DIGITAL FORENSICS PROJECT REPORT

Voice Analysis Using Pitch Recognition in Forensic Investigations

Submitted by:

Arfa Rehman (IS-010/2022-23)
Maha Mustafa (IS-058/2025)
Nabiha Roshan (IS-041/2024)

TABLE OF CONTENT

1. Introduction

The human voice is a rich and dynamic biometric identifier that carries not only linguistic information but also emotional, physiological, and psychological cues. In the field of digital forensics, voice analysis is increasingly becoming a crucial tool for identification, emotion detection, and authenticity verification. This project explores the variations in vocal features across different emotional states using pitch recognition, spectrogram, formant analysis, and intensity mapping.

We employed the software PRAAT, a widely used tool for phonetic analysis, to analyze audio recordings of five distinct emotional states—Angry, Neutral, Noisy Neutral, Sad, and Fearful—and compared them against a synthetically cloned voice generated via Play.ht. The main aim was to study and document acoustic variations and understand how such analyses can be used in forensic settings for tasks like emotion recognition, speaker verification, and voice cloning detection.

2. Objectives of the Project

The primary objective of this project was to perform pitch-based voice analysis on human speech under varying emotional states, such as anger, sadness, fear, and neutrality. By utilizing PRAAT software, we aimed to extract and compare key acoustic features including spectrograms, pitch contours, formants, and intensity patterns across these emotional categories. A significant aspect of the study involved analyzing real human voices and contrasting them with AI-cloned voices to identify potential acoustic discrepancies. This comparative analysis aimed to highlight subtle variations that may assist in distinguishing between genuine and synthetic speech. Ultimately, the project sought to explore the practical implications of voice emotion analysis and voice cloning detection in the field of digital forensics—particularly in the areas of authenticity verification, speaker profiling, and the identification of deepfake audio, thereby contributing to advancements in forensic audio authentication.

3. Tools and Methodology

Tools Used:

1. PRAAT: A software for the analysis of speech in phonetics.
2. Play.ht: An online voice cloning tool used to synthesize cloned voice samples.
3. Audio Recorder: Used to record natural human voice samples in WAV format.

Procedure:

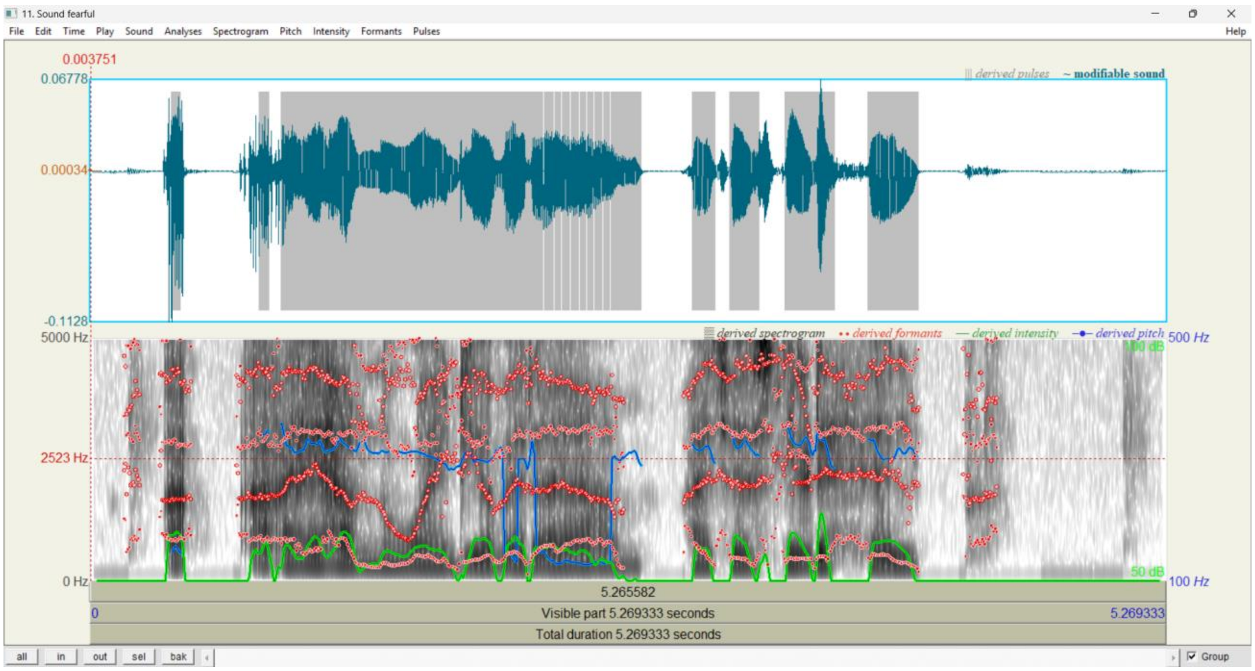
1. Voice Recording: Each group member recorded 5 samples in WAV format.
2. Cloned Voice Generation: Only One member's voice was cloned using Play.ht (because of access limitation).

3. Acoustic Feature Extraction using PRAAT: Pitch (F0), Spectrogram, Formants, and Intensity.
4. Data Analysis: Quantified key metrics such as pitch, amplitude, energy, intensity, mean power and standard deviation.

