

TECHNICAL METHODOLOGY REPORT

Acoustic Fundamental Frequency Extraction and Comparative Baseline Voice Stress Analysis

*Prepared in support of: VSA Examination Report WEI-2026-001
Craig Kootsillas, Independent Investigative Analyst | May 13, 2026*

ABSTRACT

This document describes in technical detail the methodology, toolchain, signal processing pipeline, and analytical framework used to produce Voice Stress Analysis Examination Report WEI-2026-001, concerning audio segments extracted from the August 8, 2023 Palm Coast City Council meeting. The analysis employs Praat acoustic phonetics software (Boersma & Weenink, University of Amsterdam) to extract fundamental frequency (F0) contours from two audio segments: a neutral baseline segment and a semantically loaded target segment. Quantitative comparison of pitch mean, range, variance, and anomalous event distribution is used to identify statistically notable differences between the two conditions. All tools used in this analysis are open-source, peer-reviewed in the academic literature, and freely reproducible by any qualified examiner.

1. THEORETICAL FRAMEWORK

1.1 Fundamental Frequency as a Stress Indicator

The fundamental frequency (F0) of voiced speech — commonly referred to as pitch — is produced by the vibration of the vocal folds under tension regulated by the laryngeal muscles. F0 is directly measurable as an acoustic signal and serves as the primary physiological variable examined in voice stress analysis.

The theoretical basis for using F0 as a stress indicator is grounded in the relationship between the autonomic nervous system and laryngeal muscle tone. Under psychological stress, the sympathetic nervous system activates the fight-or-flight response, producing systemic effects including elevated heart rate, increased respiratory rate, and heightened muscle tension throughout the body — including the laryngeal muscles.

Elevated laryngeal muscle tension produces measurable acoustic effects: increased mean F0, wider F0 range, greater F0 variability (as measured by jitter and standard deviation), and the characteristic micro-tremor patterns that form the basis of Computer Voice Stress Analyzer (CVSA) interpretation [1].

Additionally, the research literature documents a second stress-related pattern: effortful suppression of F0 variation, in which a speaker under psychological pressure exercises conscious or unconscious control over vocal output, producing an abnormally flat or compressed pitch contour compared to their natural baseline [2].

Both elevated and artificially compressed F0 patterns are recognized in the forensic phonetics literature as potentially stress-significant.

1.2 Baseline Comparison Methodology

The most methodologically defensible approach to voice stress analysis is the intra-speaker baseline comparison, in which the subject's own neutral speech serves as the control condition. This methodology controls for: inter-speaker variation (individual differences in baseline F0); acoustic environment (same recording setup); recording equipment characteristics; and temporal proximity (same session minimizes speaker state variation).

The present analysis employs this methodology. Both the baseline segment and the target segment were extracted from the same recording session, on the same date, in the same acoustic environment, from the same speaker, using identical recording equipment. The only independent variable is the semantic content and psychological condition of the speaker at each point in time [3].

2. SOFTWARE AND TOOLCHAIN

2.1 Praat Acoustic Analysis Software

The primary analytical instrument used in this examination is Praat (Dutch: "talk"), an open-source acoustic phonetics program developed by Paul Boersma and David Weenink at the Institute of Phonetic Sciences, University of Amsterdam, Netherlands. Praat is the de facto standard tool for acoustic analysis in academic linguistics, forensic phonetics, and speech pathology research worldwide.

Parameter	Value
Software	Praat v6.3.07
Release Date	February 6, 2023
Developers	Paul Boersma & David Weenink, University of Amsterdam
License	GNU General Public License (open source, freely available)
Platform	Linux (Ubuntu/Debian 12), installed via: <code>sudo apt install praat</code>
Official Source	https://www.praat.org

2.2 Supporting Tools

The following additional tools were used in the audio extraction and processing pipeline:

Tool	Version	Purpose
FFmpeg	5.1.8	Video demuxing and audio stream extraction

yt-dlp	2024.x	Timestamped segment extraction from YouTube source
Praat Script	Custom	Automated F0 extraction and CSV output

3. SIGNAL PROCESSING PIPELINE

3.1 Source Material Acquisition

The source audio was obtained from the City of Palm Coast's official YouTube channel, which maintains a publicly accessible archive of all city council meetings. The August 8, 2023 meeting recording (YouTube ID: ZLW6Pgc3SfM) was accessed at the following URL:

```
https://www.youtube.com/watch?v=ZLW6Pgc3SfM
```

Two discrete time segments were identified for analysis based on meeting transcript review:

```
BASELINE SEGMENT: Meeting timestamp 2:11:46 - 2:12:36 (50 seconds)
TARGET SEGMENT: Meeting timestamp 2:28:53 - 2:30:30 (97 seconds)
```

3.2 Video Segment Extraction

Each video segment was extracted using yt-dlp, an open-source media downloader, using custom shell scripts that accept meeting identifier, timestamp range, and output filename parameters. The extraction command structure is as follows:

```
yt-dlp --download-sections "*HH:MM:SS-HH:MM:SS" \  
-f 'bv*+ba/b' \  
--merge-output-format mp4 \  
"https://www.youtube.com/watch?v=VIDEO_ID" \  
-o "output_filename.mp4"
```

The resulting MP4 files contain the H.264 video stream and Opus audio stream as packaged by YouTube's CDN infrastructure. File integrity was verified by playback prior to audio extraction.

