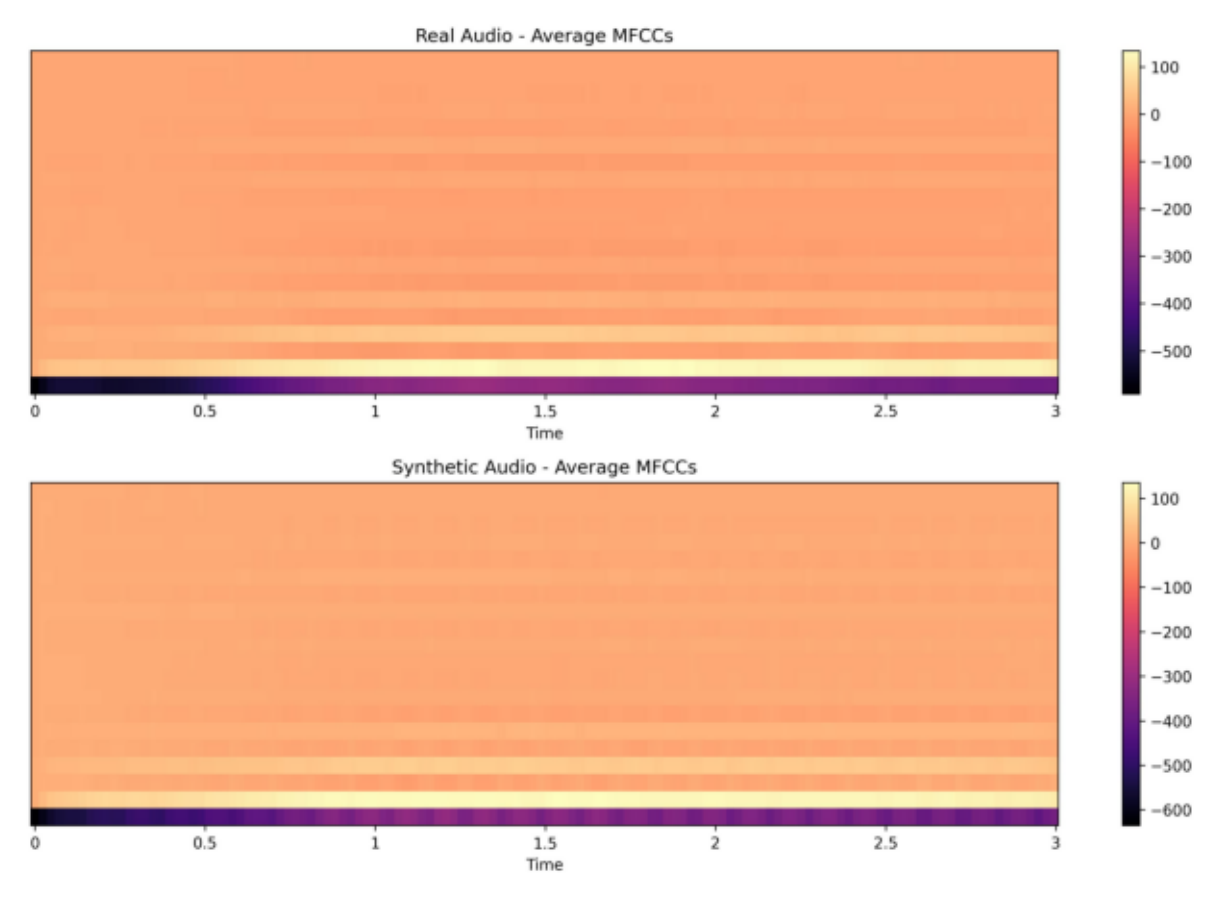
- Unnatural vertical stripes (frame boundary artifacts)

Key Differences to Observe:

1. Formant bandwidth and structure

2. High-frequency energy presence/absence

3. Harmonic-to-noise ratio patterns


Real Audio - Average MFCCs
100
0
−100
−200
−300
−400
−500
0
0.5
1
1.5
2
2.5
3
Time
Synthetic Audio - Average MFCCs
100
0
−100
−200
−300
−400
−500
−600
0
0.5
1
1.5
2
2.5
3
Time

MFCC Analysis: Real vs Synthetic Speech

Real Audio Characteristics:

- Complex, time-varying coefficient patterns

- Clear differentiation between speech sounds

- Natural trajectory of coefficient changes

Fake Audio Indicators:

