FPGA accelerated spectrogram generator for applications based on CNN techniques

Diogo Miguel Cunha Fernandes

UMinho | 2023

Universidade do Minho
Escola de Engenharia

Diogo Miguel Cunha Fernandes

FPGA accelerated spectrogram generator for applications based on CNN techniques

julho de 2023

Universidade do Minho
Escola de Engenharia

Diogo Miguel Cunha Fernandes

FPGA accelerated spectrogram generator for applications based on CNN techniques

Dissertação de Mestrado
Mestrado em Engenharia Eletrónica Industrial e Computadores
Sistemas Embebidos e Computadores

Trabalho realizado sob a orientação do
Professor Doutor Rui Pedro Oliveira Machado

junho de 2022

Contents

List of Figures

List of Tables

Acrónimos

DFT Discrete Fourier Transform.

DSP Digital Signal Processing.

FFT Fast Fourier Transform.

FT Fourier Transform.

ML Machine Learning.

RPM Revolutions per Minute.

STFT Short-Time Fourier Transform.

TFR Time-Frequency Representation.

1 Introduction

Texto

1.1 Contextualization and Problem Statement

Texto

1.2 Technical and Scientific Relevance

Texto

1.3 Motivation

Texto

1.4 Objetives

Texto

1.5 Thesis Organization

Texto

2 State-of-the-Art

It is considered that the reader has some know-how in DSP.

2.1 Time-Frequency Domain Analysis

The frequency domain analysis is an excellent method for analyzing periodic signals [1]. Its representation provides the spectral components of the signals and the phase characteristic often includes information about the time distribution information. Still, in the frequency domain, it is difficult to infer the time distribution of a signal, which isn't a problem for periodic signals since the spectral components are stationary components. However, not all signals have periodic components, and some signals may have a complex or random pattern over time. This way, the presence or absence of periodic components affects how a signal is analyzed and processed, so it is important to consider this aspect when performing signal processing, as it may be relevant to study the non-stationary components of the signal. For example, to analyze the condition of rotating machines is used a microphone to capture data during a speed sweep. The motor can either be started from a low Revolutions per Minute (RPM) and increased to a high RPM, or it can be started from a high RPM and decreased to a low RPM. In this case, it will produce non-stationary signals, as the frequency content of the signal is changing over time [2].

The time-frequency domain analysis has been one of the most effective approaches to solving signal-processing problems, considering it allows the identification and qualification of oscillatory components present in non-stationary waveforms, which are prevalent in real-life signals. It is useful to study a signal's structure in time and frequency simultaneously, through specifically constructed projections of the signal in the time-frequency plane. The combination of these projections is called Time-Frequency Representation (TFR)s. TFRs are used extensively in many areas of science and are known for their powerful ability to analyze signals in the time-frequency domain. They are consistently applied to image processing, finance, geophysics, and biological science, among other fields. The widespread use of TFRs to analyze non-stationary signals has made them an essential tool for studying the time-varying properties of complex systems. A TFR can be used by other algorithms, like Machine Learning (ML) algorithms, that detect patterns in the time-frequency domain representation of the signal, in order to identify it [3, 4, 5]. Avery Wang [6], responsible for developing the Shazam algorithm, that identifies songs using a mobile phone microphone, uses TFRs to extract a fingerprint of a sampled audio file. The fingerprint consists of a "constellation map" that is obtained by drawing out the peak points present in the TFR of the audio file and associating them combinatorial, through hash tokens. To identify the music, the extracted fingerprint is compared to all fingerprints in a database.

Many works have been dedicated to finding appropriate TFR which can extract meaningful information from sound signals. However, the most used ones are....

WRITE AN INTRO TO WHAT WILL BE TALKED NEXT

Before the advent of modern digital signal processing, was used a filter bank, resulting from a series of band-pass filters, to divide the input signal into frequency bands [7]. Then a transducer is controlled by the magnitude of the filter bank outputs to record the TFR as an image on a paper. This method is an analog processing solution, making it an inefficient method to be applied to a digital system.

2.1.1 Short-Time Fourier Transform

The Short-Time Fourier Transform (STFT), first introduced by Gabor [8], is an extension of the Fourier Transform (FT) [9], being a widely used method for studying non-stationary signals. The STFT is a method that analyses a signal by breaking it up into shorter frames and computing the FT of each segment, which can be the Discrete Fourier Transform (DFT) for a discrete-time signal. This allows studying the frequency components of the signal over time (time-frequency domain analysis), rather than just its overall frequency content, that may be obtained using a frequency domain analysis with the DFT. In practice, the Fast Fourier Transform (FFT) is often used to perform the DFT of the STFT [7, 10], as the FFT is a more efficient algorithm to compute the DFT, being defined in equation 2.1, applied to a signal $x(n)$.

