UNIVERSITI KEBANGSAAN MALAYSIA

The National University of Malaysia

SEMESTER 2, 2022/2023

STQD6114 – ANALITIK DATA TAK BERSTRUKTUR

UNSTRUCTURED DATA ANALYTICS

LECTURER:

DR. NOR HAMIZAH BINTI MISWAN

Name	Matric Number
An Hongyu	P120996
Hu Chonghui	P119925

Audio Data Analysis and Interpretation: A Comparative Study of Audio Signals

An hongyu, Hu chonghui

Abstract:

This study explores the characteristics and applications of various waveforms in audio data analysis. The objective is to understand the differences between waveforms and their suitability for different audio processing tasks. The study focuses on waveforms such as sine waves, noise, pulses, sawtooth waves, and square waves. It investigates their properties, including frequency, duration, and amplitude, and examines their spectral features using Fourier transform techniques such as FFT and STFT. The analysis involves generating and manipulating waveforms using the R programming language and relevant libraries such as tuneR and seewave. The results highlight the distinctive qualities of each waveform and their potential applications in audio signal processing, including noise reduction and waveform synthesis. The findings of this study contribute to a deeper understanding of waveforms and their significance in audio data analysis.

1. Introduction:

Waveforms are fundamental components of audio signals, playing a crucial role in the field of audio data analysis. They represent the temporal variation of sound and provide valuable information about the characteristics of audio signals. Understanding waveforms and their properties is essential for a wide range of applications, including audio synthesis, speech recognition, music analysis, and noise reduction.

Waveforms can be generated using various techniques, such as mathematical functions and signal processing algorithms. Common types of waveforms include sine waves, square waves, sawtooth waves, and pulse waves, each with its unique characteristics. Sine waves, for example, represent pure tones with a single frequency, while square waves consist of alternating high and low states, resembling a square shape.

In audio data analysis, the Fourier transform plays a central role in analyzing waveforms. The Fourier transform allows us to decompose a waveform into its constituent frequencies, revealing the spectral content of the signal. Techniques such as the Fast Fourier Transform (FFT) and Short-Time Fourier Transform (STFT) enable the extraction of frequency-domain features, facilitating tasks such as pitch detection, spectral analysis, and audio classification.

By gaining a deeper understanding of waveforms and their relationship with Fourier transform analysis, researchers and practitioners can enhance their ability to analyze and manipulate audio data effectively. This knowledge can contribute to advancements in areas such as audio synthesis, speech processing, and sound recognition systems. Ultimately, a comprehensive understanding of waveforms and their analysis techniques opens up new possibilities for various audio-related applications, leading to advancements in fields such as music production, speech processing, and audio signal processing.

1.1 Literature review

(Halliday et al., 2011) examines the appropriateness of the Fast Fourier Transform for decomposition and reconstruction of wave records taken at fixed locations and transposed to a different temporal and spatial point. In marine renewable energy, advanced control methods based on the future prediction of waves are being developed. These methods are based on the assumption that a forward-looking prediction is available and over the years there has been a conjecture that the FFT may perform this role and that the prediction of wave behavior at any point on the sea surface should be realizable. The validity of this statement is tested using numerical wave records.

The advent of fast Fourier transform methods has greatly expanded our ability to implement Fourier methods on digital computers. (Cooley et al., 1969) gives a description of the algorithm and its programming, followed by a theorem relating its operands, finite sample sequences, and the continuous functions they usually approximate. An analysis of the error caused by discrete sampling in a finite range is given in the form of aliasing. (Cooley et al., 1969) also outlines procedures for computing Fourier integrals, convolutions, and lag products.

(Dutt, 1993) proposes two sets of algorithms that generalize the Fast Fourier Transform (FFT) to the case of non-equidistant nodes at non-integer frequencies and on the interval [-7r, 7r]. These schemes are based on certain analytical considerations combined with the classical fast Fourier transform transform and generalize forward and backward FFT. The number of arithmetic operations required for each algorithm is proportional to N-log N + N-log(I/e), where 6 is the required computational precision and N is the number of nodes. Several related algorithms are also proposed, each utilizing a similar set of analytical and linear algebra techniques. These include efficient versions of fast multipole methods in one dimension and fast algorithms for evaluation, integration, and differentiation of Lagrangian polynomial interpolation. Several numerical examples are used to illustrate the efficiency of the method and compare the performance of two sets of non-uniform FFT algorithms.

