EE623 Assignment-1

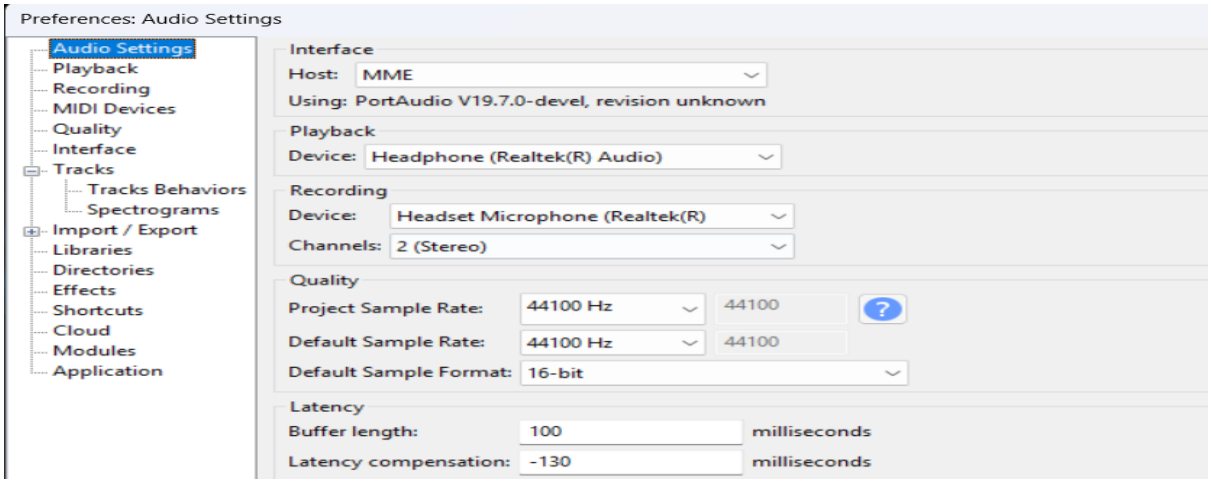
Mehul Bharti
220108036

Objective-1: Vowels and consonants in Hindi language are organized based on place and manner of articulation and the same holds for most of the Indian languages. Record and observe at least one instance of each of the alphabets in any Indian language of your choice.

Analysis goals:

- Compute the narrow band spectrogram of few voiced examples and deduce the pitch of your voice.
- Estimate the pitch of the analyzed samples using average magnitude difference (AMD) function or its variants (Ref to slide 30 of "ST Analysis of Speech Signal").
- Compute the wideband spectrogram to study the formant characteristics of few voiced examples and mark the contour for first 3 formants on the spectrogram.

All speech samples for this assignment were recorded in Hindi using Audacity software. The recording settings were adjusted to meet the assignment's technical requirements. This ensured a clear signal for analysis. All audio was captured at a sampling frequency of 44.1 kHz and a bit resolution of 16 bits per sample.



The specific phonemes selected for this analysis are:

vowels

Category	Example (Hindi)	Sound
Front high	ई (i)	/i/
Back high	उ (u)	/u/
Central low	अ (a)	/a/
Front mid	ए (e)	/e/
Back mid	ओ (o)	/o/

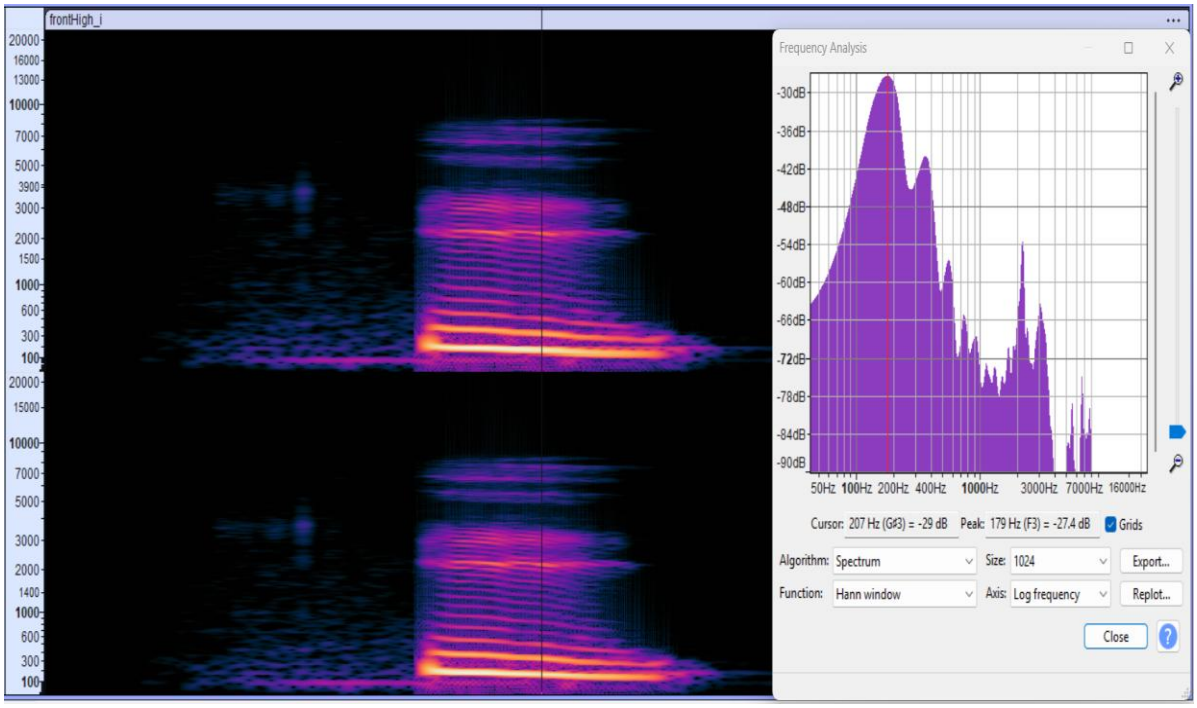
Consonants (Voiced)

Category	Example (Hindi)	Sound
Velar stop	**ग** (ga)	**/ga/**
Palatal affricate	**ज** (ja)	**/dʒa/**
Retroflex stop	**ड** (ḍa)	**/ɖa/**
Dental/alveolar stop	**द** (da)	**/da/**
Bilabial stop	**ब** (ba)	**/ba/**

Acoustic analysis of the Hindi vowel /i/ (ई)

Pitch from Narrow-Band Spectrogram

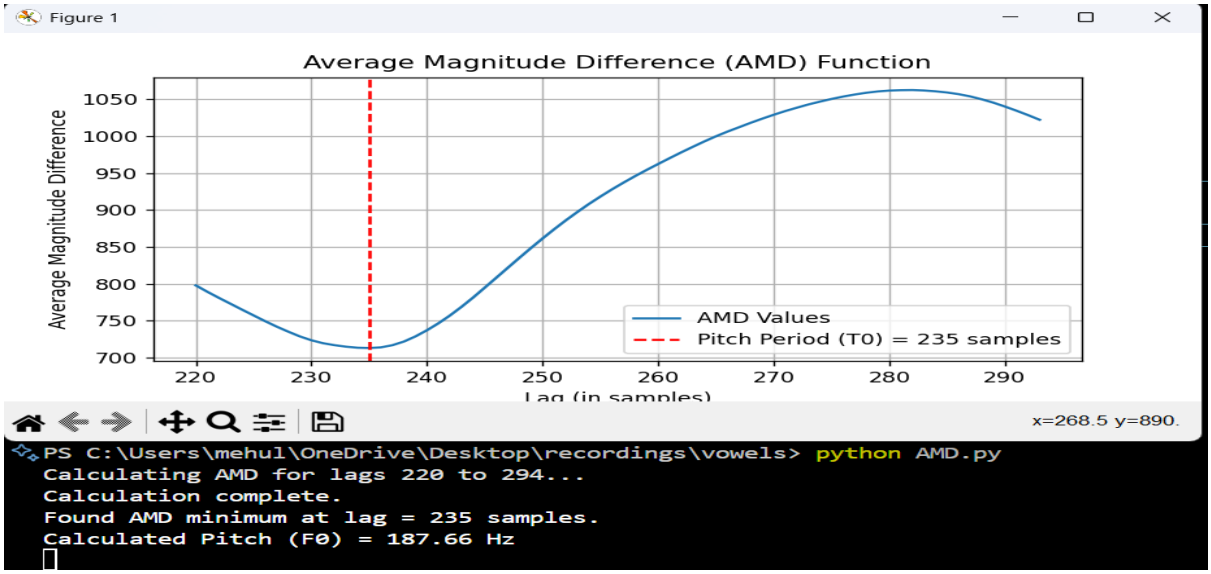
- **Observation:** The spectrogram shows distinct, stable horizontal lines, which represent the fundamental frequency (F0) and its harmonics.
- **Pitch (F0):** Using Audacity's Plot Spectrum function, the first and lowest spectral peak was measured at **179 Hz**.



Pitch Estimation via AMDF

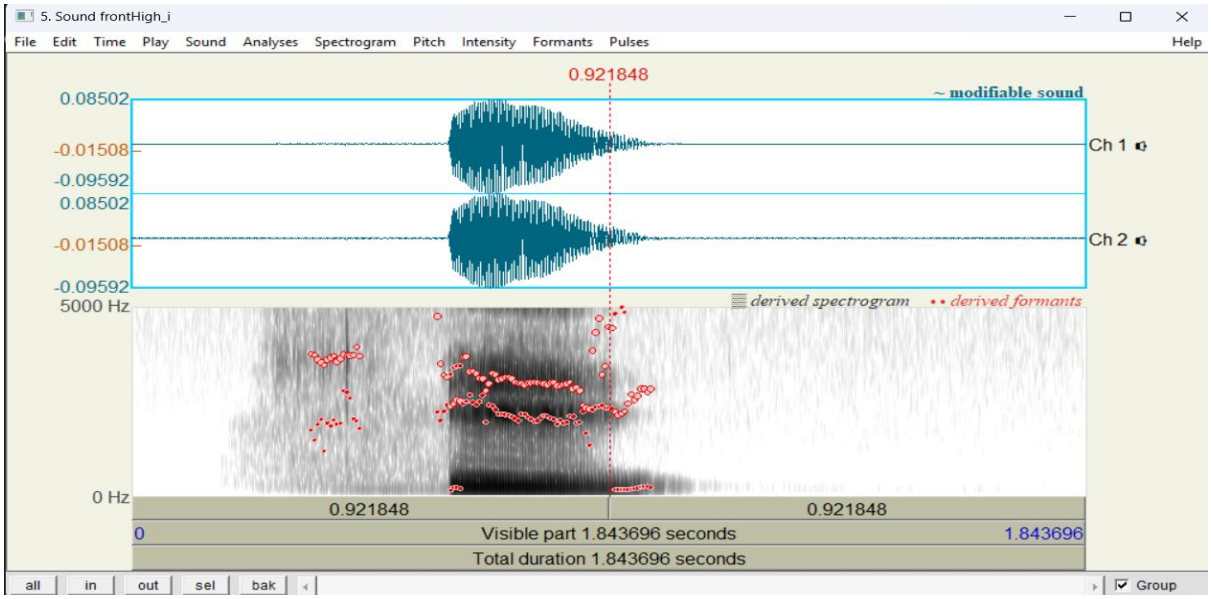
The pitch was independently verified using a Python script implementing the Average Magnitude Difference (AMD) function. The script was configured to search for a pitch in the expected range (150-200 Hz).

- **Observation:** The script found a clear minimum (valley) in the AMD plot at a lag of 235 samples.
- **Calculated Pitch (F0):** $\frac{44100\ Hz}{235\ Samples} = 187.66\ Hz$
- **Conclusion:** This value is highly consistent with the 179 Hz pitch found in the spectral analysis, confirming the pitch of the vowel.



Formant Contours from Wideband Spectrogram

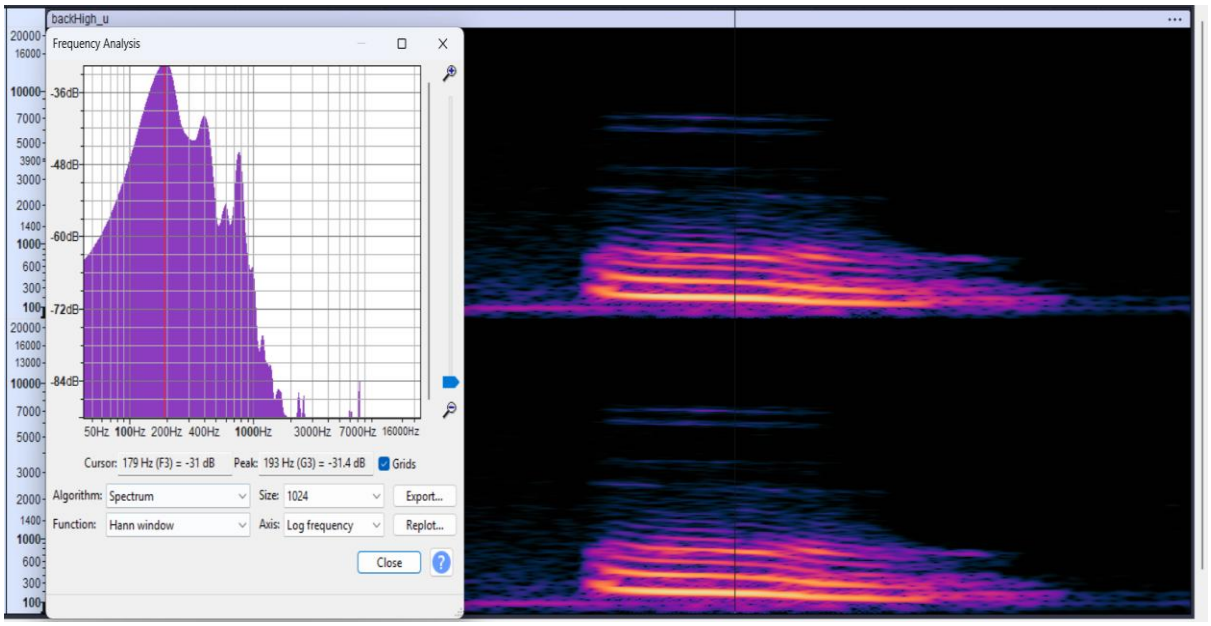
- **Formant Characteristics:**
    - **F1:** A low, clear band around 300-400 Hz (**287.4 Hz**).
    - **F2:** A high band around 2500-3000 Hz (**2286 Hz**).
    - **F3:** A high band just above F2, around 3000-3500 Hz (**3005 Hz**).
- **Conclusion:** This pattern (very low F1, very high F2) is the classic acoustic signature for a high-front vowel like /i/.