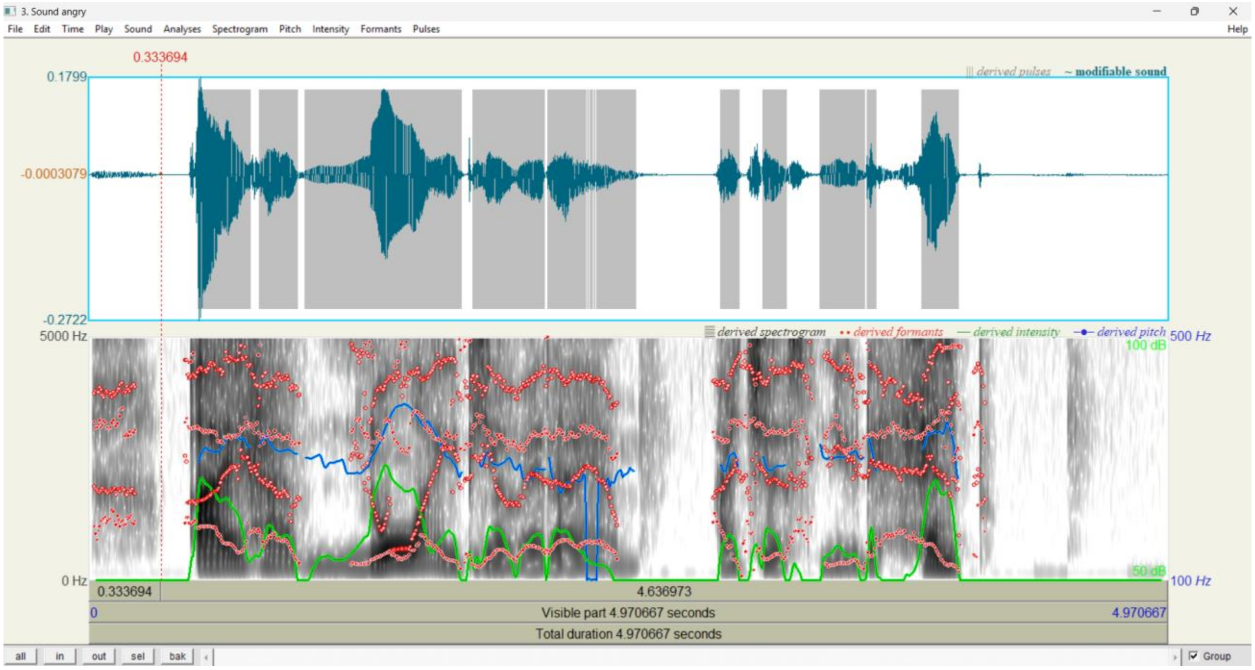
4. Analysis

Member 1: Arfa Rehman

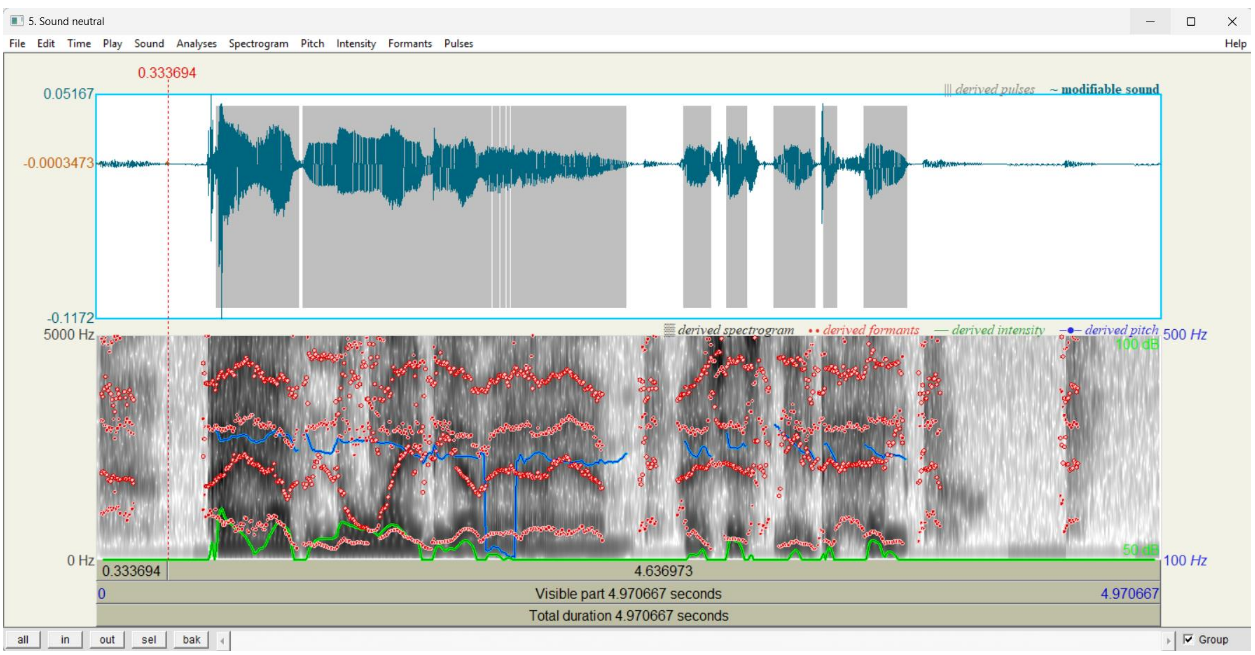
Fearful Arfa:



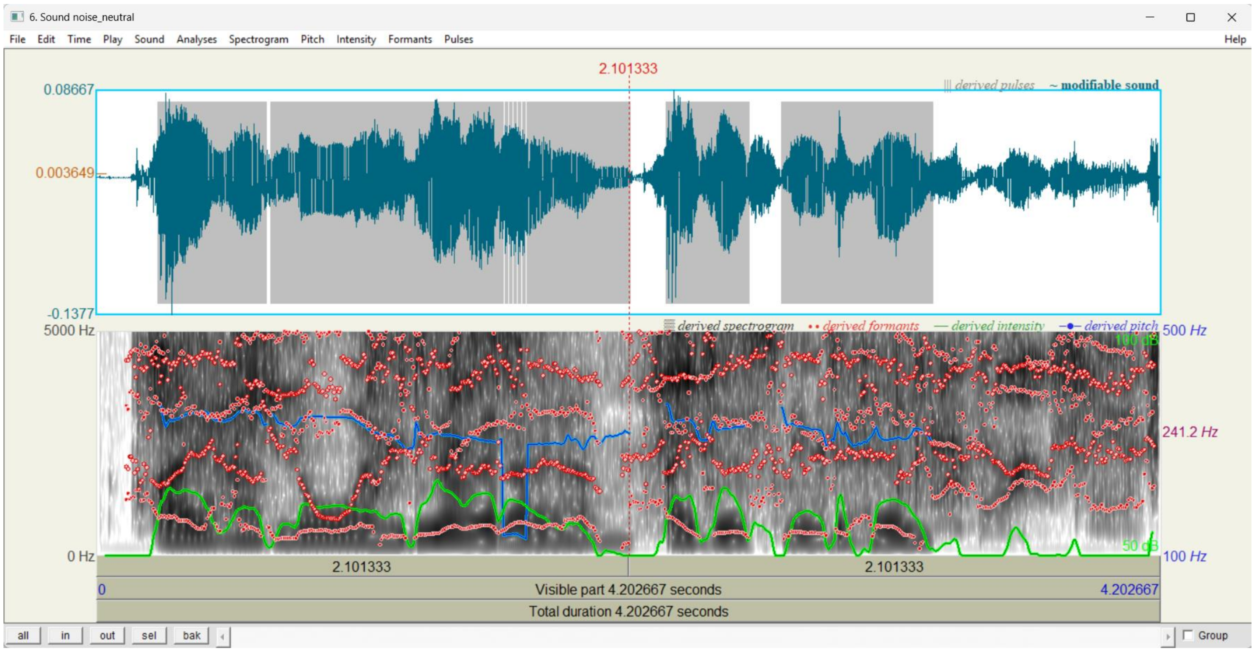
Angry Arfa:



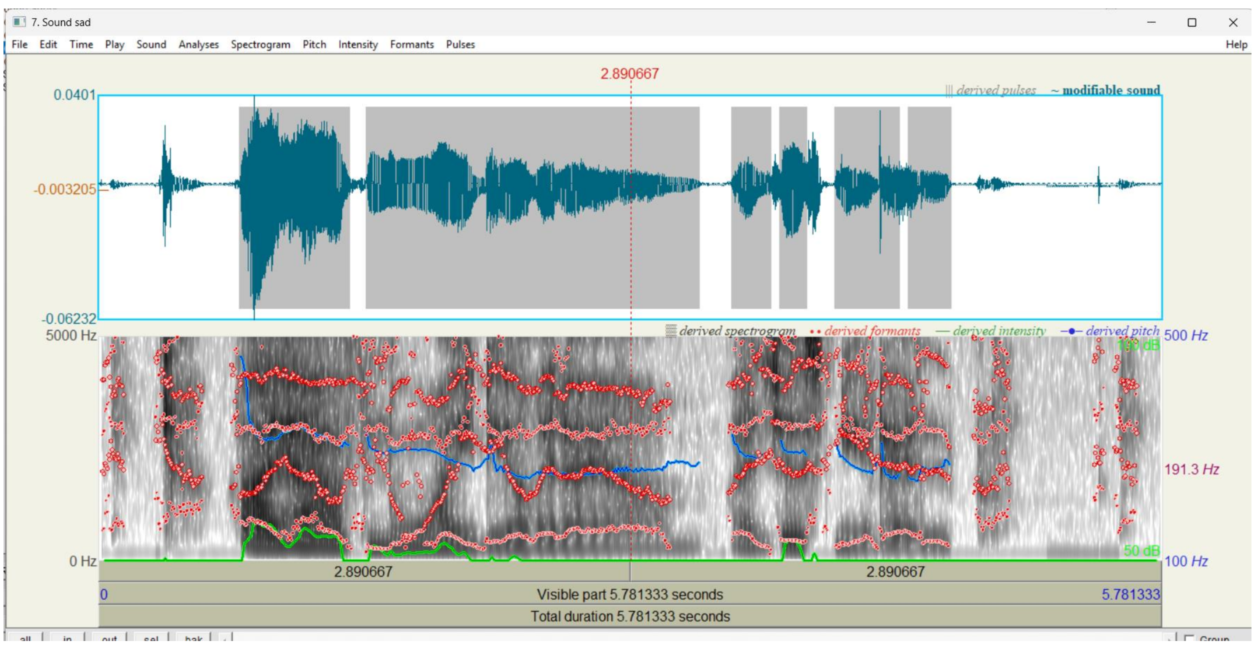
Neutral Arfa:



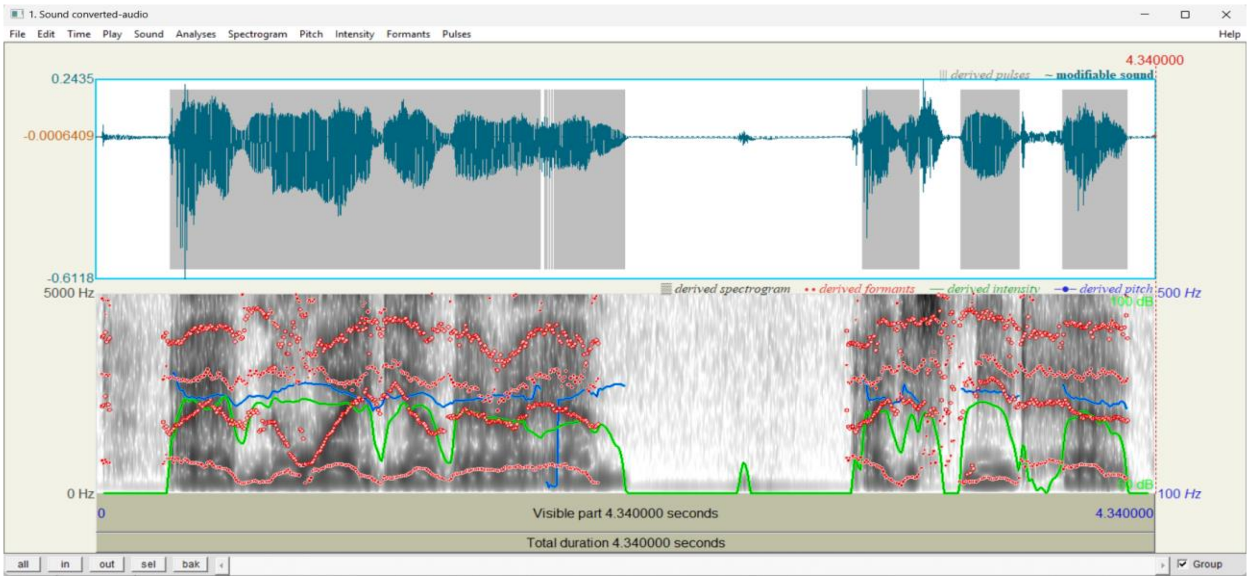
Noise-Neutral Arfa:

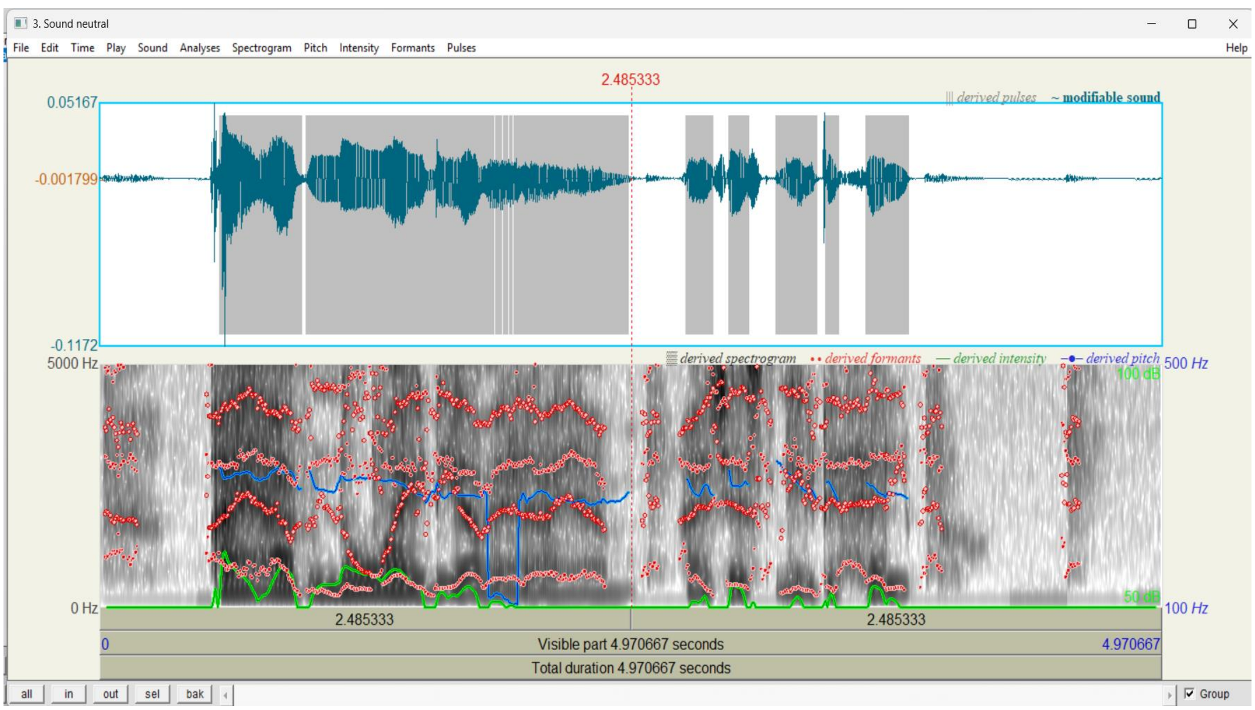


Sad Arfa:



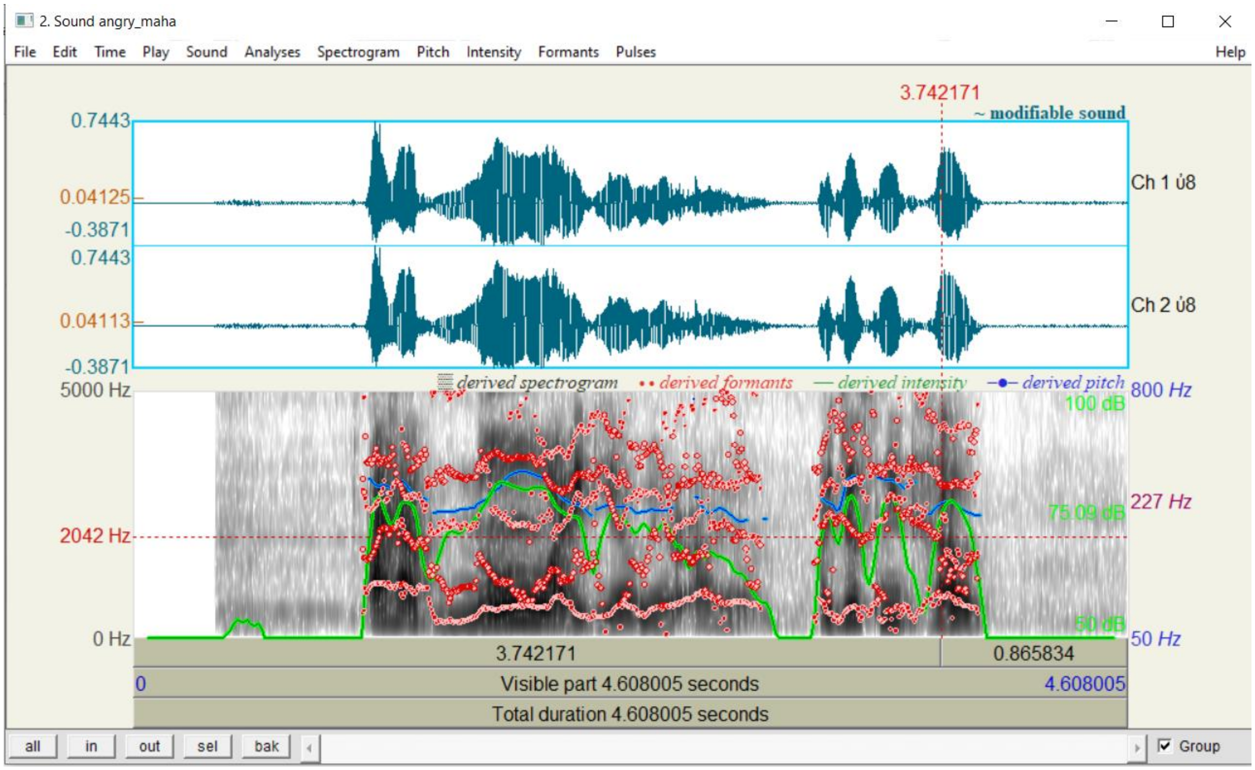
AI generated Arfa:



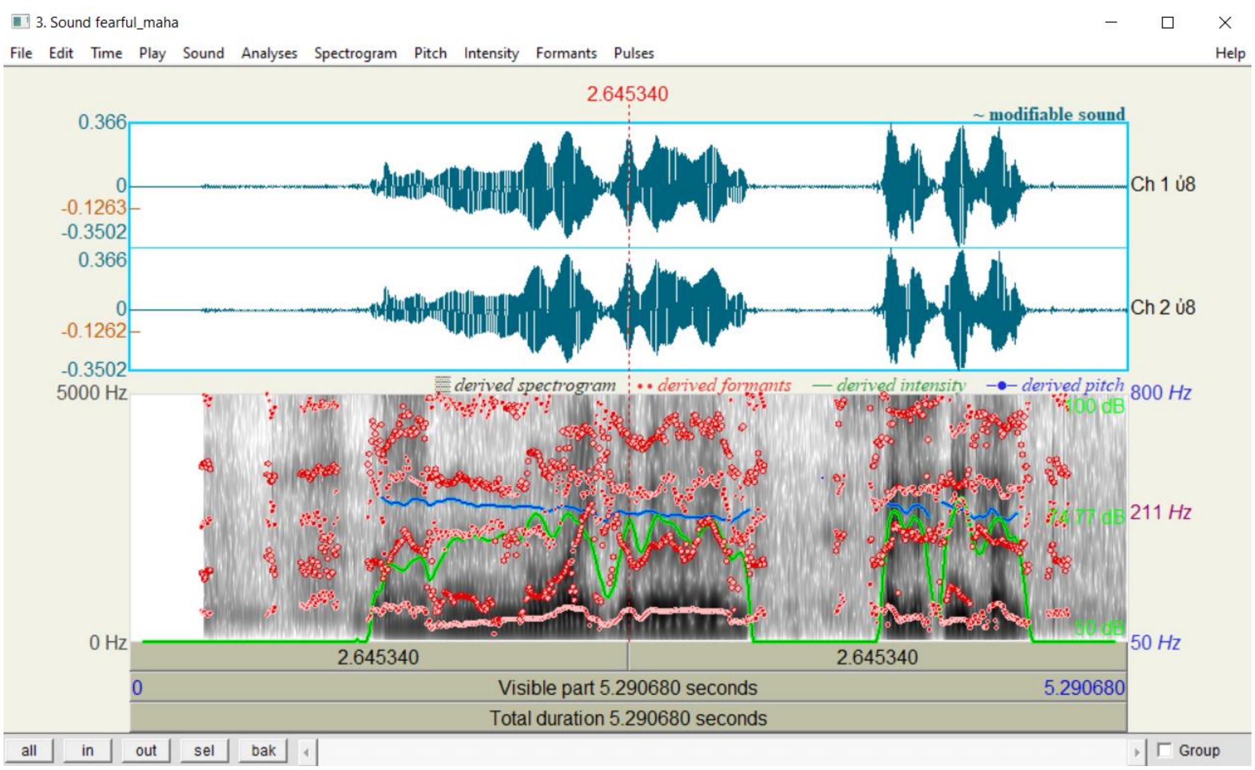


Member 2: Maha Mustafa

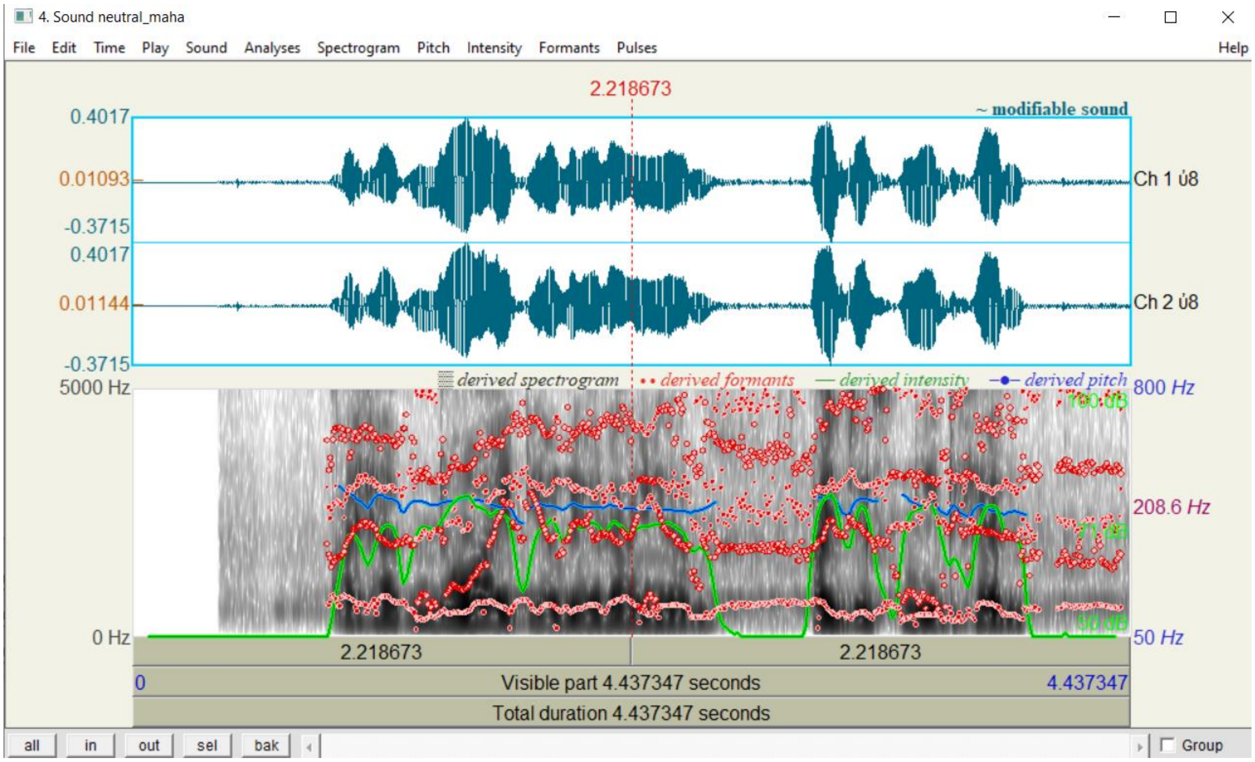
Angry Maha



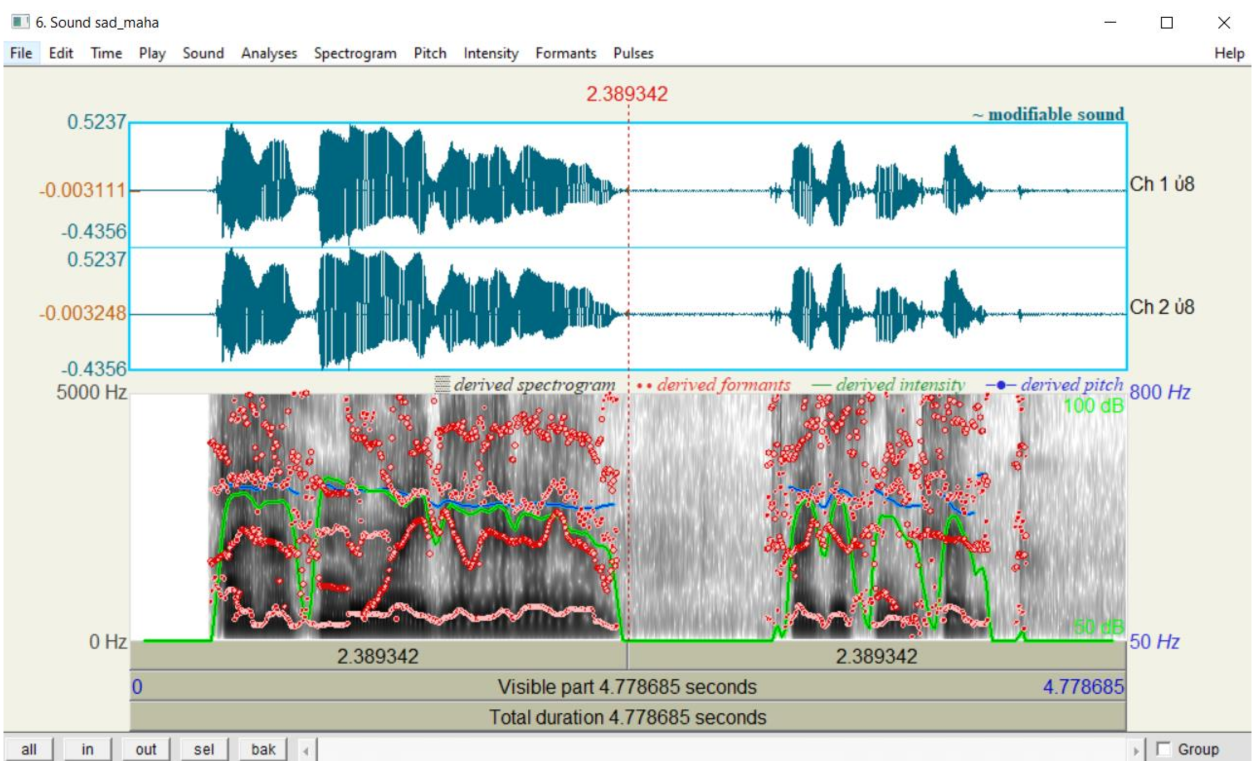
Fearful Maha



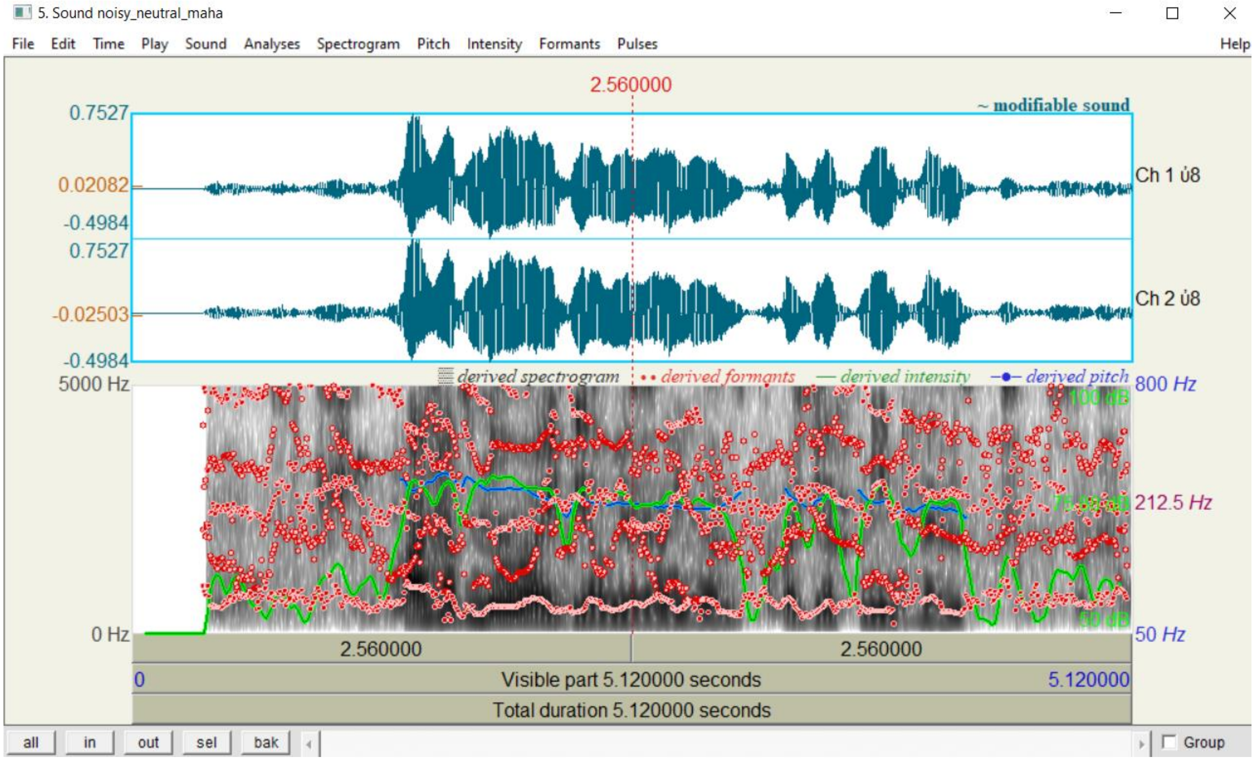
Neutral Maha



Sad Maha

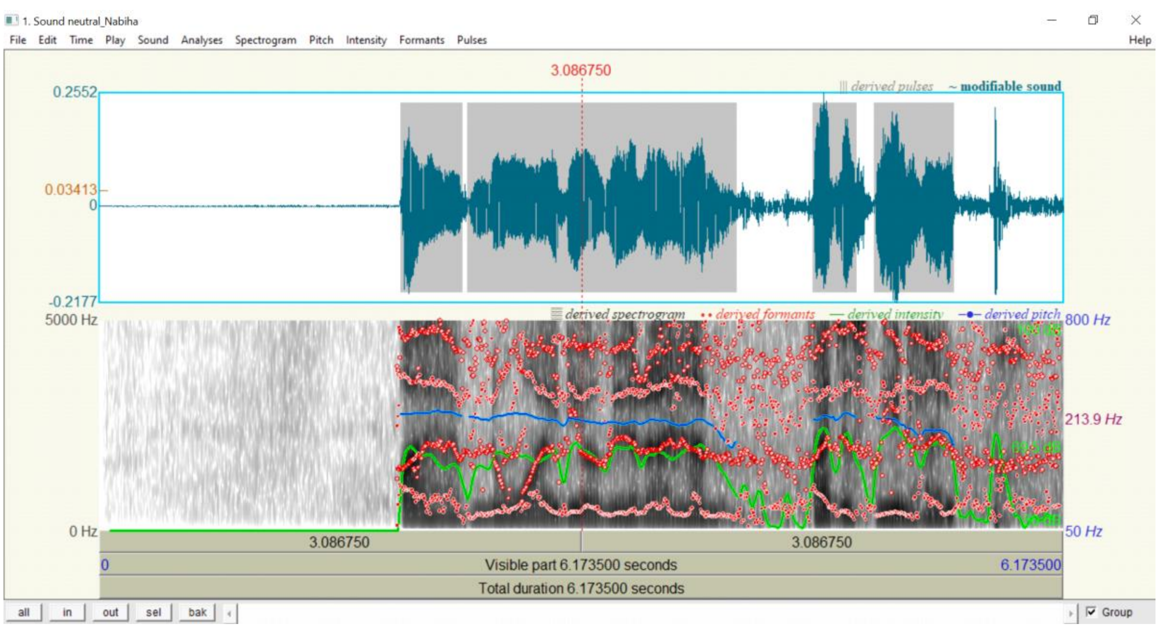


Noisy neutral Maha

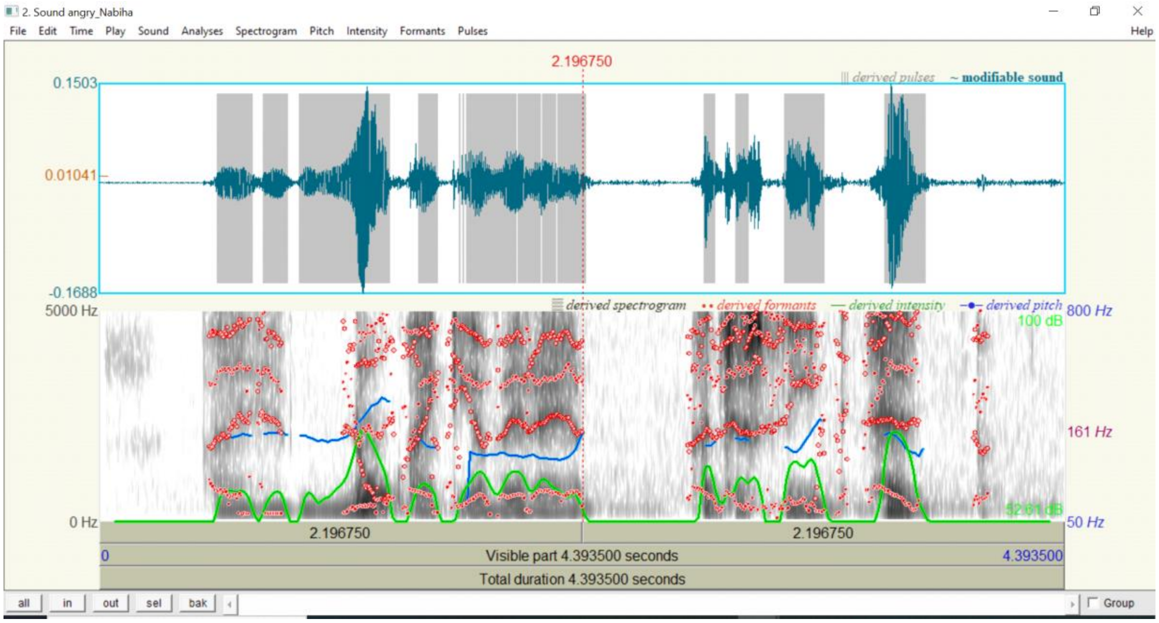


Member 3: Nabiha Roshan

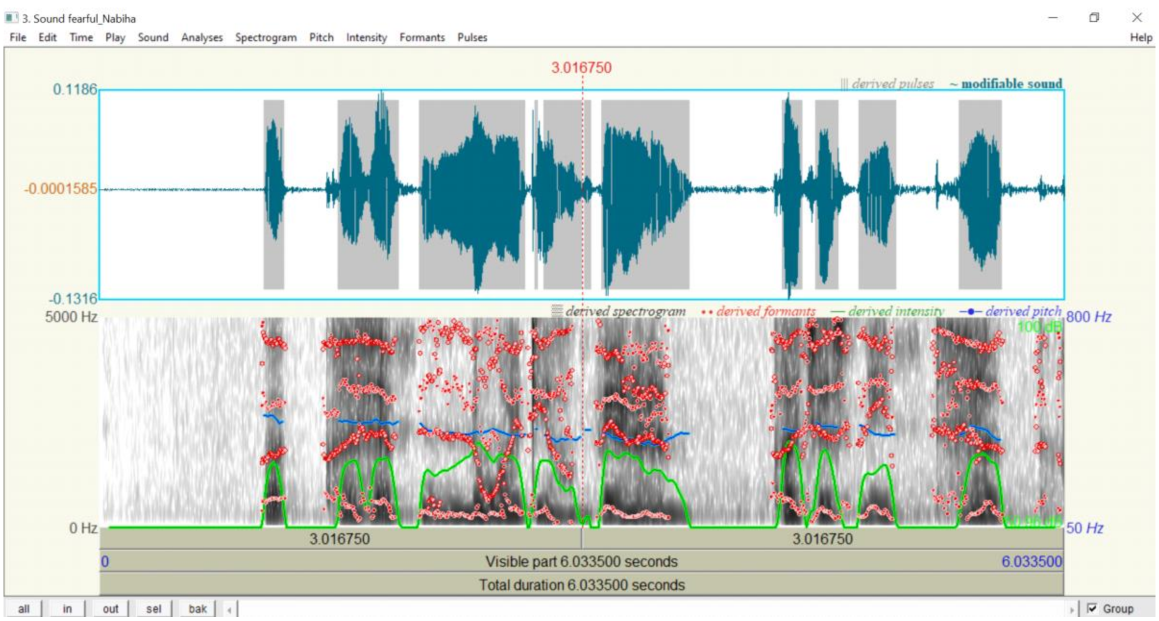
Neutral Nabiha



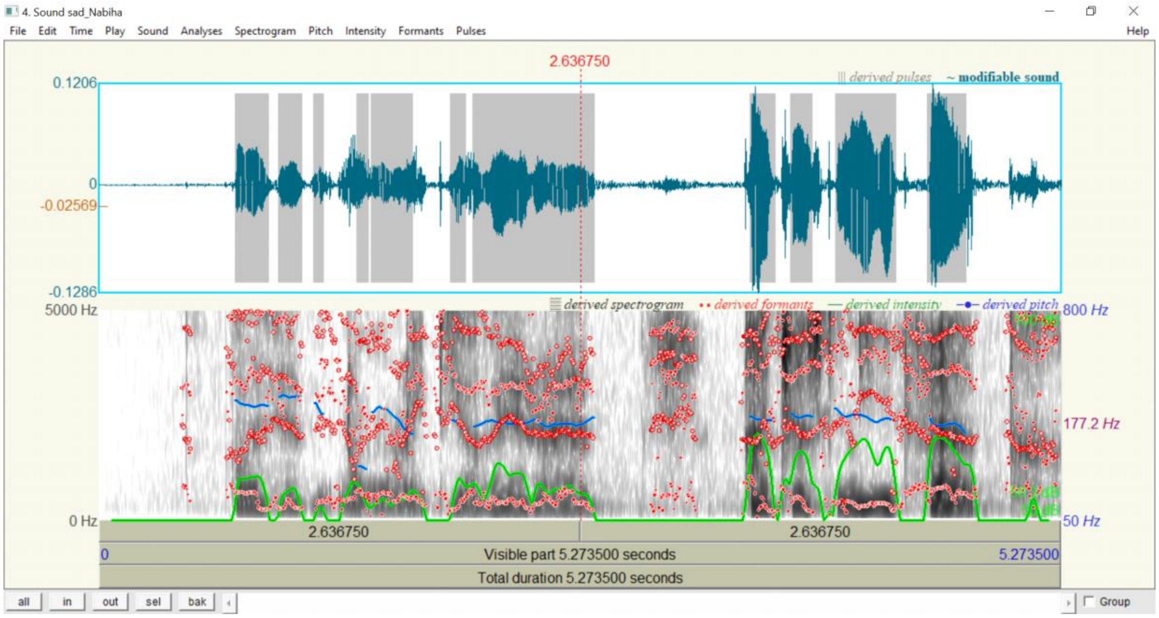
Angry Nabiha



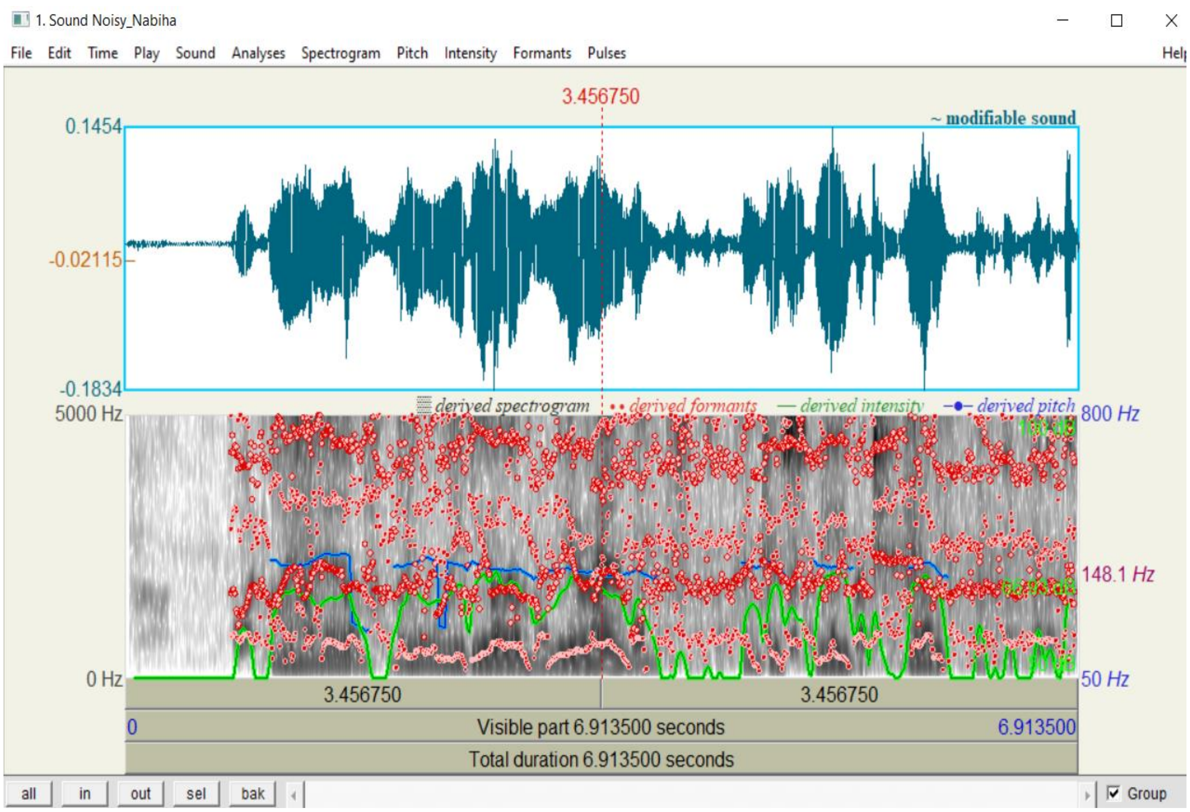
Fearful Nabiha



Sad Nabiha



Noisy Neutral Nabiha



5. Observations:

Emotion	Min Amplitude (pascal)	Max Amplitude (pascal)	Mean Amplitude (Pascal)	Root Mean Square (Pascal)	Total Energy ($Pascal^2$ sec)	Mean Power ($Watt/m^2$)	Standard Deviation (Pascal)	Intensity (dB)
Angry Arfa	-0.272	0.179	-0.00074	0.0213	0.0023	1.14E-06	0.0213	60.55
Neutral Arfa	-0.117	0.0517	-0.00072	0.00686	0.000234	1.18E-07	0.00682	50.71
Noisy Neutral Arfa	-0.137	0.0867	-0.00078	0.0175	0.00129	7.70E-07	0.0175	58.86
Sad Arfa	-0.0623	0.0401	-0.00059	0.00551	0.000175	7.60E-08	0.00548	48.81
Fearful Arfa	-0.1127	0.0678	-0.00074	0.00967	0.000493	2.34E-07	0.00965	53.69
Fake Arfa	-0.6118	0.2435	-0.00401	0.04992	0.0108	6.23E-06	0.0498	67.95
Angry Maha	-0.3870	0.744293	0.0000713	0.083923	0.03245	1.76080e-05	0.0839241	72.46
Neutral Maha	-0.3715	0.401733	-0.00000322	0.066643	0.01970	1.11032e-05	0.066643	70.45