2. Methodology

Emotion expression is an essential part of human interaction. The same text can hold different meanings when expressed with different emotions. Thus understanding the text alone is not enough for getting the meaning of an utterance. Acted and natural corpora have been used to detect emotions from speech. Many speech databases for different languages including English, German, Chinese, Japanese, Russian, Italian, Swedish and Spanish exist for modeling emotion recognition. Since there is no reported reference of an available Arabic corpus, we decided to collect the first Arabic Natural Audio Dataset (ANAD) to recognize discrete emotions.

Embedding an effective emotion detection feature in speech recognition system seems a promising solution for decreasing the obstacles faced by the deaf when communicating with the outside world. There exist several applications that allow the deaf to make and receive phone calls normally, as the hearing-impaired individual can type a message and the person on the other side hears the words spoken, and as they speak, the words are received as text by the deaf individual. However, missing the emotion part still makes these systems not hundred percent reliable. Having an effective speech to text and text to speech system installed in their everyday life starting from a very young age will hopefully replace the human ear. Such systems will aid deaf people to enroll in normal schools at very young age and will help them to adapt better in classrooms and with their classmates. It will help them experience a normal childhood and hence grow up to be able to integrate within the society without external help.

Eight videos of live calls between an anchor and a human outside the studio were downloaded from online Arabic talk shows. Each video was then divided into turns: callers and receivers. To label each video, 18 listeners were asked to listen to each video and select whether they perceive a happy, angry or surprised emotion. Silence, laughs and noisy chunks were removed. Every chunk was then automatically divided into 1 sec speech units forming our final corpus composed of 1384 records.

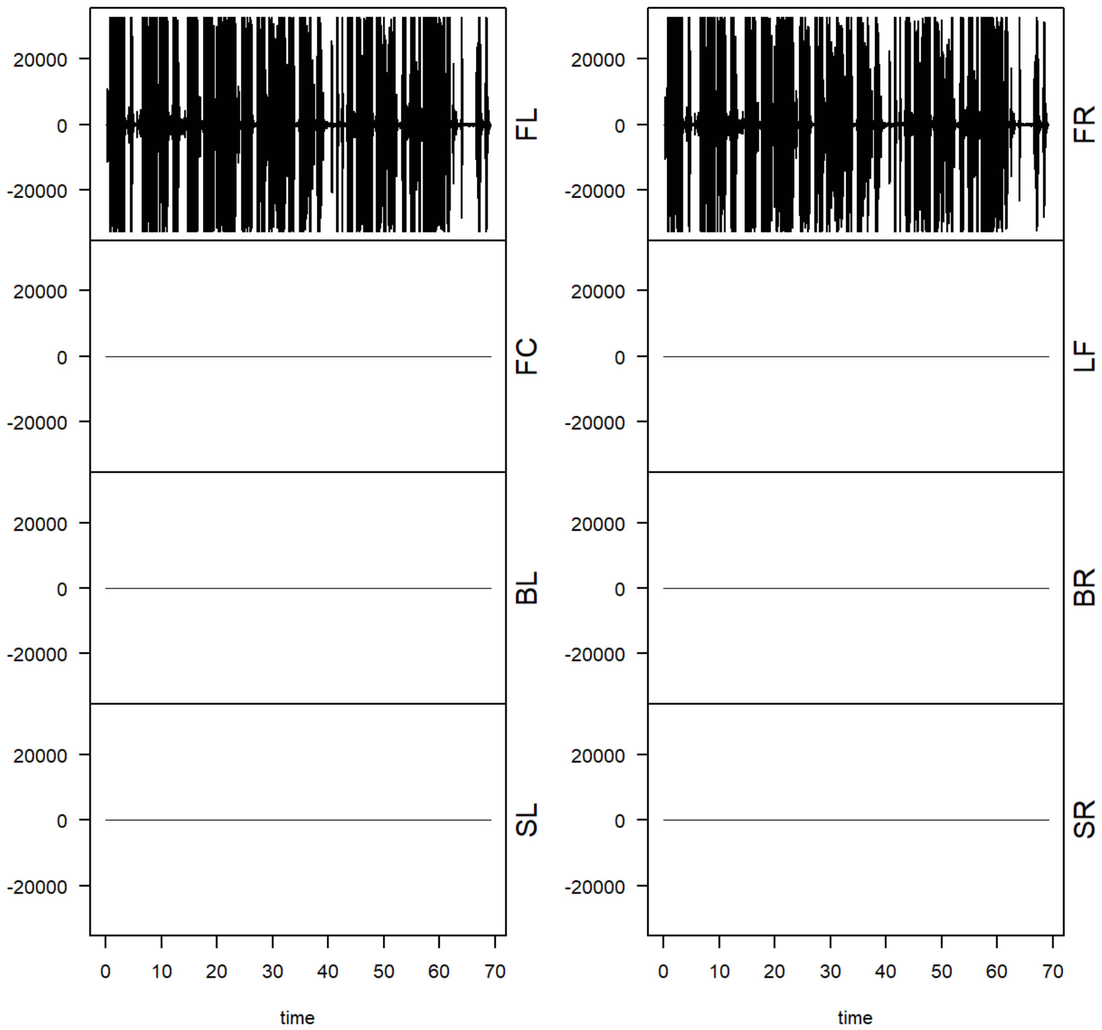
Twenty five acoustic features, also known as low-level descriptors, were extracted. These features are: intensity, zero crossing rates, MFCC 1-12 (Mel-frequency cepstral coefficients), F0 (Fundamental frequency) and F0 envelope, probability of voicing and, LSP frequency 0-7. On every feature nineteen statistical functions were applied. The functions are: maximum, minimum, range, absolute position of maximum, absolute position of minimum, arithmetic of mean, Linear Regression1, Linear Regression2, Linear RegressionA, Linear RegressionQ, standard Deviation, kurtosis, skewness, quartiles 1, 2, 3 and, inter-quartile ranges 1-2, 2-3, 1-3. The delta coefficient for every LLD is also computed as an estimate of the first derivative hence leading to a total of

