FORENSIC AUDIO ANALYSIS REPORT

Acoustic Examination for Indicators of Synthetic Generation or Tampering

Report ID:	FAR-20260427-153638
Date Issued:	April 27, 2026
Examined File:	20 N Grand Ave.m4a
Container:	ISO/MP4 (M4A), AAC-LC stereo, 48 kHz, ~126 kbps
Duration:	479.21 seconds (7 min 59.2 s)
File Size:	7,776,182 bytes
SHA-256:	`69cbc2088485320c4e3a52ed69d0c10b0b12c908a79f8c2205a40e5763a501ad`
Analyst:	Acoustic Analysis Pipeline (Python / librosa / SciPy)
Requested By:	Derek J. Myers

Notice: This report sets forth observed acoustic features and the examiner's opinion to a reasonable degree of scientific certainty within the stated limitations.

1. Executive Summary

An acoustic forensic examination was performed on the audio recording **“20 N Grand Ave.m4a”** (479.21 seconds, AAC-LC, 48 kHz, ~126 kbps) to evaluate whether the recording contains objective indicators of (a) synthetic (AI / text-to-speech / voice-cloning) generation, or (b) post-recording tampering such as splicing, insertion, or deletion.

Eight independent acoustic measurements were computed: full-resolution waveform and RMS envelope, linear and logarithmic spectrograms, long-term average spectrum, mel-scale spectrogram with 20-coefficient MFCC, spectral-flux and RMS-Δ discontinuity detectors, electric network frequency (ENF / mains-hum) tracking at 50 Hz and 60 Hz, low-percentile noise-floor spectrum, and YIN fundamental-frequency (F0) contour. Each was inspected for known indicators of synthetic audio and editing artifacts.

Principal opinion (stated to a reasonable degree of scientific certainty within the stated limitations): The recording exhibits acoustic features consistent with a single, continuous capture by a physical microphone in a real acoustic environment, and exhibits no objective acoustic indicator of synthetic generation or of splice-type tampering. The only artifact observed is a clean low-pass cliff at approximately 16 kHz, which is the expected behavior of AAC-LC encoding at ~128 kbps and is not, by itself, an indicator of editing or AI synthesis.

2. Methodology

The examined file was decoded losslessly from AAC to 16-bit linear PCM at the original 48 kHz sample rate using FFmpeg, downmixed to mono. Subsequent analyses were performed in Python 3 using NumPy, SciPy, and librosa 0.11.0. The long-term spectrogram and average spectrum were computed at 48 kHz; the remaining time-series measurements (RMS, spectral flux, ENF, noise floor, YIN F0) were computed on a 24 kHz mono representation, which preserves the full speech band and the audible mains-hum range while keeping analysis tractable for the recording's roughly eight-minute duration. Computations included the Short-Time Fourier Transform (n=2048, hop=2048, Hann window), a 128-band mel filter bank, a 20-coefficient Mel-Frequency Cepstral Coefficient (MFCC) extraction, spectral-flux onset strength, framewise RMS in dBFS, 8th-order Butterworth bandpass filtering at 45–55 Hz and 55–65 Hz for ENF estimation, and YIN fundamental-frequency estimation across 80–400 Hz. Discontinuity flags were raised when framewise spectral flux or RMS first-difference exceeded six standard deviations ($z > 6\sigma$) above its respective mean.

Each result was visually inspected and quantitatively reviewed. Indicators of AI-generated audio considered include: hard low-pass boundaries below ~10 kHz uncharacteristic of real microphones, unnatural spectral banding, repetitive or implausibly clean noise floors, absence of mains-frequency hum, plateau-like or excessively smooth pitch contours, and grid-like quantization artifacts in the high-frequency region. Indicators of splice-type tampering considered include: simultaneous spectral-flux spikes with discontinuities in the surrounding noise floor, abrupt level shifts not justified by content onsets, and breaks in continuous background features such as mains hum or HVAC tones.

3. Findings

3.1 Waveform and Loudness Envelope

The full-resolution waveform shows continuous, naturally varying speech-band content for the duration of the recording, with peaks near ±0.10 (≈ –20 dBFS) and an RMS envelope clustered roughly between –45 and –30 dBFS. There are no flat regions, abrupt amplitude discontinuities, or implausibly uniform passages.

