Time-Frequency Analysis and Filtering based on the Short-Time Fourier Transform

Hon, Tsz Kin

Awarding institution:
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENCE AGREEMENT

Unless another licence is stated on the immediately following page this work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International licence. https://creativecommons.org/licenses/by-nc-nd/4.0/

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Title: Time-Frequency Analysis and Filtering based on the Short-Time Fourier Transform

Author: Tsz Kin Hon

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENSE AGREEMENT

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. http://creativecommons.org/licenses/by-nc-nd/3.0/

You are free to:

- Share: to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Time-Frequency Analysis and Filtering based on the Short-Time Fourier Transform

Tsz-Kin Hon
Division of Engineering
King's College London

A DISSERTATION PRESENTED TO THE DIVISION OF ENGINEERING OF KING'S COLLEGE LONDON, IN THE FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy
ACKNOWLEDGEMENTS

First and foremost I would like to take the opportunity to express my sincere gratitude to my supervisor Dr. Apostolos Georgakis for providing such a dedicated learning experience, and for his limitless patience, support, and guidance. His tireless comprehensive discussions and invaluable insightful suggestions throughout the course of this research work have been a driving force and built up solid foundation to the achievement of this work.

I would also like to give thanks to my friends and colleagues at King’s College London, especially, Suba, Hui, Xiaxi, Ji, Krishna and Jason for their friendship and accompany. Special thanks to my brothers and sisters at London Huaxia Christian Church and Hong Kong Living Water Church for their prayers and warm concern.

The deepest feeling of appreciation for my parents and sister is beyond words. My parents sacrificed most of their time, effort and money to love and help me, and provided the foundation without which it would not have been possible for me to make to this stage. To my wife, Yun, despite her demanding PhD work, she has provided me endless happiness, warm care and continuous encouragement throughout my PhD. I also owe my gratitude to my parents-in-law and brother-in-law for their unreserved and continuous support.

The financial support by the King’s Continuation Research Scholarship, by King’s College London, UK is also gratefully acknowledged.

Most importantly, my utmost admiration and thanksgiving to my Father in the heaven, who is my lord, shepherd and friend, has guided, trained and supported me in this tough journey throughout the past three years in London.

“The fear of the Lord is the beginning of wisdom, and knowledge of the Holy One is understanding.”
Proverbs 9:10
ABSTRACT

The joint time-frequency (TF) domain provides a convenient platform for signal analysis by involving the dimension of time in the frequency representation of a signal. A straightforward way to acquire localized knowledge about the frequency content of the signal at different times is to perform the Fourier transform over short-time intervals rather than processing the whole signal at once. The resulting TF representation is the short-time Fourier transform (STFT), which remains to date the most widely used method for the analysis of signals whose spectral content varies with time. Recent application examples of the STFT and its variants – e.g. the squared magnitude of the STFT known as the spectrogram – include signal denoising, instantaneous frequency estimation, and speech recognition.

In this thesis, we first address the main limitation of the trade-off between time and frequency resolution for the TF analysis by proposing a novel adaptation procedure which properly adjusts the size of the analysis window over time. Our proposed approach achieves a high resolution TF representation, and can compare favorably with alternative time-adaptive spectrograms as well as with advanced quadratic representations.

Second, we propose a new scheme for the time-frequency adaptation of the STFT in order to automatically determine the size and the phase of the analysis window at each time and frequency instant. This way, we can further improve the resolution of the conventional as well as the time-adaptive spectrograms.

Finally, we focus on denoising non-stationary signals in the STFT domain. We introduced an optimized TF mask in the STFT domain, which is based on the concept of the multi-window spectrogram. Experimentation has shown that the introduced approach can effectively recover distorted signals based on a small set of representative examples of the noisy observation and the desired signal.
Table of Contents

Acknowledgements ........................................................................................................... 2

Abstract ............................................................................................................................ 3

Lists of Figures and Tables .............................................................................................. 7

Author's Publications ......................................................................................................... 10

Chapter 1 Introduction ....................................................................................................... 11
  1.1 Background and Related Work .................................................................................. 13
  1.2 Applications of the STFT ......................................................................................... 17
    1.2.1 STFT-based Time-Frequency Signal Representation .............................................. 17
    1.2.2 STFT-based Filtering ............................................................................................ 17
  1.3 Thesis Overview ......................................................................................................... 18

REFERENCES ..................................................................................................................... 19

Chapter 2 The Short-Time Fourier Transform .................................................................. 26
  2.1 The STFT Analysis, Modification and Synthesis ......................................................... 26
    2.1.1 The Continuous STFT .......................................................................................... 26
    2.1.2 The Discrete STFT ............................................................................................. 29
    2.1.3 Implementation of the Discrete STFT by FFT ...................................................... 30
    2.1.4 Vector Formulation of the Discrete STFT ............................................................ 32
  2.2 Window Size Considerations ...................................................................................... 35
  2.3 Case Study I: Filtering of Biomechanical Impact Signals .......................................... 37
    2.3.1 Methodology ....................................................................................................... 39
    2.3.2 Experimental Results .......................................................................................... 42
  2.4 Case Study II: Filtering of Ultrasonic Elastography ................................................... 47
    2.4.1 Methodology ....................................................................................................... 49
    2.4.2 Experimental Results .......................................................................................... 51

REFERENCES ..................................................................................................................... 57

Chapter 3 Time Adaptation of the STFT ........................................................................ 62
  3.1 Introduction ............................................................................................................... 62
  3.2 Definition of the Time-Adaptive STFT ..................................................................... 65
3.3 Previous Work on the Time-Adaptive STFT ............................................. 66
3.4 The Proposed Method for Localizing TF Supports .................................. 68
  3.4.1 The WD and the Geometry of its Cross-Terms .................................... 68
  3.4.2 Auto-Term Localization: Principles .................................................. 72
  3.4.3 Auto-Term Localization: Processing Steps ......................................... 73
3.5 An Alternative method for Localizing TF supports ................................... 76
  3.5.1 The Equivalent Relations between the Cross-WD and the Scaled STFT .......................................................... 77
  3.5.2 Trimming the STFT: Principles ......................................................... 78
  3.5.3 Trimming the STFT: Processing Steps ............................................... 80
3.6 Time Adaptation of the Analysis Window Width ...................................... 83
  3.6.1 Experimental Results .......................................................................... 85
3.7 An Additional Processing Step: Separation of Time-Overlapping Signal
Components ................................................................................................. 92
  3.7.1 Experimental Results .......................................................................... 93
3.8 Summary .................................................................................................. 96
REFERENCES ................................................................................................. 97

Chapter 4 Time-Frequency Adaptation of the STFT ........................................ 100
4.1 Introduction .............................................................................................. 100
4.2 The Short-Time Fourier Transform and its Variations ............................. 103
4.3 The Global Duration-Bandwidth Product for the STFT and its Variations .. 106
4.4 The Local Duration-Bandwidth Product for the Time-Frequency Adaptive
STFT ............................................................................................................. 110
4.5 Time-Frequency Adaptation of the Time-Frequency Adaptive STFT ........ 113
  4.5.1 Optimal Parameters by Minimizing the LDBP .................................... 113
  4.5.2 Estimation of the Optimal Parameters ................................................. 117
  4.5.3 Experimental Results .......................................................................... 122
4.6 Summary .................................................................................................. 130
APPENDIX A ................................................................................................. 130
APPENDIX B ................................................................................................. 132
Chapter 5 Short-Time Fourier Transform based Optimized Filter

5.1 Introduction

5.2 Fixed-Window STFT based Filtering

5.2.1 Formulation of the Optimized Mask in Least-Squares Sense

5.2.2 Experimental Results

5.3 Multi-Window STFT based Filtering

5.3.1 STFT Analysis-Modification-Synthesis Filtering

5.3.2 The Smoothing Effect of the Fixed-Window STFT on the Modification Mask

5.3.3 Minimizing the Smoothing Effect using the Weighted Sum of the Auto-Wigner Distributions of the Hermite Functions

5.3.4 Determination of the Optimized Weights

5.3.5 Trimming of the Mask

5.3.6 The Multi-Window STFT Approach

5.3.7 Experimental Results

5.4 Summary

REFERENCES

Chapter 6 Conclusions and Future Work

6.1 Summary of Main Conclusions

6.2 Future Research Ideas
LISTS OF FIGURES AND TABLES

List of Figures

Fig. 1.1 A representation of a bat echolocation sound........................................13

Fig. 2.1 Filtering functions..................................................................................41

Fig. 2.2 Processing stages of the proposed STFT filtering method......................42

Fig. 2.3 The Dowling signal................................................................................45

Fig. 2.4 The filtered acceleration of the signal by Dowling..............................45

Fig. 2.5 The running signal................................................................................46

Fig. 2.6 The filtered acceleration of the running signal..................................46

Fig. 2.7 Filtering functions................................................................................49

Fig. 2.8 An elastogram.......................................................................................52

Fig. 2.9 Filtered elastogram by different methods............................................54

Fig. 2.10 Filtered axial strain by different methods.............................................55

Fig. 3.1 An example of different window widths for the STFT.......................63

Fig. 3.2 Illustration of the cross-terms in the WD.............................................71

Fig. 3.3 Auto-term localization example...........................................................75

Fig. 3.4 Generation of target area for STFT optimisation...............................83

Fig. 3.5 Example 1............................................................................................86

Fig. 3.6 Example 2............................................................................................88

Fig. 3.7 Example 3............................................................................................89
# List of Tables

Table 1.1 Summary of some time-frequency representations………………………….16

Table 2.1 RMS and AP errors of the filtered accelerations of the Dowling signal...46

Table 2.2 RMS, 1st and 2nd AP errors of the filtered accelerations of the running signal..........................................................................................................................47

Table 2.3 $CNR_E$ of the filtered strain signal by different filtering techniques........57

Table 2.4 RMS error of the filtered strain signal by different filtering techniques...57

Table 5.1 Average MSE of the examples...............................................................155

Table 5.2 The optimized weights $c_k^{opt}$ for various $k$ and $r$............................167

Table 5.3 Average MSE of the Examples................................................................173
AUTHOR'S PUBLICATIONS


Chapter 1

INTRODUCTION

Conventional Fourier analysis pioneered by Fourier in 1807 is a powerful tool to decompose a time-domain signal into separate frequency components and the relative intensities of the individual frequencies are shown in the representation [1]. However, the temporal behavior of the signal’s frequency components is unknown in the conventional Fourier analysis. Unlike the conventional Fourier frequency domain, the joint time-frequency (TF) domain provides a convenient platform for signal analysis by involving the dimension of time in the frequency representation of a signal. A straightforward way to acquire localized knowledge about the frequency content of the signal at different times is to perform the Fourier transform over short-time intervals rather than processing the whole signal at once. The resulting TF representation is the short-time Fourier transform (STFT) [2], which remains to date
the most widely used method for the analysis of signals whose spectral content varies with time. Recent application examples of the STFT and its variants – e.g. the squared magnitude of the STFT known as the spectrogram – include signal denoising [3],[4], instantaneous frequency estimation [5],[6], and speech recognition [7]. Let’s take a look at a real-life example to illustrate the needs for the joint TF representations. Using a bat echolocation sound [8], Fig. 1.1 shows three different plots. By only looking at the time domain plot of the signal at the bottom, we can only see how the intensity or loudness varies with time. On the left of the main plot is the energy density spectrum, that is, the squared magnitude of the Fourier transform of the time signal. It indicates the relative intensity of each frequency components. The power spectrum tells us that the frequency samples mainly range from 50 to 200 samples, but it does not show when these frequencies happen. The main plot is a joint TF plot represented using the contour plot. From it, we can see the frequencies and their relative intensities happened at different time. For instance, the bat signal mainly consists of three nonlinearly FM chirps components overlapping in time.
Fig. 1.1 A representation of a bat echolocation sound. The energy density spectrum is on the left plot frequencies increasing upwards. The time domain signal is below the main figure and indicates the time duration. The main figure is a joint time-frequency plot based on the TF-adaptive STFT described in Chapter 4 and shows how the frequencies change with time.

1.1 Background and Related Work

The joint TF representations can be classified into two main categories [9]. In the first category, a signal is projected onto basis functions (or known as TF atoms) with specific time and frequency localization. The TF representation of a signal \( x(t) \) in the first category is shown as follows

\[
X(t, \omega) = \int y_{t,\omega}^*(\tau)x(\tau)d\tau
\]  

(1.1)

where \( y_{t,\omega}(\tau) \) is the basis functions and the superscript indicates the complex
conjugate. The short-time Fourier transform (STFT) [2], [10], the wavelet transform [11],[12], the adaptive STFT [13], [14] and the matching pursuit algorithms [15] are some representative members of this category. The conventional STFT applies the Fourier analysis to a short and overlapping windowed input signal. The window is fixed for the entire signal. Despite the simplicity of the STFT, its overall time-frequency resolution is constraint by the fixed window. The adaptive STFT has variable analysis window in time or time-frequency finding by some adaptation rules such as the concentration measure. The matching pursuit finds the best matching projections of a signal onto an over-complete dictionary. The second category is the Cohen class of the TF distributions [16]. Different members in this category are classified by a kernel function which also determines the properties of a specific distribution. The expression of the TF representations can be written as

\[ X(t, \omega) = \frac{1}{4\pi^2} \iiint \varphi(\theta, \tau) x^* \left( u - \frac{\tau}{2} \right) x \left( u + \frac{\tau}{2} \right) e^{-j\theta t - j\tau \omega + j\theta u} d\tau d\theta \]  

(1.2)

where \( \varphi(\theta, \tau) \) is known as the kernel by [17]. From the above expression, various distributions can be obtained by selecting different kernels. For example, the Wigner distribution (WD) [18], the spectrogram (SP) [2], and the Choi-Williams distribution (CWD) [19]. The Wigner distribution is one of the prominent members of TF representations because it satisfies many mathematics properties such as time and frequency marginal. Besides, it gives optimal TF concentration for a chirp signal. However, for multi-component or non-linearly modulated signals, it suffers from
existence of cross-term interferences and has low readability for TF analysis. By
convolving the Wigner distribution with different smoothing kernel, the cross-term
interference can be suppressed, but this also pays the price of losing some TF
concentration by spreading the signal auto-terms. Some time-frequency
representations in these two categories are summarized in TABLE 1.1.

Among the above-mentioned TF representations or other TF representation
generated by different basis or kernel functions, the short-time Fourier transform
(STFT) and its variations remain the prime method for the analysis of signals whose
spectral content varies with time because of its many merits: simplicity, linearity,
positivity, robustness to noise, efficient computation and free of cross-terms between
signal components. The STFT has been applied to analyze, modify and synthesize
non-stationary or time-varying signals. How frequency components of a signal or
system vary with time can be pictured by the three-dimensional plots of the STFT.
The STFT can also be synthesized from the time-frequency domain to its
Corresponding time domain signal, hence allowing a time-frequency implementation
of signal design, time-varying filtering, noise suppression, etc. Thanks to these
inherent advantages of the STFT, it has been successfully applied to signal
processing problems in many different areas. The squared magnitude of STFT
known as the spectrogram can be obtained by (1.2) using kernel equal to an analysis
short-time window. Applications of the STFT can be found in two main streams:
STFT-based time-frequency signal representation, and STFT-based filtering. The
details of these two applications are provided in the upcoming sections.
<table>
<thead>
<tr>
<th>TF representations</th>
<th>Basis functions $\gamma$ /Characteristic kernels $\phi$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1st Category (Inner product with basis functions)</strong></td>
<td></td>
</tr>
<tr>
<td>STFT [2], [10]</td>
<td>$\gamma(t - \tau)e^{it\omega}$, where $\gamma$ is a short-time window</td>
</tr>
<tr>
<td>Wavelets [11], [12]</td>
<td>$\frac{\omega}{\omega_0}\gamma\left(\frac{\omega}{\omega_0}(t - \tau)\right)$, where $\gamma$ is an analyzing wavelet</td>
</tr>
<tr>
<td>Matching pursuit [15]</td>
<td>$\frac{1}{\sqrt{\pi}}\gamma_{t,\omega}\left(\frac{t}{\pi}\right)e^{it\omega}$, where $\gamma$ is a normalized Gaussian window from a dictionary and $s$ is for scaling</td>
</tr>
<tr>
<td>Time-Adaptive STFT [13]</td>
<td>$\gamma_s(t - \tau)e^{it\omega}$, where $\gamma$ is a short-time window and changes with time</td>
</tr>
<tr>
<td>Time-Frequency Adaptive STFT [14]</td>
<td>$\gamma_{t,\omega}(t - \tau)e^{it\omega}$, where $\gamma$ is a short-time window and changes with time and frequency</td>
</tr>
<tr>
<td><strong>2nd Category (Cohen Class)</strong></td>
<td>$\phi = 1$</td>
</tr>
<tr>
<td>WD [18]</td>
<td></td>
</tr>
<tr>
<td>Spectrogram [2]</td>
<td>$\int \gamma^*(t - \frac{1}{2})\gamma(t + \frac{1}{2})e^{-it\theta}dt$, where $\gamma$ is a short-time window</td>
</tr>
<tr>
<td>CWD [19]</td>
<td>$e^{-\theta^2\tau^2/\sigma}$</td>
</tr>
<tr>
<td>ZAMD [44]</td>
<td>$\gamma(\tau)</td>
</tr>
<tr>
<td>SPWD [45]</td>
<td>$\gamma\left(\frac{i}{2}\right)\gamma^*\left(-\frac{i}{2}\right)\bar{G}(\theta)$, $\gamma$ and $\bar{G}$ are smoothing functions</td>
</tr>
</tbody>
</table>
1.2 Applications of the STFT

1.2.1 STFT-based Time-Frequency Signal Representation

The STFT and its squared magnitude known as the spectrogram have been successfully applied to time-varying signal analysis for wide range of signals such as human speech [7],[20],[21], music [22], [23], seismic signals [24], biomedical signals [25], radar signals [26], etc. Other applications include estimation of the instantaneous frequency of a signal from the TF domain of the STFT [5], [6], [27], speech coding [28], and signal detection and parameter estimation [29], [30].

The resolution of TF representations based on the STFT, however, highly depends on the window function in use, and the detailed analysis of this issue is given in Section 1.2 and Section 1.3. Choosing an appropriate window which matches the characteristics of the signal being analyzed is crucial. STFT with inappropriate analysis windows leads to erroneous TF representations with signal components either smearing in time or frequency. Therefore, many data-adaptive approaches based on the STFT have been proposed to solve this problem [31]-[36]. Two novel data-adaptive TF representation based on the STFT have been proposed in this thesis and presentation in Chapter 3 and Chapter 4.

1.2.2 STFT-based Filtering

As explained in the previous section, the STFT is available for reconstruction or
synthesis. It has been used for filtering or denoising non-stationary signals such as ultrasound elastography [4], biomechanical impact signals [37], and speech enhancement [38].

In most applications mentioned above, the modification or filtering mask for the STFT is normally unavailable at hand and it has to be either user-defined based on the characteristics of specific signal or noise, or automatically adapted based on the analyzed signals. Recently, the optimally determined modification of the STFT for high-resolution electrocardiogram has been proposed in [39]-[41]. Some other optimal approach based on the STFT can be found in [42]-[43]. In this thesis, two cases studies for the STFT filtering including biomechanical impact signals and ultrasonic elastography are provided in Chapter 2. A novel approach for the optimized STFT filter is proposed in Chapter 5.

1.3 Thesis Overview

In this thesis, we focus on two major areas: (i) STFT-based TF signal representation and (ii) STFT-based filtering. We develop two data-dependent adaptive schemes to select appropriate analysis windows for the STFT TF representation. We apply the STFT based filters to deal with specific problems in real-world applications, such as filtering of biomechanical impact signals, and ultrasound elastography. Moreover, we propose novel derivations for optimized STFT-based filtering using fixed-window and multi-window approaches.

The outline of the thesis is as follows. Chapter 2 describes the essential
Chapter 1. Introduction

theoretical background on the short-time Fourier transform and some of its necessary properties with respect to signal representation and filtering. Chapter 2 also includes a specific type of low-pass filters with time-varying cutoff thresholds, which can be realized by operating in the short-time Fourier transform domain, to be applied in real-world problems including denoising biomechanical impact signals, and ultrasound elastography. Comparisons will be shown between our suggested filters and current techniques for validation purposes. Chapter 3 and Chapter 4 present two adaptive schemes for the STFT-based TF signal representation. Chapter 3 introduces a simple and automatic adaptive scheme for selecting an optimal window at each time for the STFT, in which signal energy is maximized within a localized TF area determined by a trimmed Wigner distribution mask or by a trimmed fixed-window STFT mask. In addition to this, we further develop another adaptive scheme in Chapter 4 which selects an optimal window at each time-frequency point based on the optimal analytic solution, and further improve the resolution of resulting TF signal representation. Chapter 5 presents the application of short-time transform in the context of filtering. In particular, it introduces an alternative solution for formulating optimized STFT filters on both fixed-window and multi-window approaches, which has particular advantages over existing ones.

REFERENCES


tool for time-frequency signal analysis; part III: relations with other
1980.

the time-frequency plane," *Philips Journal of Research*, vol. 37, pp. 79–110,
1982.

[19] H. I. Choi and W. J. Williams, "Improved time-frequency representation of

processing and spectral analysis,” *IEEE Trans. on Acoust., Speech, and Signal


music analysis," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5,
no. 6, pp. 1088-1110, 2011.

short-time Fourier transform,” *Journal of the Audio Engineering Society*, vol. 35,
no. 1-2, pp. 3-14, Jan.-Feb. 1987.


Chapter 1. Introduction


CHAPTER 2

THE SHORT-TIME FOURIER TRANSFORM

2.1 The STFT Analysis, Modification and Synthesis

2.1.1 The Continuous STFT

The continuous short-term Fourier transform (STFT) analysis of a signal \( x(t) \) can be obtained as [1]:

\[
X_{STFT}(t, \omega; h) = \int h^*(t - \tau)x(\tau)e^{-j\omega\tau}d\tau \tag{2.1}
\]

where \( h(t) \) is the analysis window which determines the portion of \( x(n) \) being analyzed. The analysis window may be chosen to be real or imaginary, but it is typically chosen to be a real and symmetric function centered at zero, tapering off to zero away from its centre. Hence, at each time instant, (2.1) computes the Fourier
transform of a short portion of the signal around $t$, and thus the STFT can be regarded as a local spectrum of the signal $x(\tau)$ around time $t$. Accordingly, the short-time energy-density spectrum $S(t, \omega; h)$ can be obtained as the squared magnitude of (1), i.e. $S_x(t, \omega; h) = |X_{STFT}(t, \omega; h)|^2$, and is commonly called the spectrogram. When a unit-energy window is used then the total energy of the spectrogram equals that of the signal.

There exists an elemental relationship between the spectrogram and the Wigner distribution (WD)

$$W_x(t, \omega) = \int x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) e^{-j\omega\tau} d\tau.$$  \hspace{1cm} (2.2)

Namely, the convolution of the WD of the signal with the WD of the STFT window renders the spectrogram, i.e.

$$S_x(t, \omega; h) = \iint W_x(t', \omega')W_h(t - t', \omega - \omega')dt' d\omega'.$$  \hspace{1cm} (2.3)

The WD is widely recognized for its high TF concentration [2],[3]. On the other hand, it suffers from the presence of cross-term interference which limits its readability, and prevents it from being strictly positive. Although the process in (2.3) is known to eliminate interference and restore positivity, it also smears the signal in the TF plane. So, despite its relative advantages, the spectrogram yields inferior TF signal localization compared to the WD. This deterioration depends on the
smoothing kernel $W_h(t, \omega)$ of the two-dimensional convolution operation in (2.3).

Since the Gaussian function exhibits the least amount of spread in the TF plane [4], it often is the preferred window type for STFT-based signal analysis [5].

The synthesis or reconstruction formula for the STFT can be derived from the signal expansion proposed by [6],[7], in which a signal $x(t)$ is considered to be represented by a superposition of time-frequency shifted versions of an elementary signal $f(t)$.

$$x(t) = \frac{1}{2\pi} \int \int T_x(\tau, \omega) f(t - \tau) e^{j\omega t} d\tau d\omega . \quad (2.4)$$

The above expression may also be interpreted as an expansion of the signal $x(t)$ into the elementary signals $f(t - \tau)e^{-j\omega t}$ continuously indexed by $\tau$ and $\omega$. $T_x(\tau, \omega)$ is coefficient function indicating the amount of contribution of the region round time-frequency point $(\tau, \omega)$ to the signal $x(t)$. Choosing the STFT as the coefficient function $T_x(\tau, \omega)$ results in the synthesis formula of the STFT. By substituting (2.1) into (2.4), we get

$$x(t) = \frac{1}{2\pi} \int \int X_{STFT}(\tau, \omega) f(t - \tau)e^{j\omega t} d\tau d\omega . \quad (2.5)$$

The above synthesis formula holds for prefect reconstruction of the signal $x(t)$ if

satisfying the following condition [8]
\[
\int f(-t)h^*(t) dt = 1.
\] (2.6)

Therefore, (2.5) shows how to synthesize or reconstruct a signal from its STFT.