950 features.

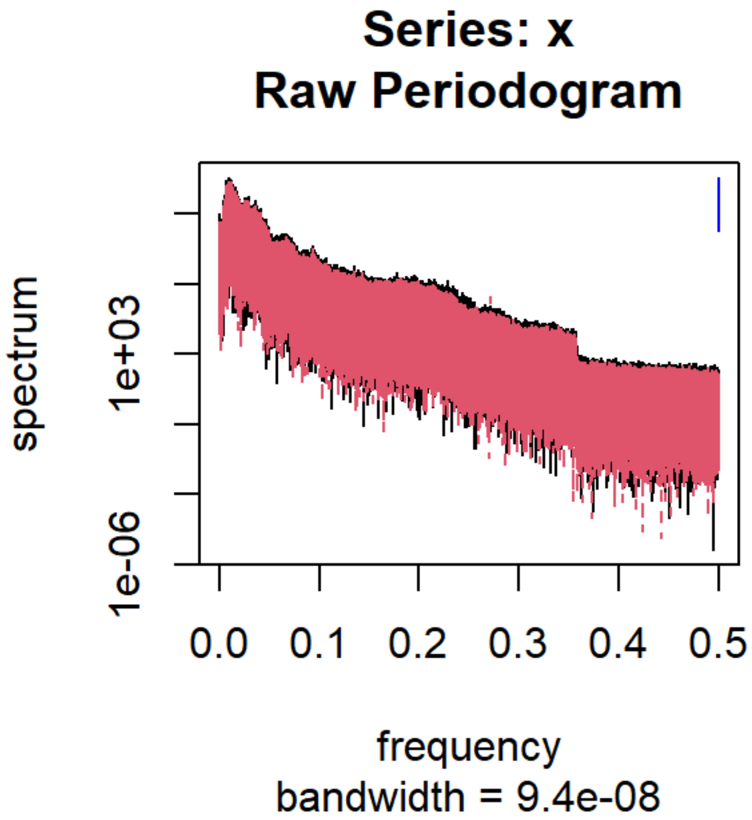
In this study, we explore the generation, manipulation, and analysis of waveforms using the R programming language and relevant libraries, such as tuneR and seewave. We investigate the characteristics of different waveform types, examine their spectral properties, and explore their applications in audio data analysis. Through practical examples and demonstrations, we aim to provide insights into the role of waveforms and Fourier transform techniques in understanding and processing audio signals.