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j\dfrac{2\pi}{N}nk} \tag{2.1}$$

The equation 2.1 gives only the computation of one spectral section, which size corresponds to the size of the sampled signal, N. In order to study the spectral components over the time domain, one has to divide the signal into small frames, applying a window function, such as a Gaussian, Hann, or Hamming window, to each segment of the signal to emphasize the time-localized frequency components of the signal. Ideally, the window sould be very small, so that the signal, in that small chunk, would be periodic. This would avoid the spectral leakage, caused by the discontinuities between consecutive windows. The window functions topologies are discussed in a topic below [REF]. Now, one can determine the running spectrum by adapting the equation 2.1, to the equation 2.2.

$$STFT\{x(n)\} = STFT_{x(n)}(m,k) = X(m,k) = \sum_{n=0}^{N-1} x(n+mH)w(n)e^{-j\dfrac{2\pi}{N}nk} \tag{2.2}$$

The equation 2.2 computes the Fourier coefficients for the k^{th} frequency at the m^{th} frame of the signal. As shown by the figure [REF], m represents the frame number to be analyzed, and k is the frequency component. Furthermore, H represents the hop size, i.e., the overlapping of the consecutive window frames, so the mH can be seen as the beginning of a window frame. This way, it is applied the DFT to each frame of the signal, with size N, which is obtained by the windowing function. Thus, the range of each frame is $[mH; mH + (N-1)]$.

IMAGEM COM A REPRESENTAÇÃO DA ONDA E ETC

The output of the computation of the STFT is a two-dimensional spectral matrix with the x and y axis representing the number of frames (N_{frames}) and the number of frequency bins ($N_{freqbins}$), respectively. Each element of the matrix contains complex Fourier coefficients. The number of frequency bins can be calculated using the equation 2.3 and the number of frames is derived from the equation 2.4. The $N_{freqbins}$ may be explained by the symmetrical property of the DFT that introduces redundancy in the computed coefficients above the central frequency, which corresponds to the Nyquist frequency, $\dfrac{f_s}{2}$ [9], so one may regard only the first part of the coefficients and discard the other part. As the DFT is applied over a frame, the first part of the coefficients matches to $\dfrac{N}{2}$. Regarding to the frames, one can predict that the greater the value of N_{frames}, the more computation time will be needed.

$$N_{freqbins} = \frac{N}{2} + 1 \tag{2.3}$$

$$N_{frames} = \frac{SamplesNum - N}{H} + 1 \tag{2.4}$$

As an example, consider a signal with $30\ kSamples$, which STFT is computed with a frame size of 2000 and hop size of 500. Substituting in the equation 2.3, the $N_{freqbins}$ will be 1001, so the frequency range, which is between $0\ Hz$ and $\frac{f_s}{2}\ Hz$, is divided into 1001 equal frequency bins. Resolving the equation 2.4, the number of frames, N_{frames}, will be equal to 57. This means that m ranges from 0 to 57. Hence, the output shape of the STFT is a two-dimensional array with size $(1001, 57)$.

Window

If the signal is not periodic, discontinuities towards the edges of the signal are created when applying the DFT, provoking spectral leakage. This happens because the DFT interprets the portion of the analyzed signal as a continuous repetition of that signal, to $+\infty$. When the signal is an aperiodic signal, which happens in most real-world situations, it will create a discontinuity in the consequent repetitions of the signal, as seen in the figure [REF]. These discontinuities add frequency components that were not present in the original signal, leading to spectral leakage, which isn't desirable in a time-frequency domain analysis system.

FIGURE

In order to avoid discontinuities, one has to attenuate both signal edges, applying a window function. The window function to choose from depends on the type of application, but the most common window functions include the Hann window, the Hamming window, and the Kaiser window. The Hamming and Hann windows are most suited for audio and vibration analysis systems [11, 10, 12], due to their noise performance and lower side lobes. Although they are two similar window functions, the Hamming window performs better at canceling the nearest side lobe, whereas the Hann window performs better at attenuating the more distant side lobes. The Kaiser window is a general-purpose window, as it allows more customization of the window response for the type of application [12]. The rectangular window, shown in the figure [REF], is not appropriate for computing the STFT because it creates discontinuities on the edges of the window, as seen in the figure [REF]. The equation 2.5 represents mathematically the Hamming window response in the time domain. In figure 2.1 is shown the time and frequency domain response of the Hamming window with size 50. One can confirm that the second side lobe is more attenuated and the attenuation through the rest of the side lobes is approximately constant.