There are infinitely many combinations of the analysis window \( h^*(t) \) and the synthesis window \( f(t) \) such that the condition in (2.6) is satisfied. A common choice of the analysis and synthesis window is \( f(t) = h(t) \). When there is modification between the STFT analysis and synthesis, the synthesis formula in (2.5) with the choice \( f(t) = h(t) \) performs the least squares reconstruction of a signal from its modified STFT analysis, which determines the signal minimizing the squared difference between the STFT and the modified STFT of the signal [9].

Inserting a multiplicative modification (a time-frequency weighting function) between the STFT analysis and synthesis is a way to implement a time-varying filter in the joint time-frequency domain.

### 2.1.2 The Discrete STFT

The discrete version of the STFT of a signal \( x(n) \) is obtained as [10]

\[
X_{STFT}[n,k] = \sum_{m=-\infty}^{\infty} h(n - m)x(m)e^{-j\Omega_M km} \quad \text{for } k=0, 1, ..., M-1
\] (2.7)

where \( h(n) \) is the analysis window. The analysis is performed at \( M \) uniformly spaced frequency bins at step \( \Omega_M = 2\pi/M \); \( n \) and \( k \) are the space and frequency indices, respectively. The STFT can also be computed by using the frequency-domain
representation of the analysis window and the signal at hand:

\[
X_{\text{STFT}}[n, k] = e^{-j\Omega_M kn} \sum_{s=-\infty}^{\infty} H(k - s)X(s)e^{j\Omega_M ns},
\]

(2.8)

where \(H(k)\) and \(X(k)\) are the Fourier transforms of the window function \(h(n)\) and the input signal \(x(n)\), respectively.

To reconstruct the signal from the STFT domain back to the time domain we employ the method for discrete short-time Fourier synthesis based on the overlap addition (OLA) algorithm [1]. This method can achieve perfect reconstruction as long as the following constraint on the analysis and synthesis windows is satisfied [8],

\[
\sum_{m=-\infty}^{\infty} f(-m)h^*(m) = 1,
\]

(2.9)

where \(f(n)\) is the synthesis window. The discrete short-time Fourier OLA synthesis can be mathematically described as:

\[
x(n) = \frac{1}{M} \sum_{k=0}^{M-1} \sum_{m=-\infty}^{\infty} f(n - m)X[m, k] e^{j\Omega_M kn},
\]

\[
= \sum_{m=-\infty}^{\infty} f(n - m) \frac{1}{M} \sum_{k=0}^{M-1} X[m, k] e^{j\Omega_M kn}.
\]

(2.10)

### 2.1.3 Implementation of the Discrete STFT by FFT

According to the previous section, the definition of the discrete STFT of a signal...
\( x(n) \) in (2.7) can be rewritten as

\[
X_k[n] = \sum_{m=-\infty}^{\infty} h(n - m)x(m)W_M^{-mk} \tag{2.11}
\]

where \( W_M^k = e^{j\frac{2\pi n k}{M}} \). The limits given in (2.11) goes to infinite, but in practice are finite and determined by the effective length of \( h(n) \). By recognizing \( X_k[n] \) for fixed \( n \), as samples equally spaced in frequency, \( X_k[n] \) can be obtained by time-domain aliasing \( h(n - m)x(m) \) and then computing the DFT of the aliased sequence [11].

The expression of the discrete STFT in (2.11) can be written as

\[
X_k[n] = \sum_{m=-\infty}^{\infty} \sum_{r=-\left(m + 1\right)M+1}^{n-mM} h(n - r) x(r)W_M^{-rk} \tag{2.12}
\]

where the infinite sum in (2.11) is rewritten as infinite sum of blocks of \( M \) samples.

By changing of variables, the inner sum can be written as

\[
X_k[n] = \sum_{m=-\infty}^{\infty} \sum_{r=0}^{N-1} h(r + mM)x(n - r - mM)W_M^{(r-n+mM)k} \tag{2.13}
\]

as \( W_M^k = \exp\left[\frac{j2\pi n k}{M}\right] \), frequencies are equally spaced in \( \Delta W = \exp\left[\frac{j2\pi}{M}\right] \). Then, the advantages of the periodicity of the complex exponential to write

\[
X_k[n] = W_M^{-nk} \sum_{r=0}^{M-1} \tilde{x}_r[n]W_M^{rk} \tag{2.14}
\]
where $\tilde{x}_m[n]$ is the aliased short-time sequence

$$\tilde{x}_r[n] = \sum_{m=-\infty}^{\infty} x(n - r - mM)h(r + mM)$$

(2.15)

In addition to the computational savings gained by computing the discrete STFT using the FFT algorithm (using $M$ is an integral power of 2). By circularly shifting $\tilde{x}_r[n]$ in $m$, prior to computing its DFT, the multiplications by $W_{M^{-nk}}$ can be avoided and (2.14) can be rewritten as [12]

$$X_k[n] = \sum_{r=0}^{M-1} \tilde{x}_{((r-n))_M}[n]W_{M^{-rk}}$$

(2.16)

and $((n))_M$ denotes $n$ reduced modulo $M$. The computational complexity can be estimated as $O(NK\log_2K)$, where $K$ and $N$ is the frequency size and signal length, respectively.

### 2.1.4 Vector Formulation of the Discrete STFT

For the sake of formulating the optimized STFT filtering in Chapter 5, we first express the STFT in the vector form proposed by [13] in this section.

Let an input real/complex signal $x(n)$ having $J$ samples be the following vector:
The input signal $\mathbf{x}$ is segmented into $M$ overlapping short sections. Each section has the length $L$, for $L$ is odd and equals to the size of the short window applied later. The shift length from section to section is 1 sample. The $m$-th section ($0 \leq m \leq J-1$) of $x(n)$ is denoted as

$$
\mathbf{x}_m = [x(m), x(m+1), \ldots, x(m+L-1)]^T \in \mathbb{C}^N. \quad (2.18)
$$

The relationship between $\mathbf{x}$ and $\mathbf{x}_m$ can be described by

$$
[\mathbf{x}_0^T, \mathbf{x}_1^T, \ldots, \mathbf{x}_{M-1}^T]^T = \mathbf{Ox} \quad (2.19)
$$

where

$$
\mathbf{O} = \begin{bmatrix}
\mathbf{I}_L \\
1 \\
\leftrightarrow \quad \\
\leftrightarrow \quad \\
\leftrightarrow \quad \ddots \\
\leftrightarrow \quad \\
\mathbf{I}_L
\end{bmatrix} \in \mathbb{R}^{ML \times J} \quad (2.20)
$$

is called the overlap matrix. It consists of $M$ identity matrices and each identity matrix is shifted 1 sample to the right with respect to the above one. Each section $\mathbf{x}_m$ is then weighted by a real valued window $h(n)$ with the same length $L$ as the short section of the signal. Then, this windowed sequence is appended by $K-L$ zeros.
by \( P \) before the discrete Fourier transform \( F \) is calculated in \( K \) uniformly spaced frequency points, i.e. \( k=0, 1, ..., K-1 \), for \( K\geq L \), with the frequency step being equal to \( \Omega = 2\pi/K \). The STFT of \( x \) can be formulated in a vector form:

\[
x_{STFT} = F_{STFT}^h L x
\]

\[
= \left( I_M \otimes (FPH) \right) O x
\]

\[
= \begin{bmatrix}
FPH \\
FPH \\
\vdots \\
FPH
\end{bmatrix}
\begin{bmatrix}
I_L \\
I_L \\
\vdots \\
I_L
\end{bmatrix}
\begin{bmatrix}
x(0) \\
x(1) \\
\vdots \\
x(J - 1)
\end{bmatrix}
\in \mathbb{C}^{MK}
\tag{2.21}
\]

where

\[
H = diag(h(0), h(1), ..., h(L - 1)) \in \mathbb{R}^{L \times L}
\tag{2.22}
\]

\[
P = \begin{bmatrix}
I_L \\
0_{(K - L) \times L}
\end{bmatrix}
\in \mathbb{R}^{K \times L}, K \geq L
\tag{2.23}
\]

\[
F = \begin{pmatrix}
e^{-j\Omega \cdot 0} & e^{-j\Omega \cdot 1} & \cdots & e^{-j\Omega \cdot (K-1)} \\
e^{-j\Omega \cdot 1} & e^{-j\Omega \cdot 1} & \cdots & e^{-j\Omega \cdot (K-1)} \\
\vdots & \vdots & \ddots & \vdots \\
e^{-j\Omega \cdot (K-1)} & e^{-j\Omega \cdot (K-1)} & \cdots & e^{-j\Omega \cdot (K-1)}
\end{pmatrix}
\in \mathbb{C}^{K \times K}
\tag{2.24}
\]

Similarly, according to [13], the vector formulation for the least-squares inverse STFT estimator \( F_{ISTFT}^h L \) for synthesizing a signal from the modified STFT is shown as follows and the recovered signal \( \hat{x} \) is given by

\[
\hat{x} = F_{ISTFT}^h L x_{STFT}
\]

\[
= \left[ O^H (I_M \otimes H^2) O \right]^{-1} O^H (I_M \otimes HP^H F^{-1}) x_{STFT}
\tag{2.25}
\]
where

\[(I_M \otimes H^2) = \begin{bmatrix} H^2 & \dots & H^2 \\ \end{bmatrix} \in \mathbb{R}^{ML \times ML} \quad (2.26)\]

\[(I_M \otimes H^{PF^{-1}}) = \begin{bmatrix} H^{PF^{-1}} & \dots & H^{PF^{-1}} \\ \end{bmatrix} \in \mathbb{C}^{ML \times MK} \quad (2.27)\]

2.2 Window Size Considerations

In signal processing, the uncertainty principle or the duration-bandwidth product theorem means that “a narrow waveform yields a wide spectrum and a wide waveform yields a narrow spectrum and both the time waveform and frequency spectrum cannot be made arbitrarily small simultaneously” by [14]. Since, in the procedure of performing the STFT, a signal is modified by a windowing function and some properties of the original signal is influenced by the windowing function. For instance, if we employ a very short window to analyze the signal, the resulting bandwidth of the modified signal becomes very high, and this bandwidth trivially does not represent or equal to the bandwidth of the original signal [15]. In fact, when we deploy the technique of the STFT for analyzing a signal, the resolution of the resulting representation is constrained and there exists inherent tradeoff between time and frequency resolution due to the windowing effect (i.e. narrowing the window in time reduces the frequency resolution, and vice versa).
The duration of the analysis window is a critical parameter of the spectrogram. While a suitable length can optimize its concentration, a badly chosen one can adversely affect the representation accuracy. To illustrate this, let us consider the Gaussian signal \( x(t) = e^{-\alpha t^2/2} \), and contrast its WD to its spectrogram representation. The WD of this signal is \( W_x(t, \omega) = \sqrt{\frac{2\pi}{\alpha}} e^{-(\alpha t^2 + \frac{1}{\alpha} \omega^2)} \). At a fixed level \( p \) – a percentage of its maximum value – the resulting contour line of this WD is an ellipse centered at zero with horizontal (time) and vertical (frequency) semi-axes equal to:

\[
\begin{align*}
  r_{t,p}^W &= \sqrt{\frac{1}{\alpha} (-\ln p)} \\
  r_{\omega,p}^W &= \sqrt{\alpha (-\ln p)} .
\end{align*}
\]  

(2.28)

Consistent with the time and frequency characteristics of the signal, when \( \alpha > 1 \) the frequency bandwidth of the distribution is greater than its time duration. The opposite happens when \( \alpha < 1 \), whereas a circle is formed for \( \alpha = 1 \). The TF spread of the WD at level \( p \) is reflected in the area of the ellipse, which is equal to \( \pi r_{t,p}^W r_{\omega,p}^W = \pi (-\ln p) \) for any value of \( \alpha \).

The spectrogram of \( x(t) \) – calculated using the unit-energy Gaussian window \( g(t) = \frac{4}{\sqrt{\pi}} e^{-\beta t^2/2} \) – is \( S_x(t, \omega; g) = \sqrt{\frac{2\sqrt{\pi \beta}}{\alpha + \beta}} e^{-(\alpha \beta t^2 + \frac{1}{\alpha + \beta} \omega^2)} \). Thus, at the percentage \( p \) of its maximum value, the spectrogram also forms an ellipse centered at zero, with time and frequency semi-axes equal to:

\[
\begin{align*}
  r_{t,p}^S &= \sqrt{\frac{\alpha + \beta}{\alpha \beta} (-\ln p)} \\
  r_{\omega,p}^S &= \sqrt{(\alpha + \beta) (-\ln p)} .
\end{align*}
\]  

(2.29)
It can be observed that the aspect ratio of this ellipse depends on the product \( \alpha \beta \) instead of the actual size of the signal; if \( \alpha \beta > 1 \) the spectrogram is stretched in frequency and squeezed in time, and vice versa. As an example, assuming a very small value of \( \alpha \) one would expect the time duration of the distribution to be much greater than its bandwidth. However, if \( \beta > 1/\alpha \), the resulting spectrogram will appear more prolonged in frequency than it is in time, yielding an erroneous representation of the signal. The correct TF proportion is achieved for \( \beta = \alpha \) whereby both semi-axes are equal to those of the WD times \( \sqrt{2} \). Furthermore, when \( \beta = \alpha \), the spectrogram reaches its smallest possible size, which is \( 2\pi (-\ln p) \).

The above conclusion does not relate only to the Gaussian signal but applies as a general rule. The spectrogram performs best when the duration of the analysis window matches that of the local signal components [16]. Thus, the fixed-window STFT cannot accurately represent the entire signal, especially when this consists of elements that have very different time-frequency characteristics. In the following Chapters, we are going to propose two data-driven adaptive schemes to deal with this issue.

### 2.3 Case Study I: Filtering of Biomechanical Impact Signals

In the following two sections, we focus on the application of filtering using the STFT. In particular, we present a specific type of low-pass filters with time-varying
cut-off thresholds, which can be realized by operating in the STFT domain, to be applied in two real-world problems: denoising biomechanical impact signals and ultrasonic elastography images. The work introduced in this chapter has been reported in [17] and [18].

The analysis of biomechanical signals is crucial for the study of human motion and finds important applications in areas related to Health and Sports. Human motion is captured by optical or magnetic sensors and the resulting displacement signals are recorded for further analysis. The velocity and acceleration can then be calculated by taking the first and second derivatives of these signals. However, during the measuring process, the displacement signals are typically contaminated with noise – and even worse, noise is amplified during differentiation – so the calculated accelerations are severely obscured by noise. It is therefore necessary that denoising is applied prior to differentiation. Denoising biomechanical signals is not a trivial issue since the frequency content of these signals may vary dramatically as a result of impact events. Thus, conventional low-pass filters are unable to cope with this problem since their cutoff thresholds are constant over time and consequently, schemes such as the popular second-order bi-directional Butterworth low-pass filter [19] either under-smooth or over-smooth the results [20].

In this section, the proposed denoising method is based on the short-time Fourier transform (STFT) [19], [21] in the effort to include the dimension of time in the frequency representation of biomechanical signals. An appropriate time-varying filtering function is designed to modify the STFT so that noise is rejected while
Chapter 2. The Short-Time Fourier Transform

preserving the actual signal frequencies. Comparisons are carried out between the proposed algorithm and the bi-directional Butterworth filter on experimental signals involving a single impact as well as multiple ones.

2.3.1 Methodology

The proposed method is based on filtering a signal in the time-frequency domain. The method is employing a time-varying filter of a low-pass characteristic. Since the time-frequency representation used is the STFT we will refer to the method as STFT filtering.

The noisy signal is first extrapolated so as to mitigate the well-known end-point problem in filtering. The signal is then transformed into the joint time-frequency domain by the STFT as described in the previous section. The Hamming window is chosen as the STFT analysis window in this application due to its favorably low side-lobe magnitude [22].

A suitable low-pass filtering function is implemented by first designing its time-varying cutoff frequency \( f_c(n) \) in the time-frequency plane so that it matches the time-frequency characteristics of the signal at hand. This means that the designed cutoff frequency should maintain an overall low bandpass area except for those times corresponding to impact events where the cutoff threshold should extend toward higher frequencies. Suitable functions for the implementation of such boundaries can readily be identified in the likes of such distributions as the Laplacian (2.30), Cauchy (2.31) and Gaussian (2.32):
where $X_L$ is the low cutoff frequency corresponding to the smooth parts of the signal, $X_H$ is the cutoff frequency increase at the time of impact $n_o$ required to accommodate the range of higher frequencies induced by the sudden change in the signal due to the impact. Parameter $\alpha$ controls the widths of each function and relates to the duration of the impact event. The 2-D filter is then constructed as a time-varying Butterworth function:

$$F[n, k] = \sqrt{\frac{1}{1 + \left(\frac{k}{f_c(n)}\right)^{2N(n)}}}$$

where $N(n)$ is the order of the Butterworth function which determines its steepness. The order is adapted so that the roll-off of the filtering function maintains the same cutoff rate at all times, e.g.:

$$N(n): \quad F[n, k] \leq 0.1 \quad \text{for} \quad k \geq f_c(n) + B$$

where $B$ is a user-defined value that corresponds to the desired transition bandwidth of the filtering function. Fig. 2.1a illustrates an example of a filtering function.
designed for single-impact signals. On the other hand, Fig. 2.1b shows a filtering function with two Laplacian boundaries which would suit a signal with two distinct impacts. In contrast, Fig. 2.1c depicts the fixed cut-off frequency of a conventional Butterworth low-pass filter.

![Fig. 2.1 Filtering functions. (a) Proposed time-varying filter for single-impact signals. (b) Proposed time-varying filter for double-impacts signals. (c) The conventional Butterworth filter with fixed cut-off frequency.](image)

The designed filtering function is subsequently multiplied with the representation of the signal in the joint time-frequency domain. The resulting
modified STFT is then reconstructed back into the time domain by the synthesis method described in the previous section. Finally, the additional points due to extrapolation are removed from the edges of the filtered signal. The basic processing stages of the algorithm are illustrated in Fig. 2.2.

Fig. 2.2 Processing stages of the proposed STFT filtering method

2.3.2 Experimental Results

In this section, the proposed algorithm is compared to the popular bi-directional Butterworth filter. The root mean square (RMS) error of the calculated acceleration and the absolute error of the acceleration peak (AP) were used to quantify the performance of the two algorithms:

\[
RMS\ error = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (\hat{a}(n) - a(n))^2} \quad (2.35)
\]
\[
AP\ error = \left|\frac{|\hat{a}(n_o)| - |a(n_o)|}{|a(n_o)|}\right| \quad (2.36)
\]

where \(\hat{a}(n_o)\) and \(a(n_o)\) are the filtered acceleration and the reference acceleration at the time of impact \(n_o\), respectively.

Two commonly used test signals in biomechanics were employed for our
comparisons. The first is the displacement data experimentally acquired by Dowling [23] which involved a single impact event region. The corresponding acceleration signal which was measured independently by means of accelerometer devices is used in this study as the reference acceleration to gauge the performance of the two algorithms. The sampling rate of this displacement signal was 512 Hz and the point of maximum acceleration due to impact is located at time sample 197.

The second biomechanical signal, known as the running signal was simulated by van den Bogert and de Koning [24] with two distinct impact regions, and its corresponding accelerations were computed by using the second-order forward differences of the original clean displacement data. The sampling rate of the displacement signal was down-sampled from the original 10000 Hz to the more realistic sampling rate of 500 Hz. The first impact relates to a heel strike appears at time sample 164, followed by a softer forefoot impact at time sample 259.

Initially, experimentation which involved different types of functions including the Laplacian (2.30), Cauchy (2.31), and Gaussian (2.32) functions was carried out to establish which one was more suitable for the time-varying cutoff threshold of the proposed STFT filter when applied to biomechanical impact signals. The results showed that the Laplacian boundary was the most appropriate choice in general yielding lower RMS and AP errors than the other boundaries. The better performance of the Laplacian boundary could be explained by its comparatively wider shape at lower frequencies and its relative higher peakedness for larger frequency values. Thus, it avoids truncating useful low-frequency components of the
signal around the time of impact while it rejects more of the high-frequency noise.

In these experiments we examined the performance of the proposed STFT filtering algorithm in dealing with biomechanical signals with either one or multiple impacts and examined its robustness against noise. Different levels of white Gaussian noise were added to our test signals. It should be noted that the Dowling signal already contains noise since it is an experimentally acquired signal. The parameters of the Butterworth digital filter (i.e. cutoff frequency and order) as well as those of the proposed scheme (i.e. height and width of the Laplacian boundary) were determined empirically so as to minimize an equal-weight combination of the error measures of (2.35) and (2.36). Table 2.1 compares the RMS and AP errors of the proposed STFT filtering and the conventional Butterworth filtering for different SNR levels of added noise on the Dowling signal. The results show that the proposed algorithm can filter the signal more efficiently than the conventional Butterworth filter. Fig. 2.3a shows the Dowling signal with added white noise at 40dB SNR, and Fig. 2.3b shows the corresponding second derivative compared with the reference acceleration. Fig. 2.4 depicts the filtered Dowling signal by both methods. Table 2.2 presents the RMS and AP errors (for the two acceleration peaks of the running signal) obtained after STFT denoising and conventional filtering at different SNR levels. In both Tables 2.1 and Table 2.2, the listed RMS and AP errors are the average results over 100 realizations of the noisy inputs. Fig. 2.5a shows the running signal with added white Gaussian noise at 40dB SNR. Meanwhile, Fig. 2.5b depicts the corresponding acceleration compared with the reference acceleration. Fig.
Chapter 2. The Short-Time Fourier Transform

2.6 shows the filtered signal by the two methods at SNR equal to 40 dB. Again, the proposed algorithm clearly outperforms the conventional filter both in terms of overall noise reduction and acceleration-peak accuracy.

(a)                                               (b)

Fig. 2.3 The Dowling signal. (a) The Dowling signal with added white noise (SNR = 40 dB) and (b) the corresponding acceleration (dotted line) compared with the reference acceleration (solid line)

(a)                                               (b)

Fig. 2.4 The filtered acceleration of the signal by Dowling (with additional white noise of SNR = 40 dB) after applying: (a) the proposed STFT filtering (solid line) and (b) the conventional Butterworth filter (solid line). The reference acceleration measured by accelerometers is also shown (dotted line)
Chapter 2. The Short-Time Fourier Transform

Fig. 2.5 The running signal. (a) The running signal with added white noise (SNR=40dB) and (b) the corresponding acceleration (dotted line) compared with the reference acceleration (solid line).

Fig. 2.6 The filtered acceleration of the running signal (with additional white noise of SNR = 40dB) after applying: (a) the proposed STFT filtering (solid line) and (b) the conventional Butterworth filter (solid line). The reference acceleration measured by accelerometers is also shown (dotted line).

TABLE 2.1 RMS AND AP ERRORS OF THE FILTERED ACCELERATIONS OF THE DOWLING SIGNAL AT DIFFERENT SNR LEVELS OF ADDED WHITE NOISE BY THE PROPOSED STFT FILTERING AND THE CONVENTIONAL BUTTERWORTH FILTERING.

<table>
<thead>
<tr>
<th>Method</th>
<th>SNR(dB)</th>
<th>60</th>
<th>50</th>
<th>40</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed STFT Filtering</td>
<td>RMSE</td>
<td>16.45</td>
<td>16.80</td>
<td>17.54</td>
<td>23.72</td>
</tr>
<tr>
<td></td>
<td>APE (%)</td>
<td>0.77</td>
<td>2.09</td>
<td>6.07</td>
<td>13.85</td>
</tr>
<tr>
<td>Butterworth Filtering</td>
<td>RMSE</td>
<td>34.92</td>
<td>35.76</td>
<td>36.15</td>
<td>37.67</td>
</tr>
<tr>
<td></td>
<td>APE (%)</td>
<td>11.45</td>
<td>11.30</td>
<td>15.95</td>
<td>28.19</td>
</tr>
</tbody>
</table>
TABLE 2.2 RMS, FIRST AP AND SECOND AP ERRORS OF THE FILTERED ACCELERATIONS OF THE RUNNING SIGNAL AT DIFFERENT SNR LEVELS OF ADDED WHITE NOISE BY THE PROPOSED STFT FILTERING AND THE CONVENTIONAL BUTTERWORTH FILTERING.

