Interpretation of Music Signals Using Time-Frequency Analysis

Kaushani Egodawele
Department of Mathematics
Faculty of Science
University of Peradeniya
Kandy, Sri Lanka
kkdegodawele@gmail.com

Rajitha Ranasinghe
Department of Mathematics
Faculty of Science
University of Peradeniya
Kandy, Sri Lanka
rajithamath@sci.pdn.ac.lk

***Abstract*—Time-frequency analysis serves as a pivotal tool in unraveling the intricate dynamics of signals evolving over time. Through the meticulous dissection of signals into their frequency components at specific time intervals, this method provides a nuanced understanding of transient events. This capability empowers researchers and practitioners to discern temporal patterns and unveil hidden structures within complex data, enhancing their ability to extract valuable insights from dynamic signal variations. In the present study, the methods of Time- frequency analysis is used for a qualitative analysis of music signals. The Short Time Fourier Transformation (STFT) and Continuous Wavelet Transformation (CWT) were investigated to discover music signal interpretations. The results of frequency approaches with time were implemented using Python Programming Language as Spectrograms and Scalograms.**

Keywords—Time-frequency analysis, STFT, spectrogram, CWT, scalogram

I. Introduction

Signal processing has been completely transformed by the use of time-frequency approaches, particularly in the analysis of musical signals. They provide a special method for analyzing and interpreting signals' complex and dynamic properties in both the time and frequency domains. The Short-Time Fourier Transform (STFT) was first introduced by Gabor in the early 20th century, which is when time frequency analysis began to take shape. Lately, wavelet transforms, spectrograms, and more sophisticated methods like the Continuous Wavelet Transform (CWT) and the Discrete Wavelet Transform (DWT) contributed to the field's substantial developments. Time-frequency techniques are used in a variety of industries, including wireless technology, seismic data analysis, voice processing, biological signal processing, and image processing. In terms of thorough representation, transcription and score following, feature extraction for pitch, rhythm, and harmony, and music classification, time-frequency approaches have benefits in the study of music signals. In addition to giving detailed information on musical content, they aid in detecting musical variations, distinguishing between instruments, transcribing music, extracting features for information retrieval and automatic music recommendation systems, and etc. The present study uses time-frequency methods to uncover such characteristics of music transmissions.

II. Time Frequency Analysis

Time-Frequency Analysis is the study of the frequency variations of signals with respect to time. Various types of time-frequency analysis techniques can be used in real-world situations in order to discover the information that lies within non-stationary signals. For the analytical modification of various sorts of signal processing applications, the Fourier Transformation was developed further. As an approach for time-frequency analysis, the Short-Time Fourier Transformation (STFT) was initially established. The Continuous Wavelet Transform (CWT) was a technique that was later developed and is capable of being used in a variety of signal processing applications. [4] Both techniques are beneficial for musical signal processing since finding information results in inventive musical solutions.

A. Short-Time Fourier Transformation

The Short-Time Fourier Transform (STFT), introduced by D. Gabor in 1971, is a signal processing method that examines a signal's time-frequency properties. It is also known as Gabor Transformation which uses a window function to split a larger signal into shorter segments, calculates the Fourier Transform for each windowed segment, and outputs a spectrogram, which is a representation of the signal's time and frequency. The STFT is frequently used in vibration analysis, voice analysis, and audio signal processing.

B. Continuous Wavelet Transformation

A time-frequency analysis method known as the Continuous Wavelet Transform (CWT) analyzes the localized frequency content of a signal over time. It is also known as Multi-resolution Analysis and involves visualizing a signal as a scalogram and is based on wavelets, localized functions in both time and frequency. CWT is the convolution of the signal with a scaled and translated version of the wavelets and finds applications in scientific research, image analysis, and signal processing.