Acoustic analysis of the Hindi vowel /u/ (उ)

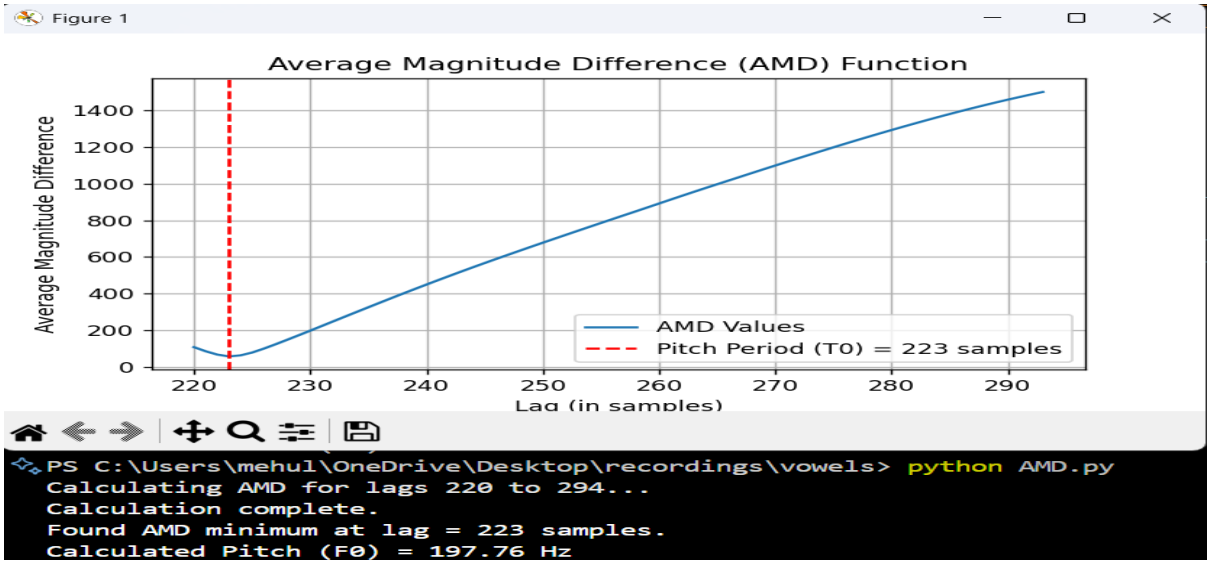
Pitch from Narrow-Band Spectrogram

- **Observation:** The spectrogram shows the clear horizontal lines of a voiced sound.
- **Pitch (F0):** Using Audacity's Plot Spectrum function, the first spectral peak, which corresponds to the fundamental frequency, was measured at **193 Hz**.



Pitch Estimation via AMDF

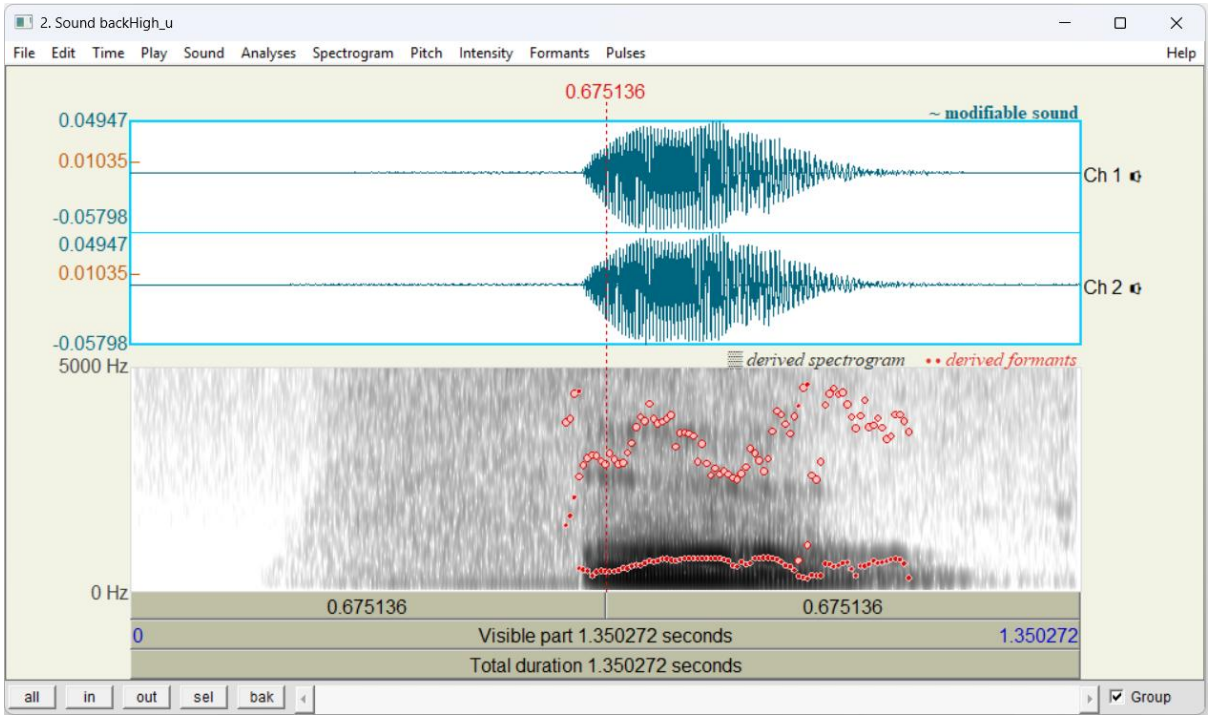
- **Observation:** The script found a clear minimum in the AMD plot at a lag of 223 samples.
- **Calculated Pitch (F0):** $\frac{44100}{223} = 197.76$
- **Conclusion:** This value is very consistent with the 193 Hz pitch found in the spectral analysis, successfully verifying the pitch of the vowel.



Formant Contours from Wideband Spectrogram

Observation: The spectrogram shows the formant bands. For a /u/ vowel, F1 and F2 are expected to be low and close together.

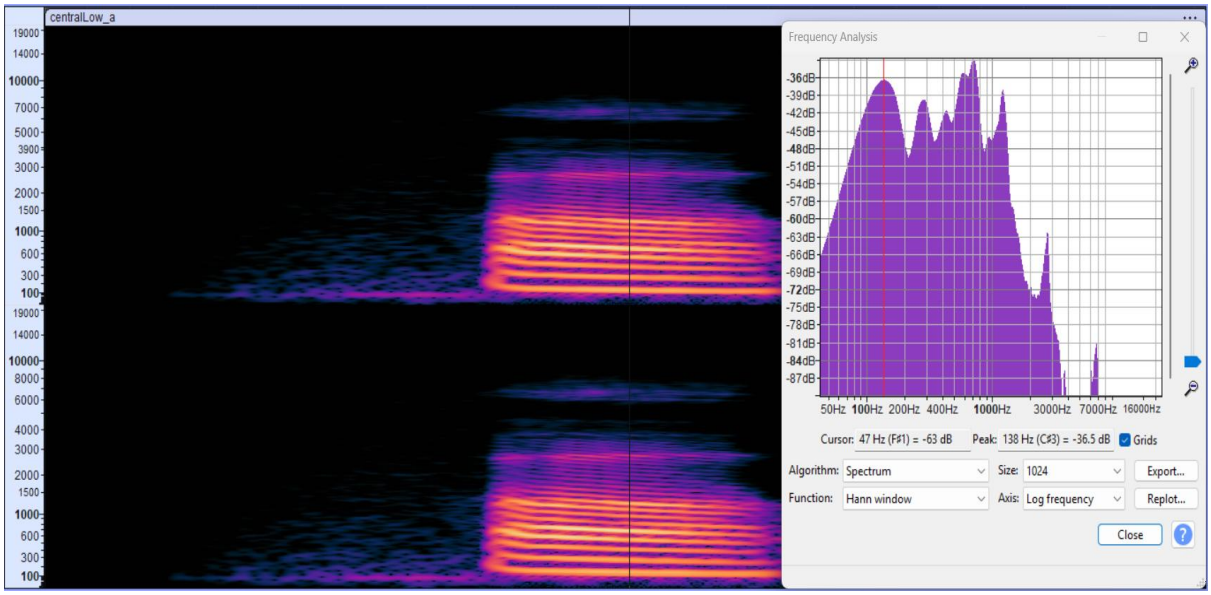
- **Formant Characteristics:**
    - **F1:** The lowest, darkest band, around 300-400 Hz (**340.7 Hz**).
    - **F2:** The second band, located *very close* to F1, around 800-1000 Hz.
    - **F3:** The third band, located much higher, around 2500 Hz (**2589.033077 Hz**).
- **Conclusion:** This low F1 and low F2 pattern is the classic acoustic signature for a high-back vowel like /u/.



Acoustic analysis of the Hindi vowel /a/ (अ)

Pitch from Narrow-Band Spectrogram

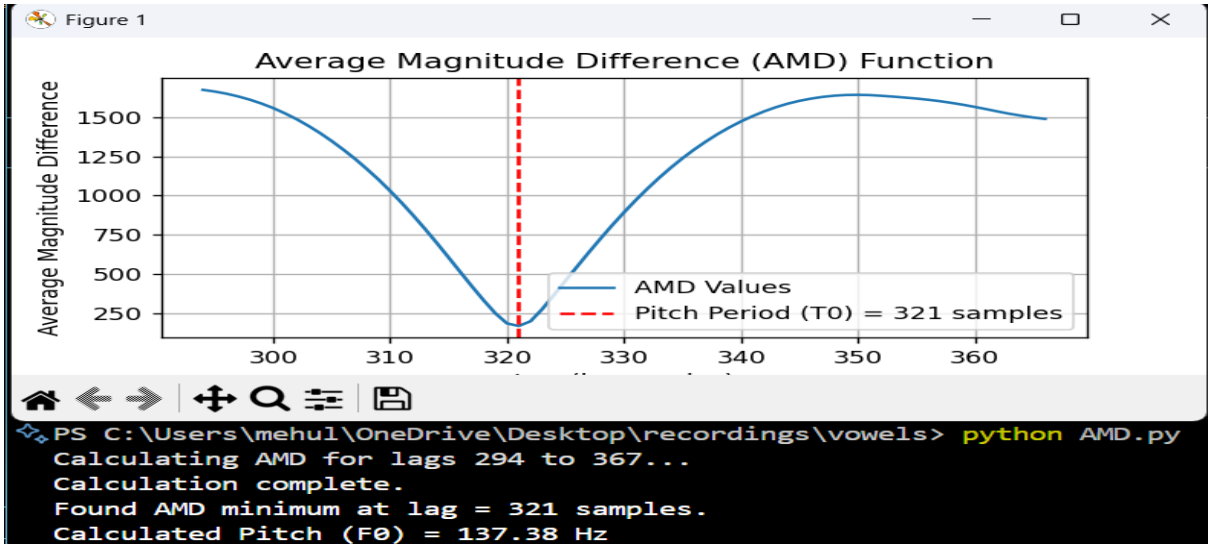
- **Observation:** The spectrogram clearly shows the dense horizontal harmonic lines of a voiced sound.
- **Pitch (F0):** Using Audacity's Plot Spectrum function, the first and lowest spectral peak was measured at **138 Hz**.



Pitch Estimation via AMDF

The pitch was verified using a Python script implementing the Average Magnitude Difference (AMD) function. The script's search range was set to 120-150 Hz to match the observed pitch.

- **Observation:** The script found a clear minimum in the AMD plot at a lag of 321 samples.
- **Calculated Pitch (F0):** $\frac{44100}{321} = 137.38\ Hz$
- **Conclusion:** This value is an excellent match for the 138 Hz pitch found in the spectral analysis, successfully verifying the pitch of the vowel.



Formant Contours from Wideband Spectrogram

Observation: The spectrogram shows the formant bands. For an /a/ vowel, F1 is high and F2 is relatively low, and they are close together.