<table>
<thead>
<tr>
<th>Method</th>
<th>SNR(dB)</th>
<th>60</th>
<th>50</th>
<th>40</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed STFT</td>
<td>RMSE</td>
<td>3.50</td>
<td>5.25</td>
<td>8.96</td>
<td>15.89</td>
</tr>
<tr>
<td></td>
<td>1st APE (%)</td>
<td>5.56</td>
<td>8.82</td>
<td>11.58</td>
<td>14.59</td>
</tr>
<tr>
<td></td>
<td>2nd APE (%)</td>
<td>1.40</td>
<td>1.19</td>
<td>3.11</td>
<td>13.86</td>
</tr>
<tr>
<td>Butterworth</td>
<td>RMSE</td>
<td>10.31</td>
<td>11.24</td>
<td>16.98</td>
<td>22.70</td>
</tr>
<tr>
<td>Filtering</td>
<td>1st APE (%)</td>
<td>8.64</td>
<td>10.59</td>
<td>19.43</td>
<td>30.99</td>
</tr>
<tr>
<td></td>
<td>2nd APE (%)</td>
<td>14.95</td>
<td>13.41</td>
<td>21.77</td>
<td>27.80</td>
</tr>
</tbody>
</table>

2.4 Case Study II: Filtering of Ultrasonic Elastography

Ultrasound elastography is a promising medical imaging technique to visualize the stiffness distribution of soft tissue [25]. Potential applications of the method include the diagnosis of different types of cancerous inclusions, such as in the case of breast cancer where scirrhous carcinomas of the breast are much stiffer than fibroadenoma tumors. One way of generating a strain image is by applying small external quasi-static compressions to the tissue, and then estimating the local tissue displacements using the cross-correlation of the pre- and post-compression ultrasound signals along the axial direction. The gradient of the tissue displacements in the axial direction provides the local tissue strains, which are normally displayed as grey-scale images known as elastograms.

A fundamental problem in the estimation of tissue strains is that the noise present in the displacements is amplified by differentiation, and therefore, the resulting elastograms may become severely degraded. To tackle this problem
low-pass filtering is usually employed prior to the application of the gradient operator [26]. However, the frequency content of the displacement image undergoes abrupt changes at those points of the image corresponding to the transitions between the stiffer tissue of the inclusion and the softer background tissue. Conventional low-pass filters are unable to accommodate these non-stationarities because their cut-off threshold is constant for the whole signal. Therefore, the resulting elastogram is either under-smoothed or over-smoothed.

Alternative advanced denoising techniques such as those based on wavelet transforms (WT) [27] or quadratic time-frequency transforms like the Wigner distribution (WD) [28] have been proposed for dealing with non-stationary signals. In fact, the published results of these methods demonstrate good estimation of the derivatives of noisy signals. Nevertheless, the efficiency of wavelet denoising relies on multiple signal-specific factors, e.g. choice of basis functions, number of decomposition levels and thresholding strategy. Even worse, the truncation of wavelet coefficients often generates interference in the form of pseudo-Gibbs artifacts. On the other hand, the WD-based approach suffers from the presence of cross-terms in the transform domain and the difficulty to reconstruct the filtered signal in its original domain [29].

In this section, a novel space–spatial-frequency filtering technique based on the concept of the short-time Fourier transform (STFT) is proposed for the estimation of the axial strain in ultrasound elastography. An appropriate multiplicative function is designed to operate on the STFT in a low-pass manner but with a varying cut-off
threshold, and the denoised signal is then reconstructed from the modified STFT. We compare the performance of the proposed method with common low-pass filtering, wavelet denoising, and WD-based filtering using simulated elastograms.

### 2.4.1 Methodology

The proposed method is based on filtering a signal in the joint space–spatial-frequency domain into which the signal is transformed by means of the STFT. The method is employing a multiplicative filtering function of a low-pass nature whose cut-off varies along the spatial parameter $n$ (Fig. 2.7).

![Filtering functions](image)

**Fig. 2.7** Filtering functions. (a) Example of the 3-D filtering function $F[n,k]$ of the proposed method, and (b) representation of the fixed cut-off threshold corresponding to conventional low-pass filtering.

Fig. 2.2 illustrates the processing stages of the proposed scheme. First, the noisy signal is extrapolated so as to overcome end-point problems typically induced by filtering. The signal is then transformed into the joint space–spatial-frequency
domain by the STFT as described in the previous section. A Hamming window was chosen as the STFT analysis window here due to its well-known low side-lobes magnitude. Next, a filtering function is designed and is subsequently multiplied with the representation of the signal in the transform domain. The resulting denoised signal is finally reconstructed from the modified STFT following the synthesis method described above. The extrapolation points added at the outset are now removed.

The implementation of the filtering function is based on the design of a 2-D low-pass cut-off threshold \( f(n) \) whose value depends on the spatial variable \( n \) so that it matches the evolution of spatial frequencies across space. For the signal at hand, this means that \( f(n) \) should maintain an overall low bandpass area except for those space samples corresponding to the boundaries between the inclusion and the surrounding medium. Around those points the cutoff threshold should extend toward higher frequencies. Thus, for all axial cross-sections of the inclusion-containing part of the displacement image, \( f(n) \) should exhibit two distinct peaks. Experimentation has shown that a suitable function for the cut-off threshold is the Laplacian, which is modified here as follows:

\[
f(n) = \begin{cases} 
X_L + X_H e^{-\frac{|n-n_1|}{\alpha_1}}, & 0 \leq n \leq n_1 \\
X_L + X_H e^{-\frac{|n-n_1|}{\alpha_2}}, & n_1 < n \leq \frac{n_1+n_2}{2} \\
X_L + X_H e^{-\frac{|n-n_2|}{\alpha_2}}, & \frac{n_1+n_2}{2} < n \leq n_2 \\
X_L + X_H e^{-\frac{|n-n_2|}{\alpha_1}}, & n > n_2
\end{cases}
\]

(2.37)
where \( X_L \) is the low cutoff frequency corresponding to the linear parts of the displacement (i.e. within the medium and the inclusion) whereas \( X_H \) is the required cutoff increase around the spatial samples \( n_1 \) and \( n_2 \) which are located at the boundaries between medium and inclusion. Parameters \( \alpha_1 \) and \( \alpha_2 \) control the widths of \( f(n) \) on either side of each boundary and relate to the size of the transition from one type of tissue to the other.

The full 3-D multiplicative function is then constructed so that it possesses a smooth roll-off along the spatial-frequency variable \( k \) in order to avoid a sharp truncation of frequencies which would lead to artifacts in the filtered signal. The roll-off is controlled by the Butterworth-type function in (2.33). Fig. 2.7a illustrates an example of the filtering function \( F[n,k] \). On the other hand, Fig. 2.7b is a representation in the joint space–spatial-frequency plane of the fixed cut-off threshold corresponding to conventional low-pass filtering. The relative advantage of the proposed scheme becomes apparent in this figure.

### 2.4.2 Experimental Results

To validate the proposed method we generated simulated displacement measurements along with their corresponding ideal elastograms according to the two-dimensional analytic model, introduced by Muskhelishvili et al. [30]. The model assumes that the tissue is subjected to an uniform uniaxial compression of 314Pa in a plan-strain state. The dimensions of the simulated phantom were 50x50 mm with an inclusion which was 12 mm in diameter and 4 times stiffer than the
medium. Zero mean white Gaussian noise were also added to the simulated displacement images at 35db SNR which results in a signal-to-noise ratio in the strain image of 10dB SNR_{st}. An ideal and noisy elastogram generated by this model are shown in Fig. 2.8.

![An elastogram.](image)

(a) (b)

Fig. 2.8 An elastogram. (a) An ideal elastogram with an inclusion of 12mm diameter, and (b) a corrupted elastogram (SNR_{st}=10dB)

To quantify the performance of the algorithms we have employed two error measures commonly used in the literature. The contrast-to-noise ratio (CNR_{e}) described in [31] as:

\[
\text{CNR}_{e} \text{(dB)} = 20 \log_{10} \left( \frac{2(\mu_{s1}-\mu_{s2})^2}{\sigma_{s1}^2 + \sigma_{s2}^2} \right),
\]

(2.38)

where \(\mu_{s1}\) and \(\mu_{s2}\) represent the mean value of strain in the inclusion and the medium, and \(\sigma_{s1}\) and \(\sigma_{s2}\) denote the strain variances, respectively. The relative mean square (RMS) error of the calculated strain was also used:
\[ \text{RMS error} = \frac{1}{N} \sum_{n=1}^{N} \left( \frac{\hat{\alpha}(n) - \alpha(n)}{\hat{\alpha}(n)} \right)^2, \]  
\[ (2.39) \]

where \( \hat{\alpha}(n) \) and \( \alpha(n) \) are the filtered strain and the ideal strain, respectively.

For the sake of comparison, we implemented three alternative denoising techniques. The first one is the established Butterworth filtering method [26]. The second approach was wavelet denoising employing the two different wavelet families suggested in previous studies [27]. In particular, Symlet (sym8) and biorthogonal (bior3.3) wavelets were used as well in this study. Finally, the WD-based time-frequency filter was also implemented in a similar way as in [28].

The values of the parameters of the different filters in this comparison, namely the above presented \( X_L, X_H, \alpha_1, \) and \( \alpha_2 \) for the proposed method; the cutoff threshold for the Butterworth low-pass filter; the number of decomposition levels and the thresholding approach for the wavelet denoising method; and the height and width of the filtering mask for the WD-based filter were all determined empirically so as to achieve a maximum \( \text{CNR}_e \) value. All signals were extrapolated prior to filtering.

For all the comparisons reported here, one hundred independent realizations of the noisy displacements were calculated. Fig. 2.9 shows the resulted elastograms obtained by the different denoising techniques whereas Fig. 2.10 illustrates their corresponding axial strain profiles along the centre of the inclusion. The \( \text{CNR}_e \) and RMS error values achieved at different \( \text{SNR}_{st} \) levels by the compared methods are presented in Tables 2.3 and 2.4. As it is evident in the above results the proposed algorithm has the potential to outperform the other techniques both in terms of
contrast and overall noise reduction. The low-pass Butterworth filter cannot maintain a clear boundary of the inclusion; neither can it achieve adequate overall noise removal. The wavelet-based algorithms suffer from spurious spikes as predicted. Lastly, the WD-based filter’s performance came close to the proposed method although it appeared particularly sensitive to end-point problems.
Fig. 2.9 Filtered elastogram by different methods. (a) The corrupted elastogram (SNR_{st}=10dB) and calculated elastogram after denoising with: (b) Butterworth low-pass filter, (c) the biorthogonal (bior3.3) wavelet-based denoising, (d) the symlet (sym8) wavelet-based denoising, (e) the WD-based filter, and (f) the proposed STFT filter.
Fig. 2.10 Filtered axial strain by different methods. (a) Ideal axial strain (SNR<sub>s</sub>=10dB) along the centre of the elastogram (dotted line) and axial strain after denoising (solid line) with: (b) Butterworth low-pass filter, (c) the biorthogonal (bior3.3) wavelet-based denoising, (d) the symlet (sym8) wavelet-based denoising, (e) the WD-based filter, and (f) the proposed STFT filter.
TABLE 2.3 CONTRAST-TO-NOISE RATIO (CNR$_{st}$) OF THE FILTERED STRAIN SIGNAL BY DIFFERENT FILTERING TECHNIQUES AT DIFFERENT SNR$_{st}$ LEVELS

<table>
<thead>
<tr>
<th>Method</th>
<th>SNR$_{st}$(dB)</th>
<th>-10</th>
<th>0</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. LP filter</td>
<td></td>
<td>19.87</td>
<td>29.54</td>
<td>31.69</td>
<td>48.86</td>
<td>49.15</td>
<td>49.22</td>
</tr>
<tr>
<td>2. WT (bior)</td>
<td></td>
<td>-2.33</td>
<td>18.57</td>
<td>35.71</td>
<td>47.41</td>
<td>58.91</td>
<td>67.03</td>
</tr>
<tr>
<td>3. WT (sym)</td>
<td></td>
<td>13.28</td>
<td>31.00</td>
<td>41.44</td>
<td>47.90</td>
<td>51.00</td>
<td>55.69</td>
</tr>
<tr>
<td>4. WD filter</td>
<td></td>
<td>11.90</td>
<td>29.36</td>
<td>42.90</td>
<td>50.66</td>
<td>55.12</td>
<td>73.96</td>
</tr>
<tr>
<td>5. STFT filter</td>
<td></td>
<td>21.13</td>
<td>32.88</td>
<td>43.22</td>
<td>51.74</td>
<td>66.37</td>
<td>75.26</td>
</tr>
</tbody>
</table>

TABLE 2.4 ROOT MEAN SQUARE (RMS) ERROR ($10^{-2}$) OF THE FILTERED STRAIN SIGNAL BY DIFFERENT FILTERING TECHNIQUES AT DIFFERENT SNR$_{st}$ LEVELS

<table>
<thead>
<tr>
<th>Method</th>
<th>SNR$_{st}$(dB)</th>
<th>-10</th>
<th>0</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. LP filter</td>
<td></td>
<td>34.08</td>
<td>33.80</td>
<td>33.76</td>
<td>3.30</td>
<td>3.12</td>
<td>2.98</td>
</tr>
<tr>
<td>2. WT (bior)</td>
<td></td>
<td>81.79</td>
<td>9.55</td>
<td>2.75</td>
<td>1.11</td>
<td>0.43</td>
<td>0.19</td>
</tr>
<tr>
<td>3. WT (sym)</td>
<td></td>
<td>14.30</td>
<td>8.81</td>
<td>4.30</td>
<td>2.41</td>
<td>1.13</td>
<td>0.41</td>
</tr>
<tr>
<td>4. WD filter</td>
<td></td>
<td>90.43</td>
<td>11.72</td>
<td>2.04</td>
<td>2.46</td>
<td>0.36</td>
<td>0.15</td>
</tr>
<tr>
<td>5. STFT filter</td>
<td></td>
<td>12.74</td>
<td>5.29</td>
<td>1.36</td>
<td>0.59</td>
<td>0.22</td>
<td>0.13</td>
</tr>
</tbody>
</table>

REFERENCES


3.1 Introduction

As discussed and analyzed using the Gaussian modal in the Chapter 2, STFT’s main limitation is its inherent tradeoff between time and frequency resolution. Specifically, decreasing the short-time interval in order to locate events in time reduces the frequency resolution, in accordance with the relevant properties of the underlying Fourier transform. The theoretical concentration bounds of TF representations are specified by the uncertainty principle. In STFT’s case, reaching these limits relies on the proper selection of the analysis window. Indeed, a study of the spectrogram [2] has revealed that its resolution can increase dramatically when the window width is matched to the content of the signal at hand. Traditionally, in communities like speech and radar processing the choice of the short-time interval’s length has been guided by experience but in general the window is arbitrarily
selected. This is why STFT has been commonly regarded as a heuristic or *ad hoc* method. To make matters worse, the standard STFT employs a fixed window for the short-term analysis of the entire signal. However, this practice may produce poor results since the signal can vary significantly over its lifespan. The inadequacy of the standard fixed-length STFT can easily be demonstrated using an example signal such as the one in Fig. 3.1a. This is a sum of a sinusoid (narrow-band component), a modulated Gaussian, and a linear frequency-modulated (chirp) signal with a steep slope (wide-band component). The three components overlap in frequency. Fig. 3.1b presents the STFT of the test signal when a long window is employed. Then the sinusoid exhibits good concentration as expected; however, all other components are significantly smeared in the time direction. The overall TF resolution seems to be satisfactory for this example when a medium-sized window is used (Fig. 3.1c), although there is clear compromise regarding the concentration of all three components. Finally, when the STFT is based on a relatively short analysis interval then only the chirp can be accurately represented (Fig. 3.1d).
Dealing with the window problem has been greatly facilitated by the fact that the STFT can be optimized locally. This has encouraged the development of methods for adapting the STFT to the local structure of the analyzed signal [3]-[8]. The associated methodologies can broadly be classified into two categories; those based on the conventional STFT definition, where the duration of the analysis window is usually the sole parameter optimised at each time instant, and those introducing modified versions of the STFT in which the window can be adjusted at every time-frequency location. Adapting the STFT both in time and frequency generally leads to computationally expensive algorithms. In addition, this degree of sophistication may be unnecessary in certain practical situations [3]. In such cases, the former class of methodologies is preferable because they combine good performance and efficient implementation.

The main contribution of this chapter is that a novel method for locally adjusting the window width in order to enhance the resolution of the spectrogram is
developed. This approach iteratively searches for the window size that maximizes the concentration of the signal components within their specified supports. In particular, we present a new method using the information about the geometrical positions of auto-terms and cross-terms of the Wigner distribution (WD) to detect the required TF supports for the adaptation. We also find a more efficient way to obtain the TF supports by deriving an equivalent relationship between the cross-WD and the scaled STFT. The proposed method is simple, does not require a priori knowledge of the signal or any user intervention, and admits of efficient implementation. To evaluate the validity of the proposed method, we implement and evaluate some existing time-adaptive spectrograms. Experimentation on different simulated signals and the experimental bat signal has shown that the proposed method is able to achieve high TF concentration, and can outperform previously reported time-adaptive spectrograms. In addition, we solve the shortcoming of the existing adaptive methods, as well as the proposed scheme, on signals with components overlapping in time, since the adaptation of the STFT is performed only with respect to time. We introduce an additional processing step which consists in dividing the signal into parts possessing a single frequency element at each point in time based on the STFT analysis-modification-synthesis method.

3.2 Definition of the Time-Adaptive STFT
To mitigate the resolution problem of the classical fixed-length STFT, the analysis window can be adapted in order to match the different components of the signal at
hand. In this chapter, we follow the route previously taken in [6]-[8] for producing simple and efficient adaptive schemes. Therefore, we similarly restrict the adaptation process to the adjustment of the window width in time only. The adaptive STFT can then be formulated as follows.

\[ X(t, \omega; w_{L(t)}) = \int w_{L(t)}(t - \tau)x(\tau)e^{-j\omega \tau}d\tau, \quad (3.1) \]

where \( w_{L(t)} \) is the appropriately sized window at the time instant \( t \). Thus, (3.1) employs analysis windows of different lengths at different times so that a TF representation with a finer resolution than that of the standard STFT can be achieved. Our objective is then to develop a method which can automatically determine suitable window sizes based on the local characteristics of the analyzed signal.

### 3.3 Previous Work on the Time-Adaptive STFT

Notable examples of time-adaptive STFT schemes include the works in [5]-[8]. In [5] a simple technique aiming to enhance the STFT-based visualization of radar target resonances iteratively increased the window width until a reflection was encountered – as that could be determined by a slope detector. Although its performance was satisfactory, this method was tailor-made for the specific application and lacks general applicability.
A generic time-adaptive spectrogram was proposed in [6]. The central idea was to evaluate TF concentration over user-specified short intervals centered at each time instant. The employed local concentration measure (LCM) was defined as the ratio of the L4 norm to the L2 norm of the windowed STFT. At each point in time, the optimal window length of the STFT was selected to be the one that maximized the proposed LCM. It was shown that this method performed well on a variety of signals at a cost only slightly higher than that of a fixed-window spectrogram. On the other hand, the STFT adaptation was necessarily based on the dominant component within the same measurement interval. As a result, the representation was poor for signal terms lying close to components of different bandwidth characteristics and higher energy concentrations.

A different approach which was based on the time adaptation of the cone-kernel distribution (CKD) emerged later [7]. In particular, at each time instant, the STFT window was assumed to be twice as long as the cone used in the ambiguity domain to minimise the cross-terms of the CKD. The time adaptation of the CKD was very efficient in computational terms, and the estimated STFT window sizes were reasonable in general. However, the mismatch between the STFT window and the CKD cone was pronounced during time transitions between different signal components. As a result, the method generated spurious values at those instances.

Recently, an alternative method was introduced [8], according to which the STFT window length was adapted to the evaluated degree of local signal stationarity. This was achieved by monitoring the gradient of the instantaneous
frequency (IF) curves of the signal components. The measured IF gradient (IFG) was used to determine segments of stationary behavior around each time instant. The STFT window was subsequently matched to the duration of those segments. The IF was estimated by detecting the ridges of the continuous wavelet transform computed using a complex Morlet wavelet. Clearly, the effectiveness of this method depends on the quality of the IF estimation, and the accuracy of the stationarity intervals approximation.

3.4 The Proposed Method for Localizing TF Supports

Our proposed method in this work follows the idea of the previous work in [6] in which the optimal length for the time-adaptive STFT is adapted in a condition to maximize the LCM at each time instant in a short interval of the TF plane. Instead, we adapt the optimal length for the time-adaptive STFT so that the energy within localized TF supports is maximized. The localized TF supports are formulated based on exploiting the geometry of the cross-terms of the WD. The details of the formulation of the localized TF support are explained in this section.

3.4.1 The WD and the Geometry of its Cross-Terms

The Wigner distribution (WD) \( WD_x(t, \omega) = \int_{-\infty}^{\infty} x \left( t + \frac{\tau}{2} \right) \overline{x} \left( t - \frac{\tau}{2} \right) e^{-j\omega\tau} d\tau \) is widely recognized as a prominent TF representation [9] because it satisfies a large number of mathematical properties and achieves high TF concentration of the signal components [10]. For real-valued signals, the WD of the their analytic signals are
practically deployed, known as the Wigner-Ville distribution (WVD), in order to eliminate the cross-terms between the positive and negative frequency components. Let us consider, for the sake of argument, the unit-energy Gaussian signal

\[ g(t) = \sqrt{\frac{2a}{\pi}} e^{-at^2} . \]  \hfill (3.2)

The WD of \( g(t) \) is equal to \( WD_g(t, \omega) = 2e^{-(2at^2+\frac{1}{2a}\omega^2)} \). For any fixed level of its magnitude, e.g. \( WD_g(t, \omega) = p \ WD_g(t, \omega)_{max} = p 2, \ 0 < p < 1 \), this distribution forms an ellipse in the TF plane, which is centered at zero with horizontal and vertical semi-axes equal to:

\[ \alpha = \sqrt{\frac{-lnp}{2a}} \quad \text{and} \quad \beta = \sqrt{2a (-lnp)} . \]  \hfill (3.3)

Therefore, the TF area required by the WD to represent the strongest values of \( g(t) \) down to level \( p WD_g(t, \omega)_{max} = \pi (-lnp) \). The corresponding area occupied by the spectrogram is twice as large at best, with the optimal analysis window being the identical Gaussian signal [11].

The WD can be expressed in discrete form as follows [12]:

\[ WD_x[n, k] = \sum_{m=-N+1}^{N-1} x[n + m]x^*[n - m]e^{-j2\omega km} , \]  \hfill (3.4)
where the frequency step $\Omega_W$ is equal to $\pi/M$. The WD is bilinear in the signal and although this bilinearity increases the sharpness of the representation, it is also responsible for the generation of cross-terms between individual signal elements [13], [14]. For example, by assuming $x[n] = ax_1[n] + bx_2[n]$ we obtain,

$$WD_x[n, k] = |a|^2 WD_{x_1}[n, k] + |b|^2 WD_{x_2}[n, k] + 2Re\{WD_{x_1, x_2}[n, k]\},$$  \hspace{1cm} (3.5)$$

where $WD_{x_1, x_2}[n, k]$ is the cross-WD,

$$WD_{x_1, x_2}[n, k] = \sum_{m=-N+1}^{N-1} x_1[n + m] x_2^*[n - m] e^{-j2\Omega_W km}.$$  \hspace{1cm} (3.6)$$

Therefore, the WD includes an extra component along with the two anticipated auto-terms $WD_{x_1}$ and $WD_{x_2}$. In general, for $L$ separate signal components there will exist $L(L-1)/2$ spurious terms which have an oscillatory nature and their amplitude can peak at a value twice as high as that of the actual signal terms. This type of interference significantly impairs the WD by reducing its readability. In fact, the resolution of the WD of multi-component signals may be lower than that of the spectrogram, provided that the latter uses suitably adjusted analysis windows [2]. Cross-terms also force the WD to locally assume negative values, a fact preventing its interpretation as a true energy density function [9].

The geometry of the cross-terms of the WD in the TF plane has been well-studied [13],[14]. In particular, it is known that a cross-term will appear
midway between any two TF points of actual signal presence. This is true whether the signal is the superposition of multiple waveforms or it just consists of a single component – in the second case the cross-terms are often termed the inner interference. Fig. 3.2 illustrates both types of WD interference by way of two example signals. Consequently, if the positions of the auto-terms were available the cross-components could easily be specified in the TF plane by considering every possible pair of signal coordinates, e.g. \((t_1, f_1)\) and \((t_2, f_2)\), and then registering interference elements in the middle of the imaginary lines connecting such points, i.e. \(\left(\frac{t_1 + t_2}{2}, \frac{f_1 + f_2}{2}\right)\). Inversely, based on the same geometric law, if the cross-terms generated between a known component and an arbitrary signal could be identified then it would be possible to determine the TF locations of the unknown signal. We will draw on this idea to estimate the TF support of the given signal as described in the following section.

Fig. 3.2 Illustration of the cross-terms in the WD: (a) cross-terms due to the interactions between each pair of four Gaussian atoms located at the corners of a square, and (b) inner interference of the WD of a mono-component quadratic frequency-modulated signal.
3.4.2 Auto-Term Localization: Principles

To specify the TF locations of the given signal we propose to take advantage of the geometry of the cross-terms of the WD as this will be explained in the ensuing paragraphs. Let us assume that the WD of a given signal \( x[n] \) resides in an unknown compact area \( S \) in the TF plane, and that a second signal \( c[n] \) could exist such that its WD would occupy a single TF point \( (t_c, f_c) \). Then, in line with the construction rule for the cross-terms, the cross-WD \( WD_{xc}[n,k] \) between the two signals would occupy an area \( S' \) consisting of the TF points \( (t'_i, f'_i) \) obtained as follows:

\[
\begin{pmatrix}
  t'_i \\
  f'_i
\end{pmatrix} = \begin{pmatrix}
  \frac{1}{2} & 0 \\
  0 & \frac{1}{2}
\end{pmatrix} \begin{pmatrix}
  t_i \\
  f_i
\end{pmatrix} + \begin{pmatrix}
  t_c/2 \\
  f_c/2
\end{pmatrix},
\]  

(3.7)

for all TF points \( (t_i, f_i) \) belonging to \( S \). Therefore, (3.7) can be interpreted as a mapping \( F: \mathbb{R}^2 \rightarrow \mathbb{R}^2 \) which results in a simple affine transformation of the area \( S \) into a new area \( S' \). Consequently, by detecting the points \( (t'_i, f'_i) \) of the cross-WD one could use the inverse transformation to precisely determine the region \( S \).