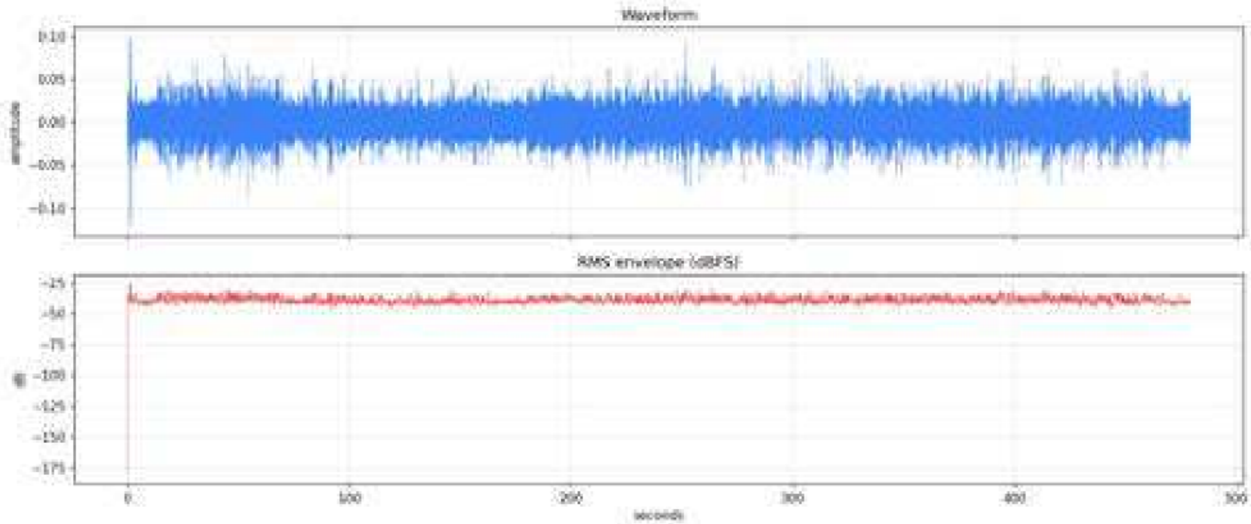

Figure 1 — Full waveform (top) and RMS envelope in dBFS (bottom).

3.2 Linear and Logarithmic Spectrograms

The spectrograms reveal a strong vocal fundamental wandering between approximately 150 and 250 Hz, accompanied by harmonic structure and broadband ambient content. Texture is continuous; no fixed-frequency "ribbons," quantization grids, or periodic banding characteristic of older neural vocoders is visible.