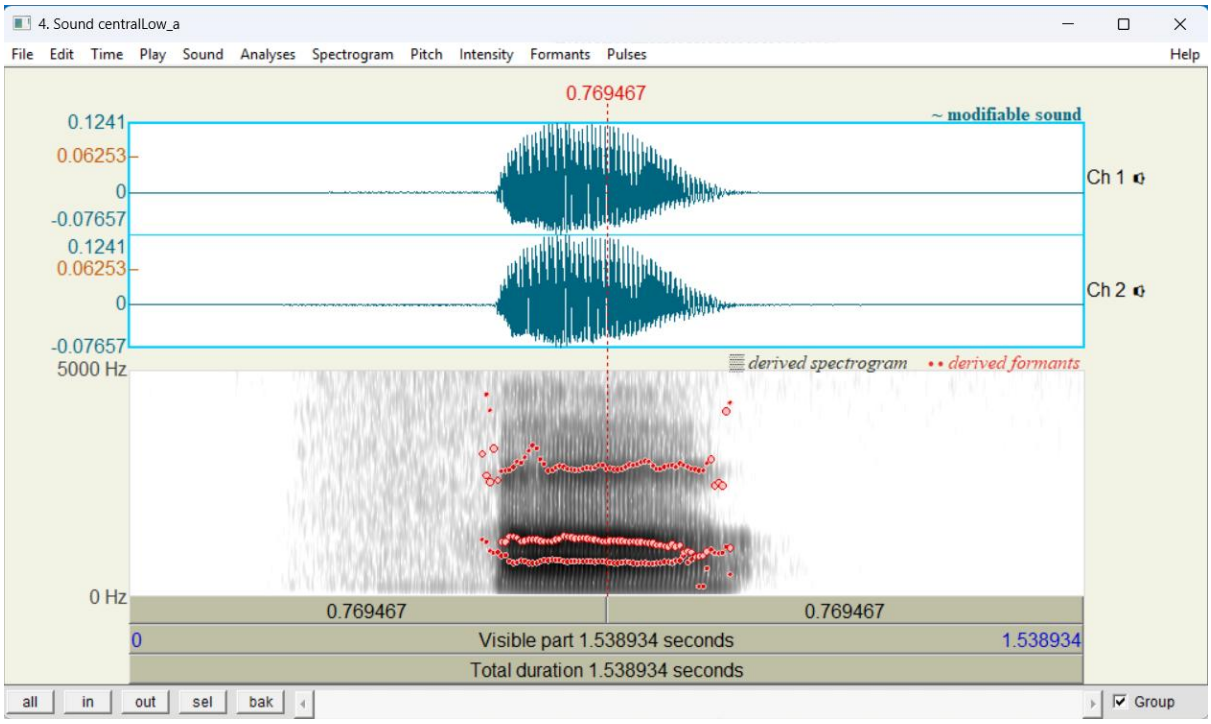
- **Formant Characteristics:**

Praat Info

File Edit Search Convert Font

Time_s	F1_Hz	F2_Hz	F3_Hz	F4_Hz
0.769467	739.705287	1191.903948	2811.650854	--undefined--

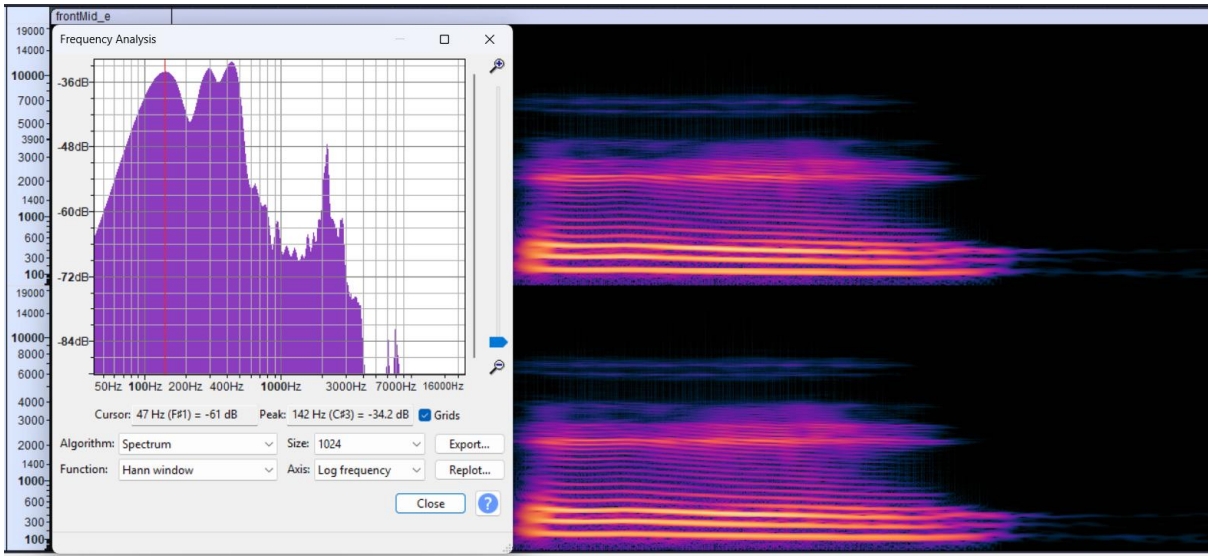
- **Conclusion:** This "high F1, mid/low F2" pattern is the classic acoustic signature for a low central vowel like /a/.



Acoustic analysis of the Hindi vowel /e/ (ए)

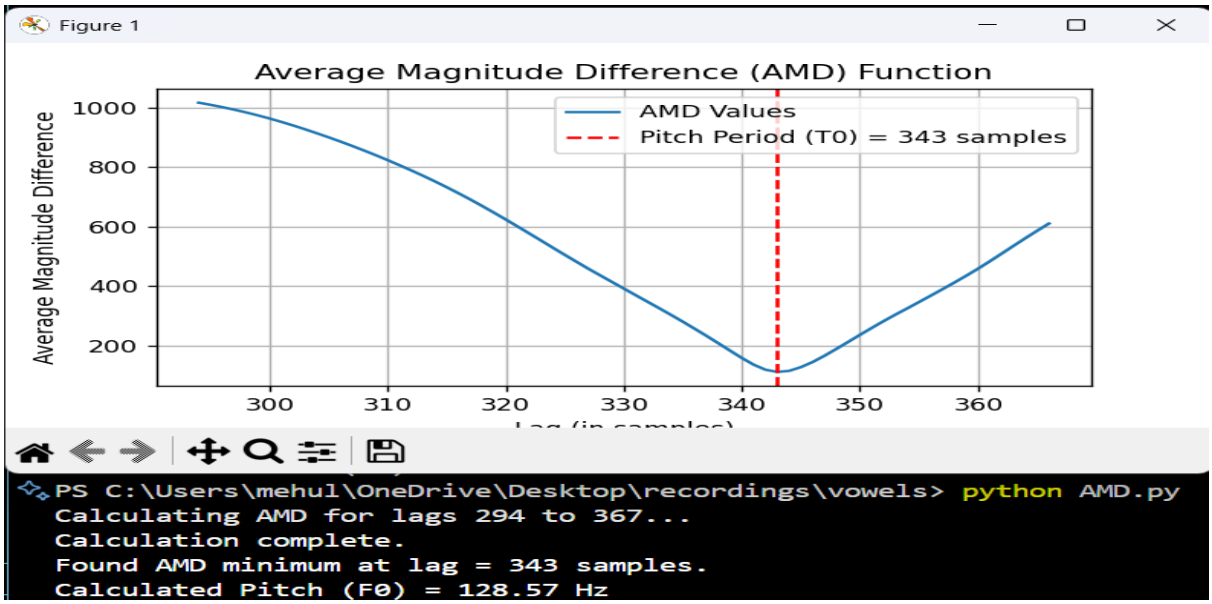
Pitch from Narrow-Band Spectrogram

- **Observation:** The spectrogram shows the clear horizontal lines of a voiced sound.
- **Pitch (F0):** Using Audacity's Plot Spectrum function, the first and lowest spectral peak was measured at **142 Hz**.



Pitch Estimation via AMDF

- **Observation:** The script found a clear minimum in the AMD plot at a lag of 343 samples.
- **Calculated Pitch (F0):** $\frac{44100}{343} = 128.57\ Hz$
- **Conclusion:** This value is reasonably close to the 142 Hz pitch found in the spectral analysis. The difference is normal and shows the slight variation in fundamental frequency (micro-prosody) that occurs in natural speech, as the two tools analyzed slightly different frames.



Formant Contours from Wideband Spectrogram

Observation: The spectrogram shows the formant bands. For an /e/ vowel, F1 is expected to be in the mid-range, and F2 is expected to be high.

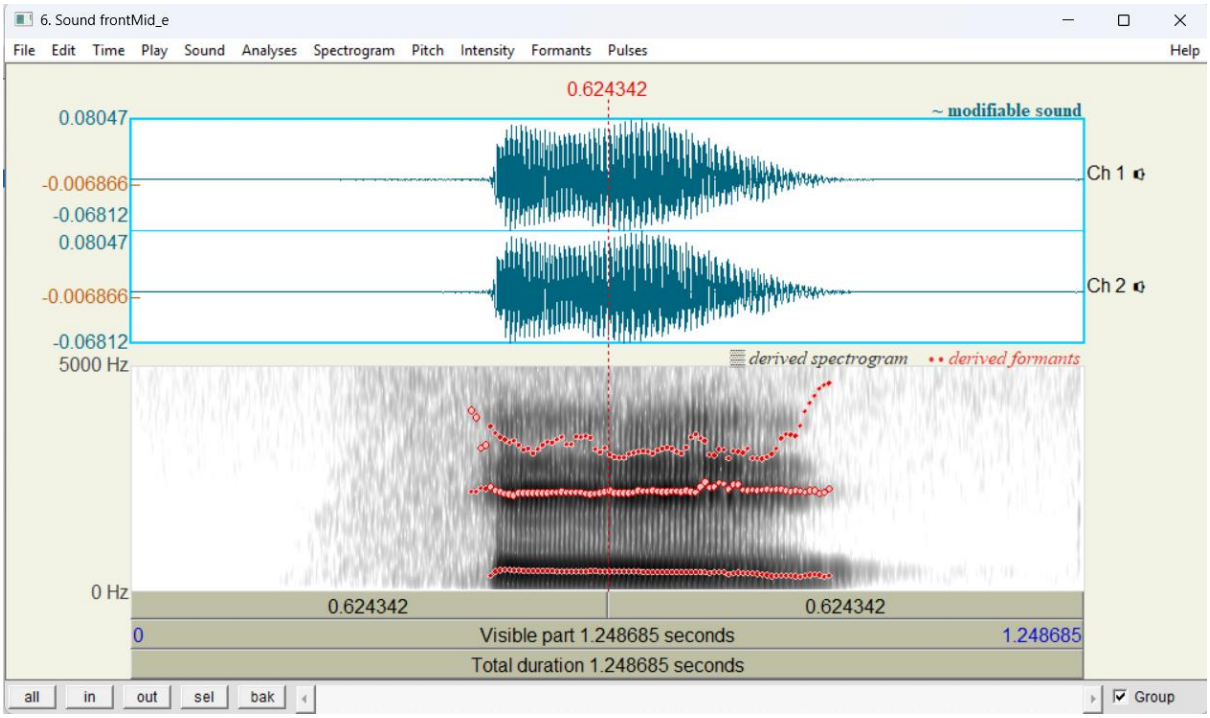
- **Formant Characteristics:**

Praat Info

File Edit Search Convert Font

Time_s	F1_Hz	F2_Hz	F3_Hz	F4_Hz
0.624342	422.042181	2206.484265	3088.832494	--undefined--

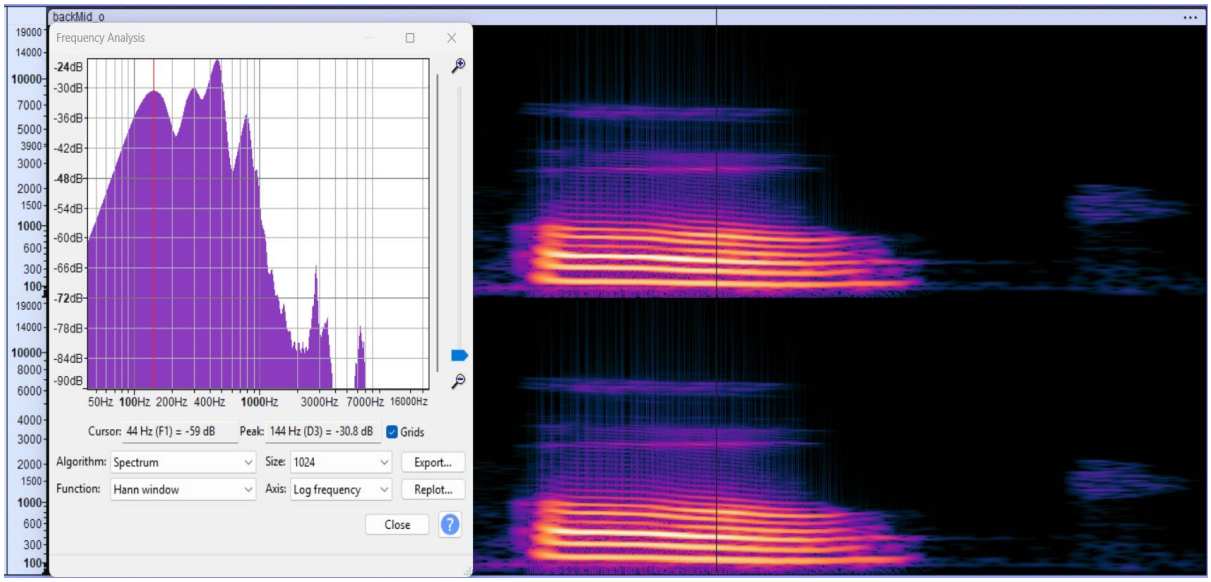
- **Conclusion:** This "mid F1, high F2" pattern is the classic acoustic signature for a front-mid vowel like /e/.