3. Result and discussion

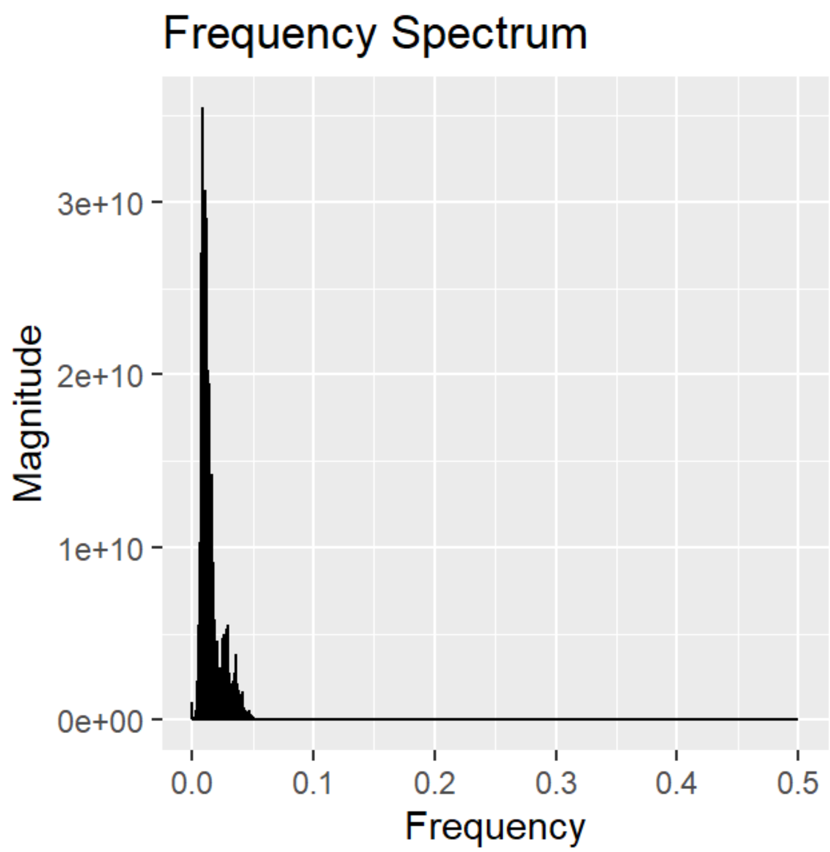
3.1 Audio1



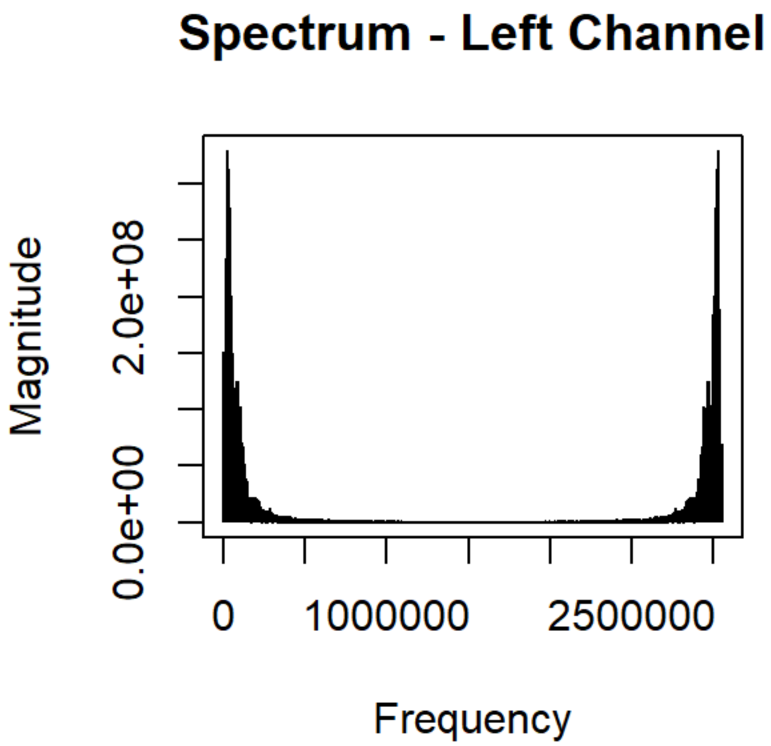
We can see that the amplitude varies between 22000 and -22000 and louder. Waveform diagram of an audio signal shaped like a square wave. Periodic variation is stable and amplitude variation is stable.