III. Gabor Transformation (STFT)

A. Fourier Transformation (FT)

Fourier Transformation is a mathematical technique introduced by J.B. Joseph Fourier in the 19^{th} century. It decomposes complex signals into sine and cosine waves, revealing the signal's frequency components. The transformation is represented mathematically converting signal functions from time domain to frequency domain which are absolutely integrable and piecewise smooth.

i.e. taking the signal function over time domain to be $f(t)$, the function is well defined and integral of the function is not divergent and clearly exists which implies the integral is convergent over time domain. [2] One continuous function is transformed at a time by the operation, which was the primary driver behind continued development of the transformation.

B. Discrete Fourier Transformation (DFT)

Discrete Fourier Transformation is a modification of FT which can be used to transform set of stationary functions at a time to its frequency components. The method was introduced by Carl Friedrich Gauss (1777-1855) which involved with computations with matrices to obtain the Fourier Transform of a signal. In DFT, if the number of data functions to be considered is *n*, then n^2 number of operations is conducted in order to output *n* number of frequency components as a sum of FTs.

Initially, the Discrete Fourier Transform (DFT) held prominence in signal processing studies. However, as the demand arose for the analysis of extensive datasets, exemplified by instances involving a considerable number of data points (such as 10^{10}), there emerged a necessity for a more sophisticated approach.

C. Fast Fourier Transformation (FFT)

FFT, introduced by J. Cooley and J. Tuckey around 1965 is a convenient computational algorithm of DFT which involves *(n log n) number* of operations for *n* number of data points [1]. The transformation classifies the number of data points continuously into odd and even functions and reduces calculations further and output computational finite sum of Fourier Transformations.

D. Windowing and STFT

The Short-Time Fourier Transformation is used for non-stationary signals with the aid of a suitable windowing function. [1] Applying a windowing function that is compactly supported and overlapped to the signal considers small chunks of the original signal for the transformation. Different types of windowing functions which changes their applications used according to the characteristics are:

- Hann Window (J.V. Hann, 1928)
- Hamming Window (R.W. Hamming, 20th cent.)
- Gaussian Window (C.F. Gauss, 19th cent.)
- Shannon Window (C. Shannon, 19th cent.)

The convolution theorem to the original signal and window function results a sequence of functions of FFTs. The convolution theorem in time domain results a product of two frequency components; signal component and shifted window function.

i.e. a finite sequence of $\mathcal{F}(f(t_k)).\mathcal{F}(w(t_k - \tau_k))$

E. Spectrogram Analysis

The sequential implementation obtained in STFT results the time- frequency variations of the signal in one frame which is known as, **Spectrogram.** [6] It is a representation of the intensity plot of the STFT magnitude showing at which time durations the respective frequencies occurred at in the signal. The variations of frequencies (Y- axis) with respect to the time (X- axis) can be plotted using a computer programming software. In this study, Python is used to generate spectrograms with corresponding arguments.

IV. Interpreting a Music Signal Using STFT

A. Music Signal

As the first attempt of interpretation, analysis of a portion of a 14-second sound signal is used to identify major characteristics of a spectrogram, which is a raw piano input. The sound signal was extracted from the chorus of the song, **"Mary Had a Little Lamb".**

B. Solution by Python

The Python solution contained libraries for the plot of spectrogram and arguments of sending the audio file as an input, arguments of windowing and FFT process and arguments of plotting were feed into the code. The different values for nperseg, noverlap, nFFT arguments output different spectrograms with visible changes. In Python, the default window in use is Hann Window which automatically aid in implementing the spectrogram. Use of any window that is not Hann has to be defined in the programme.

The nperseg value corresponds to the number of segments per windowing at one time or window width and noverlap is giving the number of segments that are going to be overlapped with windowing before and after the current consideration. nFFT value corresponds to the number of data points that the audio signal is chunked into for applying FFT.