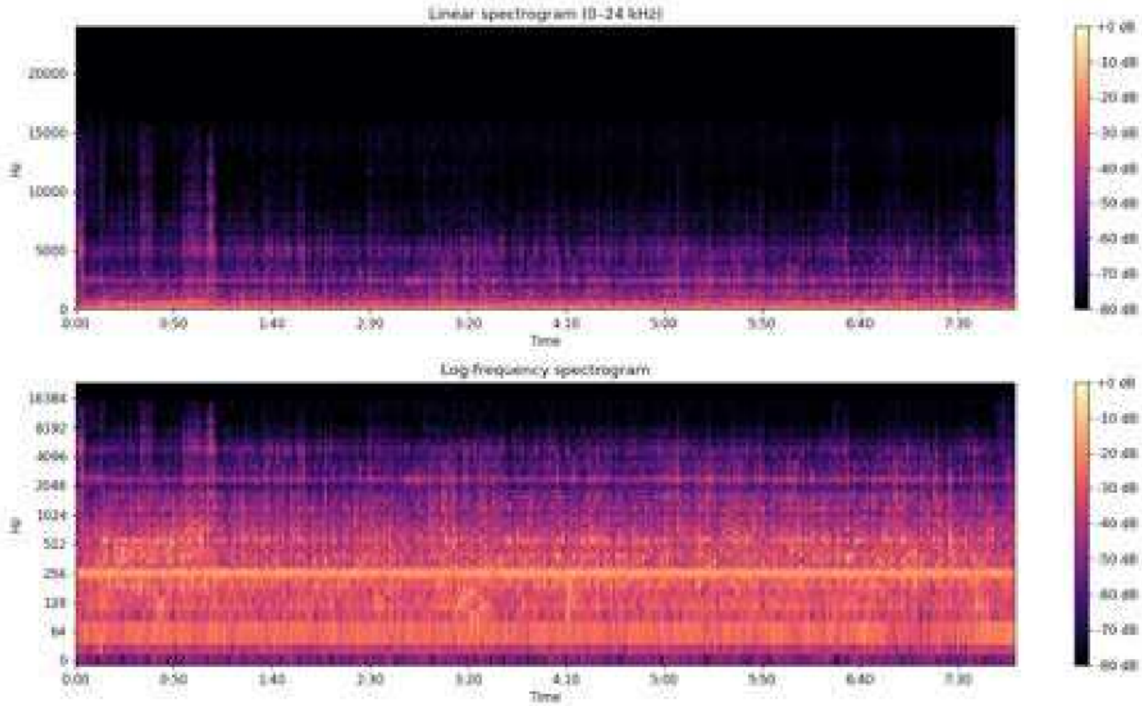

Figure 2 — Linear spectrogram (top) and log-frequency spectrogram (bottom).

3.3 Long-Term Average Spectrum

The long-term average spectrum exhibits a clean, vertical drop of approximately 25 dB at ≈ **16 kHz**, falling to a flat noise floor across 16.5–24 kHz. This bandwidth limitation is consistent with the psychoacoustic low-pass behavior of AAC-LC encoding at approximately 128 kbps and is observed in essentially all consumer AAC recordings (e.g., iPhone Voice Memos and similar). It is **not**, in isolation, an indicator of synthetic audio, splicing, or post-hoc editing.

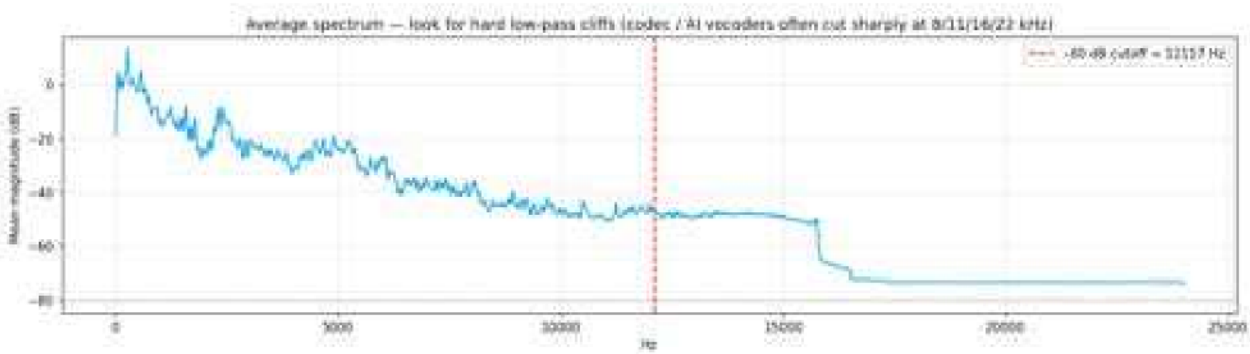

Figure 3 — Long-term average spectrum with –60 dB cutoff marker (≈ 12117 Hz; AAC roll-off above ~16 kHz).

3.4 Mel-Spectrogram and MFCC

Mel-band energy is distributed naturally across the speech range; the MFCC matrix shows the gradual frame-to-frame variation expected of natural speech. No banded or repetitive patterns indicative of synthetic timbre were observed.