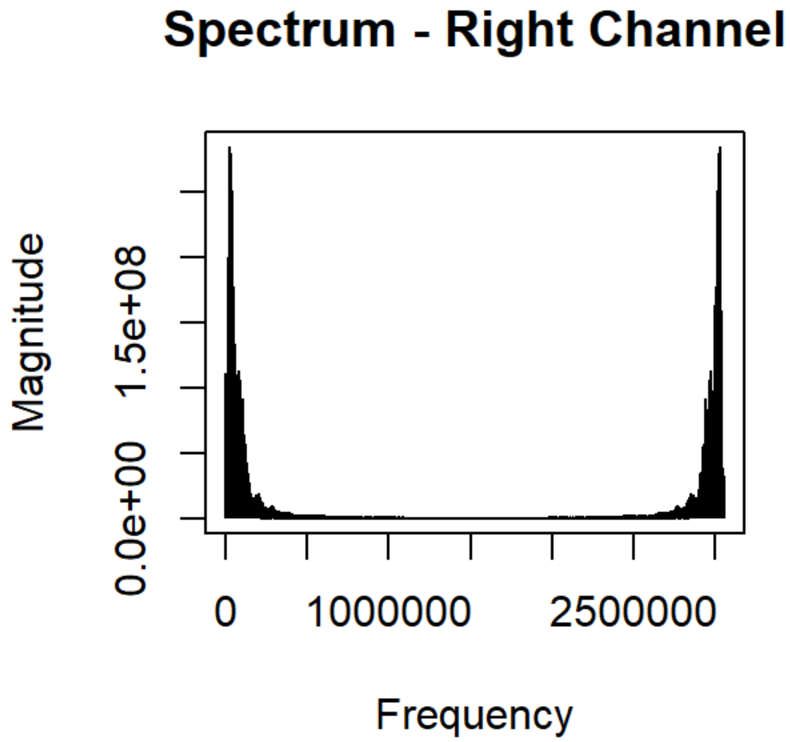
This is the spectrogram of the audio.



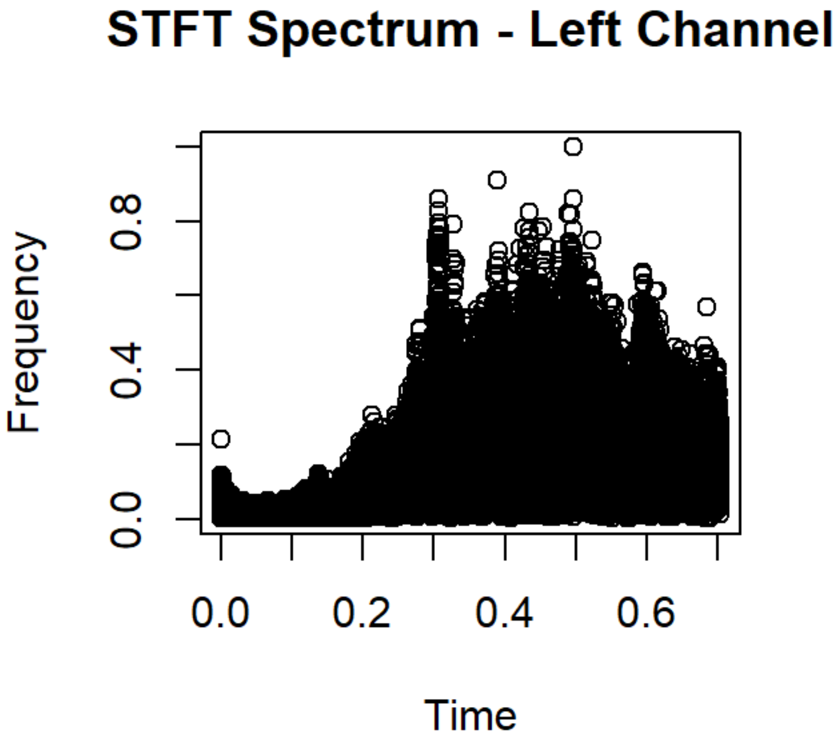
We can see that the magnitude with the highest frequency is lower, and those with higher magnitude have fewer frequencies, indicating that the audio energy is lower, and the part with low volume accounts for the majority.



Spectrogram of the Fourier transform results of the left channels. "magnitude low" means that the magnitude (or absolute value) of each complex number is small. We can find that the magnitude with the most frequency and the least frequency are relatively large, and the rest are very small.