Since our aim is to localize the areas where the actual signal elements lie, the above reasoning could be employed. However, the uncertainty principle prohibits the existence of signals whose TF representation is an impulse – in other words, there is no signal whose energy can be concentrated in a single TF point. According to the same principle, the Gaussian signal exhibits minimum uncertainty in the TF plane, thus the WD of this type of signal yields the most compact TF representation.
available. Hence, in this work we use the Gaussian signal – centered at \((t_c, f_c)\) and designed such that its WD is circularly symmetric – as the closest approximation to an impulse. We then compute the cross-WD between the Gaussian and the signal with the unknown TF support. Clearly, the TF area specified by this cross-WD is not equal to the desired \(S'_c\) so further processing is required before the inverse of (3.7) can be applied to obtain an estimate of \(S\).

### 3.4.3 Auto-Term Localization: Processing Steps

The semi-axes of the WD-based representation of the Gaussian signal (3.2) can be expressed in a Cartesian grid as \(\alpha = N_\alpha T\) and \(\beta = N_\beta \Omega_W\), where \(T(= 1)\) and \(\Omega_W(= \frac{\pi}{M})\) are the discrete steps in time and frequency, respectively. To achieve a circular Gaussian in the discrete TF plane, we require \(N_\alpha = N_\beta\), thus, \(\frac{\alpha}{\beta} = \frac{M}{\pi}\). Taking (3.3) into account, we further obtain that \(\frac{\alpha}{\beta} = \frac{1}{2a}\) and solving for \(a\), we find that \(a = \frac{\pi}{2M}\). The next step is to compute the cross-WD \(WD_{x,g}[n,k]\) between the analyzed signal and the designed circular Gaussian. The cross-WD following the shape of the auto-terms is complex-valued and has an oscillatory pattern. We remove the oscillatory pattern of \(WD_{x,g}[n,k]\) by taking its squared magnitude. Next, a low threshold \(p\) is applied to the output \(\hat{WD}_{x,g}[n,k]\) of the filter in order to eliminate any computational distortions and small values of \(\hat{WD}_{x,g}[n,k]\) with the aim of producing a uniform function \(D[n,k]\) whose support reflects the shape of the cross-terms,
\[ D[n,k] = \begin{cases} 1, & \text{if } \hat{W}D_{x,g}[n,k] \geq p\hat{W}D_{x,g}[n,k] \\ 0, & \text{otherwise} \end{cases} \] (3.8)

Since the Gaussian signal resides in a TF area much larger than that of an impulse, the support of \( WD_{x,g}[n,k] \) is the result of the combination of individual areas formed according to (3.8) between the signal points \((t_i,f_i)\) and each of the points within the circular support \( S_g \) of the Gaussian. In theory, the Gaussian is nonzero everywhere; however, as its values are effectively negligible at about three standard deviations away from its centre, a compact support can be considered in practice. Assuming that the radius of \( S_g \) is \( r \), then the area occupied by \( WD_{x,g}[n,k] \) is circularly expanded in relation to \( S'_c \) by an amount equal to \( r/2 \) on all sides. To mitigate this we manipulate \( D[n,k] \) as follows:

\[
\tilde{D}[n,k] = \frac{1}{L} \sum_{l_t} \sum_{l_f} D[n,k] D[n - l_t, k - l_f],
\] (3.9)

with \(-r/2 \leq l_t \leq r/2, -r/2 \leq l_f \leq r/2\), and the constraint \( l_t^2 + l_f^2 = r^2/4 \), where \( L \) is the number of points on the circumference of a circle centered at \((0,0)\) with radius \( r/2 \). The above operation essentially trims off the edges of \( D[n,k] \) such that its support \( \hat{S}'_c \) provides an estimate of the ideal \( S'_c \).

The outlined area \( \hat{S}'_c \) can be utilized to produce an estimate \( \hat{S} \) of the location of the signal \( x[n] \) via the inverse transformation of (3.7).
for all points \( (t'_i, f'_i) \) belonging to \( \hat{S}_c \). Clearly, (3.10) scales (magnifies) the area \( \hat{S}_c \) with respect to the point \( (t_c, f_c) \) over a discrete grid. As a consequence, there will be gaps in the computed region, which need to be filled in order to create a uniform area \( \hat{S} \). This can be achieved by different image processing methods; in this work, we have opted for the combination of the morphological operations of dilation and erosion in a joint process called ‘closing’ [15]. The two-dimensional mask \( G[n,k] \) is finally created as:

\[
G[n,k] = \begin{cases} 
1, & (n,k) \in \hat{S} \\
0, & \text{otherwise}
\end{cases},
\]  
(3.11)

The function \( G[n,k] \) is used in the following selection to delineate the area in the TF plane where the STFT is expected to exhibit its maximum energy concentration.

The above steps are illustrated in Fig. 3.3 using the example signal of Fig. 3.1a.

(a)                           (b)
3.5 An Alternative method for Localizing TF supports

In the previous section, our proposed method to find localized TF supports by exploiting the geometry information of the cross-terms in the WD has been clearly described and illustrated through the example illustrated in Fig. 3.3. This method is robust to find localized TF support for any arbitrary signals without acquiring any prior knowledge of the signals. However, this method includes some processing steps like back-projection and ‘closing’ which might not be appreciated. For
instance, the ‘closing’ step, in some signals, will deviate the shape of the obtained TF supports from the ideal one since the shape of the kernel used in the dilation and the erosion operations should be similar to that of the analyzed signals. In this section, we try to avoid some processing steps in the previous proposed method by realizing the equivalent relationship between the cross-WD and the scaled STFT.

### 3.5.1 The Equivalent Relations between the Cross-WD and the Scaled STFT

The continuous cross-WD of \(x(t)\) and \(h(t)\) can be written as

\[
WD_{x,h}(t,\omega) = \int_{-\infty}^{\infty} x\left(t + \frac{\tau}{2}\right) h^*(t - \frac{\tau}{2}) e^{-j\omega \tau} d\tau .
\]  

(3.12)

By changing of variables in the above expression, we substitute \(\varphi = \pi/2 + t\) and \(d\varphi = d\pi/2\) into (3.12). We can get

\[
WD_{x,h}(t,\omega) = 2 \int_{-\infty}^{\infty} x(t + \varphi - t) h^*(t - \varphi + t) e^{-2j\omega(\varphi - t)} d\varphi ,
\]

\[
= 2e^{2j\omega t} \int_{-\infty}^{\infty} x(\varphi) h^*(2t - \varphi) e^{-2j\omega \varphi} d\varphi .
\]  

(3.13)

The above expression become similar to the definition the STFT in (2.1) using \(h(t)\) as an analysis window apart from the TF scaling by the factor of two and the modulation. i.e.
Chapter 3. Time Adaptation of the STFT

\[ WD_{x,h}(t, \omega) = 2e^{2j\omega t}X_{STFT}(2t, 2\omega; h). \quad (3.14) \]

It can then be seen that the squared magnitude of the cross-WD is equivalent to the scaled spectrogram such that

\[ |WD_{x,h}(t, \omega)|^2 = 4|X(2t, 2\omega; h)|^2. \quad (3.15) \]

The above relationship between the squared magnitude of the cross-WD and the scaled spectrogram allows us to avoid some processing steps of the proposed method in the previous section. The affine transformation to produce an estimate \( \hat{S} \) of the location of the signal \( x(t) \) in (3.10) can be avoided since the estimate \( \hat{S} \) is equivalent to the spectrogram using a circular Gaussian window placing in the origin of the TF plane apart from the spreading by the window. The ‘closing’ step to fill the gaps caused by the affine transform can therefore be avoided. The sole processing step to find the localized TF supports remains to properly trim the spreading in the supports caused by the circular Gaussian window. In the following section, the analysis of the spreading in the STFT by the circular Gaussian window, and the formulation of the localized TF supports based on the trimmed version of the STFT will be described in details.

### 3.5.2 Trimming the STFT: Principles

There exists an elemental relationship between the spectrogram and the Wigner
distribution. Namely, the convolution of the WD of the signal with the WD of the STFT window renders the spectrogram, i.e.

\[ S_x(t, \omega; W_L) = \int \int W_x(t', \omega') W_{W_L}(t - t', \omega - \omega') dt' d\omega'. \]  

(3.16)

(3.16) shows that the spectrogram yields inferior TF signal localization compared to the WD. This deterioration depends on the smoothing kernel \( W_{W_L}(t, \omega) \) of the two-dimensional convolution operation (3.16). Since the Gaussian function exhibits the least amount of spread in the TF plane [16], it often is the preferred window type for STFT-based signal analysis [3].

It is clear from (3.16) that the duration of the analysis window is a critical parameter for the STFT. While a suitable length can optimize its concentration, a badly chosen one can adversely affect the representation accuracy. To illustrate this, we recall the window consideration in Section 2.3 using a Gaussian model \( x(t) = e^{-\alpha t^2 / 2} \), and contrast its WD to its spectrogram. The WD of this signal is \( W_x(t, \omega) = \sqrt{2\pi / \alpha} e^{-(\alpha t^2 + 1 / \alpha \omega^2)} \), and the spectrogram of \( x(t) \) – calculated using the unit-energy Gaussian window \( g(t) = \frac{\sqrt{\beta}}{\sqrt{\pi} \alpha} e^{-\beta t^2 / 2} \) – is \( S_x(t, \omega; g) = \sqrt{2\pi / (\alpha + \beta)} e^{-(\alpha / (\alpha + \beta) t^2 + \beta / 2 \omega^2)} \). The total spread of the WD at level \( p \) is reflected in the area of the ellipse, which is equal to \( \pi t_{t,p}^W t_{\omega,p}^W = \pi (-ln p) \) for any value of \( \alpha \). Furthermore, when \( \beta = \alpha \), the spectrogram achieves its smallest possible size, which is \( 2\pi (-ln p) \), whereby both semi-axes are equal to those of the WD times \( \sqrt{2} \).
The above conclusion does not relate only to the Gaussian signal but applies as a general rule. The spectrogram performs best when the duration of the analysis window matches that of the local signal components [3]. Thus, the fixed-window STFT cannot accurately represent the entire signal, especially when this consists of elements that have very different time-frequency characteristics.

### 3.5.3 Trimming the STFT: Processing Steps

In the light of (3.16), the spectrogram can be viewed as an enlarged version of the WD, exclusive of cross-term interference. Therefore, by properly eroding the edges of the spectrogram, a well-localised map of the actual signal structure could be obtained. It would then be possible to use this as a target area within which one could seek to condense the spectrogram by suitably adjusting its analysis window. In the following, we explore the practical application of the above idea.

The spectrogram in (3.16) employs a fixed window which needs to be specified first. Using a unit-energy Gaussian window is a reasonable choice, and in the lack of any particular information to determine its length, this can be set equal to unity, i.e. $\beta = 1$. In fact, the circularly symmetric Gaussian can be deemed appropriate for general-purpose STFT-based analysis. As shown in Section 3.5.2, when the window matches the duration of the signal, the semi-axes of the spectrogram are both prolonged by the constant factor $\sqrt{2}$. In any other case though, the broadening ratios for the time and frequency semi-axes are $\rho_t = \sqrt{\frac{\alpha + \beta}{\beta}}$ and $\rho_\omega = \sqrt{\frac{\alpha + \beta}{\alpha}}$, respectively. This means that in general, the spectrogram will not expand...
symmetrically around the WD. In particular, when $\beta = 1$, it can be observed that
the broadening ratios are proportional to the distribution steepness along each
semi-axis. For example, assuming $\alpha > 1$ – for which the roll-off of the WD is
sharper in time than it is in frequency – it holds that $\rho_t = \sqrt{\frac{\alpha+1}{\alpha}} > \rho_\omega = \sqrt{\frac{\alpha+1}{\alpha}}$.

We can approximate the average expansion size of the spectrogram at level $p$
by computing the mean between the longest and shortest possible extensions. In line
with the above observation, these can be found along the direction of the highest and
the lowest level of WD steepness, respectively. By considering the Gaussian signal
with a very large value of $\alpha$, we can have both extreme roll-off cases at once – the
former along the time axis and the latter along the frequency axis. The associated
spreads are as follows:

$$\Delta t_p = r_{t,p}^S - r_{t,p}^W = \sqrt{\frac{\alpha+1}{\alpha}} (-\ln p) - \frac{1}{\alpha} (-\ln p) \approx \sqrt{(-\ln p)} , \quad (3.17)$$

which is the radius of the WD of the circular Gaussian at level $p$. Similarly,

$$\Delta \omega_p = r_{\omega,p}^S - r_{\omega,p}^W = \sqrt{(\alpha + 1)(-\ln p)} - \sqrt{\alpha(-\ln p)} \approx 0 . \quad (3.18)$$

The average expansion width of the spectrogram – at level $p$ – along any single
direction can then be estimated as $\Delta_p = \frac{\Delta t_p + \Delta \omega_p}{2} = \frac{\sqrt{(-\ln p)}}{2}$.

The target area can now be determined as follows. First, an appropriate
threshold $p$ is applied to the spectrogram so that the signal terms can be extracted
from a background of unimportant small values and computational inaccuracies. Referring back to the Gaussian, it is known that its energy is well-contained within few standard deviations, so it would be straightforward to determine $p$ in that case. To find an appropriate threshold for an unknown signal, one could choose from a variety of techniques available for image segmentation. To automate the choice of $p$ here, we have relied on the optimum global thresholding approach known as Otsu’s method [17]. Nonetheless, experimentation has shown that good performance can also be achieved by simply setting a low threshold value, e.g. 10% of the normalized spectrogram.

The segmented spectrogram can be denoted as:

$$M(t, \omega) = \begin{cases} 1, & S_x(t, \omega; g) \geq p S_x(t, \omega; g) \\ 0, & \text{otherwise} \end{cases}.$$  \hspace{1cm} (3.19)

Next, we omni-directionally trim an amount equal to $\Delta_p$ from the segmented spectrogram by way of the following manipulation,

$$\tilde{M}(t, \omega) = \frac{1}{2\pi \Delta_p} \int_{-\Delta_p}^{\Delta_p} \int_{-\Delta_p}^{\Delta_p} M(t, \omega) M(t - \tau, \omega - \theta) d\tau d\theta , \hspace{1cm} (3.20)$$

under the constraint $\tau^2 + \theta^2 = \Delta_p^2$. The overall process is illustrated in Fig. 3.4, where the spectrogram $S_x(t, \omega; g)$ of an example waveform (depicted in Fig. 3.4a) is shown next to its segmented version $M(t, \omega)$ and the resulting trimmed area $\tilde{M}(t, \omega)$. 82
3.6 Time Adaptation of the Analysis Window Width

By adapting the size of the STFT window, we aim to maximize the energy of the spectrogram within $\tilde{M}(t, \omega)$ so that it achieves a finer resolution than its fixed-window counterpart. The rationale behind this approach is that the TF region defined by the eroded mask $\tilde{M}(t, \omega)$ can be regarded as an area within which a highly-localized spectrogram should be able to place a significant part of its total energy. This is a good assumption as long as the area of $\tilde{M}(t, \omega)$ is smaller than (or equal to) the minimum area achievable by the spectrogram. To test this, we turn to the Gaussian signal model, and calculate the area of $\tilde{M}(t, \omega)$ at level $p$,

$$\pi \left( \sqrt{\frac{\alpha+1}{\alpha} \left(-\ln p\right) - \frac{\sqrt{-\ln p}}{2}} \right) \left( \sqrt{(\alpha + 1)\left(-\ln p\right)} - \frac{\sqrt{-\ln p}}{2} \right)$$
Chapter 3. Time Adaptation of the STFT

\[
= \pi \left( \sqrt{\frac{\alpha + 1}{\alpha}} - \frac{1}{2} \right) \left( \sqrt{\alpha + 1} - \frac{1}{2} \right) (-\ln p).
\]  

(3.21)

It can be shown that the above area is indeed smaller than \(2\pi(-\ln p)\) for any value of \(\alpha\) between 0.058 and 17.360. Hence, \(\tilde{M}(t, \omega)\) is a generally acceptable target area for the adaptation of the spectrogram, except for the cases of signals with extremely smooth edges or abrupt (pulse-type) behavior.

The time-adaptive STFT can be expressed as follows,

\[
X(t, \omega; w_{L(t)}) = \int w_{L(t)}(t - \tau)x(\tau)e^{-j\omega t}d\tau,
\]

(3.22)

and its energy within \(\tilde{M}(t, \omega)\) at time \(t\) is:

\[
E(t; L) = \int |X(t, \omega; w_{L})|^2 \tilde{M}(t, \omega)d\omega.
\]

(3.23)

Therefore, the objective is to find the optimum length \(L^*(t)\) such that,

\[
L^*(t) = \arg \max_L E(t; L).
\]

(3.24)

The process involves the computation of fixed-window STFTs (trial STFTs) for different window sizes. Similar to [6], the number of assessed window sizes can be kept small, and the trial STFTs can be coarsely sampled in order to keep the computational cost low. By doing so, both adaptive methods reach a complexity of
only few times that of the conventional STFT.

### 3.6.1 Experimental Results

To illustrate the proposed method, we employ five different examples involving computer-simulated multicomponent waveforms of varied magnitudes and wide-ranging time-frequency characteristics and a real-life signal. All the contour plots used in this chapter for the representations of the signals in the TF plane depict a range of values from 0dB down to -10dB at intervals of 2dB, where 0dB is the normalized maximum value of each representation. For comparison, the state-of-the-art time-adaptive schemes of [6] and [8] have also been considered. The different parameters and thresholds required by the LCM-based algorithm and the IFG-based approach were set according to the recommendations in [6] and [8], respectively. In addition, two advanced quadratic TF techniques, namely the S-method [18], and the modified-B distribution [19], [20] have also been used. These were adjusted manually with the aim of providing the best achievable representations according to the visual inspection of the resulting TF plots. It should be stressed that in none of the following examples was it possible to completely suppress the cross-terms without severely compromising the overall TF concentration. This effect was more pronounced on the wide-band portions of the signals, which would undergo excessive smearing along the time direction.

**First Example: Frequency-Overlapping Signal Terms.** The signal in the first example (Fig. 3.5a) comprises the linear frequency-modulated Gaussian $x_1(t)$ that
Chapter 3. Time Adaptation of the STFT

has a steep TF slope (wide-band component), the constant frequency-modulated Gaussian \( x_2(t) \), and the sinusoid (narrow-band component) \( x_3(t) \):

\[
x_1(t) = \begin{cases} 
2.7e^{-\frac{9(t-102)^2}{6272} + j\frac{2\pi 1000(t-70)^2}{511^2}} & 75 \leq t \leq 130 \\
0 & \text{otherwise}
\end{cases}
\]

\[
x_2(t) = 1.4e^{-\frac{(t-185)^2}{338} + j\frac{2\pi 125t}{511}} & 0 \leq t \leq 511
\]

\[
x_3(t) = \begin{cases} 
e^{-\frac{(t-344)^2}{8000} + j\frac{2\pi 125t}{511}} & 245 \leq t \leq 445 \\
0 & \text{otherwise}
\end{cases}
\]

As shown in Fig.3.5, the proposed method exhibits good TF concentration. The LCM-based approach also achieves good concentration in general; however, it is not well optimized at points in time where neighboring signal elements fall inside the same concentration measurement interval. For example, the Gaussian appears smeared in time near the sinusoid, while it expands frequency-wise on its side closer to the wideband linear chirp. The performance of the IFG-based algorithm is adequate overall, although the linear chirp appears slightly distorted. The two quadratic methods offer low resolution due to existence of cross-terms.

![Graphs](image-url)
Second Example: Time-Overlapping Signal Terms. The signal considered here (Fig. 3.6a) consists of the sinusoid \( s_1(t) \), the quadratic chirp \( s_2(t) \), and the constant frequency-modulated Gaussian \( s_3(t) \), with the three components overlapping in time. Their exact expressions are as follows:

\[
\begin{align*}
    s_1(t) &= \begin{cases} 
        0.5e^{-\frac{(t-250)^2}{125000}} + \frac{2\pi t}{511}, & 100 \leq t \leq 400 \\
        0, & \text{otherwise}
    \end{cases} \\
    s_2(t) &= \begin{cases} 
        0.8e^{-\frac{9(t-250)^2}{781250}} - \frac{\pi 4000(t-250)^3}{3 \times 511^3} + \frac{2\pi 70(t-250)}{511}, & 125 \leq t \leq 375 \\
        0, & \text{otherwise}
    \end{cases} \\
    s_3(t) &= e^{-\frac{(t-250)^2}{450}} + \frac{2\pi 70(t-250)}{511}, & 200 \leq t \leq 300
\end{align*}
\]  

(3.26)

Although the TF analysis of time-overlapping components is a difficult task in
general for time-adaptive-only spectrograms [6], the proposed method provides adequate overall performance. The IFG-based algorithm yields a moderate result, whereas the LCM-based approach cannot deal with this signal. As expected, the S-method and the modified-B distribution deliver higher signal localization at the expense of also admitting interference into the representation.

Fig. 3.6 [Example 2] (a) The test signal. (b) The proposed method. (c) The LCM-based method. (d) The IFG-based method. (e) The S-method (L=8). (f) The Modified B distribution (beta=0.07).

Third Example: Intersecting Signal Terms. The scenario in the third example
involves four components – two sinusoids and two pulses – arranged in a grid formation with four intersection points (Fig. 3.7). The equation of the overall signal is:

\[ y(t) = 0.7e^{j\frac{2\pi t}{511}} + 0.7e^{j\frac{2\pi t}{175}} + 15\delta(t - 192) + 15\delta(t - 315). \] (3.27)

The proposed approach can handle this type of signal sufficiently well. The alternative spectrograms fail to adequately represent the signal in the neighborhoods of the pulses since they cannot properly adjust their windows in those areas. As seen in Fig. 3.7c and Fig. 3.7d, parts of the signal appear missing as their excessive spreading has suppressed their energy below the minimum used contour level of -10dB. This is a challenging test signal for the quadratic representations as well because of the significant amount of generated interference (Figs. 3.7e,f).
Fourth Example: Intersecting Time-Overlapping Signal Terms. The fourth example is concerned with a nonlinear (sinusoidal) frequency-modulated component which intersects with a sinusoid at multiple points in the TF plane,

\[ z(t) = e^{j\pi 16 \sin\left(\frac{\pi t}{511}\right)}+e^{j \frac{2\pi 125t}{511}} + 0.5e^{j \frac{2\pi 125t}{511}}. \]  

(3.28)

Again, the proposed scheme yields a highly readable TF representation of \( z(t) \) (Fig. 3.8b). This is not the case for either of the quadratic methods whose superior TF concentration is offset by the increased levels of interference. The LCM-based adaptation causes the overspreading of parts of the nonlinear component because it optimizes the performance for the sinusoid which is the largest-energy component within the corresponding measurement windows. Finally, the IFG-based algorithm generates a somewhat disfigured representation.
Fifth Example: Digitized Bat Echolocation. The signal pulse is emitted by the Large Brown Bat, Eptesicus Fuscus. The signal consists of 400 samples and the sampling period was 7 microseconds. The results obtained after using the three time-adaptation techniques are shown in Fig. 3.9. Visual inspection of the three spectrograms suggests that their performance is similar in this case. The SM and the MBD on the other hand deliver higher TF concentration at the expense of also admitting cross-terms into the representations in Fig. 3.9e,f.
Chapter 3. Time Adaptation of the STFT

Fig. 3.9 [Example 5] (a) The experimental bat signal. (b) The proposed method. (c) The LCM-based method. (d) The IFG-based method. (e) The S-method (L=9). (f) The Modified B distribution (beta=0.1).

3.7 An Additional Processing Step: Separation of Time-Overlapping Signal Components

A common shortcoming of the methods in [5]-[8], as well as the proposed scheme, is that since the optimisation of the STFT is performed only with respect to time, the window size will match the average signal content along the frequency direction. Therefore, if at a time instant there exist more than one signal components then – even with an adjusted window – the STFT may not represent all of them equally
well. To deal with this problem, we introduce an additional processing step which consists in dividing the signal into parts possessing a single frequency element at each point in time. The STFT is then adapted separately for each component, and the TF representation of the overall signal can be obtained as the superposition of the optimised STFTs of the individual components.

