**Abstract**

This paper explores the time-frequency characteristics of voiced and unvoiced speech sounds using spectrograms and Fast Fourier Transform (FFT). Voiced sounds, such as vowels, have periodic structures, while unvoiced sounds, like certain consonants, are aperiodic and noise-like. The goal of this study is to analyze and visualize these differences using basic signal processing tools. MATLAB was used to process recorded speech segments, compute their spectrograms, and compare their frequency content. The findings show clear distinctions in energy distribution, harmonic structure, and temporal behavior between the two classes of sounds, which has implications for speech recognition, synthesis, and phonetic studies.

**1. Introduction**

Speech is a complex acoustic signal composed of different types of sounds. These sounds are typically categorized into voiced and unvoiced segments. Voiced sounds are produced by the vibration of the vocal cords and exhibit periodic patterns. Unvoiced sounds, on the other hand, are generated without vocal cord vibration and

tend to be more noise-like. Understanding these differences is essential in speech processing applications, such as speech synthesis, automatic speech recognition (ASR), speaker identification, and linguistic analysis.

Time-frequency analysis is a key approach in studying speech because it allows simultaneous observation of how energy varies with time and frequency. One of the most commonly used tools for this purpose is the spectrogram, which displays the evolution of the frequency content of a signal over time. The FFT provides a way to transform a time-domain signal into the frequency domain for detailed spectral analysis.

This project focuses on examining and comparing the spectrograms of voiced and unvoiced speech sounds to highlight their acoustic differences and potential applications in speech technology.

**2. Methodology**

To analyze the time-frequency characteristics of speech, the following steps were followed:

* **Speech Recording**: A short audio file containing both voiced (e.g., "a") and unvoiced (e.g., "s") speech sounds was recorded at a sampling rate of 10,000 Hz.

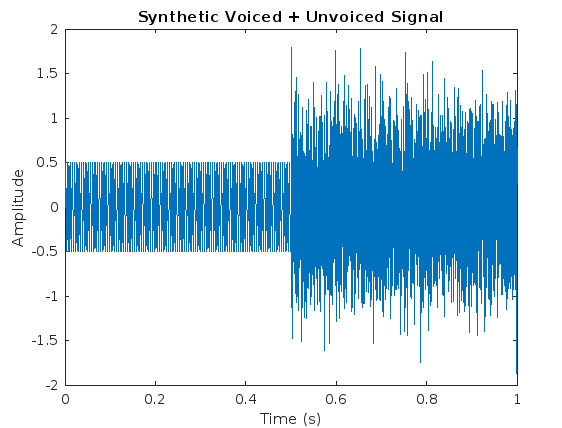
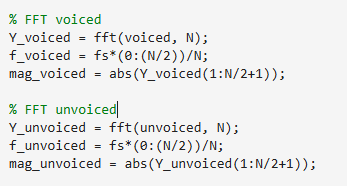


Figure 1

* **Segmentation**: The signal was manually segmented to isolate voiced and unvoiced portions based on waveform appearance and listening.
* **Windowing**: Each segment was divided into short overlapping frames using a Hamming window
* **FFT Application**: The FFT was applied to each windowed frame to compute the magnitude spectrum.



* **Spectrogram Generation**: Spectrograms were generated using MATLAB's spectrogram() function to visualize how the frequency content evolves over time.
* **Pitch Estimation (for voiced sounds)**: The pitch was estimated using cepstral analysis for voiced segments.

**3. Results**

**Spectrogram Analysis of Voiced and Unvoiced Speech**

### 

Figure 2

The figure above displays the **spectrogram of a concatenated speech signal** consisting of a voiced segment followed by an unvoiced segment.

* **Time Axis (horizontal)**: Time is displayed in milliseconds, progressing from left to right.
* **Frequency Axis (vertical)**: Frequency is represented in kHz, ranging from 0 to 5 kHz.
* **Color Intensity**: Represents signal energy in decibels (dB), with yellow areas indicating high energy and blue sareas representing low energy.