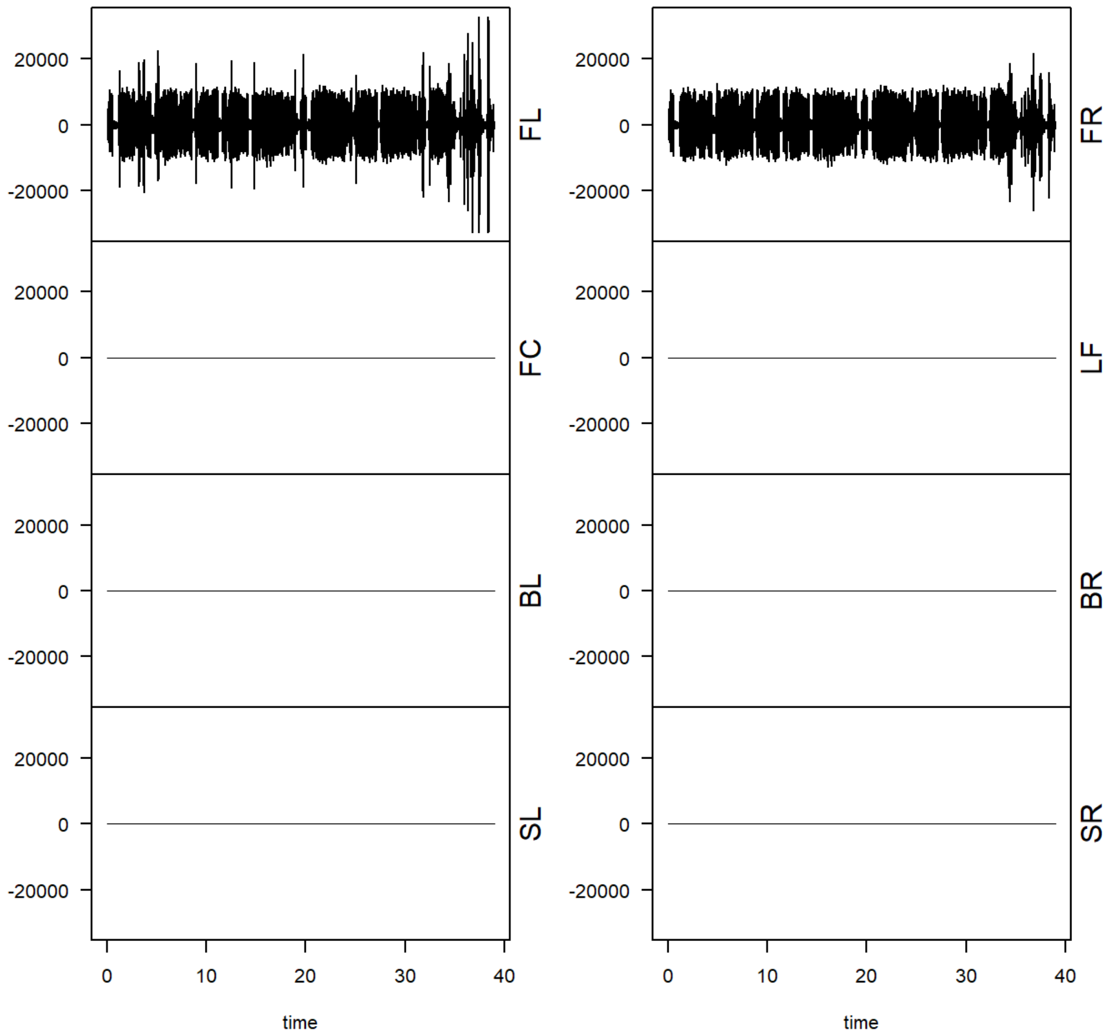


Spectrogram of the Fourier transform results of the left channels. We can find that the magnitudes of the right channel with the most frequency and the least frequency are relatively large, and the rest are very small. But at the same frequency, the magnitude is smaller than the left channel.