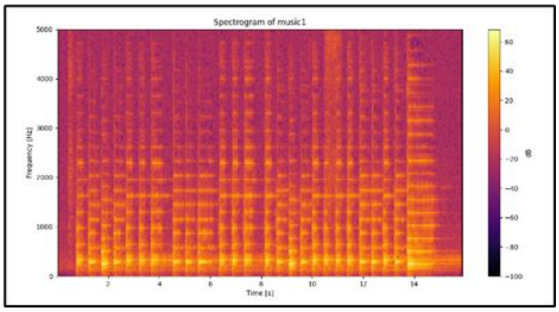
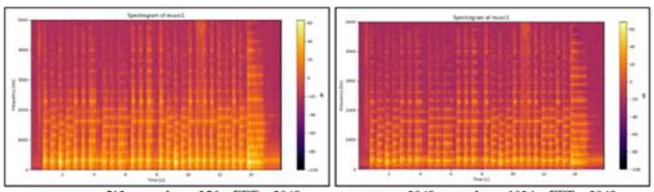


Fig. 1. Resulting Spectrogram with nperseg-2048, noverlap-1024, nfft-2048

C. Spectrogram Analysis–Varying Window width

The spectrogram observed for a large value for nperseg implements with better frequency resolution but lower time resolution which is known as **"Under-sampling"** of a signal. In contrast, **"Over- Sampling"** occurs when nperseg is lower results with higher time resolution but poor frequency resolution. (Fig. 2)



nperseg = 512, noverlap = 256, nFFT = 2048 **nperseg= 2048, noverlap = 1024, nFFT = 2048**

Fig. 2. Over- Sampling and Under- Sampling

D. Spectrogram Analysis – Varying Number of Data Points

The spectrogram obtained for higher values of nFFT implements spectrograms with finer frequency details of the signal but with poor time resolution. The lower nFFT values outputs a spectrogram with rapid time changes but with poor frequency resolution. (Fig. 3)

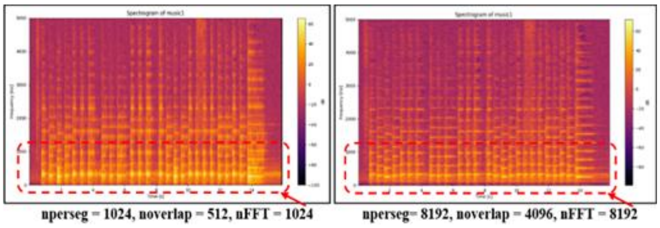
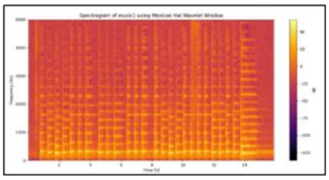


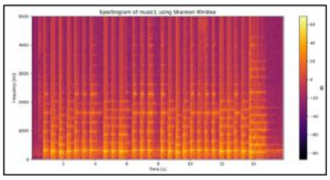
Fig. 3. Rapid Time changes and Finer Frequency Details with the change of nFFT

E. *Spectrogram Analysis – Varying the type of window*

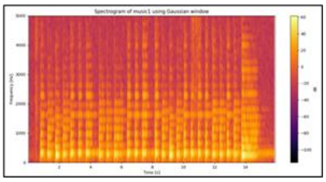
The Mexican Hat Wavelet window, Shannon Window and Gaussian Window were defined and applied to the sound signal to investigate the features of resulting spectrograms. (Fig. 4)



(a)



(b)



(c)

Fig. 4. Spectrograms implemented using (a) Mexican hat wavelet window, (b) Shannon window and, (c) Gaussian window

V. Continuous Wavelet Transformation

According to Weiner Heisenberg's (1901–1976) uncertainty principle [8], [12] encoding time and frequency data of a signal simultaneously is essentially challenging.

The Heisenberg Uncertainty Principle: A function f and its Plancherel transform $\mathcal{P}(f)$ cannot both have arbitrarily small support.