3.3 Audio Stream Extraction

Audio was extracted from each MP4 segment using FFmpeg, demuxing the audio stream and converting from the Opus compressed codec to uncompressed PCM 16-bit signed little-endian format (pcm_s16le) at the source sample rate of 48,000 Hz. Uncompressed PCM was required for accurate fundamental frequency analysis, as Opus codec compression introduces perceptual artifacts that can distort F0 estimation algorithms.

```
ffmpeg -i input_segment.mp4 \  
-vn \  
-acodec pcm_s16le \  
output_audio.wav \  
-y
```

Extraction parameters:

```
Input codec:    Opus (48 kHz, stereo, 103 kb/s)  
Output codec:  PCM signed 16-bit little-endian  
Sample rate:   48,000 Hz  
Channels:      Stereo  
Bit depth:     16-bit  
Output format: WAV (RIFF)
```

Baseline audio file size: approximately 9.4 MB (50 seconds at 1,536 kb/s uncompressed).

Target audio file size: approximately 18.3 MB (97 seconds at 1,536 kb/s uncompressed).

3.4 Fundamental Frequency Extraction — Praat Script

F0 extraction was performed using a custom Praat Script that reads the WAV audio file, applies the autocorrelation-based pitch detection algorithm, and outputs a timestamped CSV file of detected pitch values. The complete script is reproduced below:

```
# Praat F0 Extraction Script  
# WEI-2026-001 | Craig Kootsillas | May 13, 2026  
  
input_file$ = "/tmp/input.wav"  
output_file$ = "/tmp/praat-output.txt"  
  
# Read WAV file into Praat object buffer  
Read from file: input_file$  
  
# Apply autocorrelation pitch detection algorithm  
# Parameters: time step (0=auto), floor 75 Hz, ceiling 600 Hz  
To Pitch: 0, 75, 600  
  
# Write CSV header  
writeFileLine: output_file$, "time,pitch"  
  
# Extract pitch value at each analysis frame  
frames = Get number of frames  
for i from 1 to frames  
  t = Get time from frame number: i  
  p = Get value in frame: i, "Hertz"  
  # Write only voiced frames (undefined = unvoiced/silence)  
  if p <> undefined  
    appendFileLine: output_file$, fixed$(t, 3) + "," + fixed$(p, 2)  
  endif  
endfor
```

3.5 Pitch Detection Algorithm Parameters

The Praat pitch detection algorithm uses an autocorrelation method (SHR — Subharmonic-to-Harmonic Ratio) with the following configured parameters:

Parameter	Value	Rationale
Time step	0 (auto)	Auto-set to 75% of analysis window (approximately 10ms)
Pitch floor	75 Hz	Below normal adult male F0 range (85-155 Hz typical)
Pitch ceiling	600 Hz	Above typical falsetto range; captures all physiological voicing
Voiced frame threshold	Default	Praat standard 0.45 autocorrelation threshold
Output unit	Hertz (Hz)	Linear frequency scale; standard for F0 research

The autocorrelation method used by Praat is well-validated in the phonetics literature and is considered more robust against noise contamination than simple zero-crossing or cepstrum-based methods [4].

3.6 Output Data Format

The Praat script produces a comma-separated values (CSV) file with two columns: time (seconds, three decimal places) and pitch (Hz, two decimal places). Only voiced frames are included; silence and unvoiced consonants produce no output row. Sample output:

```
time,pitch
5.310,206.74
5.320,197.45
5.330,190.43
5.340,185.57
...
10.290,529.57  <- anomalous artifact (likely table tap or microphone bump)
...
14.090,438.94  <- anomalous cluster at semantically loaded phrase
14.100,441.87
14.110,456.16
14.120,472.41
```

4. ANALYTICAL FRAMEWORK AND ARTIFACT HANDLING

4.1 Artifact Identification

The acoustic environment of a city council chamber introduces several potential sources of non-vocal pitch artifacts: microphone handling noise, table vibrations, paper shuffling, and background HVAC systems. The Praat pitch detection algorithm may incorrectly identify the fundamental frequency of these non-vocal sounds as voiced speech.

Artifacts are typically identifiable by the following characteristics: abrupt onset without preceding pitch glide; values significantly above or below the speaker's established pitch range; very short duration (1-2 frames, approximately 10-20ms); and lack of formant structure in adjacent spectral analysis.

For the purposes of this preliminary analysis, values above 400 Hz are treated as potentially artifactual given the speaker's established normal speech range of approximately 75-250 Hz. However, this threshold is applied conservatively — values above 400 Hz are noted and flagged rather than silently discarded, as their temporal location relative to key speech events is itself analytically relevant.