$$w(n) = \begin{cases} 0,54 - 0,46\cos(\frac{2\pi n}{N})\,, & 0 \leq n \leq N-1 \\ 0\,, & otherwise. \end{cases} \tag{2.5}$$

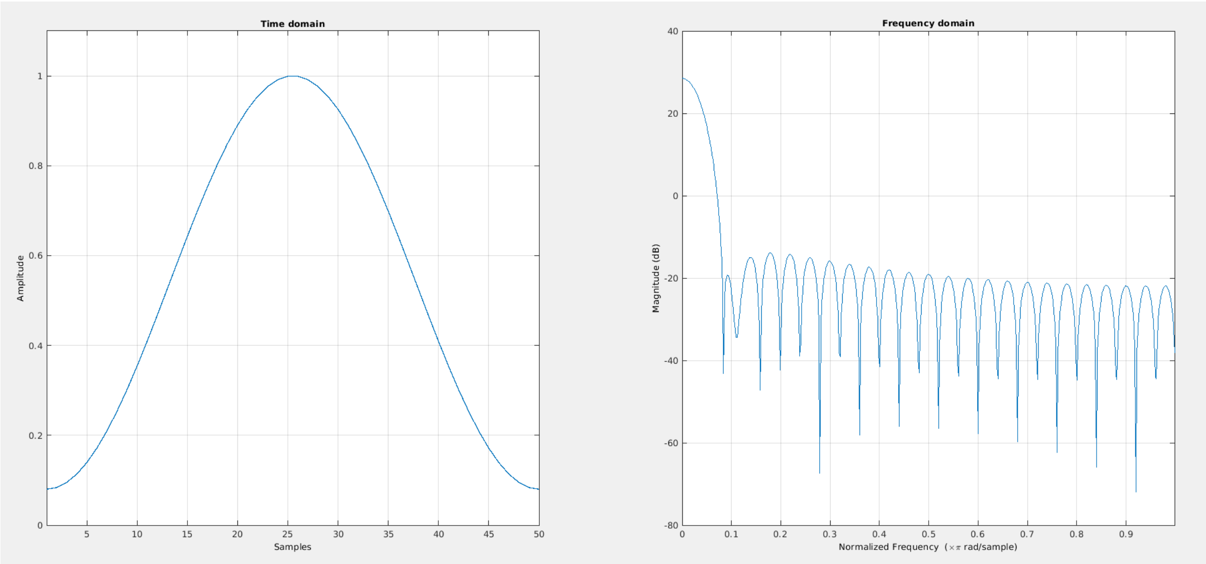

Figure 2.1: Hamming Window: Time and Frequency Domain Response with $N = 50$.

In figure 2.2 is represented a sampled signal to be analyzed, $x(n)$, and the result of the application of the Hamming window function, in the time domain, $x_w(n)$. As one can see, the signal gets modulated towards the ends, avoiding spectral leakage, as seen above.

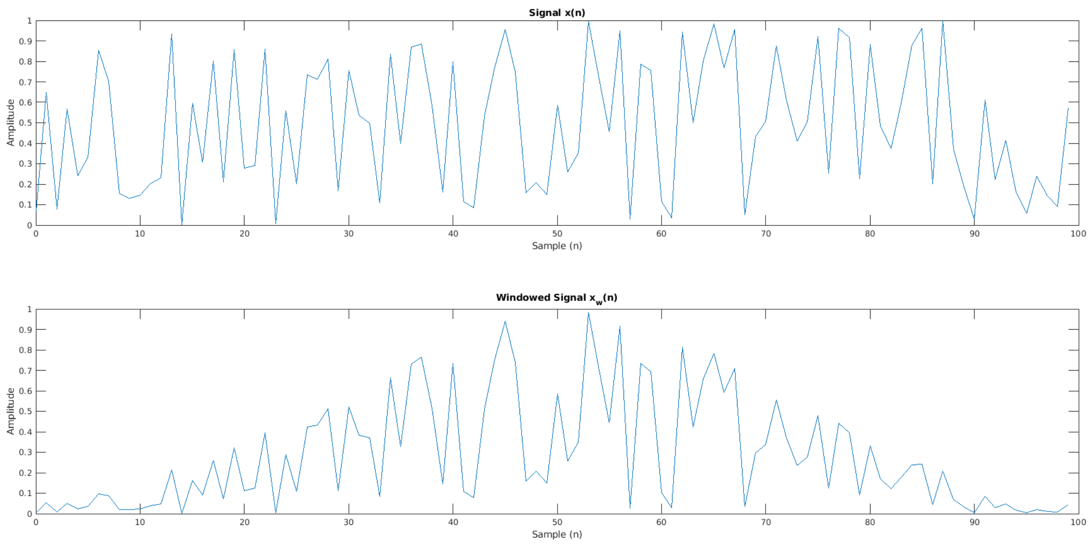

Figure 2.2: Signal, $x(n)$, and result of application of Hamming window, $x_w(n)$, with $N = 50$, in time domain.