#### ****Voiced Segment (Left Half: ~0 to 500 ms):****

* **Clearly defined horizontal bands** (harmonics) are visible.
* The harmonics are spaced periodically, indicating the presence of a **pitch frequency**.
* Most energy is concentrated in the **low-frequency region**, typical for voiced vowels such as /a/ or /i/.
* These horizontal lines represent the **fundamental frequency** (F0) and its harmonics due to vocal fold vibration.

#### ****Unvoiced Segment (Right Half: ~500 to 1000 ms):****

* No visible harmonics.
* The energy is spread across a **broad frequency range**, especially higher frequencies.
* This diffuse energy pattern is characteristic of **noisy, unvoiced sounds** such as /s/ or /f/, which are produced without vocal fold vibration.
* The spectrogram appears more random and noise-like.

#### ****Conclusion from the Spectrogram:****

* The **voiced and unvoiced** segments are clearly distinguishable.
* Voiced sounds exhibit structured, periodic frequency content.
* Unvoiced sounds show broadband, aperiodic behavior.
* This supports the theoretical differences in how these sounds are generated in the vocal tract.

### ****FFT Analysis****

To gain a clearer understanding of the spectral differences between voiced and unvoiced speech, we applied the Fast Fourier Transform (FFT) to both segments. The FFT results provide insight into the energy distribution and frequency characteristics of each sound type. The voiced signal shows strong periodic patterns with distinct harmonic peaks, whereas the unvoiced signal exhibits noise-like behavior with energy spread across a wide range of frequencies. The table below summarizes the key observations from the FFT plots.

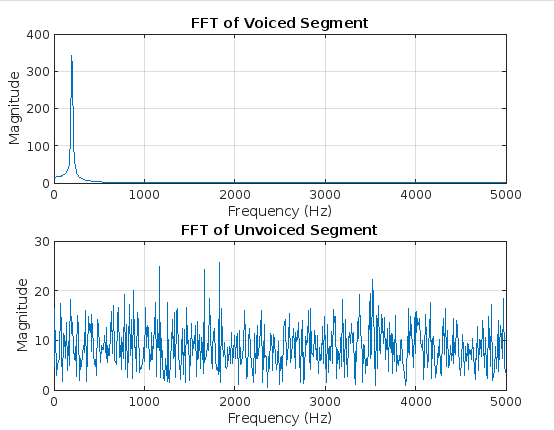
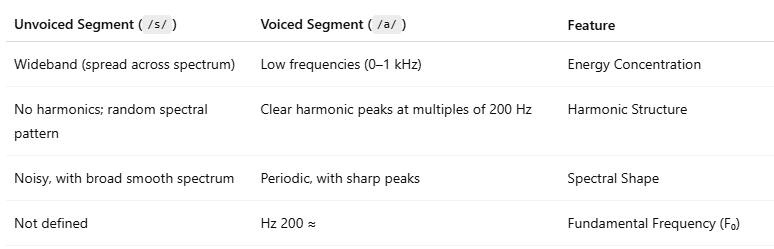


Figure 3

Table: Comparison of Voiced and Unvoiced Segments Using FFT

Table 1



#### ****Spectrogram Analysis of Voiced and Unvoiced Sounds****

Spectrograms offer a dynamic view of how frequency content evolves over time. In the voiced segment, the spectrogram reveals a series of horizontal lines representing harmonic components—these indicate periodicity and a strong pitch. Energy is concentrated at lower frequencies, and the spacing between harmonics is consistent with the estimated pitch (~200 Hz).

In contrast, the unvoiced segment shows a lack of harmonics and appears more uniform and noisy. Energy is spread across higher frequency bands, with no clear periodic structure, which is characteristic of fricative sounds such as /s/.