Noisy Neutral Maha	-0.4984	0.752747	0.00010597	0.105927	0.05744	2.80515e-05	0.105928	74.48
Sad Maha	-0.4356	0.523712	3.33×10^{-6}	0.104194	0.05187	2.71408e-05	0.104194	74.34
Fearful Maha	-0.3502	0.36596	1.54647e-05	0.05738	0.01742	8.2322e-06	0.057383	69.16
Angry Nabiha	-0.16864	0.15152	0.000171	0.01864	0.001527	8.69e-07	0.0186	59.39
Neutral Nabiha	-0.2178	0.25696	-2.8e-05	0.0391	0.009439	3.82e-06	0.0391	65.85
Noisy Neutral Nabiha	-0.18344	0.14545	-7.1e-05	0.02939	0.005917	2.16e-06	0.02939	63.34
Sad Nabiha	-0.12949	0.12265	-1.7e-05	0.01963	0.002032	9.63e-07	0.01963	59.84
Fearful Nabiha	-0.13235	0.1181	-1.3e-05	0.02412	0.003511	1.45e-06	0.02412	61.63

This table presents a detailed acoustic analysis of speech signals recorded from three speakers—Arfa, Maha, and Nabiha—under various emotional states: Angry, Neutral, Noisy Neutral, Sad, Fearful, and Fake. It includes technical parameters such as minimum and maximum amplitude (in Pascals), mean amplitude, root mean square (RMS) amplitude, total energy (Pascal²·sec), mean power (Watt/m²), standard deviation of amplitude, and sound intensity (in decibels). These features help quantify how emotions affect voice characteristics. For instance, emotions like Angry and Fake consistently show higher RMS values, energy, power, and dB levels, indicating louder, more forceful speech. In contrast, Sad and Fearful emotions exhibit lower RMS, energy, and intensity, reflecting softer, more subdued vocal