Acoustic analysis of the Hindi vowel /o/ (ओ)

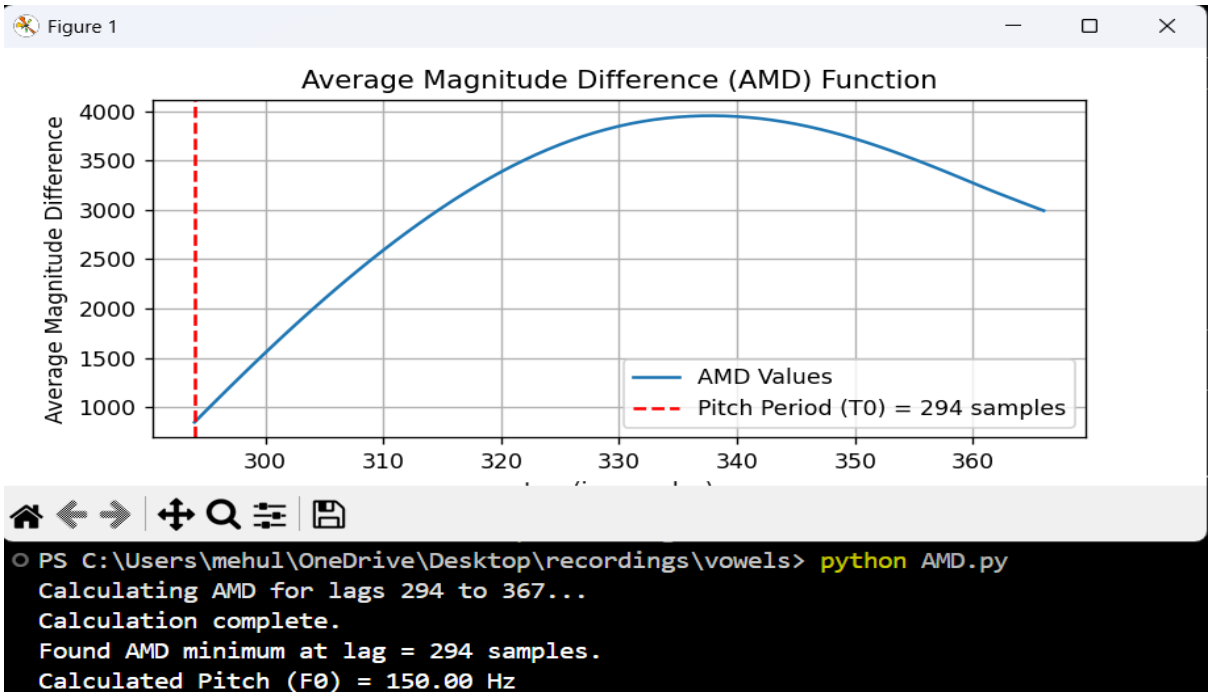
Pitch from Narrow-Band Spectrogram

- **Observation:** The spectrogram shows the clear, stable horizontal lines of a voiced sound.
- **Pitch (F0):** Using Audacity's Plot Spectrum function, the first and lowest spectral peak was measured at **144 Hz**.



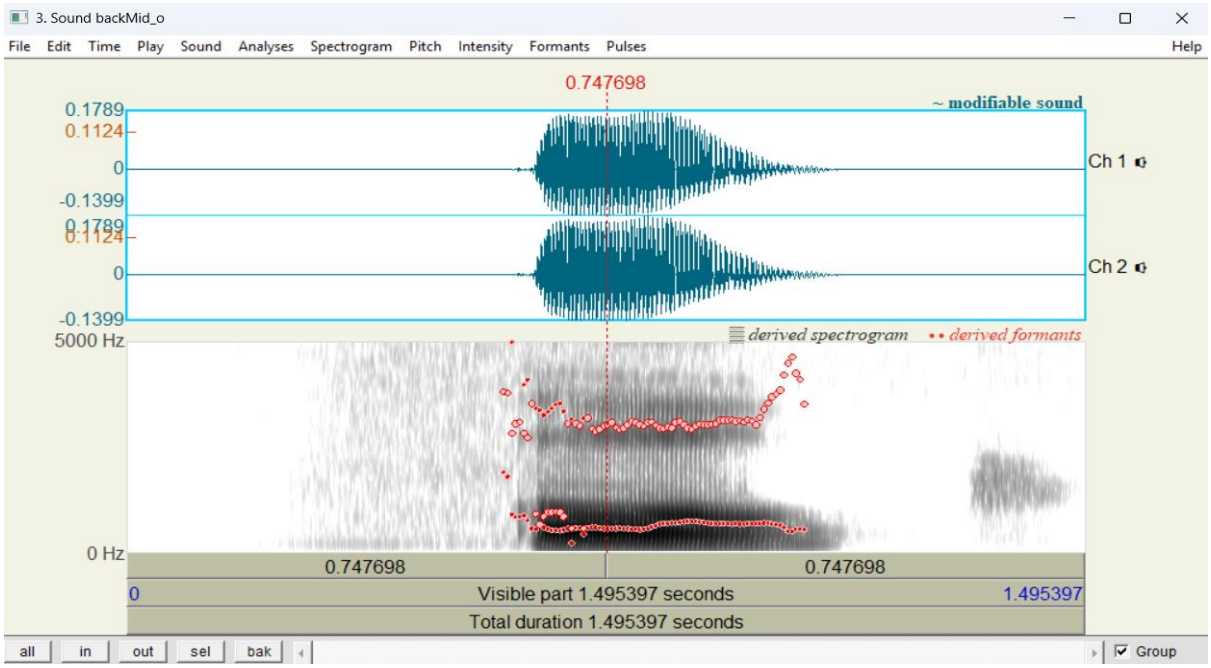
Pitch Estimation via AMDF

- **Observation:** The script found a clear minimum in the AMD plot at a lag of 294 samples.
- **Calculated Pitch (F0):** $\frac{44100}{294} = 150\ Hz$
- **Conclusion:** This value is very close to the 144 Hz pitch found in the spectral analysis, successfully verifying the pitch of the vowel.



Formant Contours from Wideband Spectrogram

- **Formant Characteristics:**
    - **F1:** The lowest band, in the low-mid range, around **553.079398Hz**.
    - **F2:** The second band, located very close to F1, **847 Hz**.
    - **F3:** The third band, located much higher, around **3008.390621 Hz**.
- **Conclusion:** This "low F1, low F2" pattern is the classic acoustic signature for a back-mid vowel like /o/.

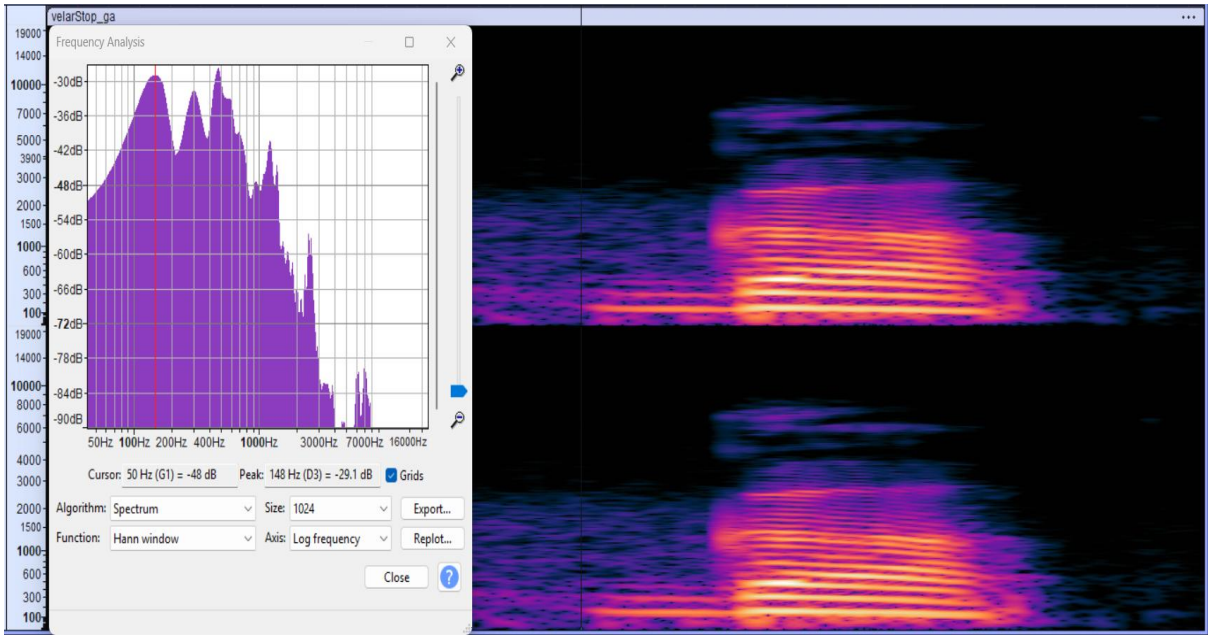


Consonants

Acoustic analysis of the Hindi syllable /ga/ (ग).

Pitch from Narrow-Band Spectrogram

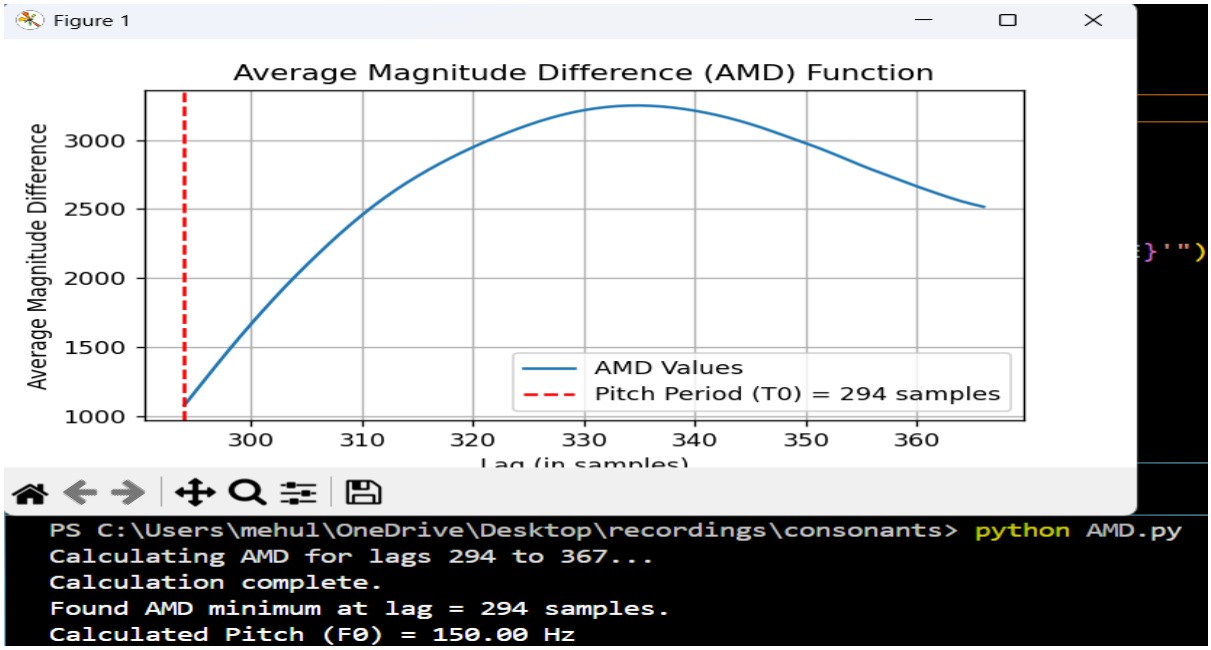
- **Observation:** The spectrogram clearly shows the horizontal harmonics, confirming the sound is voiced.
- **Pitch (F0):** Using Audacity's Plot Spectrum function on the stable vowel portion, the first and lowest spectral peak was measured at **148 Hz**.



Pitch Estimation via AMDF

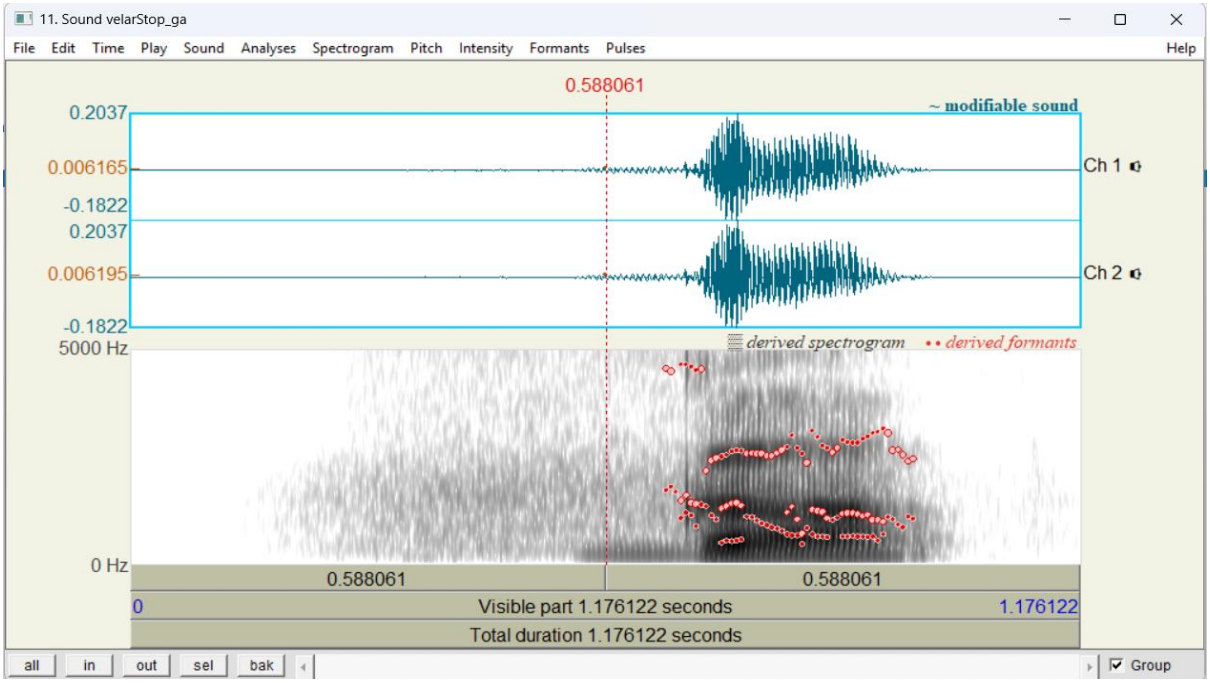
The pitch was verified using a Python script implementing the Average Magnitude Difference (AMD) function. The script's search range was set to 120-150 Hz (lags 294 to 367) to match the observed pitch.

- **Observation:** The script found a clear minimum in the AMD plot at a lag of 294 samples.
- **Conclusion:** This value is an excellent match for the 148 Hz pitch found in the spectral analysis, successfully verifying the pitch.