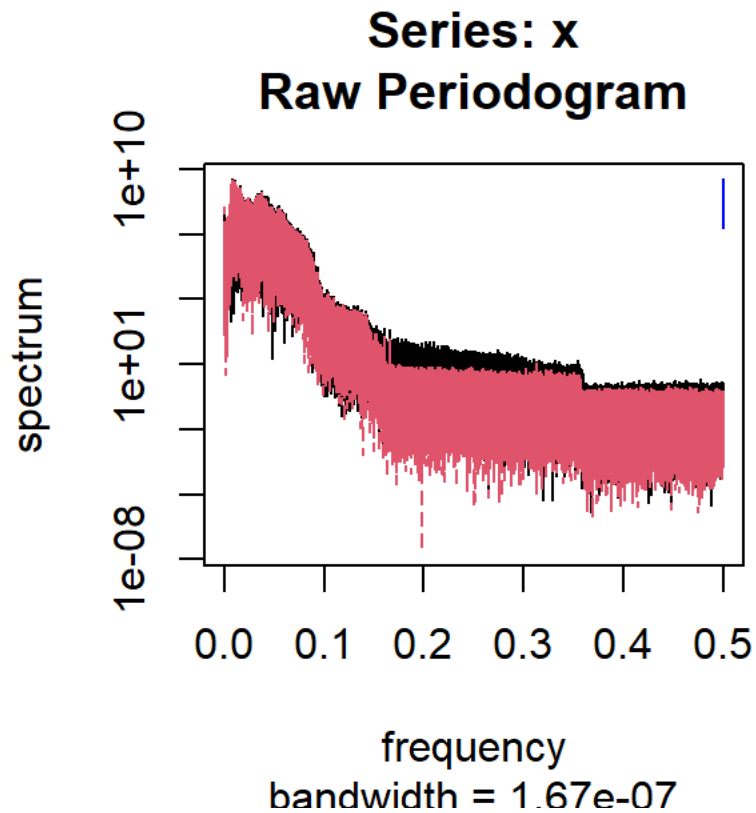


Spectrogram of left channel STFT results. We can see that the frequency increases with time. There is a peak between 0.3 and 0.6.