To separate the time-overlapping components of a given signal we first transform it in the joint TF domain in order to unfold its frequency content at each time instant. We utilise the STFT because of its simplicity and the ease with which the isolated TF components can be reconstructed back into time. At this stage, a long window is employed for the analysis since our goal is to maximally resolve the different components along the frequency direction. Once the STFT is obtained, a threshold is applied to it in order to form a TF mask. Any value below some predetermined level is reduced to zero, whereas all other values are set to one. The produced mask is subsequently divided into sub-masks along the frequency direction so that each sub-mask contains only one compact non-zero segment at each point in time. Finally, each sub-mask is multiplied with the STFT to isolate the underlying component which is then synthesised into the time domain by means of (2.10). The process is repeated for all identified components.

3.7.1 Experimental Results

In cases of time-overlapping auto-terms in the second example in the previous section, the analysis can be enhanced by incorporating the abovementioned simple
signal separation procedure. Fig. 3.10 illustrates the three generated STFT sub-masks on the left column, and the corresponding synthesized components on the right one. It should be noted that the three waveforms were precisely recovered. The reconstruction process is also effective under noisy conditions as it will be the case in the following examples. This is because the synthesis takes place within the particular areas enclosing each component while parts of the noise energy normally spread outside these regions. The final result (Fig. 3.10g) is clearly superior to those in Fig. 3.6.
Fig. 3.10 Signal separation into mono-component elements: (a), (c) and (e) show the generated STFT sub-masks, (b), (d) and (f) show the associated recovered components; (g) the superposition of the corresponding three independently adapted spectrograms.

We also deploy the same bat echolocation signal in the fifth example in the previous section. The results obtained after using the proposed separation techniques are shown in Fig.3.11a. The signal separation step was originally introduced with the aim of improving the time adaptation of the spectrogram in cases of multiple auto-terms overlapping in time. Compared with Fig. 3.5b, although Fig. 3.11a have not shown extra improvement in resolution since all the components in the bat signal are chirp-like components with similar rate of changes in instantaneous frequencies, an additional advantage of this process is that it allows the user to further manipulate the magnitudes of the isolated components according to his application needs. For
instance, the upper two components of the bat signal are weaker than the lower ones, and as a result they are not clearly visible in TF plots (e.g. Fig. 3.11a). To alleviate this, one could equalize the magnitudes of the corresponding spectrograms before superimposing them (Fig. 3.11b) to produce a potentially more useful representation for the researcher.

![Fig. 3.11 The spectrogram of the bat signal according to the proposed method (a) with signal separation and (b) after equalization of the energy of its four components.](image)

### 3.8 Summary

In this chapter, a new scheme for the time adaptation of the STFT has been introduced. The aim is to automatically determine the size of the analysis window at each time instant in order to improve on the resolution of the conventional spectrogram. The proposed method is simple, does not require a priori knowledge of the signal or any user intervention, and admits of efficient implementation. The proposed scheme exploits the simple geometric relationship between the locations of the signal components and the interference terms in the WD in order to detect the support of the signal in the TF plane. We also present an equivalent expression to detect the support by trimming a fixed-window spectrogram. Then, drawing on the
acquired knowledge of the signal support it is straightforward to search for the window size which – at each time instant – maximizes the energy of the spectrogram within the prescribed areas. Experimentation has shown that the proposed method is able to achieve good TF resolution, and can outperform previously reported time-adaptive spectrograms. In addition, the method has also compared favorably with two reputable quadratic TF representations. Finally, we developed an additional processing step to deal with signals with components overlapping in time, and this step deployed the STFT analysis-modification-synthesis method to separate the signal components overlapping in time and superimposed them to form the resulting improved representation.

REFERENCES


CHAPTER 4

TIME-FREQUENCY ADAPTATION OF THE STFT

4.1 Introduction

A considerable amount of research has been carried out with the aim of developing modified versions of the WD with suppressed cross-terms. Many of these methods were concerned with the design of interference-attenuating kernels in [2]-[6]. The common shortcoming of the above methods is that the attenuation of interference generally comes at the cost of increasing the TF spread of the auto-terms, which reduces the accuracy of the representation.

To improve the concentration properties of the STFT, its analysis window should be matched with the analyzed signal. Gaussian-type windows are commonly chosen for computing the STFT since the Gaussian function is the most concentrated
in TF. However, remains as to how the width and orientation of the window could be
designed to match the analyzed signal. Early attempts to deal with the above
problem mainly relied on the user to manually set the required parameters based on
his/her prior knowledge of the signal. Later, one of the main developments to choose
the most suitable window for the STFT was to utilize the multi-window approach
[7-15], which was based on an idea pioneered by Thomson for spectrum estimation
[16]. The first multi-window method for the STFT [7] was proposed in 1994 of
which prolate spheroidal sequences were chosen to be the windows of the STFT, and
the resulting representation was the weighted sum of the corresponding
spectrograms. Two years later, a similar approach but using Hermite functions as the
window set was proposed [8-9] because of their optimal TF concentration in the TF
domain. Then, determination of the optimal weights for the multi-window
spectrogram based on various constraints was found in [10-12] such as the constraint
of minimizing: (a) the instantaneous bandwidth, and (b) the instantaneous frequency
of the spectrogram [12]. Recently, the method combining the multi-window
spectrogram and the S-method was introduced to deal with polynomial phase signals
with multiple signal components [14-15].

The other main development for seeking the appropriate window for the STFT
is to use the data-adaptive method. In the early nineties, data-adaptive STFT
methods were developed in which the analysis window could be determined in an
automatic way. Notable examples of such schemes are the works in [17]-[23]. The
corresponding methodologies can broadly be classified into two categories; those
based on the conventional STFT definition where the duration of the analysis window is usually the sole parameter optimised in time in [17]-[21], and those introducing modified versions of the STFT in which the duration and the orientation of the window can be adjusted at every time-frequency location in [22]-[23]. In the previous chapter, we have proposed a novel method for the time-adaptive STFT in which we have shown that our method outperforms other existing time-adaptive approaches in the experimentation. However, adaptation of the STFT only in time restricts the method to signals containing a single frequency component at each time instant. In addition, it cannot provide good concentration for chirps or non-linearly FM signals. However, signals in many natural phenomena such as radar signals [24] and seismic signals [25] can all be modeled as chirp-like functions and most likely with multiple components in time.

In this chapter, we will thus focus on the adaptation of the STFT in both time and frequency. The algorithms in [22]-[23] iteratively find both window duration and orientation at each TF point from a set of finite predetermined windows so as to, in some sense, optimally match the window with the signal. In [22], a window matching algorithm was presented by locally adapting the window functions through measuring the local moments of the spectrogram. Another adaptation algorithm was built in [23] using the measure of the local signal concentration in the TF plane. Both literatures provide descriptions that the TF concentration of the spectrogram depends on matching the chirp rate of the components with the window. Although both methods provide fully automated determination of the optimal window
parameters, these methods are very expensive computationally and may be ineffective if the set of predetermined windows is not large enough. To overcome this problem, a novel adaptation scheme is developed in this chapter, which is not only considerably fast but also able to optimally find the window duration in an analytical way. As a result, our method needs only iteratively search for the most favourable window orientation for the time-frequency adaptive STFT.

4.2 The Short-Time Fourier Transform and its Variations

The short-time Fourier transform (STFT) of a signal $x(t)$ can be obtained as [26]

$$X_{STFT}(t, \omega) = \int x(\tau)g^*(t - \tau)e^{-j\omega \tau}d\tau.$$  

The operator $*$ represents the complex conjugate. Hence, at each time instant $t$, it computes the Fourier transform of a signal portion around $t$. This typical definition of the STFT, the analysis window is predetermined by the user and remains unchanged for all values of $t$.

Another point of view of the STFT by Gabor [27] is the inner product or projection of the signal with the window function $g(\tau - t)e^{j\omega \tau}$ concentrated around the time-frequency location $(t, \omega)$. Ideally, to achieve the highest localized time-frequency representation, the projection function should be an impulse in TF. However, the uncertainty principle prohibits the existence of a signal whose energy can be concentrated in a single TF point [27]. In other words, the time duration and the bandwidth of a signal cannot be both arbitrarily small. According to the same principle, the Gaussian signal exhibits minimum uncertainty in the TF plane, and
implies that the time-shifted and frequency-modulated Gaussian functions would be the most suitable basis for the STFT. Let us consider, for the sake of argument, the unit-energy time-shifted and frequency-modulated Gaussian window, with \( \alpha \) determining its time-width,

\[
g_{t,\omega,\alpha}(\tau) = \sqrt{\frac{\alpha}{\pi}} e^{-\frac{\alpha}{2} (\tau-t)^2 + j\omega \tau}.
\]  

(4.1)

The STFT can then be expressed as the inner product of the signal and the window with fixed \( \alpha_o \) for all time and frequency:

\[
X_{STFT}(t, \omega) = \langle g_{t,\omega,\alpha_o}(\tau), x(\tau) \rangle
\]

(4.2)

The magnitude square of the STFT known as the spectrogram can be written as

\[
|X_{STFT}(t, \omega)|^2 = |\langle g_{t,\omega,\alpha_o}(\tau), x(\tau) \rangle|^2.
\]  

(4.3)

Mann and Haykin [28] in 1995 generalized Gabor’s use of the Gaussian function for the STFT and introduced the use of the Gaussian chirplet as the basis function, in which the modified basis function keeps the same Gaussian envelope as in (4.3) and replaces the modulation by a harmonic oscillation with a linear frequency-modulated chirp:
Chapter 4. Time-Frequency Adaptation of the STFT

\[ g_{t,\omega,\alpha,\beta}(\tau) = \sqrt{4\pi} \frac{(-\alpha+j\beta)}{\tau-\omega} e^{\frac{(-\alpha+j\beta)(\tau-\omega)^2}{2}}. \]  

(4.4)

This generalization, instead of keeping \( \alpha \) constant in (4.4), allows \( \alpha \) and \( \beta \) to be parameters so that the width and the chirp-rate of the basis function both vary. The new mapping becomes a mapping from the one-dimensional function (time) to a four-dimensional function (time, frequency, width and chirp rate), and it is referred to the Gaussian chirplet transform (GCT) in [28]. i.e.

\[ X_{GCT}(t, \omega, \alpha, \beta) = \langle g_{t,\omega,\alpha,\beta}(\tau), x(\tau) \rangle. \]  

(4.5)

The authors in [28] also interpret this four-dimensional parameter space whose coordinate axes correspond to the affine coordinate transforms in the time-frequency plane. Such that the parameter \( t \) corresponds to time-translation, the parameter \( f \) corresponds to frequency-translation, the parameter \( \alpha \) corresponds to time-dilation or frequency-contraction and the parameter \( \beta \) corresponds to frequency-shearing in the time-frequency plane. For a situation using both fixed \( \alpha_o \) and \( \beta_o \) in (4.5), it becomes the conventional STFT using a linear Gaussian-enveloped chirp and it can be also thought of as a two-dimensional plane in the four-dimensional parameter space of the GCT such that

\[ X_{STFT}(t, \omega) = \langle g_{t,\omega,\alpha_o,\beta_o}(\tau), x(\tau) \rangle. \]  

(4.6)
Chapter 4. Time-Frequency Adaptation of the STFT

The fact that the different windows are appropriate for different signal components in the TF plane, suggests the use of a signal-dependent, time-and-frequency varying window. That means the parameters $\alpha_o$ and $\beta_o$ should be properly chosen at each TF location. The resulting transformation we refer to the time-frequency adaptive STFT:

$$X_{TFA-STFT}(t, \omega) = \langle g_{t,\omega,\alpha(t,\omega),\beta(\tau)}(\tau), x(\tau) \rangle$$

$$= \int x(\tau)g^*(\tau)d\tau$$

$$= \left(\frac{\alpha(t,\omega)}{\pi}\right)^{\frac{1}{2}} \int x(\tau)e^{-\frac{[\alpha(t,\omega)+j\beta(t,\omega)](\tau-t)^2-j\omega\tau}{2}}d\tau. \quad (4.7)$$

This adaptive time-frequency representation with the basis function $g_{t,\omega,\alpha(t,\omega),\beta(\tau)}(\tau)$, which has one more degree of freedom about frequency-shearing in the time-frequency plane than the conventional STFT, is able to give a potentially better projection between the basis function and various kinds of signals. The performance of the time-frequency adaptive STFT lies in the choice of the parameters $\alpha$ and $\beta$.

4.3 The Global Duration-Bandwidth Product for the STFT and its Variations

According to [29], the global duration and the global bandwidth of the spectrogram computed by the STFT in (4.3) have been proved to be
Chapter 4. Time-Frequency Adaptation of the STFT

\[ T_{SP} = \left[ \frac{\int (t-\langle t \rangle)^2 |X_{STFT}(t,\omega)|^2 dt d\omega/2\pi}{\int |X_{STFT}(t,\omega)|^2 dt d\omega/2\pi} \right]^{1/2} \]  \hspace{1cm} (4.8)

and

\[ B_{SP} = \left[ \frac{\int (\omega-\langle \omega \rangle)^2 |X_{STFT}(t,\omega)|^2 dt d\omega/2\pi}{\int |X_{STFT}(t,\omega)|^2 dt d\omega/2\pi} \right]^{1/2} \]. \hspace{1cm} (4.9)

\( \langle t \rangle \) and \( \langle \omega \rangle \) are the mean time and the mean frequency of a signal \( x(t) \). The global duration-bandwidth product (GDBP) of the spectrogram is then given by [30]

\[ \text{GDBP}_{SP(\alpha)} = T_{SP} \times B_{SP} = \sqrt{T_x^2 + \frac{1}{2\alpha}} (B_x^2 + \frac{\alpha}{2}) \]  \hspace{1cm} (4.10)

\( T_x \) and \( B_x \) are the duration and the frequency bandwidth of \( x(t) \) respectively, which are defined as:

\[ T_x = \frac{\int (t-\langle t \rangle)^2 |x(t)|^2 dt}{\|x\|^2} \] \hspace{1cm} \text{and} \hspace{1cm} \[ B_x = \frac{\int (\omega-\langle \omega \rangle)^2 |X(\omega)|^2 d\omega/2\pi}{\|x\|^2} \], \hspace{1cm} (4.11)

where

\[ \langle t \rangle = \frac{\int t |x(t)|^2 dt}{\|x\|^2} \] \hspace{1cm} \text{and} \hspace{1cm} \[ \langle \omega \rangle = \frac{\int \omega |X(\omega)|^2 d\omega/2\pi}{\|x\|^2} \]. \hspace{1cm} (4.12)

\( X(\omega) \) is the Fourier transform of \( x(t) \). Considering minimizing the global duration-bandwidth product, for a mono-component and stationary signal, the optimal \( \alpha \) of the Gaussian window can therefore be obtained as [29], [30]:

\[ \alpha_{opt} = \frac{B_x}{T_x} \] \hspace{1cm} (4.13)

and the corresponding optimal Gaussian window is
Chapter 4. Time-Frequency Adaptation of the STFT

\[
g_{opt}(t) = \frac{B_x}{\pi T_x} e^{\frac{-B_x t^2}{2 T_x^2}}. \quad (4.14)
\]

However, this well-defined and convenient optimal STFT window is inadequate to analyze non-linear signals or multi-component signals.

Similarly, following the definition of the GDBP in (4.10), the GDBP of the spectrogram computed by the STFT in (4.6) can be written as:

\[
GDBP_{SP(\alpha,\beta)} = \sqrt{(T_x^2 + \frac{1}{2\alpha})(B_x^2 + \frac{\alpha}{2} + \frac{\beta^2}{2\alpha})}. \quad (4.15)
\]

The detailed proof of (4.15) can be found in APPENDIX A. According to the above definition, the chirp rate of the window \( \beta \) always contributes to the GDBP in (4.15) when \( \beta \) is not equal to zero. This implies that when considering minimizing the GDBP, the optimal window of the STFT in (4.15) should not have any chirp rate (i.e. \( \beta \) should be equal to zero). However, this is only true for mono-component and stationary signals. Let us consider the following example of the spectrogram of a linear chirp signal which is analyzed with two STFTs. The first spectrogram uses a chirp window with \( \alpha \) which is optimally determined and \( \beta \) which is equal to zero. (See Fig. 4.1a) The second one employs a chirp window with \( \alpha \) is same as that of the first spectrogram and \( \beta \) which matches the chirp rate of the input signal. (See Fig. 4.1b) Although the GDBP of the first spectrogram is smaller than that of the second one, it is obviously that the second spectrogram gives a more localized and concentrated time-frequency representation. This example illustrates the deficit of
the GDBP for characterizing non-stationarities of signals, and thus a more
generalized measure is needed to be formulated in order to describe the
non-stationary characteristics of any arbitrary signals. Another problem with the
global measure is that for signals composed of several different components, this
global measure is unable to portray each component. The local duration-bandwidth
product (LDBP) for time-frequency distributions was introduced in [31] in 2004. We
will later make use of this local measure in this work to tackle the above problems of
the global measure.

Fig. 4.1 The spectrogram of a linear chirp with (a) optimally chosen $\alpha$ and $\beta$ equal to zero,
and (b) with the same $\alpha$ and $\beta$ matched the chirp rate of the linear chirp.

The local duration-bandwidth product (LDBP) for time-frequency distributions
is defined as [31]:

$$\langle \sigma_{\omega|t}^2 \rangle \langle \sigma_{t|\omega}^2 \rangle = \int \sigma_{\omega|t}^2 P(t) dt \times \int \sigma_{t|\omega}^2 P(\omega) d\omega$$  \hspace{1cm} (4.16)$$

where $P(t)$ and $P(\omega)$ are the time and frequency marginals of a time-frequency
distribution, respectively. $\sigma_{\omega|t}^2$ and $\sigma_{t|\omega}^2$ are the frequency variances at time $t$ and
the time variances at frequency \( \omega \), respectively, and they are defined in (4.21) and (4.22). Although the global duration-bandwidth product or commonly known as the standard uncertainty product is upper-bounded by 1/2, it has been proven that the LDBP is always less than or equal to the standard uncertainty product, and for some signals, the LDBP can be arbitrarily small for a large class of time-frequency distributions such as the WD, the Choi-Williams distribution and the reduced interference distribution [31]. For the classical fixed-window STFT, the LDBP cannot approach an arbitrarily small value since the windowing operation places a limitation to this technique. In this work, we deploy this local measure which is able to portray the local characteristics of a signal to search for the optimal window parameters for the time-frequency adaptive STFT so as to generate a high resolution TF representation.

### 4.4 The Local Duration-Bandwidth Product for the Time-Frequency Adaptive STFT

In the previous section, we addressed the problems with the measure by the GDBP to portray non-stationary or local characteristics of a linear chirp signal. To solve the deficit of the global measure, the LDBP is deployed and the LDBP for the time-frequency adaptive STFT using the Gaussian linear chirp as its basis function is going to be formulated in this section.

This work is based on Cohen’s work in [1] and the results obtained are revealing when the signal and the window of the spectrogram are expressed in terms
of their amplitudes and phases. The notations of the signal and the window used to formulate the measure are

\[ x(t) = A(t)e^{j\varphi(t)} \quad \text{and} \quad g(t) = A_g(t)e^{j\varphi_g(t)} , \quad (4.17) \]

and the signal is assumed to be a mono-component linear chirp (i.e. \( \varphi(t) \) is a quadratic function.) Their Fourier transform pairs are respectively expressed as

\[ X(\omega) = B(\omega)e^{j\varphi(\omega)} \quad \text{and} \quad G(\omega) = B_g(\omega)e^{j\varphi_g(\omega)} . \quad (4.18) \]

The corresponding spectrogram is defined as:

\[ |X(t, \omega)|^2 = \left| \int A(\tau)A_g(\tau - t)e^{j\varphi(\tau) + j\varphi_g(\tau - t) - j\omega\tau} d\tau \right|^2 . \quad (4.19) \]

The spectrogram may also be expressed in terms of the Fourier transform of the signal and the window:

\[ |X(t, \omega)|^2 = \left| \frac{1}{2\pi} e^{-j\omega t} \int B(\omega')B_g(\omega' - \omega)e^{j\varphi(\omega') + j\varphi_g(\omega' - \omega)} dt \right|^2 . \quad (4.20) \]

Following the similar way as the global measure of the spectrogram defined in (4.8) and (4.9), instead of measuring the time duration and the frequency bandwidth in the whole time-frequency plane, the LDBP considers one-dimensional slices in the time-frequency plane at specific times \( t_o \) or frequencies \( \omega_o \) to be
Chapter 4. Time-Frequency Adaptation of the STFT

\[ \sigma_{t|\omega_o}^2 = \frac{\int (t-(t)_{\omega_o})^2 |X(t,\omega_o)|^2 dt}{\int B^2(\omega)B_\omega^2(\omega-\omega_o) d\omega} \]  

(4.21)

and

\[ \sigma_{\omega|t_o}^2 = \frac{\int (\omega-(\omega)_{t_o})^2 |X(t_o,\omega)|^2 d\omega}{\int A^2(\tau)A_\omega^2(\tau-t_o) d\tau}. \]  

(4.22)

\( \sigma_{t|\omega_o}^2 \) and \( \sigma_{\omega|t_o}^2 \) are defined as the time variance of the spectrogram for a given frequency \( \omega_o \) and the frequency variance for a given time \( t_o \), respectively, while \( \langle t \rangle_{\omega_o} \) and \( \langle \omega \rangle_{t_o} \) are the mean time at frequency \( \omega_o \) and the mean frequency at time \( t_o \) of \( x(t) \), respectively. \( \sigma_{t|\omega_o}^2 \) and \( \sigma_{\omega|t_o}^2 \) could be explicitly expressed up to the second derivatives of the phases of the signal since the signal is assumed to be a linear chirp:

\[ \sigma_{t|\omega_o}^2 = \langle t^2 \rangle_{\omega_o}^0 + \{N_2 - N_1^2 \} \left\{ \varphi''(\omega_o) + \varphi''(0) \right\}^2 \]  

(4.23)

and

\[ \sigma_{\omega|t_o}^2 = \langle \omega^2 \rangle_{t_o}^0 + \{M_2 - M_1^2 \} \left\{ \varphi''(t_o) + \varphi''(0) \right\}^2 \]  

(4.24)

where

\[ \langle t^2 \rangle_{\omega_o}^0 = \frac{\int \frac{d}{d\omega} B(\omega)B_\omega(\omega-\omega_o))^2 d\omega}{\int B^2(\omega)B_\omega^2(\omega-\omega_o) d\omega}, \quad \langle \omega^2 \rangle_{t_o}^0 = \frac{\int \frac{d}{d\tau} A^2(\tau)A_\omega^2(\tau-t_o))^2 d\tau}{\int A^2(\tau)A_\omega^2(\tau-t_o) d\tau} \]  

(4.25)

and

\[ N_n(\omega_o) = \frac{\int B^2(\omega)B_\omega^2(\omega-\omega_o) \omega^n d\omega}{\int B^2(\omega)B_\omega^2(\omega-\omega_o) d\omega}, \quad M_n(t_o) = \frac{\int A^2(\tau)A_\omega^2(\tau-t_o) \tau^n d\tau}{\int A^2(\tau)A_\omega^2(\tau-t_o) d\tau}. \]  

(4.26)

Therefore, the LDBP at time \( t_o \) and frequency \( \omega_o \) for the spectrogram can be defined as

112
4.5 Time-Frequency Adaptation of the Time-Frequency Adaptive STFT

4.5.1 Optimal Parameters by Minimizing the LDBP

As explained in the previous sections, we deploy a better measure called the LDBP to characterize the local and non-stationary behavior of a signal. Although the formulation in Section 4.4 is limited to mono-component and linear chirp signals for which the WD could simply give an optimal result without any modification, we will make use of this formulation in cases of multi-component signals. In this section, we explain how we determine the optimal parameters for the time–frequency adaptive STFT through minimizing its LDBP.