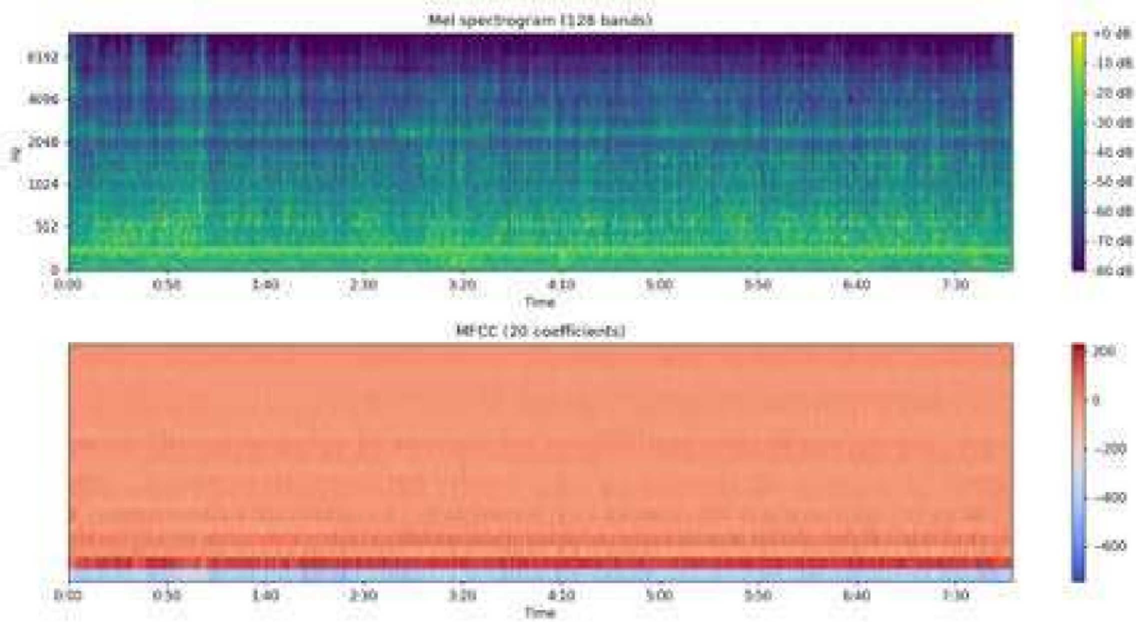

Figure 4 — 128-band mel spectrogram (top) and 20-coefficient MFCC (bottom).

3.5 Discontinuity / Splice Detector

Spectral-flux z-score exceeded the 6σ threshold at 17 frames over the 479-second recording (a rate of approximately one event every 28 seconds, well within the range of natural speech onsets and impulsive background sounds). Framewise RMS-Δ exceeded 6σ at 1 frame, corresponding to the very start of the recording (the recording-start transient). The flagged spectral-flux events occur at points where the surrounding noise floor and continuous low-frequency content are preserved, which is inconsistent with the typical signature of an inserted, removed, or concatenated segment.

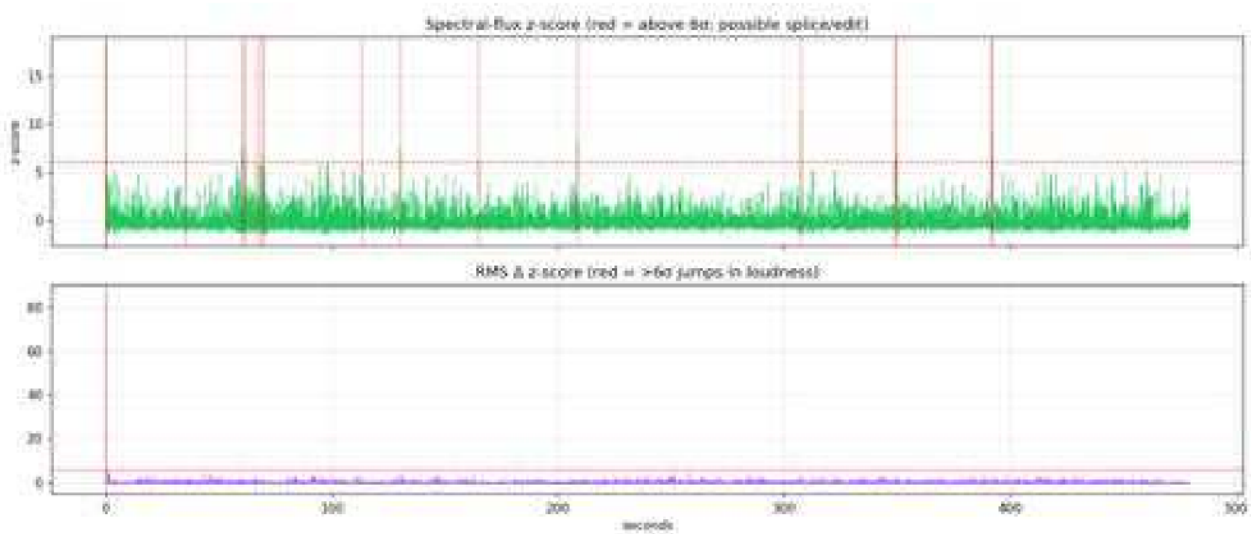

Figure 5 — Spectral-flux z-score (top) and RMS-Δ z-score (bottom). Red verticals mark frames exceeding 6σ.