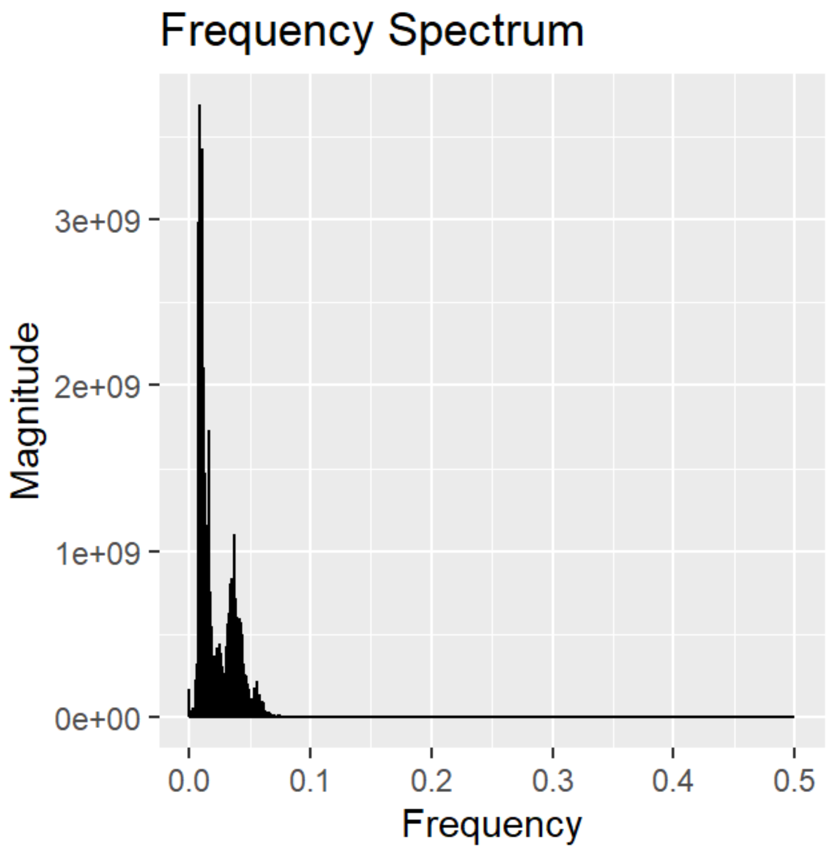
3.2 Audio2



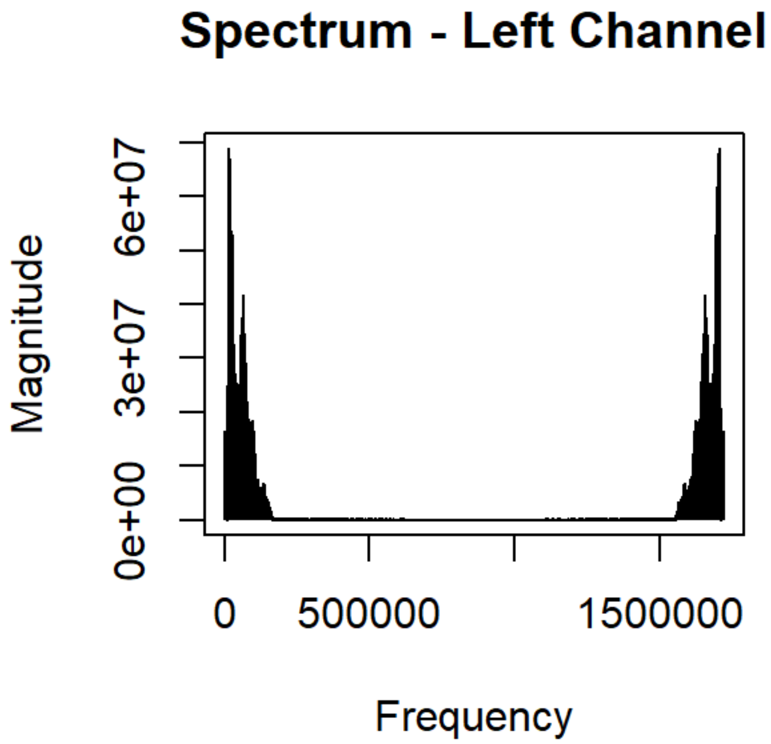
Waveform diagram of an audio signal shaped like a square wave. The change in amplitude is small in the first period of time, and the amplitude becomes larger as time goes on. The volume is small in the early stage, and it becomes louder in the later stage.



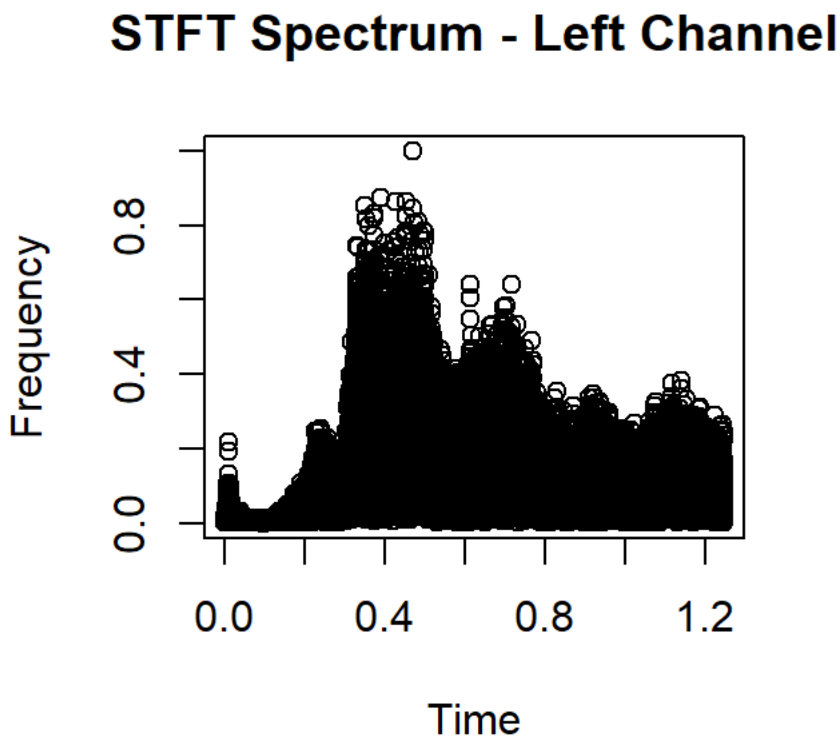
This is the spectrogram of the audio1.



Compared with audio, the interval between two peaks in audio1 is wider.



Compared with audio, the frequency interval between two peaks in audio1 becomes smaller and the maximum amplitude becomes smaller.



It can be seen from the figure that the frequency increases from time 0.2, and four peaks appear, the two largest peaks are at 0.4 and 0.7, and the frequency after 0.7 decreases to a certain extent.

4. Conclusion

In conclusion, the study of waveforms and their analysis using Fourier transform techniques in the context of audio data has provided valuable insights into the characteristics and properties of audio signals. By exploring different waveform types, such as sine waves, square waves, and pulse waves, we have gained a deeper understanding of their unique features and applications.

5. References

Cooley, J. W., Lewis, P. A. W., & Welch, P. D. (1969). The Fast Fourier Transform and Its Applications. In *IEEE TRANSACTIONS ON EDUCATION* (Vol. 12, Issue 1).

Dutt, A. (1993). *A, Fast • FuirTansforms for Nonequispaced Data*.

Halliday, J. R., Dorrell, D. G., & Wood, A. R. (2011). An application of the Fast Fourier Transform to the short-term prediction of sea wave behaviour. *Renewable Energy*, *36*(6), 1685–1692. https://doi.org/10.1016/j.renene.2010.11.035