Limitations

The computation of the STFT, as all Digital Signal Processing (DSP) techniques, has its own advantages and disadvantages. With this technique, there is a limit on the resolution plotted about the time and frequency domains. A time series can be accurately measured in terms of when events occur but provide no information about what frequency components are present. The Fourier transform, on the other hand, can provide precise frequency components but fails to supply information about when these frequencies occur. The STFT offers a balance between time and frequency resolution, at the cost of lower resolution in each domain. This principle is called the uncertainty principle [8, 13] and demonstrates that there is a fundamental trade-off between time and frequency domain resolution. The parameter that defines the time-frequency domain resolution is the width of the window (frame size N) [14]. A large frame size window gives poor time resolution but relatively good frequency resolution. A smaller frame size window gives poor frequency resolution but relatively good time resolution. This trade-off depends on the type of problem to be studied.

This is motivated by the fact that since a computation of the discrete Fourier transform is necessarily restricted to a computation on a finite lenght of data, there is implicit a time window imposed in x(t), that is, x(t) is multiplied be a rectangular window with a width equal to NT.

2.2 Digital Filtering

Texto

2.3 FFT

Texto

2.4 Floating Point Arithmetic

Texto

2.5 Design Flow FPGA

Texto

3 System Specification and Design

Texto

3.1 Analysis

Texto

3.1.1 Requirements

Texto

3.1.2 Block Diagram

Texto

3.2 HW Specification

Texto

3.2.1 Development Board

Texto

3.2.2 Microphone

Texto

3.2.3 Accelerometer

Texto

3.3 Design

Texto

3.3.1 Data Sampling

Texto

3.3.2 Spectrogram Filter

Texto

3.3.3 Spectrogram Setup

Texto

3.3.4 Processor System

Texto

3.4 High Level Spectrogram Generator

Texto

4 System Implementation

Text

5 Tests and Results

Text

6 Conclusion

Text

Bibliography

[1] H. Ahmed and A. Nandi, *Frequency Domain Analysis*. John Wiley & Sons, Ltd, 2019, ch. 4, pp. 63–77.

[2] A. Brandt, *Noise and Vibration Analysis (Signal Analysis and Experimental Procedures)*. Wiley, 2011, ch. 12.

[3] C. Lee, C. Yoon, H.-j. Kong, H. C. Kim, and Y. Kim, "Heart rate tracking using a doppler radar with the reassigned joint time-frequency transform," *IEEE Antennas and Wireless Propagation Letters*, vol. 10, pp. 1096–1099, 2011.

[4] Z. Chi, Y. Li, and C. Chen, "Deep convolutional neural network combined with concatenated spectrogram for environmental sound classification," in *2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT)*, 2019, pp. 251–254.

[5] H. Liu, L. Li, and J. Ma, "Rolling bearing fault diagnosis based on stft-deep learning and sound signals," *Shock and Vibration*, vol. 2016, p. 12, 01 2016.

[6] A. Wang, "An industrial strength audio search algorithm." 2003.

[7] A. V. Oppenheim, "Speech spectrograms using the fast fourier transform," *IEEE Spectr.*, vol. 7, no. 8, p. 57–62, aug 1970.

[8] D. Gabor, "Theory of communication." 2003, pp. 429–439.

[9] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, 2nd ed. Prentice-hall Englewood Cliffs, 1999, ch. 8.

[10] H. Ahmed and A. Nandi, *Time-Frequency Domain Analysis*. John Wiley & Sons, Ltd, 2019, ch. 5, pp. 79–114.

[11] J. O. Smith, *Spectral Audio Signal Processing*. Stanford University, 2011, online book.

[12] Tektronix, "Window functions in spectrum analyzers," acedido em 16 December 2022. [Online]. Available: https://www.tek.com/en/blog/window-functions-spectrum-analyzers

[13] N. Bey, "Multi-resolution fourier analysis: Time-frequency resolution in excess of gabor-heisenberg limit," *Signal, Image and Video Processing*, vol. 8, 05 2014.

[14] E. F. Shair, S. A. Ahmad, A. R. Abdullah, M. H. Marhaban, and S. B. M. Tamrin, "Determining best window size for an improved gabor transform in emg signal analysis." *Telkomnika*, vol. 16, no. 4, pp. 1650 – 1658, 2018.

Aliasing