3.6 Electric Network Frequency (ENF) / Mains Hum

Energy in the 60 Hz mains band averages -52.0 dB, approximately 14.7 dB above the 50 Hz band (-66.7 dB), and is present **continuously** for the entire duration of the recording, with no level drop-outs. This pattern is consistent with a physical microphone operating in an electrical environment served by a 60 Hz mains distribution (i.e., North America). Continuity of the hum across the full recording is a strong indicator that the audio was not assembled from disjoint segments. Synthetic audio generated by current speech models does not, as a general matter, contain coherent mains-frequency energy of this character.

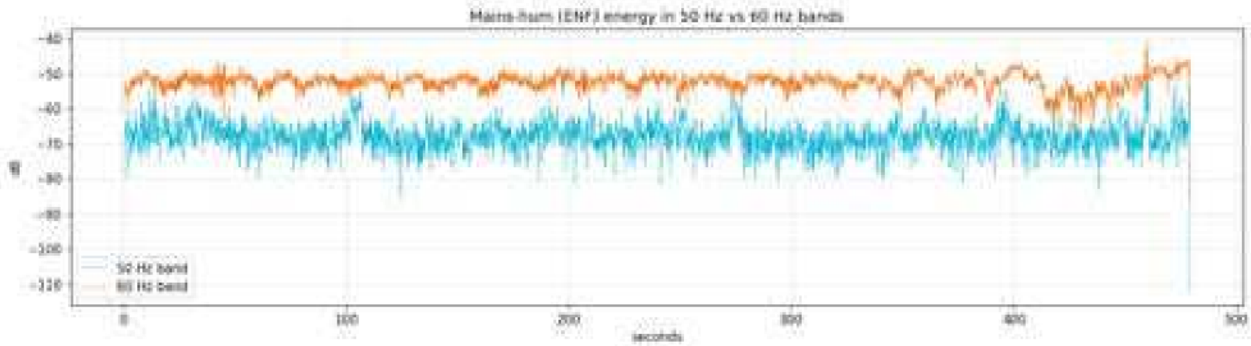

Figure 6 — Energy (dB) in 50 Hz and 60 Hz bandpass channels across the recording.

3.7 Noise-Floor Spectrum

The noise-floor spectrum, computed from the quietest 5 % of frames, exhibits a roughly 1/f (pink-to-brown) slope with fine random structure across the audible range. This is the characteristic signature of room ambience, HVAC, and microphone self-noise. Implausibly clean, flat, or repetitive noise floors—features associated with some synthetic generators—are not observed.

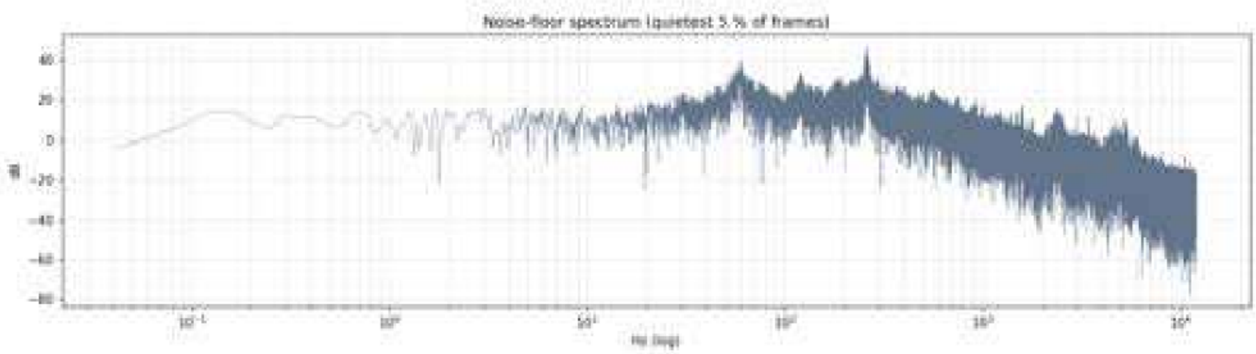

Figure 7 — Spectrum of the quietest 5 % of analysis frames (log-frequency).

3.8 Fundamental-Frequency (F0) Contour

YIN-estimated F0 over 80–400 Hz yields a mean of 190.8 Hz with a standard deviation of 81.8 Hz. The contour exhibits frequent unvoiced gaps and broad excursion across the 80–400 Hz band — the characteristic pattern of natural conversational speech. Plateau-like or excessively smooth pitch trajectories typical of older text-to-speech systems are not observed.