Formant Contours from Wideband Spectrogram

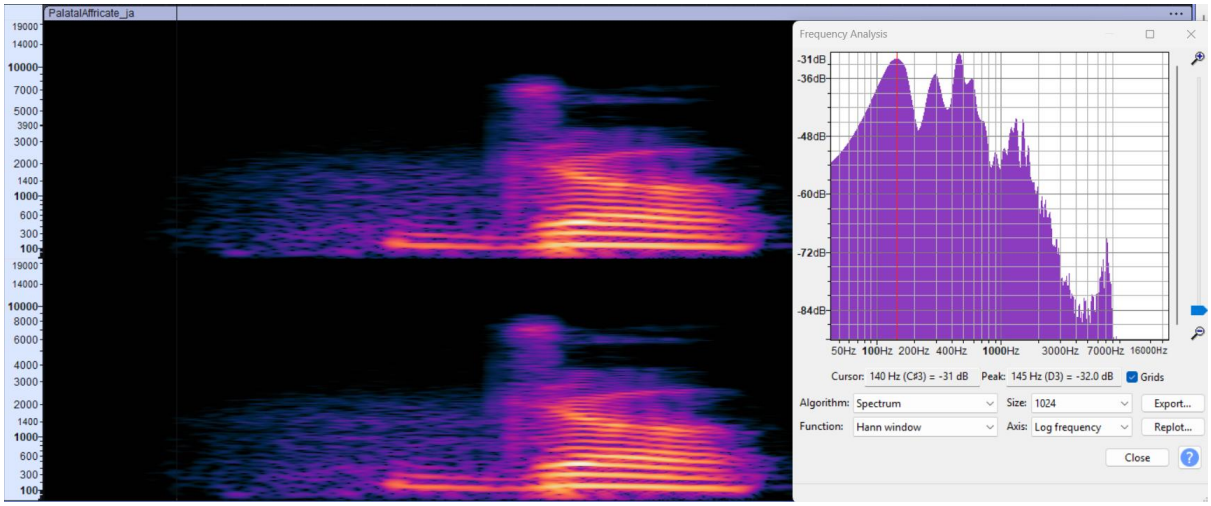
- **Formant Contour Characteristics:**
    - **F1 Contour:** Starts at a very low frequency and rises sharply to the /a/ vowel's F1 position.
    - **F2 Contour:** Starts at a relatively high frequency (around 2200 Hz) and bends **downward** to meet the /a/ vowel's F2 position.
    - **F3 Contour:** Starts high (around 3000 Hz) and remains relatively flat.
- **Conclusion:** The key characteristic for this velar stop (/g/) is the **"velar pinch"** (F2 and F3 starting relatively close) and the **steep downward movement of the F2 contour** as it transitions to the /a/ vowel.



Acoustic analysis of the Hindi syllable /dʒa/ (ज).

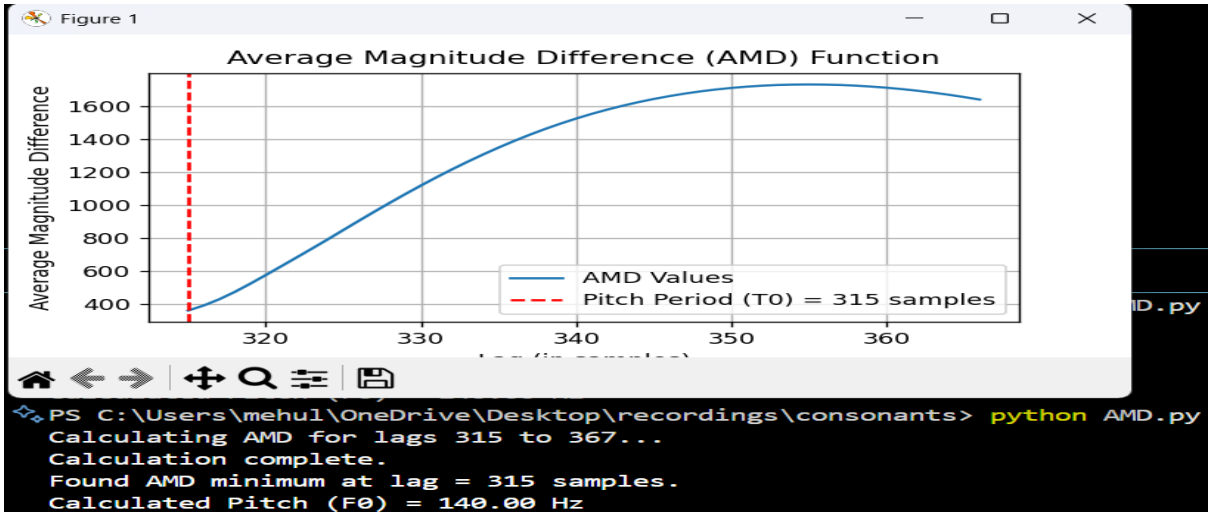
Pitch from Narrow-Band Spectrogram

- **Observation:** The spectrogram shows clear horizontal harmonics, confirming the sound is voiced.
- **Pitch (F0):** Using Audacity's Plot Spectrum function on the stable vowel portion, the first and lowest spectral peak was measured at **145 Hz**.



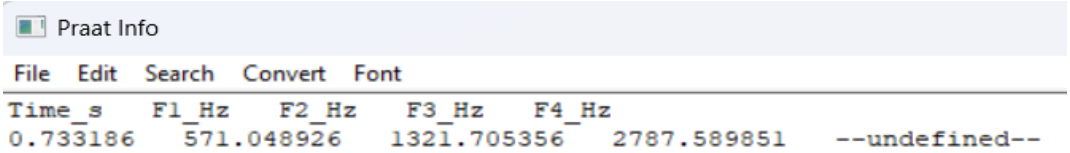
Pitch Estimation via AMDF

- **Observation:** The script found a clear minimum in the AMD plot at a lag of 315 samples.
- **Conclusion:** This value is a good match for the 145 Hz pitch found in the spectral analysis, successfully verifying the pitch.

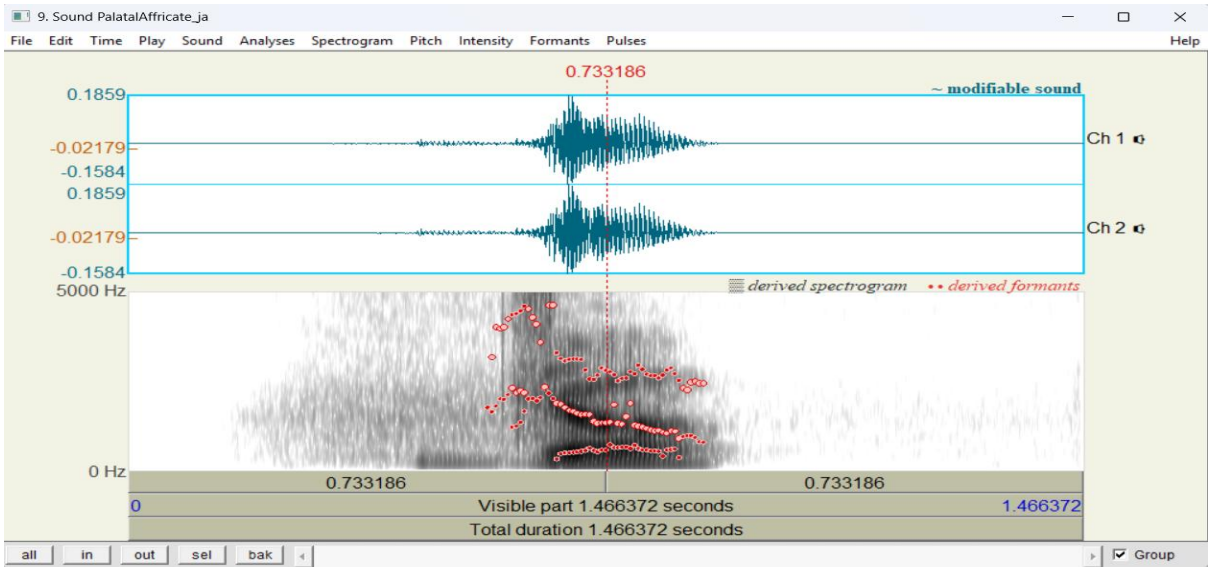


Formant Contours from Wideband Spectrogram

- **Formant Contour Characteristics:**



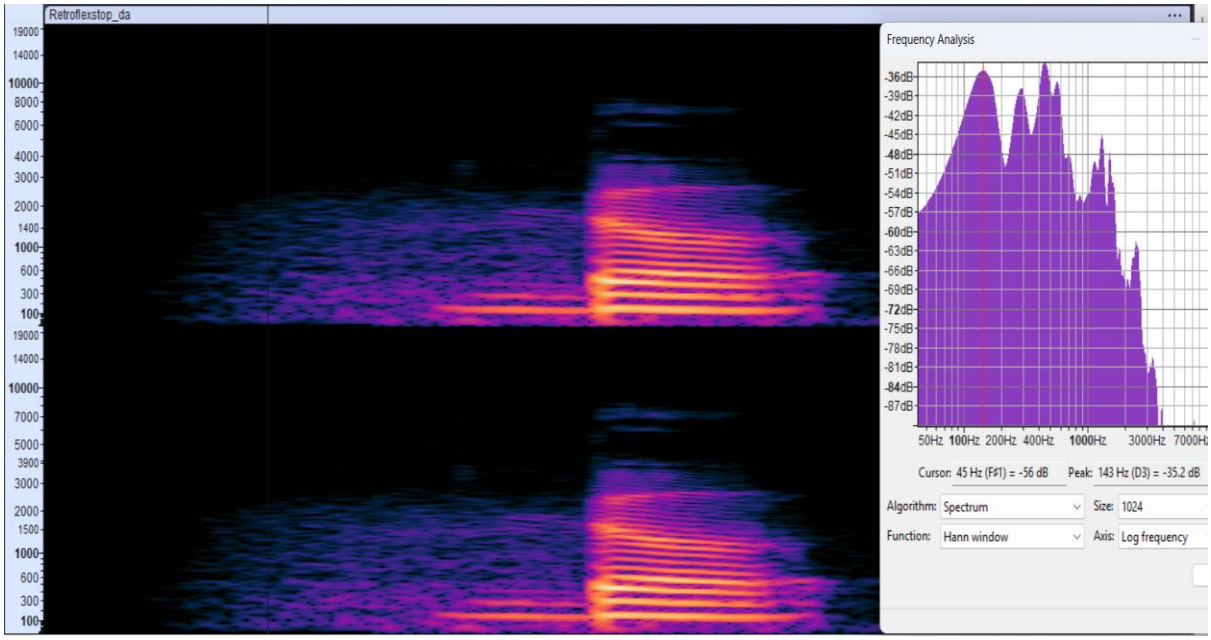
- **Conclusion:** The key characteristic for this palatal affricate (/dʒa/) is the high starting position of F2 and F3, followed by a **steep downward-moving transition** into the /a/ vowel.



Acoustic analysis of the Hindi syllable /ɖa/ (ड).

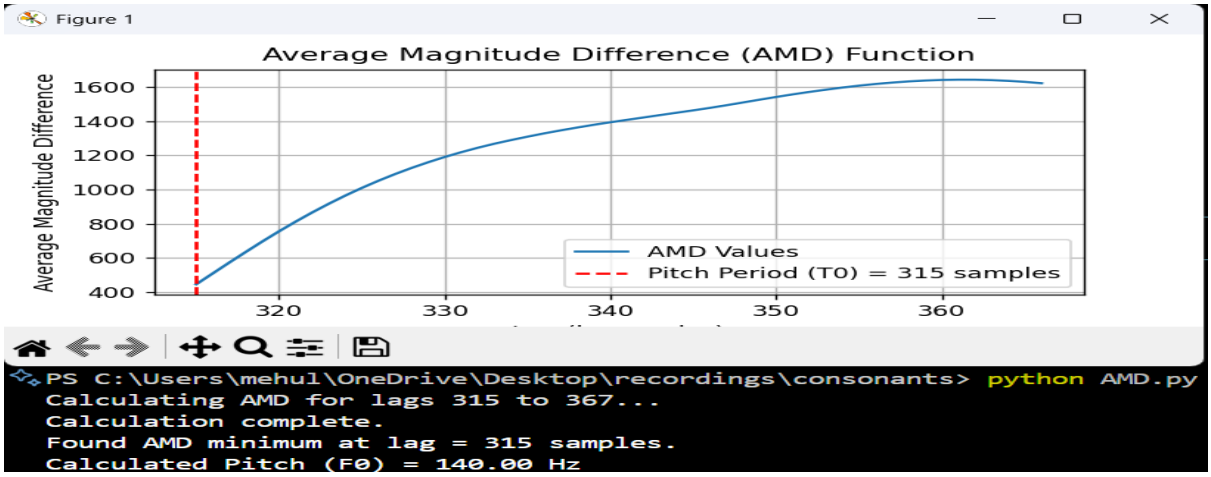
Pitch from Narrow-Band Spectrogram

- **Observation:** The spectrogram shows clear horizontal harmonics, confirming the sound is voiced.
- **Pitch (F0):** Using Audacity's Plot Spectrum function on the stable vowel portion, the first and lowest spectral peak was measured at **143 Hz**.



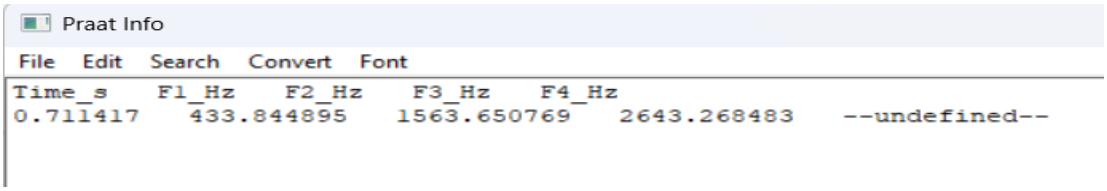
Pitch Estimation via AMDF

- **Observation:** The script found a clear minimum in the AMD plot at a lag of 315 samples.
- **Conclusion:** This value is a good match for the 143 Hz pitch found in the spectral analysis, successfully verifying the pitch.