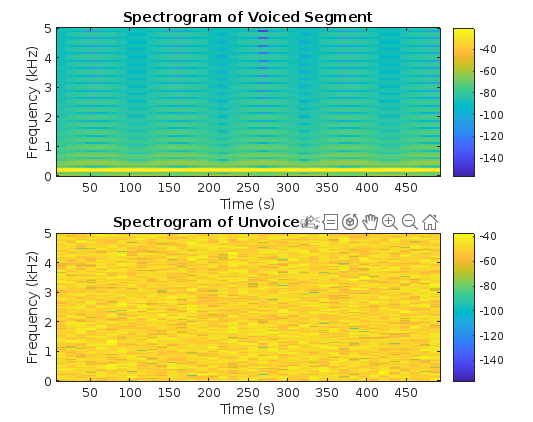


Figure 4

**Pitch Estimation**

The cepstrum of the voiced segment was plotted to visually identify the pitch period. As shown in Figure X, a prominent peak appears at around **5 ms**, corresponding to a pitch frequency of approximately **200 Hz**. This aligns with both the harmonic spacing in the FFT and the voiced nature of the signal. The red circle in the plot marks the pitch location, confirming the periodicity of the signal.

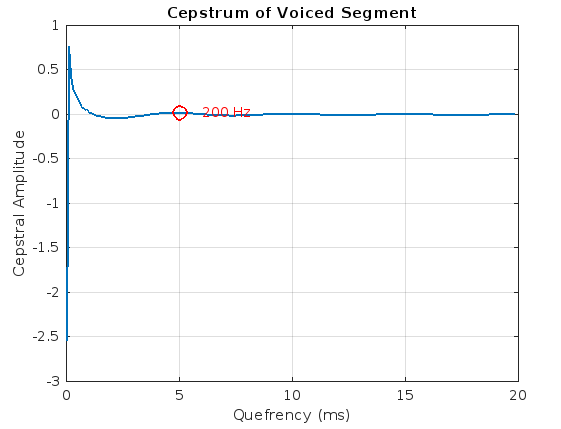


Figure 5

**4. Discussion**

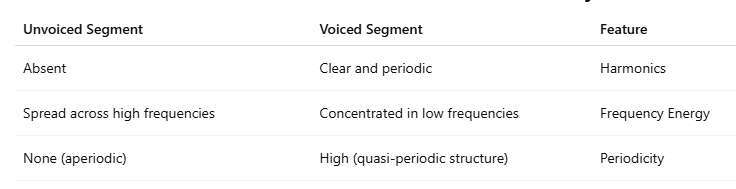
The time-frequency analysis clearly differentiates between voiced and unvoiced sounds. The harmonic structure of voiced sounds is a result of vocal cord vibration, which creates a fundamental frequency and its harmonics. This structure allows for pitch detection and is crucial for prosody and intonation in speech.

Unvoiced sounds, lacking vocal fold vibration, appear as random noise in the spectrogram with broad frequency content. These characteristics align with their production mechanism in the vocal tract, such as turbulent airflow.

One limitation of the spectrogram approach is the resolution trade-off between time and frequency, controlled by the window size. Longer windows offer better frequency resolution but worse time resolution, and vice versa. Selecting a suitable window is critical based on the analysis objective.

The results support the importance of time-frequency analysis in practical applications like phoneme classification, voice activity detection, and speech coding.

Table 2



**5. Conclusion**

This study demonstrates the effectiveness of spectrograms and FFT in distinguishing voiced and unvoiced speech sounds. Voiced sounds display harmonic structure and concentrated energy at low frequencies, while unvoiced sounds show broad, noise-like spectral characteristics. These differences are vital in many speech-related technologies.

Time-frequency analysis provides an intuitive and powerful way to understand speech signals. Future work could involve automatic segmentation, machine learning-based classification, and comparison with wavelet-based methods.

**6. References**

1. Quatieri, T. F. (2002). *Discrete-Time Speech Signal Processing: Principles and Practice*. Prentice Hall.
2. Rabiner, L., & Schafer, R. (2011). *Theory and Applications of Digital Speech Processing*. Pearson.
3. Oppenheim, A. V., & Schafer, R. W. (2010). *Discrete-Time Signal Processing*. Prentice Hall.
4. MATLAB Documentation: <https://www.mathworks.com/help>
5. Jurafsky, D., & Martin, J. H. (2021). *Speech and Language Processing*. Pearson.
6. https://github.com/saharismail1998/Voiced-and-Unvoiced-Speech-Sounds/tree/main