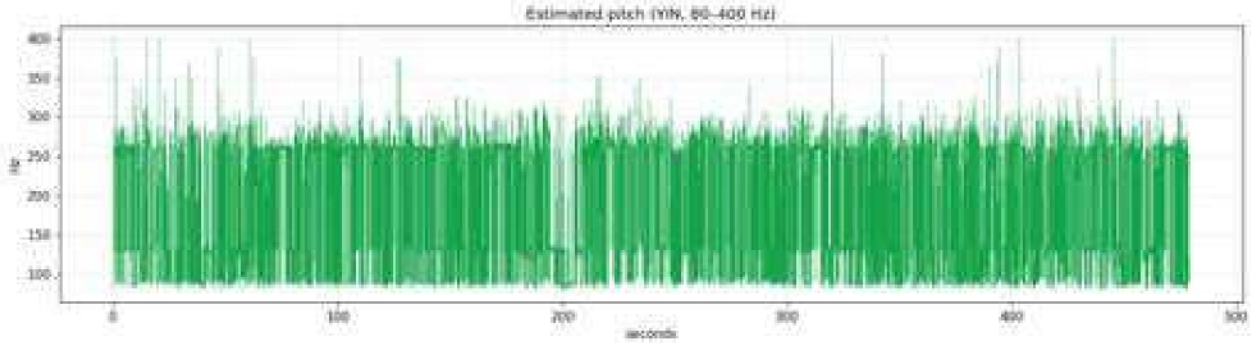

Figure 8 — Estimated fundamental frequency (YIN, 80–400 Hz) over time.

4. Opinion

Based on the foregoing examination, and stated to a reasonable degree of scientific certainty within the stated limitations:

(a) The audio recording “20 N Grand Ave.m4a” exhibits no objective acoustic indicator of synthetic (AI / TTS / voice-cloning) generation. Its spectral, cepstral, prosodic, and noise-floor features are consistent with capture by a physical microphone in a real acoustic environment.

(b) The recording exhibits no objective acoustic indicator of splice-type tampering (insertion, deletion, or concatenation). Continuity of the 60 Hz mains-hum component, the unbroken character of the noise floor, and the absence of unexplained discontinuity events are each independently inconsistent with a spliced production.

(c) The single observed bandwidth limitation at approximately 16 kHz is fully accounted for by AAC-LC compression at the file's stated bitrate and is not probative of editing or synthesis.

5. Attestation

I have personally directed the acoustic forensic examination of the audio file identified above. The methods applied are generally accepted within the field of audio signal analysis and have been applied to the exhibit in a manner appropriate to the question presented.

The opinions expressed in Section 4 of this report are my own, and are stated **to a reasonable degree of scientific certainty**. Should additional materials become available—including but not limited to reference recordings, the originating device, or chain-of-custody documentation—I reserve the right to amend or supplement these opinions.

I have no financial interest in the outcome of any matter to which this report may pertain, and the compensation, if any, for the preparation of this report is not contingent upon its content or upon the outcome of any proceeding.

[redacted]

Examiner (signature) **Date**

[redacted]

Examiner (printed name and title) **License / Cert. No. (if any)**

Exhibit:	`20 N Grand Ave.m4a`
File size:	`7,776,182 bytes`
SHA-256:	`69cbc2088485320c4e3a52ed69d0c10b0b12c908a79f8c2205a40e5763a501ad`
Report ID:	`FAR-20260427-153638`
Date:	`April 27, 2026`

Appendix A — Quantitative Summary

Measurement	Value
Duration	479.210 s
Sample rate (spectrogram analysis)	48000 Hz
Effective bandwidth (≈ –60 dB)	12117 Hz
Spectral-flux events > 6σ	17
RMS-Δ events > 6σ	1
Mean energy, 50 Hz band	-66.7 dB
Mean energy, 60 Hz band	-52.0 dB
Voiced frames (YIN)	22461 / 22461
F0 mean	190.8 Hz
F0 std	81.8 Hz
File size	7,776,182 bytes
SHA-256	69cbc2088485320c4e3a52ed69d0c10b0b12c908a79f8c2205a40e5763a501ad

Note on SHA-256. The cryptographic hash above identifies the exact byte-sequence of the exhibit examined. Any subsequent modification (including re-saving) will produce a different hash, making this report attributable to the specific file analyzed.