The local time variance and the local frequency variance of the spectrogram are analyzed and expressed in terms of the amplitudes and phases of the input signal and the analysis window in (4.23) and (4.24). Therefore, the correlation between the phase of the signal and that of the window is revealed. In (4.24), it can be observed that when the second derivative of the phase of the signal at time \( t_o \) equals to the
negative of that of the analysis window (positioned at the centre of the time domain),
the term \( \{ M_2 - M_1 \} \{ \varphi''(t_o) + \varphi''(0) \}^2 \) can be cancelled out. In a dual way,
when the second derivative of the phase of the signal at frequency \( \omega_o \) equals to the
negative of that of the analysis window at the centre in the frequency domain, the
term \( \{ N_2 - N_1 \} (\varphi''(\omega_o) + \varphi''(0))^2 \) can also be deleted in (4.23). Since both
terms are always positive and contribute to the LDBP, cancelling these two terms
minimizes the LDBP. If this is the case, the LDBP can be minimized to

\[
LDBP_{SP} = \int (t^2)_{t=0}^{\omega_o} P_{SP}(\omega_o) d\omega_o \times \int (\omega^2)_{\omega_o=0}^{\omega_o} P_{SP}(t_o) dt_o.
\] (4.28)

Clearly, the above minimization occurs when the orientation of the analysis window
at time \( t_o \) and frequency \( \omega_o \) in the time-frequency plane of the spectrogram
matches the orientation of the signal component at the same point, the LDBP can be
minimized to (4.28). The parameter \( \beta \) of the window in (4.4) controls the shearing
effect along the frequency axis of the window in the time-frequency plane, thus
parameter \( \beta \) controls the orientation of the window in the time-frequency plane.
Hence, when parameter \( \beta \) matches the orientation of the signal in the time-frequency
plane, the LTBP can be minimized to (4.28). The optimal \( \beta_{opt}(t_o) \) can be written as

\[
\beta_{opt}(t_o) = \varphi''(t_o)
\] (4.29)

After the parameter \( \beta \) have been optimally selected, the remaining problem is to
optimally determine the parameter \( \alpha \). The minimized LDBP in (4.28) can be further expanded into

\[
LDBP_{SP} = \int \left[ \frac{d}{d\tau} A(t + \tau) A_g(\tau) \right]^2 d\tau \int A^2(t + \tau) A_g^2(\tau) d\tau \, dt \times \\
\int \frac{d}{d\omega} B(\omega + \omega_o) B_g(\omega)^2 d\omega \int B^2(\omega + \omega_o) B_g^2(\omega) d\omega \, d\omega_o
\]

\[
= \int \left[ A'(t + \tau) A_g(\tau) + A(t + \tau) A' \right]^2 d\tau \, dt \times \\
\int \int \{ B'(\omega + \omega_o) B_g(\omega) + B(\omega + \omega_o) B_g'(\omega) \}^2 d\omega \, d\omega_o
\]

\[
= \int \{ A^1(t + \tau) A_g^2(\tau) + A^2(t + \tau) A_g' \} d\tau \, dt \times \\
\int \int \{ B^1(\omega + \omega_o) B_g^2(\omega) + B^2(\omega + \omega_o) B_g'^2(\omega) \} d\omega \, d\omega_o
\]

(4.30)

Since the window is a symmetric function then \( \int B_g(\omega) B_g'(\omega) d\omega = 0 \). This also applies for their time-domain counterparts, and the LDBP of the corresponding spectrogram can be simplified to

\[
LDBP_{SP} = \int \left[ A^1(t + \tau) A_g^2(\tau) + A^2(t + \tau) A_g' \right] d\tau \, dt \times \\
\int \int \{ B^1(\omega + \omega_o) B_g^2(\omega) + B^2(\omega + \omega_o) B_g'^2(\omega) \} d\omega \, d\omega_o
\]

(4.31)

As both of the signal and the window are normalized, we can apply \( \int B^2(\omega + \omega_o) d\omega_o = 1 \), and \( \int B_g^2 d\omega = 1 \), and these also apply for their time domain counterparts. Therefore,

\[
LDBP_{SP} = \left\{ \int A^2(t + \tau) dt + \int A_g^2(\tau) d\tau \right\} \times \left\{ \int B^2(\omega + \omega_o) d\omega_o + \int B_g^2(\omega) d\omega \right\}
\]
where $B_{x AM}$ and $T_{x AM}$ are known as the AM contributions of the signal $x(t)$ to the bandwidth and to the time duration according to [1]. When the signal is merely amplitude modulated or stationary, $B_{x AM}$ and $T_{x AM}$ would become close to the bandwidth of the signal $B_x$ and the time duration of the signal $T_x$, respectively. When the signal is highly frequency modulated or non-stationary, $B_{x AM}$ and $T_{x AM}$ would become very small and can be neglected.

In the following, we consider the orientation of a signal component in two different situations so as to optimize the parameter $\alpha$. One is the signal component having no orientation. (i.e. the signal component is stationary in time or frequency in nature.) The other is the signal component having orientation. (i.e. the signal component is non-stationary in nature.)

For stationary signals in time or frequency: When the signal component is stationary in time or frequency in nature, in other words, having no orientation in the time-frequency plane, $\beta_{opt}(t_o)$ should be zero. The terms in (4.32) $B_{x AM}$ and $T_{x AM}$ would become close to the bandwidth $B_x$ and the time duration $T_x$ of the signal. In this case, the LDBP can be shown as:

$$LDBP_{sp} \sim (B_x^2 + \frac{\alpha}{2}) \left( T_x^2 + \frac{1}{2 \alpha} \right).$$  \hspace{1cm} (4.33)

It can be seen that the LDBP in (4.33) is in the same form of the GTBP in (4.10), and thus the GTBP is one of the generalized case of the LDBP when the analyzed
signal is stationary in time or frequency. Therefore, the LDBP approaches its minimum if

\[ \alpha_{opt}(t_o) = \frac{B_x}{T_x} . \]  \hspace{1cm} (4.34)

For non-stationary linear chirp signals: When the signal component is non-stationary, \( \beta_{opt}(t_o) \) is not zero and should be equal to \( \varphi''(t_o) \). The terms in (34) \( B_{xAM} \) and \( T_{xAM} \) would become very small and can be omitted. In this case, the LTBP can be shown as

\[ LDBP_{sp} \sim \frac{a^2}{\sqrt{4(a^2 + \beta^2)}} . \]  \hspace{1cm} (4.35)

The LDBP now becomes proportional to the parameter \( a \) so the optimal \( \alpha_{opt}(t_o) \) should be chosen as small as possible to minimize the LDBP. (i.e. the optimal width of the analysis window should be long.)

\subsection*{4.5.2 Estimation of the Optimal Parameters}

From the above results, in order to obtain the optimal parameters for the corresponding STFT, some knowledge of the analyzed signal should be determined in advance. This knowledge includes the rate of change of the local frequency, and the ratio between the bandwidth and the time duration of the analyzed signal. When the rate of change of the local frequency of the signal is zero, \( \beta_{opt}(t_o) \) should also
set to zero and $\alpha_{opt}(t_o)$ should be the ratio between the bandwidth and the time duration of the analyzed signal. When the rate of change of the local frequency is not zero, $\beta_{opt}(t_o)$ should be equal to the rate of the signal and $\alpha_{opt}(t_o)$ should be set as small as possible. For the rate of change of instantaneous frequency at each time, one approach is to determine the instantaneous frequency of the signal and calculate its derivative. There has been substantial research on the instantaneous frequency (e.g. [32], [33] and references therein). However, the noise corrupted in the signal will be significantly amplified by the differentiation operator and thus the resulting calculated rate of change of instantaneous frequency is not very useful. We adopt an iterative approach to match the rate of change of the instantaneous frequency of a signal. On the other hand, it is not hard to determine the ratio between the bandwidth and the time duration of a mono-component signal using (4.13). To apply our method to multi-component signals, the prior knowledge about the signal mentioned above is determined by a mask generated from the conventional fixed-width STFT in (3.19) using a circular Gaussian window as explained in Chapter 3. The TF locations of each component are captured by the mask, and thus, we can estimate the optimal parameters for each component within this mask. The detailed process is presented by using an illustrative example in Fig. 4.2.

The signal in Fig. 4.2a has three components which are a sinusoid, a constant frequency-modulated Gaussian, and a linear chirp. The signal is first analyzed by the fixed-width STFT and the discrete STFT is defined as [34]
\[ X_{STFT}(nT, k\Omega) = \frac{1}{2\pi} \int x(\tau) g(\tau - nT)e^{-jk\Omega \tau} d\tau , \quad (4.36) \]

where \( n \) and \( k \) are integers, and \( T \) and \( \Omega \) are the sampling intervals of time and frequency. \( g(t) \) is a circular Gaussian window. Next, a low threshold \( p \) is applied to the output \( X_{STFT}(nT, m\Omega) \) in order to produce \( S[n, k] \) whose support reflects the shape of the signal components (in Fig. 4.2b). To automate the choice of \( p \) here, we have relied on the optimum global thresholding approach known as Otsu’s method [35] so that the signal components can be divided from the background in the TF plane, i.e.

\[ S[n, k] = \begin{cases} 1, & X_{STFT}(nT, m\Omega) \geq pX_{STFT}(nT, m\Omega) \\ 0, & \text{otherwise} \end{cases} . \quad (4.37) \]

Different components overlapping in time in the mask \( S[n, k] \) are separated by zeros. The optimal \( \beta_{opt}[n, k] \) for each of the components is iteratively adapted at each time using the time-adaptive STFT. The time-adaptive STFT can be expressed as follows,

\[ X[n, k; g_{\beta[n]}] = \sum_{-\infty}^{\infty} x[m] g_{\beta[n]}[n - m]e^{-j\frac{2\pi k}{M}m} , \quad (4.38) \]

and its peak value within \( S[n, k] \) at time \( n \) is:
Chapter 4. Time-Frequency Adaptation of the STFT

\[ \text{peak}[n; \beta] = \max(|X[n, k; g_{\beta}]|^2 S[n, k]) \quad \text{for } k_1 \leq k \leq k_2 \] (4.39)

Therefore, the objective is to iteratively find the estimated optimum orientation \( \beta_{\text{opt}}[n, k] \) such that,

\[ \beta_{\text{opt}}[n, k] = \arg \max_{\beta} \text{peak}[n; \beta] \quad \text{for } k_1 \leq k \leq k_2 , \] (4.40)

where \( k_1 \) and \( k_2 \) define the range of frequencies of a component in time \( n \). The estimated optimal parameter \( \beta_{\text{opt}} \) is assigned to each time-frequency point within the support of the mask \( S[n, k] \) corresponding to its signal component. The process involves the computation of fixed-window STFTs (trial STFTs) for different orientations \( \beta \). The time-frequency points outside the support of the mask are assigned \( \beta_{\text{opt}} \) equal to zero. The resulting matrix is the estimated optimal parameter \( \beta_{\text{opt}}[n, k] \) for each time-frequency point of the signal. (See Fig. 4.2c.)

According to the result in Section 4.5.1, when \( \beta_{\text{opt}} \) is not equal to zero, \( \alpha_{\text{opt}} \) should be chosen as small as possible to minimize the LDBP. For the example in Fig 4.2, only the signal component of the linear chirp in Fig.2e has \( \beta_{\text{opt}} \) which is not equal to zero. So, the \( \alpha_{\text{opt}} \) within the support of the linear chirp in the mask \( S[n, k] \) is assigned to be the small one. (\( \alpha_{\text{opt}} = 2 \) is assigned in this example, to which the effective Gaussian width is equal to half length of the whole signal.). For the case \( \beta_{\text{opt}} \) is equal to zero, \( \alpha_{\text{opt}} \) should be chosen to be the ratio between the bandwidth and the time duration of the signal. Hence, we estimate the parameter \( \alpha_{\text{opt}} \) by
\[ \alpha_{opt}[n_o, k_o] = \frac{\sum_{k_1}^{k_2} s[n_o, k]}{\sum_{n_1}^{n_2} s[n, k_o]} \]  

(4.41)

where \( k_1 \) and \( k_2 \) is the frequency range at \( n_o \) of the mask, and \( n_1 \) and \( n_2 \) is the time range at \( k_o \)- the mask contains ones over a signal component, and zeros otherwise.

The time-frequency points outside the support of the mask are assigned an \( \alpha_{opt} \) equal to that of the circular Gaussian. The resulting matrix is the estimated optimal parameter \( \alpha_{opt}[n, k] \) for each time-frequency point of the signal. (See Fig. 4.2d.)

Finally, the optimal spectrogram (in Fig. 4.2e) can be computed by the discrete time-frequency adaptive STFT using the estimated optimal parameters \( \alpha_{opt}[n, k] \) and \( \beta_{opt}[n, m] \) as:

\[ |X(nT, k\Omega)|^2 = \left| \left( \frac{\alpha_{opt}[n,k]}{8\pi^5} \right) \frac{1}{\int x(\tau) e^{-[\alpha_{opt}[n,k]+j\beta_{opt}[n,k]](\tau-nT)^2-jk\Omega t} dt \right|^2. \]  

(4.42)
4.5.3 Experimental Results

In this section we present four illustrative examples to demonstrate the performance of the proposed algorithm. These include simulated test signals typically employed in similar studies, as well as a real-life bat echolocation sound recording. Furthermore, we have also compared the proposed scheme with the TF adaptation scheme using the concentration measure (CM) in [22], three time-adaptive STFT representations described in the previous chapter [19]-[21], the multi-window spectrogram using the optimal weights proposed in [12], two advanced quadratic TF representations, namely the S-method (SM) [4], the modified-B distribution (MBD) [5]. The parameters for the SM and the MBD were manually tuned to achieve the
best representations according to the visual inspection of the resulting TF plots. It should be stressed that in none of the following examples was it possible for the quadratic representations to completely suppress the cross-terms without severely compromising the overall concentration. This effect was more pronounced on the wide-band portions of the signals, which would undergo excessive smearing along the time direction. We consider the resulting representations, normalized so that the maximum value corresponds to 1. All contour plots depict a range of values from -1 up to 1 at 2dB intervals.

First Example: The signal shown in Fig. 4.3 is composed of four signal components: one frequency modulated sinusoid, one Gaussian atom, and two linear chirps used. They are placed close to each other and overlapping in time in the TF plane. The superposition of WDs of each component is given in Fig. 4.3a. which provides a reference TF representation for comparison with the results. The representation of the input signal based on the proposed TF-adaptive STFT is shown in Fig. 4.3b. It is clear that the signal appears well-concentrated as the WD without any cross-terms. On the other hand, the other TF adaptive scheme using CM in Fig. 4.3c fails in this example and misses three components causing by incorrect adaptation affected by the nearby stronger components. As mentioned in the previous chapter, the T-adaptive STFT is not designed for signals overlapping in time, and particularly in this example, the three above-mentioned T-adaptive STFTs cannot resolve this signal in Fig. 4.3d-f. The MWSP, the SM and the MBD on the other hand deliver higher TF concentration at the expense of also admitting
cross-terms into the representations in Fig. 4.3g-i.
Second Example: This example is composed of two linear chirps intersecting with each other designed to form a cross pattern. The T-adaptive STFTs in Fig. 4.4d-e are particularly able to yield highly readable TF representations; nevertheless, since the window width is the only parameter in the adaption, the resulting TF concentration is much less than that of the WD for chirps or non-linearly modulated signals. On the other hand, the proposed TF-adaptive STFT can give high TF concentration for chirps and appear no cross-terms even in multi-component cases, which can be shown in this example in Fig. 4.4b. Although the TF adaptive STFT by CM gives comparable good result in Fig. 4.4c, it is emphasized in [22] that this method is expansive computationally, and it can be seen in other examples in this section that the adaptation is adversely affected by nearby strong components. The IFG-based method somehow generates a disfigured representation in Fig. 4.4f. The MWSP and the two quadratic methods offer low resolution due to existence of cross-terms in Fig. 4.4g-i.
Chapter 4. Time-Frequency Adaptation of the STFT

(a)

(b) (c)

(d) (e)

(f) (g)
Third Example: The waveform used in the previous chapter in Section 3.6.1 (Fig. 3.8a) is also adopted as the test signal in this example. The signal contains a nonlinear (sinusoidal) frequency-modulated component intersecting with a sinusoid. Again, the proposed scheme yields a highly concentrated representation without any cross-terms (Fig. 4.5b). Although the proposed T-adaptive STFT in Fig. 4.5d provides a less concentrated representation, its readability is much better than that of the other TF adaptive STFT and T-adaptive STFTs. The significant amount of generated interference in the MWSP and the quadratic representations makes them difficult from interpreting.
Fig. 4.5 [Example 3] (a) Superposition of the WDs of individual signal components. (b) The proposed TF adaptive STFT. (c) The TF adaptive STFT by CM. (d) The proposed Time adaptive STFT. (e) The LCM-based method. (f) The IFG-based method. (g) The MWSP (8 windows). (h) The S-method (L=8). (i) The Modified B distribution (beta=0.14).

Forth Example: The final example involves a real-life bat signal which was also described and applied in Section 3.6.1. The results for the different TF methods
are shown in Fig. 4.6. Again, the proposed scheme in this work outperforms all other compared approaches since it can provide the highest resolution of the analysed signal without any undesirable cross-terms.
Chapter 4. Time-Frequency Adaptation of the STFT

4.6 Summary

In this chapter, a new scheme for the time-frequency adaptation of the STFT has been introduced. The aim is to automatically determine the size and the chirp-rate of the analysis window at each time-frequency point in order to further improve on the resolution of the time-adaptive STFT. The proposed adaptation does not require a priori knowledge of the signal or any user intervention, and the complexity is similar to the previous proposed time-adaptive STFT in Chapter 3 by iteratively adapting for one parameter. The proposed adaptive scheme is based on the analytic optimal solution derived in this chapter. Experimentation has shown that the proposed method is able to achieve good TF resolution, and can outperform previously reported time-frequency and time adaptive spectrograms. In addition, the method has also compared favorably with the multi-window spectrogram and two reputable quadratic TF representations.

APPENDIX A

In this appendix, we prove that the global duration-bandwidth product (GDBP) of the spectrogram computed by (4.6) can be written as (4.15).

Let’s consider taking the STFT of a signal \(x(t)\) using an analysis window
Chapter 4. Time-Frequency Adaptation of the STFT

\[ g(t) = \left( \frac{a}{\pi} \right)^{\frac{1}{2}} e^{-\frac{(a+jb)t^2}{2}} \]. \( g(t) \) and \( G(\omega) \) are the Fourier pairs of the window. The global time duration and the global frequency bandwidth of the signal are notated as \( T_s \) and \( B_s \) respectively. The global mean time and the global mean square time are calculated to be

\[ \langle t_g \rangle = \int t |g(t)|^2 dt = \int t \left( \left( \frac{a}{\pi} \right)^{\frac{1}{2}} e^{-at^2} \right)^2 dt = 0 \]  \hspace{1cm} (4.43)

and

\[ \langle t_g^2 \rangle = \int t^2 |g(t)|^2 dt = \int t^2 \left( \left( \frac{a}{\pi} \right)^{\frac{1}{2}} e^{-at^2} \right)^2 dt = \frac{1}{2a} \]  \hspace{1cm} (4.44)

Hence, the global time duration of the window is

\[ T_g = \sqrt{\langle t_g^2 \rangle - \langle t_g \rangle^2} = \sqrt{\frac{1}{2a}} \]  \hspace{1cm} (4.45)

Similarly, the global mean frequency and the global mean square frequency, using the mathematical trick employed in [1], are calculated to be

\[ \langle \omega_g \rangle = \int \omega |G(\omega)|^2 d\omega \]

\[ = \int g^*(t) \frac{1}{j \partial t} g(t) dt \]

\[ = \left( \frac{a}{\pi} \right)^{\frac{1}{2}} \int e^{-\frac{(a+jb)t^2}{2}} \frac{1}{j \partial t} e^{\frac{(a+jb)t^2}{2}} dt \]

\[ = 0 \]  \hspace{1cm} (4.46)

and

\[ \langle \omega_g^2 \rangle = \int \omega^2 |G(\omega)|^2 d\omega \]

\[ = \int g^*(t) \left( \frac{1}{j \partial t} \right)^2 g(t) dt \]
\[ = \int \left( \frac{d}{dt} g(t) \right)^2 dt \]
\[ = \frac{1}{\pi} \int e^{(-\alpha + j\beta)t^2} \{(\alpha + j\beta)t\}^2 dt \]
\[ = \frac{\alpha}{2} + \frac{\beta^2}{2\alpha} . \quad (4.47) \]

Hence, the global frequency bandwidth of the window becomes

\[ B_g = \sqrt{\langle \omega_g^2 \rangle - \langle \omega_g \rangle^2} = \sqrt{\frac{\alpha}{2} + \frac{\beta^2}{2\alpha}} . \quad (4.48) \]

According to [29], the global time duration and the global frequency bandwidth of the spectrogram is given to be

\[ T_{SP} = \sqrt{T_x^2 + T_g^2} \quad \text{and} \quad B_{SP} = \sqrt{B_x^2 + B_g^2} . \quad (4.49) \]

Therefore, the global time-bandwidth product of the spectrogram computed by (4.6) can be written as (4.15).

**APPENDIX B**

In this appendix, we prove that the local duration-bandwidth product (LDBP) for can be estimated as (4.27).

**A. Expansion of Local Bandwidth**

The local mean frequency of the spectrogram at time \( t_o \) is defined as
and we use the mathematical trick employed in [1] to rewrite it as

\[
\langle \omega \rangle_{t_o} = \frac{\int \omega |X(t_o, \omega)|^2 d\omega}{\int A^2(\tau) A^2_B(\tau-t_o) d\tau} \tag{4.50}
\]

We expand \( \varphi'(\tau + t_o) \) and \( \varphi''_g(\tau) \) as a power series in \( \tau \) so that

\[
\varphi'(\tau + t_o) = \sum_{n=0}^{\infty} \varphi^{(n+1)}(t_o) \frac{\tau^n}{n!} \tag{4.53}
\]

\[
\varphi''_g(\tau) = \sum_{n=0}^{\infty} \varphi''_g^{(n+1)}(0) \frac{\tau^n}{n!} \tag{4.54}
\]

and substitute them into (4.52), we have
\[
\langle \omega \rangle_{t_o} = \frac{\int A^2(\tau + t_o)A_g^2(\tau)\tau^n \sum_{n=0}^{\infty} \frac{\varphi^{(n+1)}(\tau + t_o)}{n!}}{P(t_o) \sum_{n=0}^{\infty} \frac{\varphi'_g^{(n+1)}(0)}{n!}} d\tau \\
= \sum_{n=0}^{\infty} \frac{M_n(t_o)}{n!} \{ \varphi^{(n+1)}(\tau + t_o) + \varphi'_g^{(n+1)}(0) \} 
\]

(4.55)

where

\[
M_n = \frac{\int A^2(\tau + t_o)A_g^2(\tau)\tau^n d\tau}{P(t_o)} 
\]

(4.56)

We write out the first few terms, and the local mean frequency can be estimated as

\[
\langle \omega \rangle_{t_o} = \varphi'(t_o) + \varphi'_g(0) + M_1 \{ \varphi''(t_o) + \varphi''_g(0) \} 
\]

(4.57)

Similarly, we consider the expansion for the local mean square frequency:

\[
\langle \omega^2 \rangle_{t_o} = \frac{\int \omega^2 |X(t_o, \omega)|^2 d\omega}{\int A^2(\tau + t_o)A_g^2(\tau) d\tau} \\
= \frac{\int \frac{d}{d\tau} A(\tau + t_o)A_g(\tau)|\varphi'(\tau + t_o) + \varphi'_g(\tau)|^2 d\tau}{P(t_o)} \\
= \langle \omega^2 \rangle_{t_o}^0 + \frac{\int \frac{d}{d\tau} A(\tau + t_o)A_g(\tau)|\varphi'(\tau + t_o) + \varphi'_g(\tau)|^2 d\tau}{P(t_o)} 
\]

(4.58)

where

\[
\langle \omega^2 \rangle_{t_o}^0 = \frac{\int \frac{d}{d\tau} A(\tau + t_o)A_g(\tau)^2 d\tau}{P(t_o)} 
\]

(4.59)

\(\varphi'(\tau + t_o)\) and \(\varphi'_g(\tau)\) are then expanded as a power series in \(\tau\) and the local mean square frequency becomes
and the terms with high-order differentiation are neglected so that

\[
\langle \omega^2 \rangle_{t_o} = \langle \omega^2 \rangle_{t_o}^0 + \{\varphi'(t_o) + \varphi_g'(0)\}^2 + 2\left(\varphi'(t_o) + \varphi_g'(0)\right)
\]
\[
\{M_1(\varphi''(t_o) + \varphi_g''(0))\} + M_2\left(\varphi''(t_o) + \varphi_g''(0)\right)^2
\]  

(4.61)

The square of the local mean frequency can be estimated as

\[
\langle \omega \rangle_{t_o}^2 = \{\varphi'(t_o) + \varphi_g'(0)\}^2 + 2\{\varphi'(t_o) + \varphi_g'(0)\}\{M_1(\varphi''(t_o) + \varphi_g''(0))\} +
\]
\[
M_1^2\{\varphi''(t_o) + \varphi_g''(0)\}^2
\]  

(4.62)

Hence, the frequency variance of the spectrogram at time \( t_o \) is defined as

\[
\sigma^2_{\omega|t_o} = \langle \omega^2 \rangle_{t_o} - \langle \omega \rangle_{t_o}^2
\]
\[
= \langle \omega^2 \rangle_{t_o}^0 + \{M_2 - M_1^2\}\{\varphi''(t_o) + \varphi_g''(0)\}^2
\]  

(4.63)

B. Expansion of Time duration for a given frequency

Following the similar expansion above and expressing the signal and window in the corresponding frequency domain, the mean time and the mean square time of the spectrogram for a given frequency \( \omega_o \) are expressed as follows:
\[ \langle t \rangle_{\omega_o} = \frac{\int B^2(\omega)B^2_g(\omega-\omega_o)(\psi^'(\omega)+\psi^'_g(\omega-\omega_o))d\omega}{\int B^2(\omega)B^2_g(\omega-\omega_o)d\omega} \quad (4.64) \]

and

\[ \langle t^2 \rangle_{\omega_o} = \langle t^2 \rangle^0_{\omega_o} + \frac{\int B^2(\omega)B^2_g(\omega-\omega_o)(\psi^'(\omega)+\psi^'_g(\omega-\omega_o))^2d\omega}{\int B^2(\omega)B^2_g(\omega-\omega_o)d\omega} \quad (4.65) \]

where

\[ \langle t^2 \rangle^0_{\omega_o} = \frac{\int \frac{d}{d\omega}B(\omega)B_g(\omega-\omega_o)^2d\omega}{\int B^2(\omega)B^2_g(\omega-\omega_o)d\omega}. \quad (4.66) \]

By using the power series expansion, after eliminating the high differential terms, the time duration of the spectrogram at frequency \( \omega_o \) can be estimated as

\[ \sigma^2_{t|\omega_o} = \langle t^2 \rangle^0_{\omega_o} + \{N_2 - N_1^2\} \{\phi^''(\omega_o) + \phi^''_g(0)\}^2 \quad (4.67) \]

where

\[ N_n = \frac{\int B^2(\omega)B^2_g(\omega-\omega_o)\omega^n d\omega}{\int B^2(\omega)B^2_g(\omega-\omega_o)d\omega}. \quad (4.68) \]

Therefore, using the above time duration and bandwidth, the local duration-bandwidth product (LDBP) of the spectrogram at time \( t_o \) and frequency \( \omega_o \) can be formulated as that in (4.27).