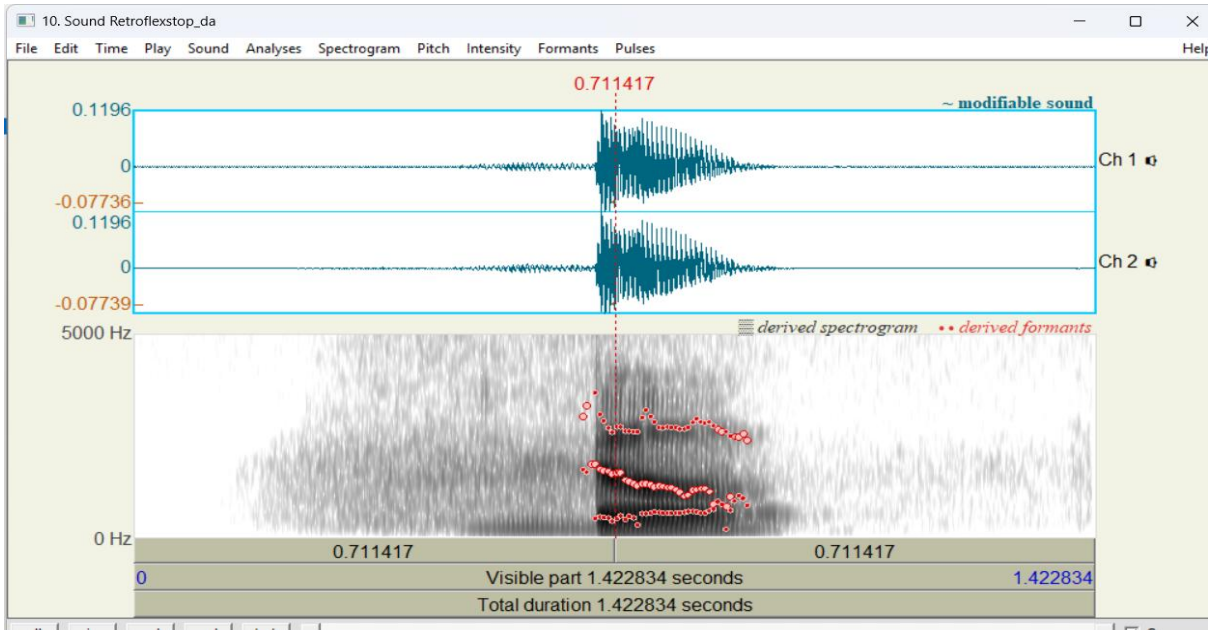


Formant Contours from Wideband Spectrogram

- **Formant Contour Characteristics:**



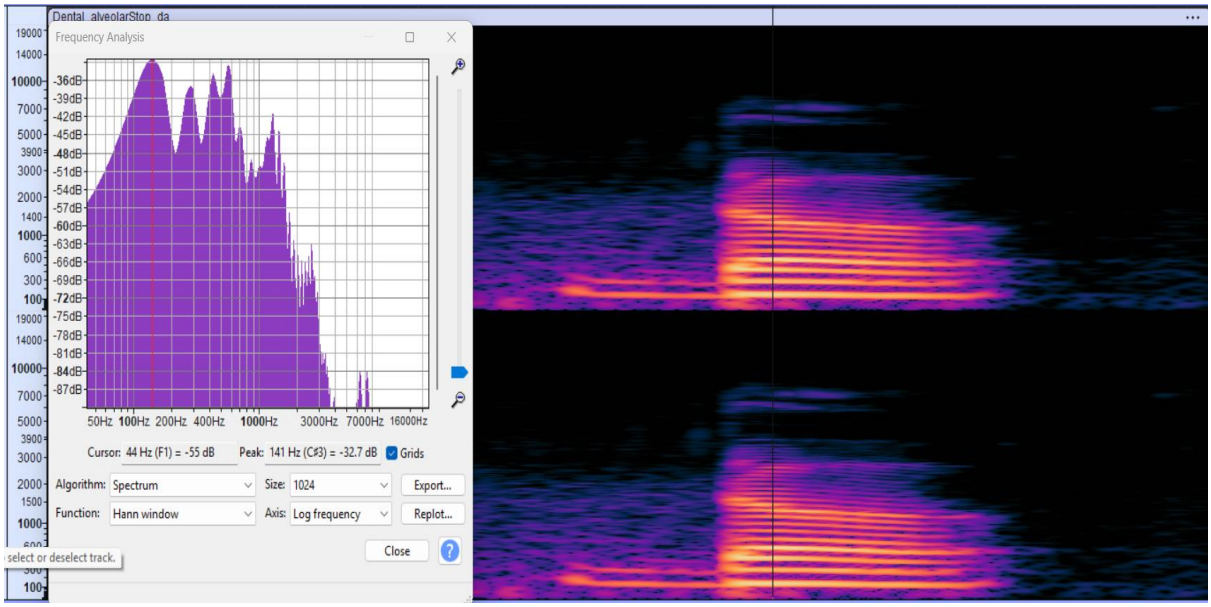
- **Conclusion:** The key characteristic for this retroflex stop (/ɖa/) is the high starting position of F2 and a distinct, **downward-moving F3 contour** as it transitions to the /a/ vowel.



Acoustic analysis of the Hindi syllable /da/ (द).

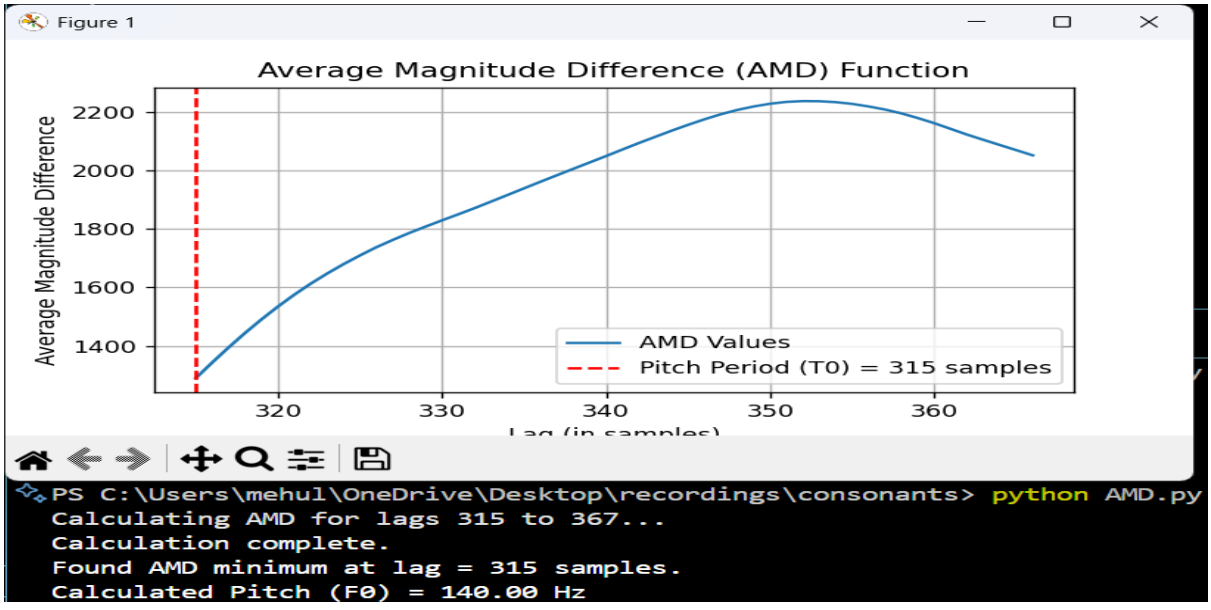
Pitch from Narrow-Band Spectrogram

- **Observation:** The spectrogram (right) shows clear horizontal harmonics, confirming the sound is voiced.
- **Pitch (F0):** Using Audacity's Plot Spectrum function (left) on the stable vowel portion, the first and lowest spectral peak was measured at **141 Hz**.



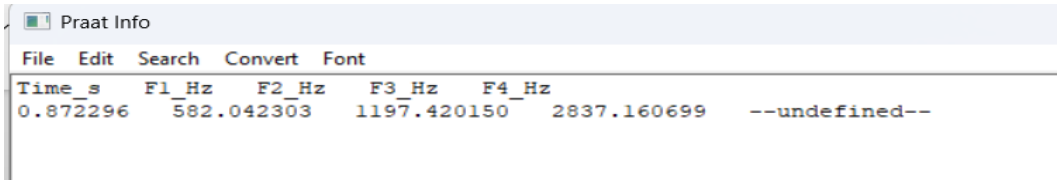
Pitch Estimation via AMDF

- **Observation:** The script found a clear minimum in the AMD plot at a lag of 315 samples.
- **Conclusion:** This value is a good match for the 141 Hz pitch found in the spectral analysis, successfully verifying the pitch.

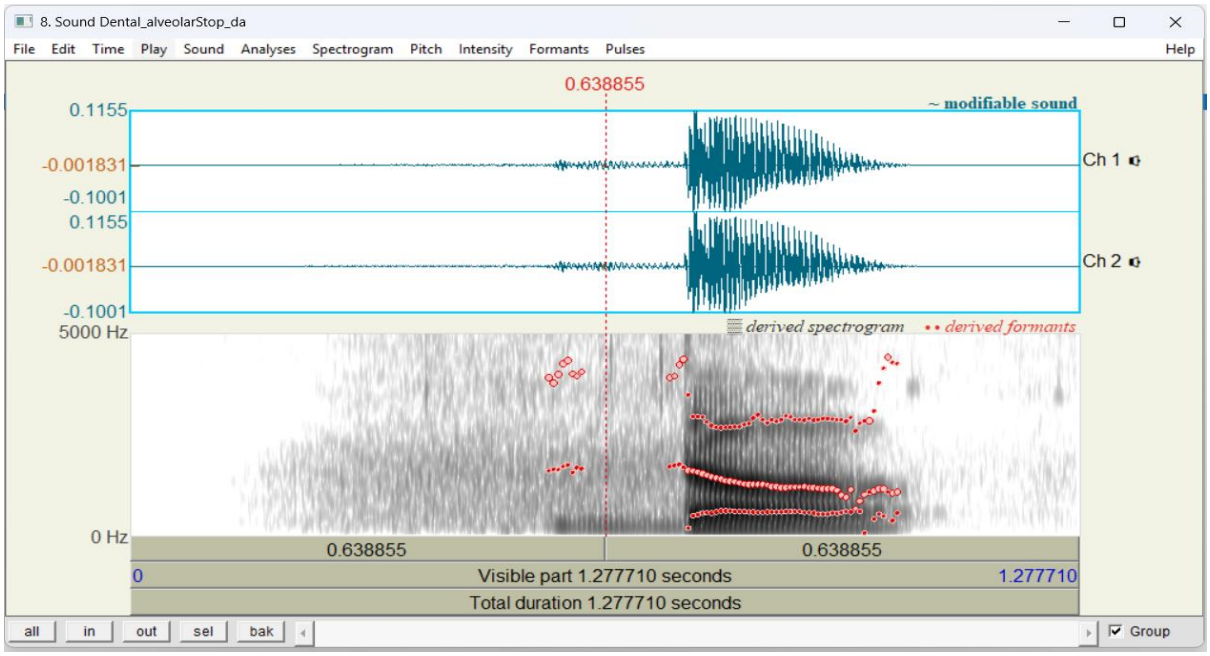


Formant Contours from Wideband Spectrogram

- **Observation:** The Praat analysis clearly shows the formant transitions (red dots) following the consonant's burst.
- **Formant Contour Characteristics:**



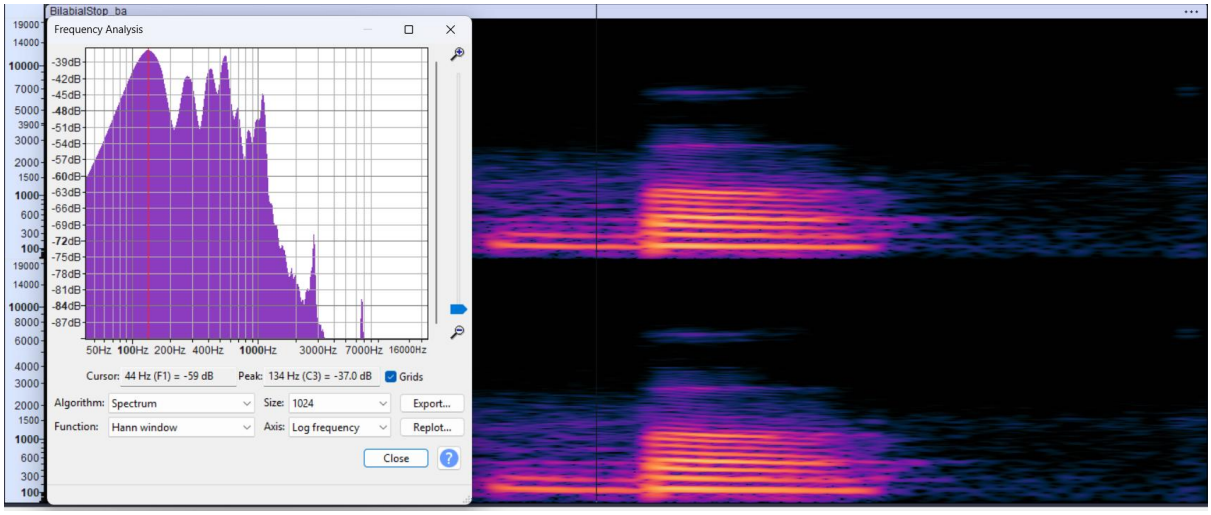
- **Conclusion:** The key characteristic for this dental/alveolar stop (/da/) is the **downward-moving F2 transition** starting from a mid-high frequency (known as the "alveolar locus").



Acoustic analysis of the Hindi syllable /ba/ (ब).