- Overly smooth coefficient transitions

- Reduced dynamic range in coefficients

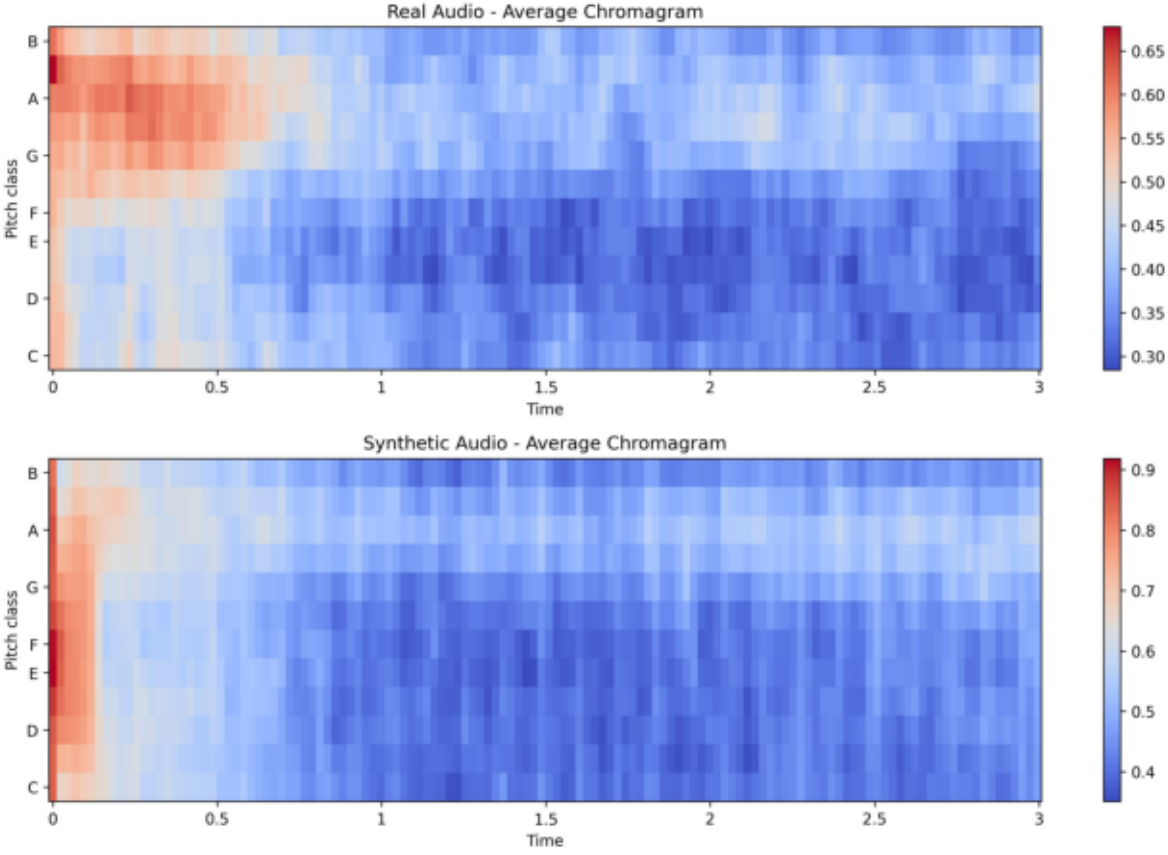
- Less differentiation between phoneme types

Key Differences to Observe:

1. Coefficient variance patterns

2. Dynamic range of MFCC values

3. Temporal evolution of features


Real Audio - Average Chromagram
Pitch class
B
A
G
F
E
D
C
0
0.5
1
1.5
2
2.5
3
Time
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0.30
Synthetic Audio - Average Chromagram
Pitch class
B
A
G
F
E
D
C
0
0.5
1
1.5
2
2.5
3
Time
0.9
0.8
0.7
0.6
0.5
0.4

Chromagram Analysis: Real vs Synthetic Speech

Real Audio Characteristics:

- Natural pitch variations (micro-intonation)

- Emotion-driven pitch contours

- Proper harmonic energy distribution

Fake Audio Indicators:

- Overly stable pitch (mechanical sounding)

- Quantized-looking pitch contours

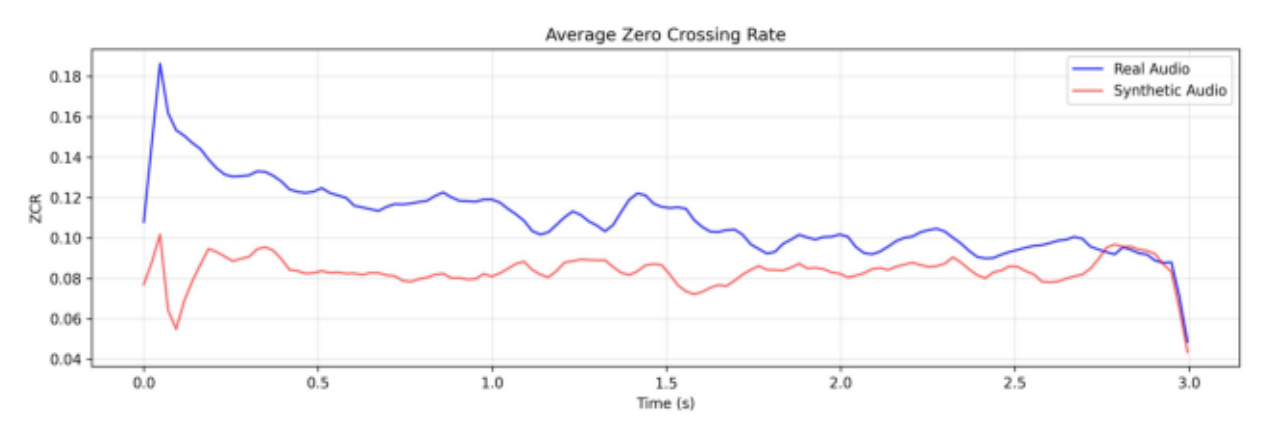
- Incorrect harmonic energy distribution

Key Differences to Observe:

1. Pitch stability and variation patterns

2. Chroma band energy distribution

3. Temporal patterns in pitch changes


Average Zero Crossing Rate
Real Audio
Synthetic Audio
ZCR
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Time (s)

Zero Crossing Rate Analysis: Real vs Synthetic Speech

Real Audio Characteristics:

- Natural alternation between voiced/unvoiced sounds

- Proper ZCR values for different phoneme types

- Gradual transitions between sound types

Fake Audio Indicators:

- Abrupt voicing state changes

- Incorrect ZCR for fricatives (s/sh sounds)

- Artificially regular temporal patterns

Key Differences to Observe:

1. Voicing transition patterns

2. Phoneme-specific ZCR values

3. Temporal distribution of ZCR values