REFERENCES


Chapter 4. Time-Frequency Adaptation of the STFT


Chapter 4. Time-Frequency Adaptation of the STFT


CHAPTER 5

SHORT-TIME FOURIER
TRANSFORM BASED OPTIMIZED FILTER

5.1 Introduction

The restoration of signals degraded by noise or interference is an important application in signal processing. Conventional optimal Wiener filters dominate in the theory of statistical signal processing due to their simple implementation and good performance [1]-[2]. But both the signal and the noise processes are restricted to be stationary ones in the formulation of Wiener filters. However, most natural signals have variable frequency contents (i.e. non-stationary signals) and the conventional Wiener filters are not suitable to deal with such kind of signals. This requires the development of time-varying optimal filters. In [3]-[5], filtering in TF
domain based on multiplication of STFT with a TF weighting function was proposed. The weighting function was formulated by the concept of a posteriori Wiener filtering (APWF) which is suitable for ensembles of unknown, repetitive, noisy signals and has successfully been applied to denoise high-resolution electrocardiogram (HRECG). Modification of STFT based on the Wiener theory has also long been deployed for speech enhancement [6].

Recently, the TF formulation of time-varying Wiener filter based on the Weyl symbol (WS) and the Wigner-Ville spectrum (WVS) has been proposed for non-stationary underspread processes [7]-[9]. The multi-window STFT was also used to approximate the TF optimal filter, yielding a simple and efficient implementation in [7]. Nevertheless, in addition to this method being limited to jointly underspread signal and noise processes only, it also requires a priori knowledge of the WVS of the ideal signal and noise processes. Also, both processes were considered to be independent from each other and zero-mean. Although in certain applications it may be possible to make the above assumptions, the majority of signals are not underspread and a degree of dependency may exist between the noise and the desired signal. Furthermore, noise may not be a strictly zero-mean process. Even worse, when the WVS of the noise and the ideal signal are unavailable altogether, the above solution cannot be used at all.

In this chapter, we provide a solution based on the STFT to the problem discussed above about optimal filtering for non-stationary signal processes such that the need for any assumptions about the noise and the signal model are eliminated:
We formulate a filtering TF mask for fixed-window STFT filtering using a small set of training signal. The most suitable window length is determined iteratively.

Based on the multi-window method originally introduced by Thomson [10] for spectrum estimation, we discuss an efficient TF implementation using the multi-window STFT.

We also develop an effective method for calculate the optimized weights for the Hermite functions.

Experimentation has shown that the introduced approach can efficiently recover wide range of distorted random signal processes based on a small set of representative examples of the noisy observation and the desired signal.

5.2 Fixed-Window STFT based Filtering

5.2.1 Formulation of the Optimized Mask in Least-Squares Sense

The vector formulations of the discrete STFT and its inverse [11] in the Section 2.1.4 facilitate, in terms of the notations and mathematical derivation, the formulation of the optimized modification for the STFT filtering in this section.

In this formulation, the input signal and the corrupting noise are considered to be discrete-time, finite-length random processes of size $J$. In measurements under additive noise, the following observation model can be assumed in discrete form as
$y = x + n$, where $y$, $x$, and $n$ are column vectors of size $J$ representing the acquired signal, the desired signal, and the noise, respectively. The goal of the filter is to find an estimate $\hat{x}$ as close as possible to the ideal $x$. A natural optimally criterion is the mean square error (MSE),

$$mse(\hat{x}_n(\omega)) = \frac{1}{J} E[\|\hat{X}^{h_L}_{STFT,n}(\omega) - X^{h_L}_{STFT,n}(\omega)\|^2] \quad (5.1)$$

where $\|\cdot\|^2$ denotes the 2-norm operator of a vector, i.e. $\|\hat{X}^{h_L}_{STFT,n}(\omega) - X^{h_L}_{STFT,n}(\omega)\|^2 = (\hat{X}^{h_L}_{STFT,n}(\omega) - X^{h_L}_{STFT,n}(\omega))^H (\hat{X}^{h_L}_{STFT,n}(\omega) - X^{h_L}_{STFT,n}(\omega))$, and $\hat{X}_{STFT,n}(\omega)$ and $X_{STFT,n}(\omega)$ is the STFT at time instant $n$, for $n$ is an integer, of $\hat{x}(n)$ and $x(n)$, respectively. $h_L$ denotes the analysis window of the STFT and its length. The estimate signal $\hat{x}(n)$ can be obtained through filtering in the STFT domain such that

$$\hat{X}_{STFT,n} = M_n f^{h_L}_{STFT,n} x \quad \text{and} \quad X_{STFT,n} = f^{h_L}_{STFT,n} x \quad (5.2)$$

where $f^{h_L}_{STFT,n(K \times J)}$ are the discrete short-time Fourier transform at time instant $n$.

According to Section 2.1.4, $f^{h_L}_{STFT,n(K \times J)}$ is given by

$$[FPH] \begin{bmatrix} I_{\binom{L-1}{2} \times \binom{L+1}{2} + n} \\ 0_{\binom{L+1}{2} \times \binom{L+1}{2} - n} \end{bmatrix} \in \mathbb{C}^{K \times J}, 0 \leq n \leq (L - 1)/2$$

$$[FPH] \begin{bmatrix} I_{L \times (n-(L-1)/2)} \\ 0_{L \times (J-(L+1)/2)} \end{bmatrix} \in \mathbb{C}^{K \times J}, (L - 1)/2 < n \leq J - (L - 1)/2$$
\[ \mathbf{F}_{\text{STFT}}[\mathbf{h}_L](\omega) = \mathbf{X}_{\text{STFT},n}(\omega) \] for \( n = 0, 1, \ldots, J \). Substituting (5.2) into (5.4) yields the following cost function:

\[ J(\mathbf{m}_n) = \frac{1}{N} \sum_{i=1}^{N} \left\| \mathbf{M}_n \mathbf{F}_{\text{STFT},n} \mathbf{y}_i - \mathbf{F}_{\text{STFT},n} \mathbf{x}_i \right\|^2. \] (5.5)

The aim is to minimize the cost function defined in (5.5) with respect to the vector \( \mathbf{m}_n \) which is constrained to be real-valued. Let \( \mathbf{z}_i = [z_{i,0}, z_{i,1}, \ldots, z_{i,K-1}]^T = \mathbf{F}_{\text{STFT},n} \mathbf{y}_i \), so that (5.5) becomes

\[ J(\mathbf{m}_n) = \frac{1}{N} \sum_{i=1}^{N} \left\| \mathbf{M}_n \mathbf{z}_i - \mathbf{F}_{\text{STFT},n} \mathbf{x}_i \right\|^2 
= \frac{1}{N} \sum_{i=1}^{N} \left( \mathbf{M}_n \mathbf{z}_i - \mathbf{F}_{\text{STFT},n} \mathbf{x}_i \right)^H \left( \mathbf{M}_n \mathbf{z}_i - \mathbf{F}_{\text{STFT},n} \mathbf{x}_i \right). \] (5.6)
It can be observed that since $M_n$ is a diagonal matrix then, $M_nz_i = Z_i m_n$, where $Z_i(K \times K)$ is a diagonal matrix such that $z_i = diag(Z_i)$. Now, (5.6) can be further simplified as:

$$J(m_n) = \frac{1}{N} \sum_{i=1}^{N} (Z_i m_n - F_{STFT,n}^{h_L} x_i)^H (Z_i m_n - F_{STFT,n}^{h_L} x_i) .$$  \hspace{1cm} (5.7)$$

Expanding (5.7) yields:

$$\frac{1}{N} \sum_{i=1}^{N} (m_n^H Z_i^H Z_i m_n - m_n^H Z_i^H F_{STFT,n}^{h_L} x_i - x_i^H F_{STFT,n}^{h_L} Z_i m_n + x_i^H F_{STFT,n}^{h_L} F_{STFT,n}^{h_L} x_i)$$

$$= \frac{1}{N} \sum_{i=1}^{N} (m_n^H Z_i^H Z_i m_n - (x_i^H F_{STFT,n}^{h_L} Z_i z_i)^H - x_i^H F_{STFT,n}^{h_L} Z_i m_n + x_i^H F_{STFT,n}^{h_L} F_{STFT,n}^{h_L} x_i)$$

$$= \frac{1}{N} \sum_{i=1}^{N} (m_n^H Z_i^H Z_i m_n - 2Re(x_i^H F_{STFT,n}^{h_L} Z_i z_i) + x_i^H F_{STFT,n}^{h_L} F_{STFT,n}^{h_L} x_i) .$$  \hspace{1cm} (5.8)$$

Since $m_n$ is real-valued, the above expression becomes:

$$J(m_n) = \frac{1}{N} \sum_{i=1}^{N} (m_n^T Q_i m_n + b_i^T m_n + c_i) .$$  \hspace{1cm} (5.9)$$

where the matrix $Q_i(K \times K)$, the column vector $b_i(1 \times K)$, and the scalar $c_i$ are:

$$Q_i = Z_i^H Z_i ,$$  \hspace{1cm} (5.10)$$

$$b_i^T = (-2Re(x_i^H F_{STFT,n}^{h_L} Z_i))^T ,$$  \hspace{1cm} (5.11)$$

$$c_i = \| F_{STFT,n}^{h_L} x_i \|^2 .$$  \hspace{1cm} (5.12)$$

It can be observed that since $Z_i$ is diagonal then $Q_i$ is also diagonal as well as
real-valued. Finally, (5.9) can be expressed as:

\[ J(g) = m_n^T Qm_n + b^T m_n + c. \]  

(5.13)

where \( Q = \frac{1}{N} \sum_{i=1}^{N} Q_i \), \( b = \frac{1}{N} \sum_{i=1}^{N} b_i \), and \( c = \frac{1}{N} \sum_{i=1}^{N} c_i \). We seek the vector \( m_n^{opt} \) that minimize (5.13),

\[ \frac{\partial J(m_n)}{\partial m_n} \bigg|_{m_n = m_n^{opt}} = 0. \]  

(5.14)

The derivative of the first term of (5.13) with respect to the vector \( m_n \) is in general equal to:

\[ \frac{\partial (m_n^T Q m_n)}{\partial m_n} = Q m_n + Q^T m_n, \]  

(5.15)

and since \( Q \) is diagonal the right-hand side becomes \( 2Qm_n \). Similarly, the derivative of the second term of (5.13) with respect to the vector \( m_n \) is \( \frac{\partial (b^T m_n)}{\partial m_n} = b \). Thus, (5.14) can be written as

\[ 2Qm_n^{opt} + b = 0. \]  

(5.16)

The optimized mask at time instant \( n \) can be analytically calculated by solving the system of \( K \) linear equations with \( K \) unknowns in (5.16) using the analysis
window of length $L$. Since $Q$ is a diagonal squared-matrix, its inverse can be easily determined. The resulting optimized mask is $M_L^{opt} = [m_0^{opt}, m_1^{opt}, \ldots, m_J^{opt}]$.

However, the problem of finding the most suitable length $L$ for the analysis window such that the cost function (5.5) can be minimized is difficult to be solved analytically. For this problem, we use the iterative approach in which $M_L^{opt}$ is consecutively computed for a finely sampled set of values within a reasonable range. The length $L$ which results in the smallest average error (5.5) is retained.

Based on the above solution, for given training sets of data (i.e. a number of realizations of the desired signal $x_i$ and their corresponding observations $y_i$), the most suitable window length and the corresponding multiplicative optimized mask can easily be calculated. Beside, a distinct merit of the proposed filter is that it does not require knowledge of the noise statistics.

5.2.2 Experimental Results

To evaluate the performance of the proposed algorithm in the previous section, three illustrative examples are presented here. In each example, a set of five realizations of the desired signal together with their corresponding noisy observations are generated as training data for analytically determining the optimized mask. The most suitable window length is iteratively measured. The resulting filter is then applied to a previously unseen realization of the distorted signal. For the sake of comparison, this particular waveform is also filtered in the conventional frequency domain, and the corresponding optimized mask in the frequency domain is determined using the
same training signals and deploying the similar formulation of the analytic solution as (5.16). For all the comparisons reported here, 25 independent realizations of the desired signals together with their corresponding noisy observations were calculated.

In the first example, we consider a Gaussian signal with randomized amplitude and time shift, i.e. \( x(t) = A e^{-\pi(t-s)^2 + j10\pi(t-s)} \), where \( A \) and \( s \) are random variables uniformly distributed in the interval [1 3] and [-1 1], respectively. The noise component which was then added to corrupt the signal was generated as follows; white noise of finite duration between [-2 2] was low-pass filtered (normalized cut-off frequency at 0.04), and subsequently modulated using the quadratic complex exponential (chirp) function \( e^{-j2\pi t^2} \) in order to tilt the noise component in the TF plane. The normalized MSE achieved for different values of \( L \) is plotted in Fig. 5.1a where it can be seen that the most appropriate window length for filtering the type of signal at hand is the one at \( L = 65 \) samples. The optimized mask of the filter is presented in Fig. 5.1b. A realization of the noisy observed \( y(t) \) signal and its spectrogram using \( L = 65 \) samples is shown in Fig. 5.1c and Fig. 5.1d, respectively. The resulting estimated signal \( \hat{x}(t) \) after processing the depicted \( y(t) \) with the proposed STFT filter using the window length at \( L = 65 \) samples are shown in Fig. 5.1e with MSE = 0.000249 and is compared with the estimated signal using the corresponding filter in the frequency domain in Fig. 5.1f with MSE = 0.0141. It can be seen that a nearly perfect recovery of the desired signal can be achieved as a result of filtering in the STFT domain using the suitable window length.
Fig. 5.1[Example 1] (a) Normalized MSE for different values of $L$; (b) Corresponding optimized mask with the most suitable $L=65$ samples; (c) A realization of the noisy signal $y(t)$ (solid) and the desired signal $x(t)$ (dotted); (d) Spectrogram of $y(t)$ using window with $L=65$ samples; The recovered signal $\hat{x}(t)$ with $L=65$ samples (solid) and the desired signal $x(t)$ (dotted) by (e), by the proposed STFT optimized filtering and (f) the frequency domain optimized filtering.

The desired signal in the second example is a non-linearly modulated sinusoidal chirp whose amplitude and time shift are random variables, that is,
\[ x(t) = A e^{j2\pi(6\sin(0.5\pi(t-s))+3(t-s))}, \]

where \( A \) and \( s \) are uniformly distributed on the intervals [1 3] and [-0.01 0.01], respectively. The interference consists of the sum of one sinusoidal chirp \( A e^{j2\pi(6\sin(0.5\pi(t-s))+3(t-s))} \) and three Gaussian atoms defined as

\[ 2e^{-8\pi(t-s)^2+j17\pi(t-s)} + 2e^{-8\pi(t-s-1)^2} + 2e^{-8\pi(t-s+1)^2}. \]

The sum of the desired signal and the interference is then quantized such that further distortion is added to it in the form of quantization noise. This can be expressed as

\[ y(t) = \frac{\text{round}(b x_c(t))}{b}, \]

where \( x_c(t) \) is the corrupted signal and \( b \) is the quantization factor which can be any positive integer. The lower the value of \( b \), the coarser the quantization effect, producing large rounding-off errors. In this example, \( b \) was set to the lowest value of 1. It should be stressed that this is a kind of distortion whose statistics is impossible to know in advance because it is a function of the input process. For such type of noise it is not feasible to collect a set of independent realizations either. The normalized MSE achieved for different values of \( L \) is depicted in Fig. 5.2a where it can be seen that the most appropriate window length for filtering in this example is the one at \( L = 41 \) samples. The corresponding optimized mask of the filter is presented in Fig. 5.2b. A realization of the noisy observed \( y(t) \) signal and its spectrogram using \( L = 41 \) samples is shown in Fig. 5.2c and Fig. 5.2d, respectively. The results \( \hat{x}(t) \) using the proposed STFT filter with the most suitable window length (\( L = 41 \) samples) is shown in Fig. 5.2e with MSE = 0.0329 and is compared with the estimated signal using the corresponding filter in the frequency domain in Fig. 5.2f with MSE = 0.3276.
Chapter 5. Short-Time Fourier Transform based Optimized Filter

![Fig. 5.2](image)

(a) Normalized MSE for different values of $L$; (b) Corresponding optimized mask with the most suitable $L=41$ samples; (c) A realization of the noisy signal $y(t)$ (solid) and the desired signal $x(t)$ (dotted); (d) Spectrogram of $y(t)$ using window with $L=41$ samples; The recovered signal $\hat{x}(t)$ with $L=41$ samples (solid) and the desired signal $x(t)$ (dotted) (e) by the proposed STFT optimized filtering, and (f) by the frequency domain optimized filtering.

In the final example, the input process is now given as a non-linearly
modulated sinusoidal chirp $x(t) = Ae^{j2\pi(6\cos(0.5\pi(t-s)))}$, which is degraded by the addition of zero-mean white Gaussian noise at 0dB SNR, where $A$ and $s$ are uniformly distributed on the intervals [1 3] and [-0.01 0.01], respectively. The normalized MSE is presented for different values of $L$ in Fig. 5.3a. The minimum MSE was obtained in the window length $L=35$ samples. The related optimized mask $M_{L}^{opt}$ is shown in Fig. 5.3b. A realization of the noisy observed $y(t)$ signal and its spectrogram using $L = 35$ samples is shown in Fig. 5.3c and Fig. 5.3d, respectively. The estimate $\hat{x}(t)$ by the proposed STFT filter using the favorable window length ($L = 35$ samples) is presented in Fig. 5.4e with MSE = 0.3403 and is compared with the estimated signal using the corresponding filter in the frequency domain in Fig. 5.5f with MSE = 1.3420. Once again, it can be seen that filtering with an appropriately window length can yield a superior performance. The mean-squared error values achieved for the above three examples by the compared methods are presented in Tables 5.1 with averaging the results of 25 independent desired and noisy signals.
Chapter 5. Short-Time Fourier Transform based Optimized Filter

Fig. 5.3 [Example 3] (a) Normalized MSE for different values of L; (b) Corresponding optimized mask with the most suitable $L=35$ samples; (c) A realization of the noisy signal $y(t)$ (solid) and the desired signal $x(t)$ (dotted); (d) Spectrogram of $y(t)$ using window with $L=35$ samples; The recovered signal $\hat{x}(t)$ with $L=35$ samples (solid) and the desired signal $x(t)$ (dotted) (e) by the proposed STFT optimized filtering, and (f) by the frequency domain optimized filtering.

<table>
<thead>
<tr>
<th>Method</th>
<th>Examples</th>
<th>Example 1</th>
<th>Example 2</th>
<th>Example 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed Fixed-Window STFT Optimized Filtering</td>
<td>AMSE (standard deviation)</td>
<td>0.000894 (0.0015)</td>
<td>0.0274 (0.0035)</td>
<td>0.2563 (0.1358)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5.3 Multi-Window STFT based Filtering

5.3.1 STFT Analysis-Modification-Synthesis Filtering

The STFT of a signal \( x(t) \) using an analysis window \( h(t) \) is given in Chapter 2 in (2.1). There exists a well-know synthesis formula in (2.5) for the STFT using a synthesis window \( f(t) \) [12], [13] \( \hat{x}(t) = \frac{1}{2\pi} \int \int X_{STFT}(\tau, \omega) f(t - \tau) e^{j\omega t} \, d\tau d\omega \), where \( \hat{x}(t) \) is the recovered signal. For the perfect reconstruction of \( x(t) \) from its STFT, the analysis and synthesis windows is needed to satisfy the following constraint: \( \int f(-t) h^*(t) \, dt = 1 \). The STFT Analysis-Modification-Synthesis based filtering is obtained by inserting a multiplicative modification \( Mask(t, \omega) \) between the analysis and synthesis procedures as follows:

\[
\hat{x}(t) = \frac{1}{2\pi} \int \int Mask(\tau, \omega) X_{STFT}(\tau, \omega) f(t - \tau) e^{j\omega t} \, d\tau d\omega . \tag{5.17}
\]

By substituting the STFT in (2.1) into the above STFT filtering equation, (5.17) becomes

\[
\hat{x}(t) = \frac{1}{2\pi} \int \int Mask(\tau, \omega) x(\tau) h^*(\tau - \tau') e^{-j\omega \tau'} \, dt' \int f(t - \tau) e^{j\omega t} \, d\tau d\omega \\
= \frac{1}{2\pi} \int x(\tau') \int \int Mask(\tau, \omega) h^*(t - \tau) e^{j\omega t} f(\tau - \tau') e^{-j\omega \tau'} \, d\tau d\omega \, dt' \tag{5.18}
\]

then

<table>
<thead>
<tr>
<th>Frequency-Domain Optimized Filtering</th>
<th>AMSE (standard deviation)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0177 (0.0090)</td>
</tr>
<tr>
<td></td>
<td>0.2357 (0.0924)</td>
</tr>
<tr>
<td></td>
<td>0.9012 (0.4634)</td>
</tr>
</tbody>
</table>

5.17
\[ \hat{x}(t) = \int T(t, \tau)x(t - \tau)d\tau' \] (5.19)

which expresses the STFT Analysis-Modification-Synthesis based filtering of \( x(t) \) into the response of the linear time-varying system \( T(t, \tau) \) to \( x(t) \), illustrated in Fig. 5.4 and 5.5.

![Fig. 5.4 The STFT Analysis-Modification-Synthesis based filtering.](image)

\[ y(t) \quad \xrightarrow{\text{STFT analysis \ with \ using \ } h} \quad M(t, \omega) \quad \xrightarrow{\text{STFT synthesis \ with \ using \ } f} \quad \hat{x}(t) \]

\[ y(t) \quad \xrightarrow{T(t, \tau; M, h, f)} \quad \hat{x}(t) \]

Fig. 5.5 The linear time-varying system representing the STFT filtering.

### 5.3.2 The Smoothing Effect of the Fixed-Window STFT on the Modification Mask

In this section, we explore the windowing effect of STFT filtering on the multiplicative modification mask. According to the above section, the STFT filtering based on the analysis-modification-synthesis approach can be thought of as passing a signal into the linear time-varying system \( T(t, \tau; M^{opt}, h, f) \) which is influenced by both the multiplicative modification mask, and the analysis and synthesis windows. The influence can be seen in the following equation:

\[ T(t, \tau; M^{opt}, h, f) = \frac{1}{2\pi} \iint M^{opt}(\tau, \omega)h^*(t - \tau)e^{j\omega t} f(\tau - \tau')e^{-j\omega \tau'}d\tau d\omega . \] (5.20)
By taking the Weyl symbol of $T(t, \tau)$, the STFT-based time-varying system is characterized as follows [14]:

$$L_T(t, \omega) = M^{opt}(t, \omega) \ast \ast WD_{h,f}(t, \omega), \quad (5.21)$$

where $\ast \ast$ denotes a 2-D convolution. From the above equation, we can see that $M^{opt}(t, \omega)$ is smoothed by the cross-Wigner distribution of the analysis and synthesis window. The smoothing effect enlarges the time-frequency size of the estimated optimized mask $M^{opt}(t, \omega)$, and thus results in accommodating more corrupted noise in the filtering process. Our aim is to minimize the smoothing effect caused by the analysis and synthesis windows so that $L_T(t, \omega) \approx M^{opt}(t, \omega)$. We therefore require that $WD_{h,f}(t, \omega) \approx \delta(t, \omega)$. In the next section, we will introduce using the weighted sum of the auto-Wigner distributions of the Hermite functions to approach a Dirac delta-like function in the time-frequency plane.