patterns. Noisy Neutral samples show higher RMS and energy compared to normal Neutral speech, likely due to background noise or vocal strain, and can help in detecting environmental disturbances or stress. The Fake Arfa entry has extreme values—especially high energy and amplitude variation—suggesting it may be an artificially produced or exaggerated emotional state. Speaker-specific differences are also observable; for example, Maha generally exhibits higher RMS and intensity across emotions compared to Arfa and Nabiha, indicating a stronger vocal presence. Overall, this table enables emotion detection, comparison of vocal effort, detection of unnatural or fake expressions, noise influence assessment, and speaker-specific vocal profiling. These insights are valuable for emotion recognition systems, psychological research, human-computer interaction, lie detection, and developing emotionally intelligent AI.

6. Forensic Relevance and Learning Outcomes:

This study demonstrates significant forensic applications through the analysis of vocal characteristics across various emotional states. Emotion detection plays a vital role in identifying stress, fear, or aggression in audio samples, which can be crucial in criminal investigations or interrogation analyses. Furthermore, the comparison between real and cloned voices supports speaker verification techniques, aiding in the differentiation of authentic human speech from AI-generated content. This becomes particularly relevant in the context of voice cloning detection, where algorithms can be trained to recognize the subtle discrepancies in artificially produced speech. Additionally, tampering detection benefits from this approach by identifying abnormal acoustic patterns, which may indicate manipulation or editing in audio evidence.

From an academic perspective, the learning outcomes of this project are multifaceted. It has deepened our understanding of how emotional states influence vocal parameters such as amplitude and intensity. We gained practical experience using PRAAT software for speech analysis, which is a widely accepted tool in both linguistic and forensic domains. The work also highlighted the current limitations of AI in replicating the full emotional range and subtleties of human speech. Ultimately, this project reinforces the importance of voice forensics in modern investigations, especially in detecting fraudulent or manipulated audio evidence.

7. Conclusion

This project provided comprehensive insights into the acoustic characteristics of emotional speech and its significance in digital forensics. The findings support the hypothesis that emotionally driven acoustic variations are strong indicators of speech authenticity and can be pivotal in forensic examinations. The study also highlights the limitations of current voice cloning technologies in mimicking genuine emotional speech.