As a consequence, CWT is also has been widely used in signal processing as a time-frequency analysis technique [8]. The CWT is particularly useful for analyzing signals that have non-stationary or time-varying characteristics. The transformation is constructed using a particular function called the **Wavelet Function**, which yields a CWT matrix (Scalogram) that visualizes the change in the frequency content of the signal over time at various scales.

A. *Wavelet Function and CWT*

The CWT is proceeded for non-stationary signals but offering insights into both high and low frequency components with their temporal evolution. [5] The transformation is carried out using a wavelet function denoted as $\psi(t)$ (mother wavelet) which can be scaled and translated prior using it to the signal unlike the window functions in STFT which had a fixed width for an instance. This function is typically a short-lived oscillation, and it is used to analyze the signal at different scales and positions in time. A wavelet function is only can be used if it satisfies two main constraints: zero mean as the admissibility condition and finite energy as the necessary condition [12].

i.e. $\int_{-\infty}^{\infty} \psi(t)dt = 0$ and $0 < \int_{-\infty}^{\infty} |\psi(t)|^2\, dt < \infty$

There are various types of wavelet functions where the type of the wavelet affects the characteristics of the analysis. [11] Types of wavelet functions can be mentioned as follows:

- Haar Wavelet (A. Haar, 1909)
- Morlet Wavelet (J. Morlet, 1983)
- Gabor Wavelet (D. Gabor, 1946)
- Daubechies Wavelet (I. Daubechies, 1960)

CWT is calculated by applying the convolution theorem to the audio signal and the complex conjugate of the scaled and translated wavelet function:

$$\mathcal{W}_{\psi}(f)(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t).\psi^{*}\left(\frac{t-b}{a}\right) dt$$

where a is the scaling parameter and b is the translation parameter. [9], [10].

B. *Scalogram Analysis*

The CWT formula's continuous integral is calculated numerically. The wavelet and the signal convolve at multiple scales and positions by varying the scale and position parameters. The magnitudes of the wavelet coefficients of the signal at various scales and time localizations is represented by the entries in the CWT matrix and visually represented in the **Scalogram** [1]. The variations of scales (Y-axis) with respect to the time (X-axis) can be plotted in here. In this study, Python is used to generate scalograms with corresponding arguments.

VI. Interpreting a Music Signal Using CWT

A. *Music Signal and the Solution by Python*

The music signal used is the chorus of **"Mary Had a Little Lamb"** which was used in spectrogram analysis in the study of STFT. (Fig. 5) In the case, the python solution contains specific libraries for CWT algorithms and arguments to use wavelets and output scalograms for audio signals. "Width scale" is the main argument that can be varied to analyze the audio signal with different scalograms.

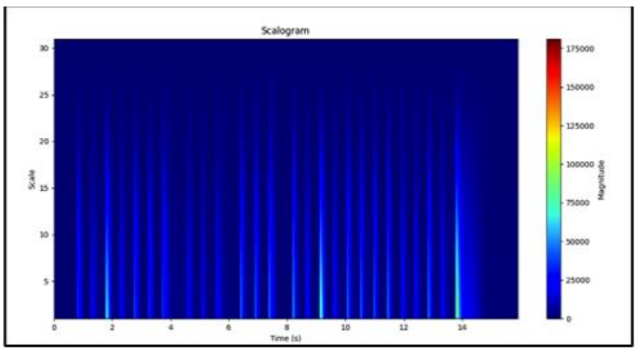


Fig. 5. Scalogram using morlet wavelet width= (1,31)

Width scale is an array of scales (or widths) that determines the range of scales for the wavelet and wavelet frequency represents the frequency of the oscillatory part of the corresponding wavelet which can be changed according to the frequency range of the signal. The wavelet that is being used has to be defined to the programme.

B. Scalogram Analysis–Varying Width Scale

The scalogram obtained for large scales is more sensitive to capture low frequencies of the signal averaging out high frequency components. The characteristic is useful to analyze slow changes and identify global patterns of the signal. The morlet wavelet is used for following observations. This property is advantageous in music signal processing to analyze musical expressions with related to the intensity of the sound and temporal changes (Fig. 6).