### 5.3.3 Minimizing the Smoothing Effect using the Weighted Sum of the Auto-Wigner Distributions of the Hermite Functions

In order to reduce the smoothing effect on the masking function described in the previous section and achieve a better filtering result, we would like to model the cross-Wigner distribution of the analysis and synthesis windows to be a very
concentrated and Dirac delta-like TFR as

\[ D(t, \omega; r) = e^{-\frac{(t^2 + \omega^2)}{r^2}}, \quad (5.22) \]

where \( D(t, \omega; r) \) is a real and radially symmetric TFR and its radius is controlled by the parameter \( r \). As \( r \) tends to zero, \( D(t, \omega; r) \) approaches a 2-D Dirac delta function in the time-frequency plane centered at (0,0). According to [14], a real and radially symmetric symbol can be expanded into a series of auto-Wigner distributions of Hermite functions. Therefore, \( D(t, \omega; r) \) can be interpreted as

\[ D(t, \omega; r) = \sum_{k=0}^{\infty} c_k WD_{h_k, h_k}(t, \omega) \quad (5.23) \]

where \( c_k \) is the weight for the corresponding WD of the \( k \)-th order Hermite function which is defined as:

\[ h_k(t) = \frac{1}{\sqrt{\sqrt{\pi} 2^k k!}} e^{-\frac{t^2}{2}} H_k(t), \text{ for } k = 0, 1, ... \quad (5.24) \]

and

\[ H_k(t) = (-1)^k e^{t^2} \left( \frac{d}{dt} \right)^k e^{-t^2}. \quad (5.25) \]

In the following, we will employ \( K \) Hermite functions to approximate \( D(t, \omega; r) \), and we will show how to obtain the optimized weights \( c_k^{opt} \) for this approximation,
\[ D(t, \omega; r) \approx \sum_{k=0}^{K-1} c_k W D_{h_k h_k}(t, \omega). \]  

### 5.3.4 Determination of the Optimized Weights

In this section, we introduce a method in the least-squares sense to determine the optimized weight \( c_k^{\text{opt}} \) to approximate a discrete \( D(n, m; r) \) with small \( r \). The weighted sum of the discrete auto-Wigner distributions of the Hermite functions \( h_k(n) \) is given by

\[ WSWD(n, m; K) = \sum_{k=0}^{K-1} c_k W D_{h_k h_k}(n, m). \]  

(5.27)

To find the optimized weights \( c_k^{\text{opt}} \) such that the \( WSWD(n, m) \) approaches the desired time-frequency representation, we then solve the following linear least-square problem

\[ J(c_k) = \min_{c_k} \sum_{n=1}^{N} \sum_{m=1}^{M} |D(n, m) - \sum_{k=0}^{K-1} c_k W D_{h_k h_k}(n, m)|^2 \]  

(5.28)

where \( D(n, m) \) is the desired time-frequency representation. (5.28) can be expressed in matrix form as

\[ J(c) = \min_{c_k} \| \mathbf{d} - \mathbf{Hc} \|^2 \]  

(5.29)

where

\[ \mathbf{d} = \sum_{n=1}^{N} D(n, m) \in \mathbb{R}^{M\times1} \]  

(5.30)
\[ \mathbf{H} = [\sum_{n=0}^{N-1} WD_{h_0, h_0}(n, m) \quad \sum_{n=0}^{N-1} WD_{h_1, h_1}(n, m) \ldots \]
\[ \sum_{n=0}^{N-1} WD_{h_{K-1}, h_{K-1}}(n, m)] \in \mathbb{R}^{M \times K} , \quad (5.31) \]

and

\[ \mathbf{c} = [c_0 \quad c_1 \quad \cdots \quad c_{K-1}]^T \in \mathbb{R}^{K \times 1} . \quad (5.32) \]

Expanding (5.29) yields

\[ J(\mathbf{c}) = (\mathbf{d} - \mathbf{H}\mathbf{c})^T (\mathbf{d} - \mathbf{H}\mathbf{c}) \]
\[ = (\mathbf{d}^T - \mathbf{c}^T \mathbf{H}^T)(\mathbf{d} - \mathbf{H}\mathbf{c}) \]
\[ = \mathbf{c}^T \mathbf{H}^T \mathbf{H} \mathbf{c} - \mathbf{c}^T \mathbf{H}^T \mathbf{d} - \mathbf{d}^T \mathbf{H} \mathbf{c} + \mathbf{d}^T \mathbf{d} \]
\[ = \mathbf{c}^T \mathbf{H}^T \mathbf{H} \mathbf{c} - 2\mathbf{H}^T \mathbf{d} \mathbf{c} + \mathbf{d}^T \mathbf{d} \quad (5.33) \]

We seek the vector \( \mathbf{c}^{opt} \) that minimizes (5.33),

\[ \frac{\partial J(\mathbf{c})}{\partial g} \bigg|_{\mathbf{c}=\mathbf{c}^{opt}} = 0. \]
\[ \mathbf{H}^T \mathbf{H} \mathbf{c} + \mathbf{H} \mathbf{H}^T \mathbf{c} - 2\mathbf{H}^T \mathbf{d} = 0 \]
\[ \mathbf{H}^T \mathbf{H} \mathbf{c} - \mathbf{H}^T \mathbf{d} = 0 \]
\[ \mathbf{c}^{opt} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{d} \quad (5.34) \]

We are going to show some examples to illustrate the determination of the optimized weights \( c_k^{opt} \) in order to approximate a Dirac delta-like TFR \( D(n, m; r) \) with small \( r \).
Fig. 5.6 The oscillatory sidelobes caused by using insufficient Hermite functions. The graphs in the left column are presented in side view and the right ones are in top view. (a)-(b) Mesh of the TFR of $D(n, m; r)$ with $r = 5.5$; The approximate by the weight sum of auto-Wigner distributions $WSWD(n, m; K)$ using (c)-(d) 4 Hermite functions and (e)-(f) 8 Hermite functions.

From Fig. 5.6, it can be noticed that in order to approximate $D(n, m; r)$ with $r = 5.5$ without oscillatory distortions and loss of concentration, at least 8 Hermite functions are required in the approximation. Indeed, the smaller the $r$, the larger the
number of Hermite functions which are required to approximate $D(n, m; r)$
yielding similar concentration and no oscillatory sidelobes. Experimentation has
shown that the approximated $D(n, m; r)$ above whose $r = 5.5$ is sufficient for our
purposes; we found that for $r < 5.5$, the smoothing effect on the modification mask
is insignificant. Therefore, 8 Hermite functions are required in our application.

### 5.3.5 Trimming of the Mask

In Section 5.2.1, we analytically found the estimated optimized mask for the STFT
filtering. However, since the cost function in (5.13) is formulated in the STFT
domain using a fixed circular Gaussian window, according to the analysis in
Section 3.5.2 and Section 3.5.3, the estimated optimized mask $M_{\text{opt}}$ in (5.16) is
spread by the window used in the STFT. The average expansion width of the $M_{\text{opt}}$
along any single direction can then be estimated as

$$
\Delta_p = \sqrt{\frac{\text{ln} p}{2}},
$$

for $p$ being the percentage of the maximum value of the $M_{\text{opt}}$. Therefore, a more precise mask can
be developed by omni-directionally trimming the $M_{\text{opt}}$ by the amount of $\Delta_p$, following the trimming processing steps in the Section 3.5.3. The $M_{\text{opt}}$ is first segmented by

$$
S(n, k) = \begin{cases} 
1, & M_{\text{opt}}(n, k) \geq p \cdot M_{\text{opt}}(n, k) \\
0, & \text{otherwise}
\end{cases}
$$

(5.35)

Next, the $M_{\text{opt}}$ is omni-directionally trimmed by the segmented mask $S(n, k)$ by
way of the following manipulation,
\[
\tilde{M}^{\text{opt}}(t, \omega) = \frac{1}{2\pi\Delta_p} \sum_{n'=-\Delta_p}^{\Delta_p} \sum_{k'=-\Delta_p}^{\Delta_p} M^{\text{opt}}(n, k) S(n - n', k - k') ,
\]  

(5.36)

under the constraint \( n'^2 + k'^2 = \Delta_p^2 \), which results in the trimmed estimated optimized mask \( \tilde{M}^{\text{opt}} \).

### 5.3.6 The Multi-Window STFT Approach

From Section 5.3.2 and Section 5.3.3, we examine the smoothing effect on the modification mask when using the fixed analysis and synthesis window in the STFT filtering, and introduce the idea to reduce the undesirable smoothing effect by employing the weighted sum of the auto-Wigner distributions of the Hermite functions. We also determine the weights in the optimized way for the auto-Wigner distributions of \( K \) Hermite functions to estimate a 2-D pulse-like function in the TF plane in Section 5.3.4. These results facilitate the introduction of the concept of the multi-window STFT (MW-STFT) based filtering in this section.

From (5.21), we can see that the modification mask is smoothed by the 2-D convolution with the cross-Wigner distribution of the analysis and synthesis windows. In the Section 5.3.3 and Section 5.3.4, we have shown that the a 2-D pulse-like function in the TF plane can be estimated by the weighted sum of the auto-Wigner distributions of the Hermite functions, and the optimized weights can be obtained.
\[
\delta(t, \omega) \approx D(t, \omega; r) \approx \sum_{k=0}^{K-1} c_k WD_{h_k, h_k}(n, m), \quad (5.37)
\]

where \( r \) is chosen to be small. Therefore, (5.21) can be rewritten as

\[
\widehat{L}_T(t, \omega) = \widehat{M}^{\text{opt}}(t, \omega) \ast \sum_{k=0}^{K-1} c_k^{\text{opt}} WD_{h_k, h_k}(n, m) \\
\approx \widehat{M}^{\text{opt}}(t, \omega) \ast \delta(t, \omega) \\
= \widehat{M}^{\text{opt}}(t, \omega) \quad (5.38)
\]

Hence, the smoothing effect caused by the analysis and synthesis window of the STFT can be minimized by employing the weighted sum of the auto-Wigner distributions of the Hermite functions. By the linearity property of the Weyl symbol, the time-varying system \( T \) in (5.20) representing the STFT filter can be expand into a weighted sum of STFT filters in parallel, i.e.

\[
\tilde{T}(t, \tau) = \int \int \tilde{M}^{\text{opt}}(\tau, \omega) \sum_{k=0}^{K-1} c_k^{\text{opt}} h_k(t - \tau) e^{j\omega t} h_k(\tau - \tau) e^{-j\omega \tau} d\tau d\omega \\
= \sum_{k=0}^{K-1} c_k^{\text{opt}} T(t, \tau; \tilde{M}^{\text{opt}}, h_k, h_k). \quad (5.39)
\]

This is known as the multi-window STFT filter [15]-[16] and the corresponding filtering process is illustrated in Fig. 5.7.
5.3.7 Experimental Results

To evaluate the performance of the proposed multi-window STFT based optimized filtering in the previous section, three illustrative examples are presented here. In each example, the optimized multiplicative mask is analytically generated in the same manner of that in the Section 5.2.1 using a set of five realizations of the desired signal together with their corresponding noisy observations. The resulting filter is then applied to a previously unseen realization of the distorted signal. For the sake of comparison, this particular waveform is also filtered both in the fixed-window STFT domain with the most favorable window length, and in the conventional frequency domain. For all the comparisons reported here, 25 independent realizations of the desired signals together with their corresponding noisy observations were calculated. The optimized weights \( c_k^{opt} \) were determined by the algorithm described in the Section 5.3.4. As mentioned before in the Section 5.3.4, the smoothing effect on the multiplicative mask with \( D(n, m; r) \) for \( r < 5.5 \)
is insignificant, and thus the multi-window STFT based optimized filtering will deploy the Hermite functions up to the first 8 orders. Table 5.2 shows the optimized weight $c_k^{opt}$ for $2 \leq k \leq 8$ with the corresponding values of $r$ approximating $D(n,m;r)$ without oscillatory distortions and loss of concentration. The optimized weight $c_k^{opt}$ is normalized so that $\sum_{k=0}^{K-1} c_k^{opt} = 1$.

<table>
<thead>
<tr>
<th>(order $k$, values $r$)</th>
<th>(2,11.6)</th>
<th>(3, 9.4)</th>
<th>(4, 8)</th>
<th>(5, 7.03)</th>
<th>(6, 6.4)</th>
<th>(7, 5.85)</th>
<th>(8, 5.5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_0^{opt}$</td>
<td>1.099221</td>
<td>1.284993</td>
<td>1.448821</td>
<td>1.519368</td>
<td>1.611973</td>
<td>1.637892</td>
<td>1.699879</td>
</tr>
<tr>
<td>$c_1^{opt}$</td>
<td>-0.099221</td>
<td>-0.383997</td>
<td>-0.627915</td>
<td>-0.813324</td>
<td>-0.961847</td>
<td>-1.069793</td>
<td>-1.165718</td>
</tr>
<tr>
<td>$c_2^{opt}$</td>
<td>0.099004</td>
<td>0.278733</td>
<td>0.430933</td>
<td>0.577141</td>
<td>0.696127</td>
<td>0.801443</td>
<td></td>
</tr>
<tr>
<td>$c_3^{opt}$</td>
<td>-0.099639</td>
<td>-0.236749</td>
<td>-0.341311</td>
<td>-0.456621</td>
<td>-0.548329</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$c_4^{opt}$</td>
<td></td>
<td></td>
<td>0.099772</td>
<td>0.211255</td>
<td>0.293871</td>
<td>0.378878</td>
<td></td>
</tr>
<tr>
<td>$c_5^{opt}$</td>
<td></td>
<td></td>
<td></td>
<td>-0.097212</td>
<td>-0.199718</td>
<td>-0.256024</td>
<td></td>
</tr>
<tr>
<td>$c_6^{opt}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.098242</td>
<td>0.183787</td>
<td></td>
</tr>
<tr>
<td>$c_7^{opt}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.093916</td>
</tr>
</tbody>
</table>

To show the multi-window approach which is better than the fixed-window one, we use the same signal as the third example of the Section 5.2.2. The optimized mask $\mathbf{M}^{opt}$ generated by a fixed circular Gaussian window and the related trimmed mask $\tilde{\mathbf{M}}^{opt}$ is shown in Fig. 5.8a and Fig. 5.8b, respectively. The resulting estimate
\( \hat{x}(t) \) by the proposed multi-window STFT filter using 2 and 8 Hermite windows with MSE equal to 0.255 and 0.239, are shown in Fig. 5.8c and Fig. 5.8d, respectively. The estimate \( \hat{x}(t) \) by the proposed fixed-window STFT filter using the favorable window length \( (L = 35 \text{ samples}) \) with MSE equal to 0.340 is presented in Fig. 5.8e, and is compared with the estimated signal using the corresponding filter in the frequency domain with MSE equal to 1.341 in Fig. 5.8f. It can be seen that filtering with multi-window approach improves the result in the fixed-window STFT domain and the frequency domain.
Chapter 5. Short-Time Fourier Transform based Optimized Filter

Fig. 5.8 [Example 1] (a) The optimized mask $\mathbf{M}^{opt}$; (b) The related trimmed mask $\tilde{\mathbf{M}}^{opt}$; The recovered signal $\hat{x}(t)$ (solid) and the desired signal $x(t)$ (dotted) by the proposed multi-window STFT optimized filtering using (c) 2 Hermite functions and (d) 8 Hermite functions, (e) by the proposed fixed-window STFT optimized filtering with $L=35$ samples, and by (f) the frequency domain optimized filtering.

We also use the second example of the Section 5.2.2 here for the sake of comparison and showing the improvement. The optimized mask $\mathbf{M}^{opt}$ generated by a fixed circular Gaussian window and the related trimmed mask $\tilde{\mathbf{M}}^{opt}$ is shown in Fig. 5.9a and Fig. 5.9b, respectively. The resulting estimate $\hat{x}(t)$ by the proposed multi-window STFT filter using 2 and 8 Hermite windows with MSE equal to 0.0216 and 0.0198, are shown in Fig. 5.9c and Fig. 5.9d, respectively. The estimate $\hat{x}(t)$ by the proposed fixed-window STFT filter using the favorable window length ($L=41$ samples) with MSE equal to 0.0329 is presented in Fig. 5.9e, and is compared with the estimated signal using the corresponding filter in the frequency domain with MSE equal to 0.3275 in Fig. 5.9f.
Chapter 5. Short-Time Fourier Transform based Optimized Filter

![Image](image.png)

Fig. 5.9 [Example 2] (a) The optimized mask $\mathbf{M}^{opt}$; (b) The related trimmed mask $\mathbf{M}^{opt}$; The recovered signal $\hat{x}(t)$ (solid) and the desired signal $x(t)$ (dotted) by the proposed multi-window STFT optimized filtering using (c) 2 Hermite functions and (d) 8 Hermite functions, (e) by the proposed fixed-window STFT optimized filtering with $L=41$ samples, by (f) the frequency domain optimized filtering.

We add one more example here to demonstrate the power of denoising and signal separation of our proposed method for non-linearly modulated signals. In this example, the input random process is a non-linearly modulated signal $x(t) = Ae^{i2\pi(3\cos(0.8\pi(t-s))+4\sin(0.4\pi(t-s))+10(t-s))}$, which is degraded by the addition of...
zero-mean white Gaussian noise at 0dB SNR together with another non-linearly modulated signal component $A e^{j2\pi (3 \cos(\pi(t-s)) + 4 \cos(0.5\pi(t-s)) - 7(t-s))}$, where $A$ and $s$ are uniformly distributed on the intervals [1 3] and [-0.01 0.01], respectively. A realization of the noisy observed $y(t)$ signal and its spectrogram using $L = 29$ samples is shown in Fig. 5.10a and Fig. 5.10b, respectively. The optimized mask $M^{opt}$ generated by a fixed circular Gaussian window and the related trimmed mask $\tilde{M}^{opt}$ is shown in Fig. 5.10c and Fig. 5.10d, respectively. The resulting estimate $\hat{x}(t)$ by the proposed multi-window STFT filter using 2 and 8 Hermite windows with MSE equal to 0.1231 and 0.1190, are shown in Fig. 5.10e and Fig. 6.10f, respectively. The estimate $\hat{x}(t)$ by the proposed fixed-window STFT filter using the favorable window length ($L = 29$ samples) with MSE equal to 0.1417 is presented in Fig. 5.10g, and is compared with the estimated signal using the corresponding filter in the frequency domain with MSE equal to 0.4068 in Fig. 5.10h. Once again, it can be seen that filtering by the multi-window approach can yield a superior performance. The average (over 25 repetitions of each experiment) mean-squared error values achieved for the above three examples by the compared methods are presented in Tables 5.3.
Chapter 5. Short-Time Fourier Transform based Optimized Filter

(a)  
(b)  
(c)  
(d)  
(e)  
(f)  
(g)  
(h)
Fig. 5.10 [Example 3] (a) A realization of the noisy signal $y(t)$ (solid) and the desired signal $x(t)$ (dotted); (b) Spectrogram of $y(t)$ using window with $L=29$ samples; (c) The optimized mask $\mathbf{M}^{opt}$; (d) The related trimmed mask $\tilde{\mathbf{M}}^{opt}$; The recovered signal $\hat{x}(t)$ (solid) and the desired signal $x(t)$ (dotted) by the proposed multi-window STFT optimized filtering using (e) 2 Hermite functions and (f) 8 Hermite functions, (g) by the proposed fixed-window STFT optimized filtering with $L=29$ samples, by (h) the frequency domain optimized filtering.

<table>
<thead>
<tr>
<th>Method</th>
<th>Examples</th>
<th>Example 1</th>
<th>Example 2</th>
<th>Example 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed multi-window STFT optimized filtering (8 windows)</td>
<td>AMSE (standard deviation)</td>
<td>0.2015 (0.1067)</td>
<td>0.0228 (0.0088)</td>
<td>0.3087 (0.1600)</td>
</tr>
<tr>
<td>Proposed multi-window STFT optimized filtering (2 windows)</td>
<td>AMSE (standard deviation)</td>
<td>0.2021 (0.1061)</td>
<td>0.0209 (0.0064)</td>
<td>0.3118 (0.1588)</td>
</tr>
<tr>
<td>Proposed fixed-window STFT optimized filtering</td>
<td>AMSE (standard deviation)</td>
<td>0.2563 (0.1358)</td>
<td>0.0274 (0.0035)</td>
<td>0.3781 (0.1901)</td>
</tr>
<tr>
<td>Frequency-domain optimized Filtering</td>
<td>AMSE (standard deviation)</td>
<td>0.9012 (0.4634)</td>
<td>0.2357 (0.0924)</td>
<td>0.9408 (0.4183)</td>
</tr>
</tbody>
</table>

### 5.4 Summary

In this chapter, we derived an estimator to produce an optimized STFT filter using a fixed Gaussian window in the mean square sense. Our approach reduces the number of parameters that are needed to be known and minimizes the relevant underlying
assumptions. Specifically, the presented formulation does not require knowledge of
the noise statistics or of any other degradation process. This in turn may make it
easier to apply the above filters in real-world signal processing problems. In addition,
we further improved our developed optimized STFT filter by using the
multi-window approach so that the computation used to iteratively select the
appropriate window size could be eliminated. Results indicated that the proposed
filtering technique was superior to than conventional filtering for non-stationary and
non-linearly modulated signals.

REFERENCES

    1981.
    high-resolution ECG: background and time-frequency representations,” IEEE
    Transactions on Biomedical Engineering, vol. 44, no. 4, pp. 247-255, Apr.
    1997.
    high-resolution ECG: development and application,” IEEE Transactions on


Chapter 5. Short-Time Fourier Transform based Optimized Filter


6.1 Summary of Main Conclusions

This thesis presented some novel methods for the short-time Fourier transform and contributed to two main areas: the STFT-based time-frequency signal representation and the STFT-based filtering. The development in the first area was presented in Chapter 3 and Chapter 4. On the other hand, the contributions in the second area were described in Chapter 5. The main achievements of this thesis can be summarized as follows:

In Chapter 3, we aimed to improve the inherent problem of the STFT about the TF resolution caused by the windowing function. We developed a new approach for the time adaptation of the STFT. It automatically determines the size of the analysis
window at each time instant similar to other existing time-adaptive methods. The proposed method is simple, does not require a priori knowledge of the signal or any user intervention, and admits an efficient implementation based on the FFT. Our proposed algorithm searches for the window size which – at each time instant – maximizes the energy of the spectrogram within localized TF areas. We have presented two ways to identify the localized TF areas: one is by exploiting the simple geometric relationship between the locations of the signal components and the interference terms in the WD; the other is through trimming of a fixed-window spectrogram. We have shown that both ways are mathematically equivalent, but the later one has fewer processing steps. Experimentation has shown that the proposed method is able to achieve good TF resolution, and can outperform previously reported time-adaptive spectrograms. In addition, the method has also compared favorably with two reputable quadratic TF representations. Moreover, we developed an additional processing step to deal with signals with components overlapping in time, and this step deployed the STFT analysis-modification-synthesis method to separate the signal components overlapping in time and superimposed them to form the resulting improved representation. However, the time-adaptive methods only adapt for the appropriate window size and search at each time each. The TF resolution is still low for chirp-like components, and additional processing step was required for signals with components overlapping in time. Therefore, a novel TF adaptive method was introduced in Chapter 4.

In Chapter 4, a new TF adaptive scheme based on the STFT aiming to
automatically determine the size and the chirp-rate of the analysis window at each
time-frequency point was introduced. The proposed adaptation does not require a
priori knowledge of the signal or any user intervention, and the complexity is similar
to the previous proposed time-adaptive STFT in Chapter 3 by iteratively adapting
for one parameter (i.e. the chirp-rate parameter). Experimentation has shown that the
proposed method is able to achieve good TF resolution, and can outperform
previously reported time-frequency and time adaptive spectrograms. In addition, the
method has also compared favorably with the multi-window spectrogram and two
reputable quadratic TF representations. The other main focus of this thesis is the
STFT-based filtering which is described in Chapter 5.

In Chapter 5, we derived an estimator to produce an optimized STFT
filtering mask using a fixed Gaussian window in the mean square sense. Our
approach reduces the number of parameters that are needed to be known and
minimizes the relevant underlying assumptions. Specifically, the optimized filtering
mask is estimated by a small set of training signals. This in turn may make it easier to
apply the above filters in real-world signal processing problems. In addition, we
further improved our developed optimized STFT filter by using the multi-window
approach so that the computation used to iteratively select the appropriate window
size could be eliminated. Results indicated that the proposed filtering technique was
superior to the existing conventional method for non-stationary and non-linearly
modulated signals.
6.2 Future Research Ideas

An immediate extension to the work proposed in Chapter 3 can be to adapt for both window width and chirp-rate of the analysis window at each time instant to improve the TF resolution for signals with chirp-like components. Although the computational complexity for iteratively searching for two parameters is high, it can be lowered by means of coarsely sampling the set of both parameters. Also, in Chapter 4, the window parameters for the adaptation can be extended to higher orders in order to match the characteristics for polynomial-phased signal components. Again, the computational cost would increase significantly as the phase order of the window increases. Other fast or efficient approaches would need to be developed so as to make it practical to real-life problems.

In Chapter 2, the windowing functions utilized for filtering the biomechanical impact signals and ultrasound elastography are fixed for the entire signal, although the optimized fixed-window length is empirically selected. In order to adequately match the characteristics of the window with the signal components, and obtain a better filter result, the window width should appropriately vary so as to match signals with spectral content slowly varying with time. To enhance the work carried out in Chapter 5, the fixed window used in the optimized STFT filter can also be further developed to be of variable width.