4.2 The Significance of Temporally Localized Anomalous Events

A critical distinction in voice stress analysis is between randomly distributed anomalous events and temporally localized anomalous event clusters. In both the baseline and target segments, high-frequency values above the speaker's normal range appear. The analytical question is whether these events cluster at semantically significant moments or are distributed randomly.

In the baseline segment, anomalous values above 400 Hz appear at timestamps 7.33, 7.34, 7.82, 7.83, 10.29-10.71, 10.58-10.71, and 11.61-11.62 seconds. Examination of the meeting transcript at these points does not reveal semantically loaded content — the events are consistent with random environmental artifact.

In the target segment, the most notable anomalous cluster appears at timestamps 14.09-14.24 seconds, corresponding to meeting timestamp 2:29:07 and the spoken phrase "these are in no particular order." Additional elevated-value clusters appear at timestamps 87.55-87.61, 89.33-89.43, 90.51-90.56, 91.38-91.45, and 93.15-93.31 seconds. These later clusters occur during the question-and-answer exchange with Vice Mayor Pontieri and may warrant further investigation.

The temporal localization of the primary cluster at the key denial phrase — rather than at a random point in the segment — is the analytically significant finding of this examination.

4.3 The Compressed Pitch Contour Finding

Independent of the anomalous spike events, the overall pitch contour of the target segment shows significantly compressed range compared to the baseline: 70 Hz spread versus 174 Hz spread in normal speech range (excluding values above 300 Hz). Mean F0 in the target segment (approximately 125 Hz) is lower than the baseline mean (approximately 145 Hz).

This finding is consistent with the effortful pitch suppression pattern documented in the forensic phonetics literature [2]. When a speaker exercises conscious or unconscious control over vocal output — as might occur when delivering a carefully managed statement — the result is a more monotone, less naturally variable pitch contour than the speaker's baseline.

The combination of compressed overall pitch contour with a localized anomalous spike event at the key phrase represents the pattern of most analytical interest: controlled speech punctuated by a moment of apparent physiological disruption at the semantic locus of the denial.

5. REPRODUCIBILITY STATEMENT

This analysis is fully reproducible by any qualified examiner with access to the following: the publicly available YouTube recording (<https://www.youtube.com/watch?v=ZLW6Pgc3SfM>); yt-dlp (<https://github.com/yt-dlp/yt-dlp>); FFmpeg (<https://ffmpeg.org>); and Praat (<https://www.praat.org>). All three software tools are open-source and freely downloadable. The Praat script used for F0 extraction is reproduced in full in Section 3.4 of this document. The raw data output files (praat-test-001.txt for the baseline and praat-test-002.txt for the target) are available upon request and can be independently verified against re-extraction from the original source.

The complete processing pipeline from YouTube source to CSV output can be executed on any Linux, macOS, or Windows system with the above tools installed, requiring no proprietary software, commercial licenses, or specialized hardware.

6. LIMITATIONS AND SCOPE OF FINDINGS

The following limitations are acknowledged:

1. This analysis has not been reviewed by a qualified forensic phonetician. The findings represent preliminary investigative analysis, not expert forensic testimony.
2. Anomalous pitch events above 400 Hz cannot be definitively classified as physiological versus artifactual without spectrographic analysis of individual frames by a qualified examiner.
3. F0 analysis alone is insufficient for deception determination. Voice stress analysis is not accepted as evidence in Florida courts.
4. The compressed pitch contour in the target segment may reflect factors other than psychological stress, including deliberate rhetorical emphasis, the speaker's awareness of being recorded, or natural variation in speaking style between formal introduction and substantive presentation contexts.
5. The source audio underwent Opus codec compression prior to this analysis. While pcm_s16le reversion was used to minimize compression artifacts in the time domain, some frequency-domain distortion from the original Opus encoding may persist.
6. This report does not constitute a determination of deception, veracity, or any legal conclusion. It is offered as one analytical instrument among many in the context of a broader investigative effort.

REFERENCES

- [1] Brenner, M., Shipp, T. (1987). Voice stress analysis. AGARD Conference Proceedings No. 311, pp. 1-1 to 1-10. NATO Advisory Group for Aerospace Research and Development.
- [2] Streeter, L.A., Krauss, R.M., Geller, V., Olson, C., Apple, W. (1977). Pitch changes during attempted

deception. *Journal of Personality and Social Psychology*, 35(5), 345-350.

- [3] Hollien, H. (1990). *The Acoustics of Crime: The New Science of Forensic Phonetics*. Plenum Press, New York.
- [4] Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences*, 17, 97-110. University of Amsterdam.
- [5] Boersma, P., Weenink, D. (2023). Praat: doing phonetics by computer [Computer program]. Version 6.3.07. Retrieved from <https://www.praat.org>
- [6] Maher, R.C. (2010). Audio forensics examination. *IEEE Signal Processing Magazine*, 26(2), 84-94.