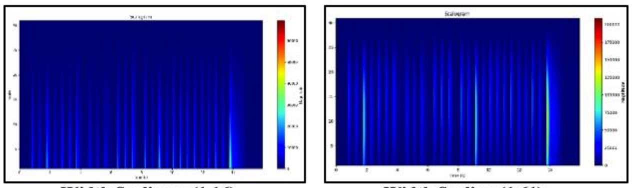


Fig. 6. Scalograms for different width scales of the same signal

VII. CONCLUSION

Implementing spectrograms with considerably higher values for nFFT with lower values for window width than nFFT value is allowed. Interpretation of higher nFFT values for sound signals with large number of instruments taking more time to implement but with fine frequency details. Bell-shaped curves, exemplified by Gaussian or Mexican Hat wavelets, prove highly effective in acute audio signal processing. In contrast, rectangular windows, such as the Shannon window, are not recommended as they fail to provide optimal data representation for this context. The use of bell-shaped wavelets ensures a more nuanced and accurate analysis of audio signals, capturing subtle variations and nuances that might be overlooked by less suitable windowing functions.

The interpretation of scalograms takes a significant leap forward with an advanced STFT model, where the flexibility of wavelets in scaling proves particularly advantageous for audio signals. In this refined approach, scaling is plotted in lieu of frequency over time, offering a more adaptive representation. The utility of wavelet transformation in audio signals lies in its capacity to capture not just frequencies but also the intensity or magnitudes at specific time instances. This characteristic becomes instrumental in analyzing the dynamic intricacies of musical expressions, such as crescendos and decrescendos, providing a nuanced understanding of sound evolution [3].

The integration of spectrogram and scalogram implementations opens up innovative possibilities, such as the creation of an interactive music instrument program capable of detecting and measuring user-initiated tempos. A promising avenue for future research involves the exploration of musical note recognition systems through spectrogram analysis, presenting opportunities to deepen our understanding of musical structures and enhance interactive music technology.

REFERENCES

[1] S. Brunton and N. Kutz, "Data Driven Science and Engineering, Machine Learning, Dynamical Systems and Control", 2022, 2nd edition, pp.53-76.

[2] E. Kreyszig, "Advanced Engineering Mathematics", 10th edition, 2011, pp.474 - 534.

[3] M. Müller, "Fundamentals of Music Processing-Audio, Analysis, Algorithms, Applications" ,2015, pp .40 -102.

[4] K. Gröchenig, "Foundations of Time- Frequency Analysis", Brikhäuser, Boston, 2001, pp.245–300.

[5] X. Cheng, J. V. Hart and J. S. Walker, "Time-frequency analysis of musical rhythm", Notices of the AMS, vol.56, no.3, 2009, pp.256–372.

[6] K. Hastuti, A. Syarif, A. Fanani, A.Mulyana, "Natural automatic musical note player using time frequency analysis on human play", Telkomnika, vol.17, No.1, 2019, pp. 235-245.

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[8] C. Torrence and G. P. Compo, "A practical guide to wavelet analysis", Bulletin of the American Meteorological Society, vol.79, No. 1, 1998, pp. 61–78.

[9] I. Daubechies, "The wavelet transforms, time–frequency localization and signal analysis", IEEE Transactions on Information Theory, vol.36, No.5, 1990, pp.961–1005.

[10] S. G. Mallat, "A Wavelet Tour of Signal Processing", Academic Press, 1999.

[11] E. Lai, "Practical Digital Signal Processing for Engineers and Technicians", Elsevier, 2003, pp. 61-78.

[12] B. Foster, P. Massopust, "Applied and Numerical Harmonic Analysis", Brikhäuser, Boston, 2010, pp.30-45.