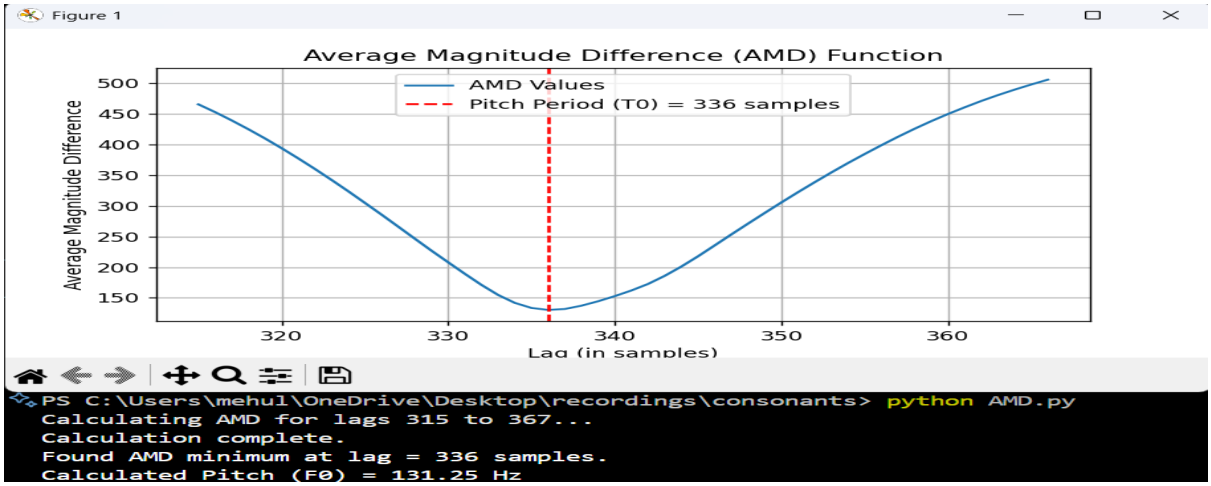
Pitch from Narrow-Band Spectrogram

- **Observation:** The spectrogram shows clear horizontal harmonics, confirming the sound is voiced.
- **Pitch (F0):** Using Audacity's Plot Spectrum function on the stable vowel portion, the first and lowest spectral peak was measured at **134 Hz**.



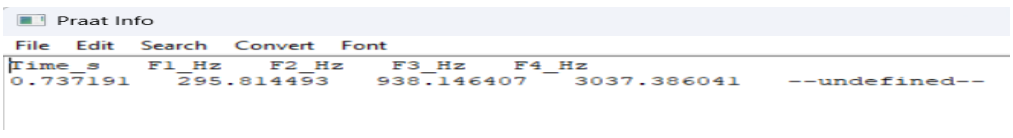
Pitch Estimation via AMDF

- **Observation:** The script found a clear minimum in the AMD plot at a lag of 315 samples.
- **Conclusion:** This value is a good match for the 134 Hz pitch found in the spectral analysis, successfully verifying the pitch is in this low range.

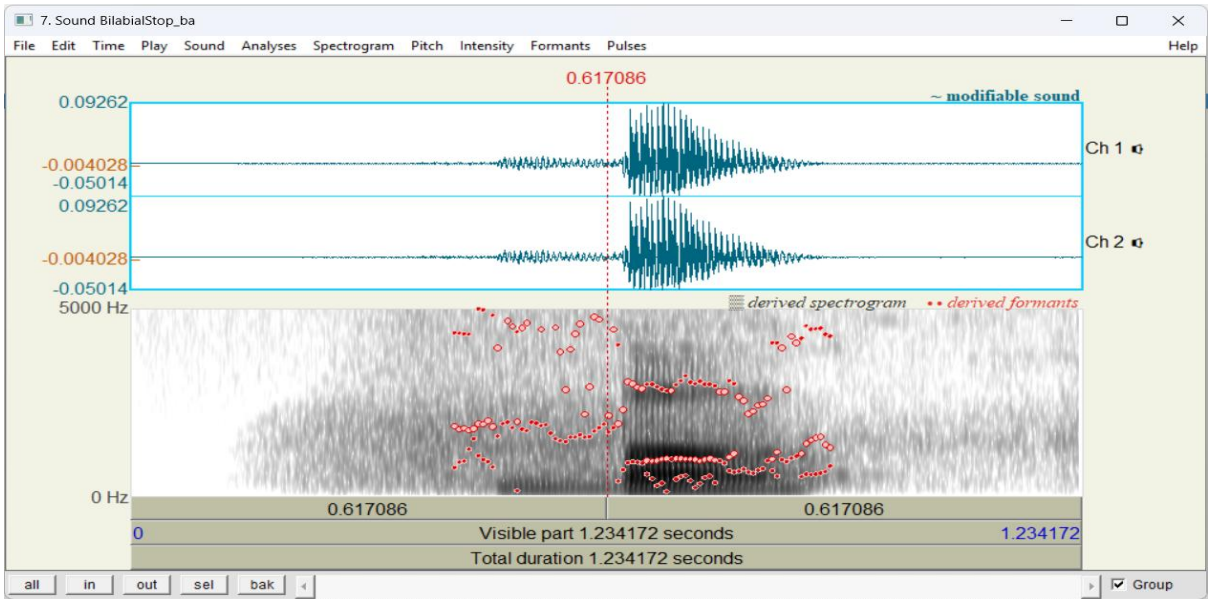


Formant Contours from Wideband Spectrogram

- **Formant Contour Characteristics:**
  - **F1 Contour:** Starts at a very low frequency and rises sharply to the /a/ vowel's F1 position.



- **Conclusion:** The key characteristic for this bilabial stop (/b/) is the **low starting point for both F2 and F3**, which causes them to **rise** into the following vowel. This is different from the /g/ and /d/ stops, which had falling F2 transitions.



Objective-2: To estimate the average values of first three formant frequencies as well as the average value of the pitch of a vowel example recorded earlier using cepstral analysis. Procedure: Considering at least 6 consecutive frames for computing these averages. Also produce framewise plots of cepstrally smoothed spectra and cepstral sequence similar to those shown on slide 30 of “Cepstral Analysis of Speech” for your example.

Vowel Example: centralLow_a.wav

Average Pitch and Formant Values

A Python script is used to perform cepstral analysis on the audio file. The script automatically identified a stable, voiced segment of 27 consecutive frames (from frame 59 to 85), which satisfies the "at least 6 consecutive frames" requirement.

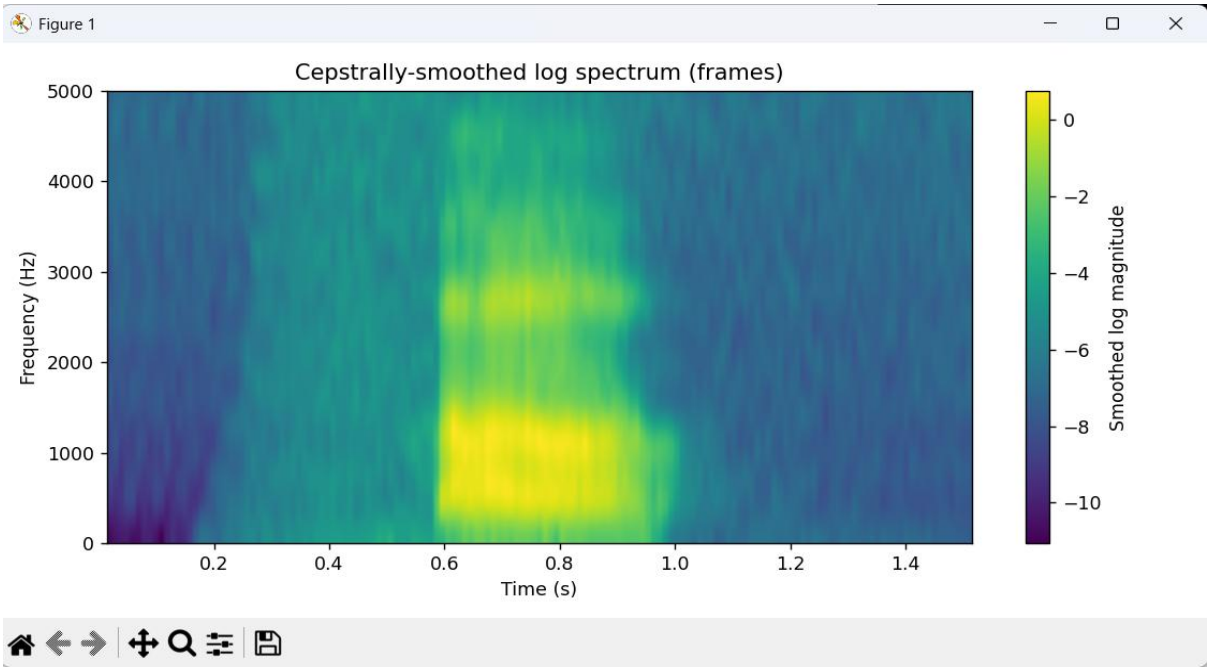
The average values for pitch and the first three formants over this segment are:

```
PS C:\Users\mehul\OneDrive\Desktop\recordings\vowels> python obj2.py

Selected frame indices: [59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 8
3, 84, 85]
Average F1: 726.1 Hz
Average F2: 1570.1 Hz
Average F3: 2959.6 Hz
Average Pitch: 142.9 Hz
```

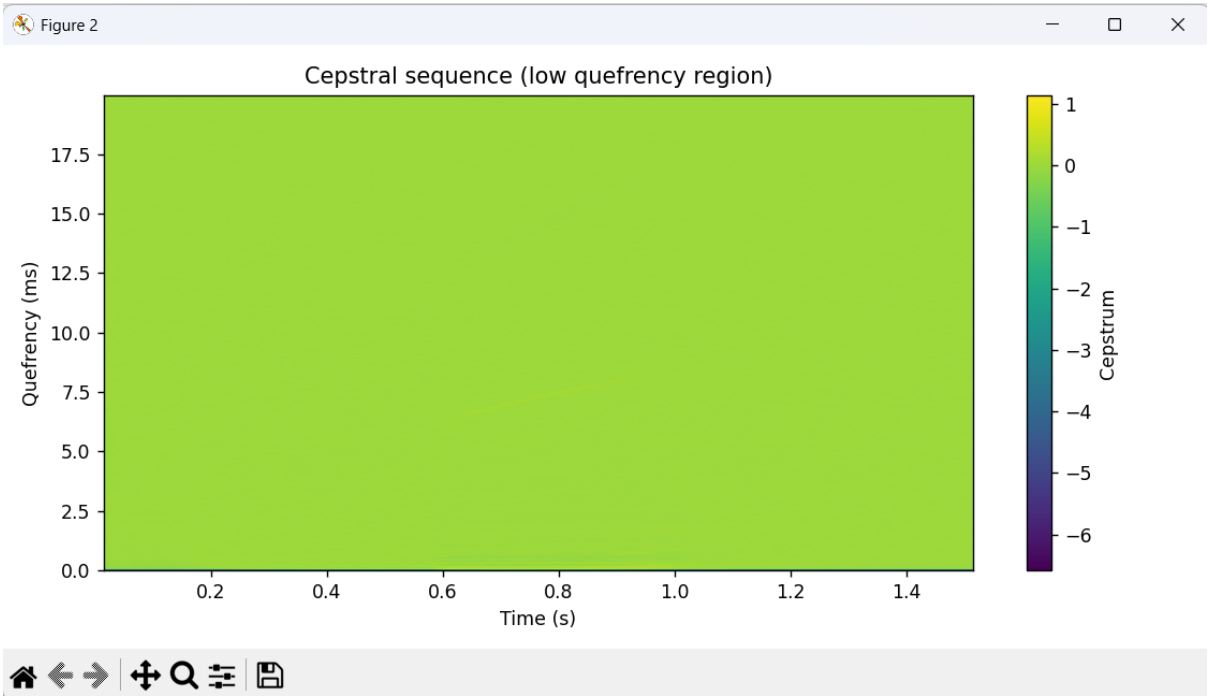
These results are consistent with the known acoustic properties of a low central /a/ vowel (high F1, mid-range F2).

Framewise Analysis Plots:

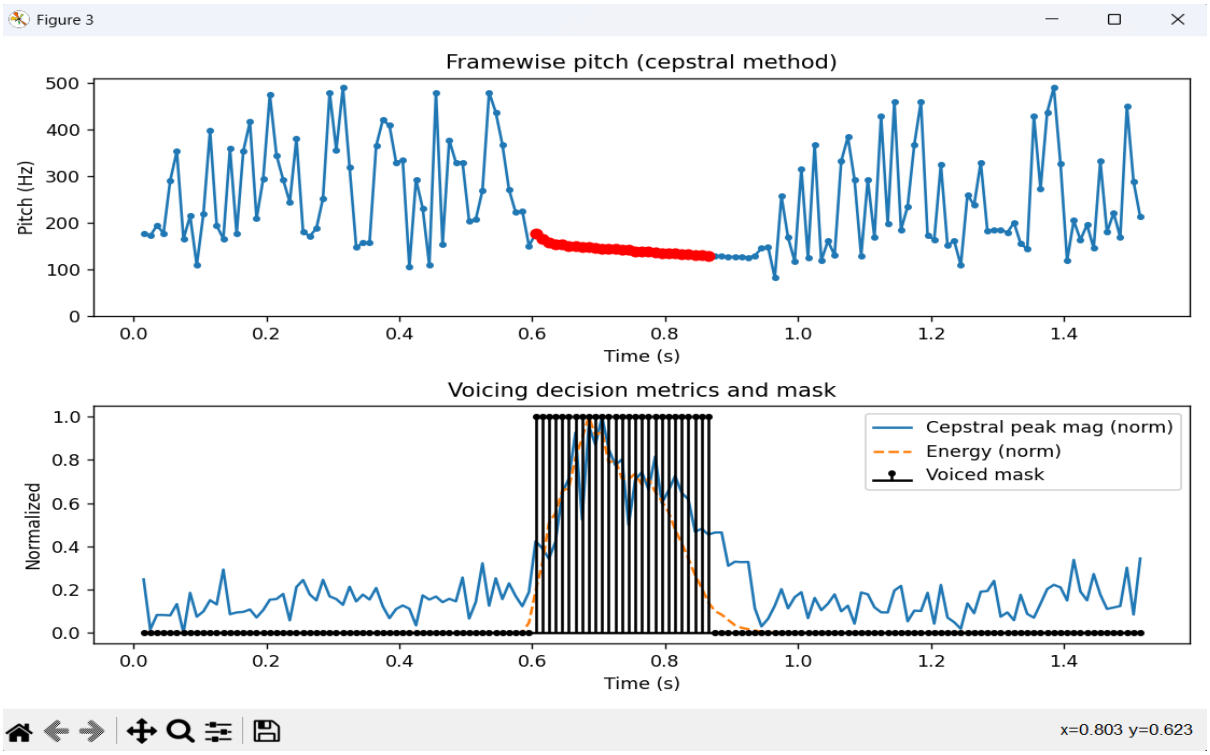
1. **Cepstrally-Smoothed Log Spectrum:** This plot shows the formant structure (F1, F2, F3) as bright horizontal bands over time. The stable vowel segment is clearly visible between 0.6 and 0.9 seconds.



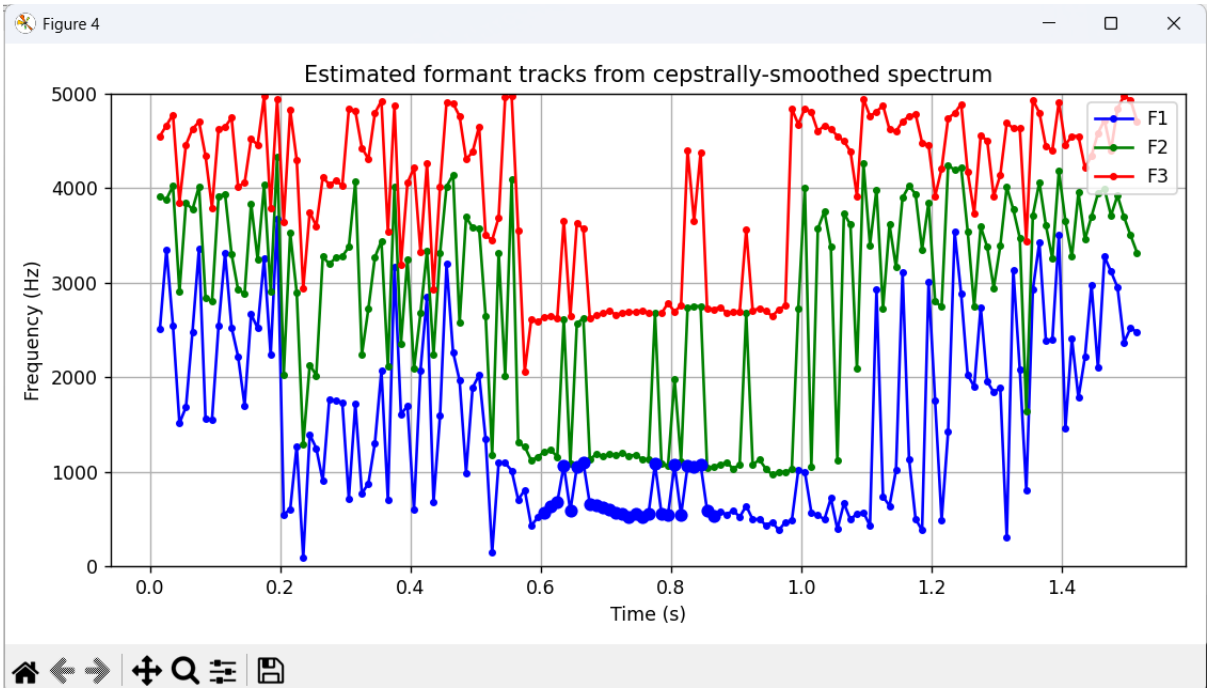
2. **Cepstral Sequence (Quefrency Plot):** This plot shows the cepstral energy over time. The faint but consistent horizontal line around 7ms (1 / 142.9 Hz ≈ 0.007s) represents the pitch period.



3. **Pitch and Voicing:** The plot below shows the framewise pitch (top) and the voicing decision logic (bottom). The red dots highlight the stable pitch of the 27 selected frames.



4. **Formant Tracks:** This plot shows the contours of F1 (blue), F2 (green), and F3 (red) over time. The tracks are stable and clear during the selected voiced segment.



Vowel Example: frontMid_e.wav

Average Pitch and Formant Values

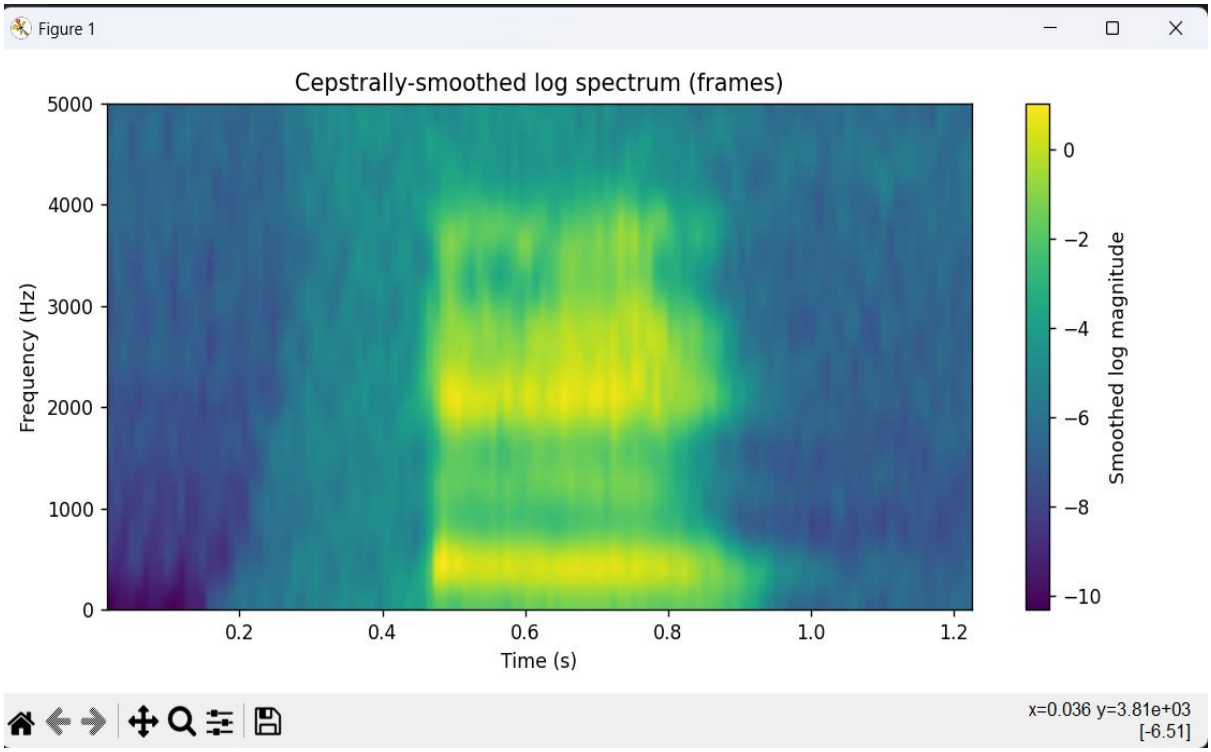
A Python script was used to perform cepstral analysis on the audio file. The script automatically identified a stable, voiced segment of 33 consecutive frames (from frame 48 to 80), satisfying the "at least 6 consecutive frames" requirement.

The average values for pitch and the first three formants over this segment are:

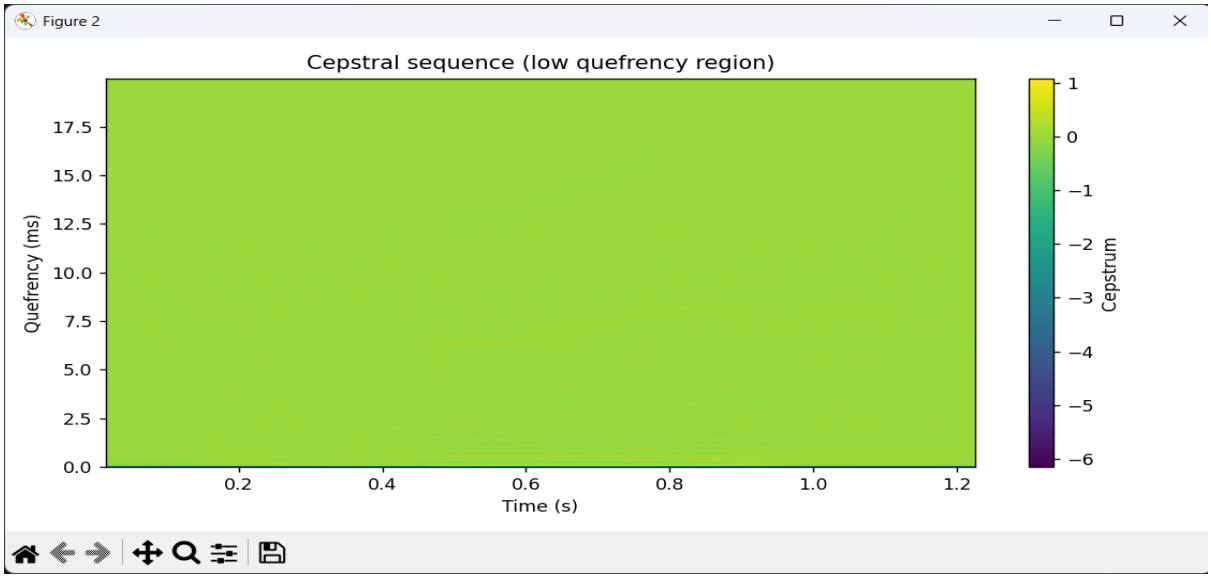
```
PS C:\Users\mehul\OneDrive\Desktop\recordings\vowels> python obj2.py

  Selected frame indices: [48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 7
  2, 73, 74, 75, 76, 77, 78, 79, 80]
  Average F1: 414.4 Hz
  Average F2: 1836.5 Hz
  Average F3: 3005.7 Hz
  Average Pitch: 142.5 Hz
```

Framewise Analysis Plots

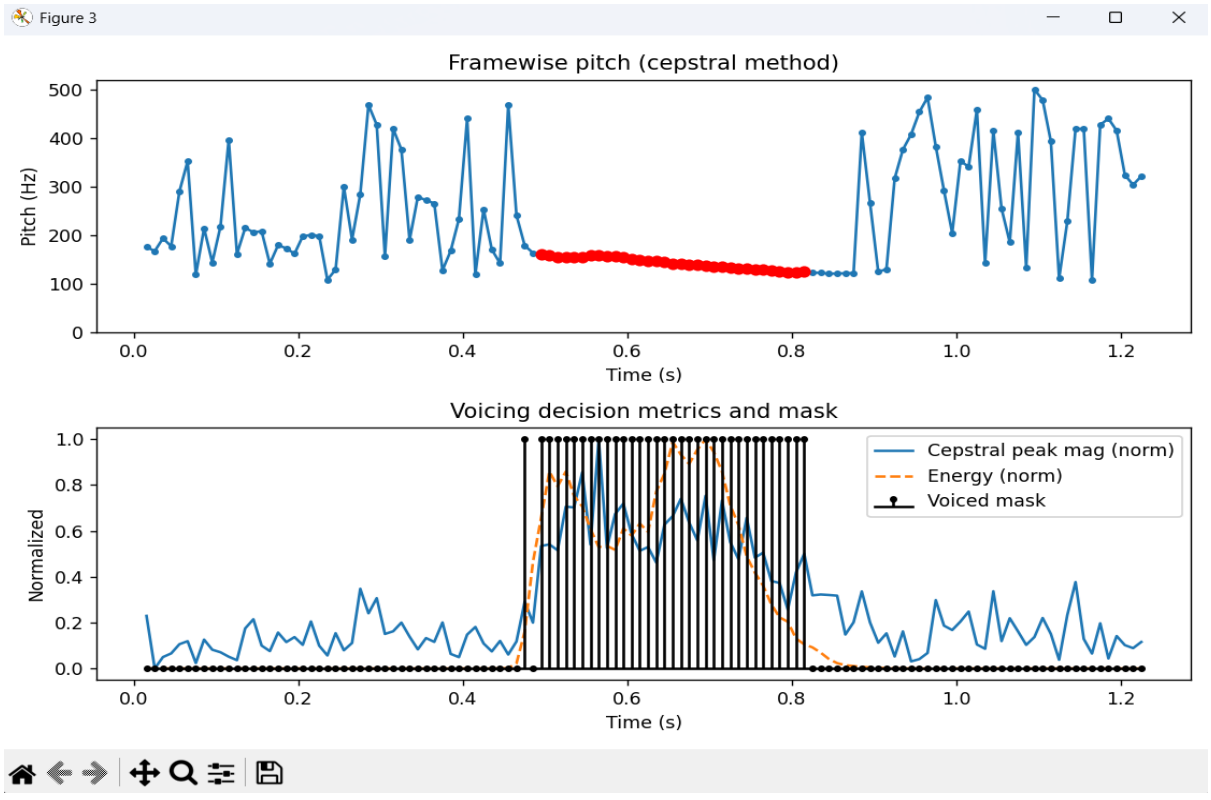
1. **Cepstrally-Smoothed Log Spectrum:** This plot shows the formant structure as bright horizontal bands over time. The stable vowel segment is clearly visible between ~0.45 and ~0.85 seconds.



2. **Cepstral Sequence (Quefrency Plot):** This plot shows the cepstral energy. The faint but consistent horizontal line around 7 ms (1 / 142.5 Hz ≈ 0.007s) during the vowel segment represents the pitch period.



3. **Pitch and Voicing:** The plot below shows the framewise pitch (top) and the voicing decision (bottom). The red dots highlight the stable pitch of the 33 selected frames.



4. **Formant Tracks:** This plot shows the contours of F1 (blue), F2 (green), and F3 (red) over time. The tracks are stable and clear during the selected voiced segment.